



**UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA  
DEPARTAMENTO DE FÍSICA**

**TESIS DOCTORAL**

**MODELIZACIÓN ESTADÍSTICA DE EVENTOS  
EXTREMOS DE OLEAJE Y NIVEL DEL MAR**

**FRANCISCO JAVIER GONZÁLEZ HERRERA**

**LAS PALMAS DE GRAN CANARIA  
JULIO 2013**



**D. SALVADOR GARCÍA GALVÁN SECRETARIO DEL  
DEPARTAMENTO DE FÍSICA DE LA UNIVERSIDAD DE LAS  
PALMAS DE GRAN CANARIA,**

**CERTIFICA,**

Que el Consejo de Doctores del Departamento en su sesión de fecha 15 de julio de 2013 tomó el acuerdo de dar el consentimiento para su tramitación, a la tesis doctoral titulada **“Modelización Estadística de eventos extremos de oleaje y nivel del mar”** presentada por el doctorando **D. Francisco Javier González Herrera** y dirigida por el Doctor Germán Rodríguez Rodríguez.

Y para que así conste, y a efectos de lo previsto en el Artº 6 del Reglamento para la elaboración, defensa, tribunal y evaluación de tesis doctorales de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a 16 de julio de dos mil trece.



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA  
DEPARTAMENTO DE FÍSICA

**UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA**

Departamento: FÍSICA

Programa de Doctorado: FÍSICA, MATEMÁTICAS, GEOLOGÍA Y CLIMA

**Título de la Tesis**

**MODELIZACIÓN ESTADÍSTICA DE EVENTOS EXTREMOS DE  
OLEAJE Y NIVEL DEL MAR**

Tesis Doctoral presentada por D. Francisco Javier González Herrera

Dirigida por el Dr. D. Germán Rodríguez Rodríguez

**El Director**

A handwritten signature in blue ink, consisting of several overlapping loops and a long horizontal stroke at the bottom.

**El Doctorando**

A handwritten signature in blue ink, featuring a large, stylized letter 'A' with a long horizontal stroke extending to the right.

Las Palmas de Gran Canaria, a 12 de julio de 2013



---

# Agradecimientos

---

Este trabajo de tesis doctoral ha sido dirigido por el Dr. Germán Rodríguez, a quien agradezco su esfuerzo y dedicación en las tareas de dirección y asesoramiento durante estos años de trabajo.

El Ministerio de Educación, Cultura y Deporte del Gobierno de España a través del programa de becas de Formación del Profesorado Universitario (FPU), ha financiado este trabajo haciendo posible la consecución de los objetivos planteados, así como la realización de dos estancias de investigación en Noruega y Portugal.

A los Profesores Dr. Arvid Naess y Dr. Laurens de Haan, por su amabilidad y dedicación durante mis estancias en la *Norwegian University of Science and Technology (NTNU)* y en la Universidad de Lisboa.

A mi familia por su apoyo a lo largo de mi carrera académica y profesional.

*This doctoral thesis has been directed by Prof. Dr. Germán Rodríguez, whom I thank for his dedication, leadership and advice during these years.*

*The Ministry of Education, Culture and Sports of the Spanish Government through the scholarship program of FPU has funded this work, making possible the achievement of the objectives established and the performance of two research stays in Norway and Portugal.*

*I would like to thank Professors Dr. Arvid Naess and Dr. Laurens de Haan, for their kindness and dedication during my research stays at the Norwegian University of Science and Technology (NTNU) and the University of Lisbon.*

*To my family for their support through my academic and professional career.*



---

# Índice general

---

<b>Agradecimientos</b>	<b>III</b>
<b>1 Introducción</b>	<b>1</b>
<b>2 Datos experimentales y descripción de la zona de estudio</b>	<b>9</b>
2.1 Oleaje . . . . .	9
2.2 Nivel del mar . . . . .	14
2.3 Características climatológicas de la zona de estudio . . . . .	16
<b>3 Análisis extremal univariante</b>	<b>25</b>
3.1 Introducción . . . . .	25
3.2 Métodos de selección de eventos extremos . . . . .	27
3.2.1 Método de muestreo basado en bloques . . . . .	28
3.2.2 Método de muestreo basado en un umbral . . . . .	28
3.3 Frecuencia de presentación de los eventos extremos . . . . .	30
3.4 Magnitud de los eventos extremos . . . . .	34
3.4.1 Distribuciones probabilísticas límite . . . . .	35
3.4.1.1 Distribución generalizada de valores extremos . . . . .	35

---

3.4.1.2	Distribución Generalizada de Pareto . . . . .	40
3.4.1.3	Distribuciones de probabilidad alternativas . . . . .	42
3.4.2	Métodos de estimación de parámetros . . . . .	46
3.4.2.1	Método de los mínimos cuadrados . . . . .	46
3.4.2.2	Método de los momentos . . . . .	47
3.4.2.3	Método de máxima verosimilitud . . . . .	47
3.4.2.4	Método de los $L$ -moments . . . . .	48
3.5	Modelos combinados Frecuencia-Magnitud . . . . .	57
3.5.1	Técnicas para la selección del umbral . . . . .	58
3.5.1.1	Métodos basados en propiedades teóricas del modelo . . . . .	58
3.5.1.2	Métodos basados en el número de excedencias . . . . .	60
3.5.2	Técnicas para estudiar la independencia estadística . . . . .	60
3.5.3	Modelo Poisson - GPD . . . . .	61
3.5.4	Modelo de Poisson - $K4P$ . . . . .	62
3.6	Pruebas de bondad de ajuste . . . . .	62
3.6.1	Estadística de Kolmogorov-Smirnov . . . . .	63
3.6.2	Raíz del error cuadrático medio . . . . .	64
3.6.3	Error cuadrático medio normalizado . . . . .	64
3.6.4	Coefficiente de determinación . . . . .	64
3.6.5	Gráficos cuantil-cuantil . . . . .	65
3.7	Periodo de Retorno . . . . .	65
3.7.1	Cálculo de $T$ a partir de la selección del máximo anual . . . . .	66
3.7.2	Cálculo de $T$ a partir del método POT . . . . .	67



---

3.8	Intervalos de confianza . . . . .	68
3.9	Estacionalidad de los eventos extremos . . . . .	70
3.10	Resultados y discusión . . . . .	78
3.10.1	Aplicación del método de máximos anuales . . . . .	78
3.10.2	Aplicación del método POT . . . . .	90
3.10.2.1	Selección de eventos extremos . . . . .	90
3.10.2.2	Análisis de la frecuencia de presentación de eventos extremos .	92
3.10.2.3	Análisis de la magnitud de los eventos extremos. Modelo Poisson-GPD. . . . .	95
3.10.2.4	Modelo Poisson-K4P. Análisis de la magnitud de las excedencias.	104
3.10.3	Análisis de la estacionalidad . . . . .	112
<b>4</b>	<b>Análisis extremal multivariante</b>	<b>135</b>
4.1	Introducción . . . . .	135
4.2	Modelo teórico . . . . .	138
4.2.1	Estructura de dependencia . . . . .	140
4.3	Estimación . . . . .	143
4.4	Coficiente de independencia asintótica . . . . .	146
4.4.1	Estimador de Hill . . . . .	148
4.4.2	Estimador de Peng . . . . .	149
4.5	Método de selección de los eventos extremos . . . . .	150
4.6	Resultados y discusión . . . . .	151
<b>5</b>	<b>Conclusiones</b>	<b>159</b>



---

## Índice de figuras

---

2.1	Localización de las boyas de oleaje LP-I, LP-II y mareógrafo del Puerto de Las Palmas . . . . .	10
2.2	Localización de la boya de La Coruña (CO) . . . . .	11
2.3	Mareógrafo SONAR ubicado en el Puerto de la Coruña. <i>Fuente: Puertos del Estado</i> . . . . .	15
2.4	Ilustración de las zonas de tormentas y costas dominadas por el swell (Adaptada de Davies, 1972) . . . . .	17
2.5	Condiciones de oleaje de viento y de fondo durante las épocas de invierno y verano en el Atlántico Norte (Adaptada de Titov, 1971) . . . . .	17
2.6	Trayectorias de tormentas tropicales y huracanes observadas durante el periodo 1851-2008. TD: Depresión tropical. TS: Tormenta tropical. 1-5: Categoría de huracán según escala de Saffir-Simpson. Fuente: National Oceanic and Atmospheric Administration (NOAA) . . . . .	18
2.7	Oleaje incidente en la zona de Canarias . . . . .	19
2.8	Valor medio horario anual de $H_s$ en la boya LP-I (1987-2001) . . . . .	20
2.9	Valor medio horario anual de $H_s$ en la boya LP-II (1991-2001) . . . . .	20
2.10	Valor medio horario anual del periodo medio en la boya LP-I (1987-2001) . . . . .	21
2.11	Valor medio horario anual del periodo medio en la boya LP-II (1991-2001) . . . . .	21

2.12	Boya de La Coruña (CO). Estadísticas anuales de $H_s$ y $T_p$ . Serie temporal 1982-2011. Fuente: <i>Puertos del Estado</i> . . . . .	23
3.1	Extremos definidos según los métodos de máximos anuales (rojo) y POT (azul) para un registro anual de $H_s$ con un intervalo de muestreo de 3 horas. . . . .	29
3.2	Método de Simiu (1996) para la selección de valores extremos . . . . .	31
3.3	Función cuantil de la distribución GEV para diferentes valores del parámetro de forma $k$ obtenida a través de (3.21) . . . . .	39
3.4	Diferentes formas de la distribución K4P en función de sus parámetros . . . . .	44
3.5	Configuraciones para muestras de tamaño 2,3 y 4 . . . . .	52
3.6	Relación entre $\tau_3$ y $\tau_4$ para diferentes distribuciones de probabilidad: Uniforme, Exponencial, Gumbel, generalizada de valores extremos (GEV) y generalizada de Pareto (GPD). . . . .	53
3.7	Boya LP-I. Valor medio horario anual de la altura de ola significativa (1987-2001)	71
3.8	Boya LP-II. Valor medio horario anual de la altura de ola significativa (1991-2001)	71
3.9	Boya LP-I. Distribución temporal de los valores extremos de la altura de ola significativa, $H_s(m)$ , durante el año sobre diferentes umbrales . . . . .	72
3.10	Boya LP-II. Distribución temporal de los valores extremos de la altura de ola significativa, $H_s(m)$ , durante el año sobre diferentes umbrales . . . . .	73
3.11	Método de máximos anuales para la boya LP-I. Gráficos cuantil-cuantil correspondientes a las funciones de distribución de <i>Gumbel</i> (superior), <i>Fréchet</i> (centro) y <i>Weibull</i> (inferior) utilizando los métodos MOM, MLE y LS en la estimación de parámetros . . . . .	81
3.12	Método de máximos anuales para la boya LP-I. Gráficos cuantil-cuantil correspondientes a las funciones de distribución GEV, K4P, GAM y WAK . . . . .	84
3.13	Método de máximos anuales para la boya LP-II. Gráficos cuantil-cuantil correspondientes a las funciones de distribución de <i>Gumbel</i> (superior), <i>Fréchet</i> (centro) y <i>Weibull</i> (inferior) utilizando los métodos MOM, MLE y LS. . . . .	86



3.14	Método de máximos anuales para la boya LP-II. Gráficos cuantil-cuantil correspondientes a las funciones de distribución GEV, K4P, GAM y WAK utilizando el método LMOM para la estimación de los parámetros . . . . .	88
3.15	Boya LP-I. (a) Función de autocorrelación. (b) Run test (Z) para la independencia sobre diferentes valores de $u$ e intervalos de confianza al 95 % (línea discontinua). . . . .	91
3.16	Boya LP-II. (a) Función de autocorrelación. (b) Run test (Z) para la independencia sobre diferentes valores de $u$ e intervalos de confianza al 95 % (línea discontinua). . . . .	91
3.17	Índice de dispersión (DI) para diferentes valores de $u$ e intervalos de confianza para $DI=1$ al 95 % (línea discontinua). (a) Boya LP-I. (b) Boya LP-II. . . . .	92
3.18	Boya LP-I. Gráficos de Poisson para diferentes valores del umbral $u$ . . . . .	93
3.19	Boya LP-II. Gráficos de Poisson para diferentes valores del umbral $u$ . . . . .	93
3.20	Boya LP-I. Ajuste de la distribución binomial negativa sobre diferentes valores del umbral $u$ . . . . .	94
3.21	Boya LP-II. Ajuste de la distribución binomial negativa sobre diferentes valores del umbral $u$ . . . . .	94
3.22	Boya LP-I. Función media de exceso para los datos observados (izquierda). Evolución del parámetro de forma, $k$ , de la distribución GPD con el valor umbral (derecha) . . . . .	95
3.23	Boya LP-II. Función media de exceso para los datos observados (izquierda). Evolución del parámetro de forma, $k$ , de la distribución GPD con el valor umbral (derecha) . . . . .	96
3.24	Boya LP-I. Gráficos cuantil-cuantil correspondientes a la función GPD sobre diferentes valores del umbral . . . . .	99
3.25	Boya LP-II. Gráficos cuantil-cuantil correspondientes a la función GPD sobre diferentes valores del umbral . . . . .	99

3.26	Boya LP-I. Variación de los cuantiles $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con diferentes valores del umbral. Intervalos de confianza al 95 % de confianza . . . . .	101
3.27	Boya LP-II. Variación de los cuantiles $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con diferentes valores del umbral. Intervalos de confianza al 95 % de confianza . . . . .	101
3.28	Relación entre los $L$ -moments de tercer $\tau_3$ y cuarto $\tau_4$ orden de la distribución GPD. La línea continua representa la relación $\tau_3 - \tau_4$ para los datos observados mientras que la línea discontinua representa idéntica relación teórica dada por la expresión (3.63) . . . . .	102
3.29	Variación de los parámetros de forma de las distribuciones GPD y K4P sobre diferentes valores umbrales . . . . .	105
3.30	Boya LP-I. Gráficos cuantil-cuantil correspondientes a los modelos <i>Poisson-GPD</i> (izquierda) y <i>Poisson-K4P</i> (derecha) sobre el rango de valores umbrales 2.6 - 2.8 metros . . . . .	107
3.31	Boya LP-I. Gráficos cuantil-cuantil correspondientes a los modelos <i>Poisson-GPD</i> (izquierda) y <i>Poisson-K4P</i> (derecha) sobre diferentes valores umbrales . . . . .	108
3.32	Boya CO. Gráficos cuantil-cuantil correspondientes a los modelos <i>Poisson-GPD</i> (izquierda) y <i>Poisson-K4P</i> (derecha) sobre diferentes valores umbrales . . . . .	109
3.33	Boya CO. Gráficos cuantil-cuantil correspondientes a los modelos <i>Poisson-GPD</i> (izquierda) y <i>Poisson-K4P</i> (derecha) sobre diferentes valores umbrales . . . . .	110
3.34	Boya LP-I. Variación de los parámetros de forma (superior), localización (medio), y escala (inferior) para las diferentes estaciones climáticas sobre diferentes valores umbrales . . . . .	117
3.35	Boya LP-II. Variación de los parámetros de forma (superior), localización (medio) y escala (inferior) para las diferentes estaciones climáticas sobre diferentes valores umbrales . . . . .	118
3.36	Boya LP-I. Variación de los cuantiles de $H_s$ asociados a periodos de retorno, $T_R$ , de 50, 100 y 500 años sobre diferentes valores umbrales . . . . .	122

3.37	Boya LP-II. Variación de los cuantiles de $H_s$ asociados a periodos de retorno, $T_R$ , de 50, 100 y 500 años sobre diferentes valores umbrales . . . . .	122
3.38	Boya LP-I. Gráficos cuantil-cuantil para cada estación climática sobre diferentes valores umbrales . . . . .	125
3.39	Boya LP-II. Gráficos cuantil-cuantil para cada estación climática sobre diferentes valores umbrales . . . . .	126
3.40	Boya LP-I. Identificación de estaciones hidroclimáticas homogéneas . . . . .	128
3.41	Boya LP-II. Identificación de estaciones hidroclimáticas homogéneas . . . . .	128
3.42	Boya LP-I. Comparación del cuantil $H_{50}$ obtenido con los diferentes métodos de análisis de la estacionalidad . . . . .	134
3.43	Boya LP-II. Comparación del cuantil $H_{50}$ obtenido con los diferentes métodos de análisis de la estacionalidad . . . . .	134
4.1	Altura de ola significativa ( $H_s$ ) frente a nivel del mar (SWL) en la costa norte de Gran Canaria con su respectiva zona de fallo . . . . .	136
4.2	Representación en coordenadas polares del conjunto $B_{\theta,r}$ . . . . .	143
4.3	Valores conjuntos de altura de ola significativa ( $H_s$ ) y nivel del mar (SWL) registrados durante 714 tormentas. . . . .	151
4.4	Estimaciones del parámetro de forma $\gamma$ en función del número de extremos $m$ para la altura de ola significativa $H_s$ y nivel del mar SWL . . . . .	152
4.5	Probabilidades de excedencia marginales. La línea continua representa la estimación paramétrica. El símbolo (+) representa las estimaciones empíricas. La línea vertical representa el número de extremos ( $n - m$ ) utilizados en el análisis, donde $n = 714$ y $m = 50$ . . . . .	152
4.6	Curvas cuantil. Superior: Boya LP-I. Inferior: Boya LP-II. Niveles para la función $\hat{l}$ (no-paramétrica)(izquierda) y $\hat{l}'$ (paramétrica) (derecha). Las curvas cuantil corresponden a $m = 20, 30, 40, 50, 60, 70, 80$ and $100$ . . . . .	153

4.7	Medida espectral. Superior: Boya LP-I; Inferior: Boya LP-II. Valores individuales de $\theta_i$ se muestran con el símbolo '+'. La línea continua representa la función de distribución $\theta$ . En la parte derecha se muestran las estimaciones usando los rangos; en la parte izquierda se muestran las estimaciones usando $(\hat{X}_i(n/m), \hat{Y}_i(n/m))$ . El eje horizontal se mueve en un rango de valores desde 0 ( $H_s$ ) a $\pi/2$ (SWL). . . . .	154
4.8	Estimador de Hill para el coeficiente de dependencia (línea continua) y límite de aceptación al 5 % para $\eta = 1$ (línea discontinua). Izquierda: Boya LP-I. Derecha: Boya LP-II . . . . .	155
4.9	Estimador de Peng para el coeficiente de dependencia (línea continua) y límite de aceptación al 5 % para $\eta = 1$ (línea discontinua). Izquierda: Boya LP-I. Derecha: Boya LP-II . . . . .	155
4.10	Relación entre $H_s$ y SWL durante 718 tormentas registradas en las costas holandesas. Extraído del artículo de <i>Ronde, J. de and de Haan, L. (1998)</i> y reproducido con el permiso de los autores. . . . .	158
4.11	Curvas cuantil y medida espectral para las tormentas representadas en la figura 4.10 . . . . .	158



---

# Índice de tablas

---

2.1	Características de las boyas Waverider . . . . .	12
3.1	Linealización de las distribuciones de probabilidad de Gumbel, Fréchet y Weibull . . . . .	47
3.2	Expresiones teóricas de $\tau_3$ y $\tau_4$ para diferentes distribuciones de probabilidad: Uniforme, Exponencial, Gumbel, generalizada de valores extremos (GEV) y generalizada de Pareto (GPD) . . . . .	52
3.3	Estaciones climáticas . . . . .	77
3.4	Distribuciones de probabilidad utilizadas en la caracterización de los valores máximos anuales de la altura del oleaje . . . . .	79
3.5	Método de máximos anuales. Boya LP-I. Parámetros de las funciones de distribución de <i>Gumbel</i> , <i>Fréchet</i> y <i>Weibull</i> , obtenidos con los métodos MOM, MLE y LS. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05 . . . . .	79
3.6	Método de máximos anuales. Boya LP-I. Cuantiles asociados a periodos de retorno $T_r = 50, 100, 500$ y $1000$ años. . . . .	82
3.7	Método de máximos anuales. Boya LP-I. Parámetros de las funciones de distribución GEV, K4P, GAM y WAK obtenidos con el método LMOM. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05 . . . . .	83
3.8	Método de máximos anuales. Boya LP-I. Cuantiles asociados a periodos de retorno $T_r = 50, 100, 500$ y $1000$ años. . . . .	83

3.9	Método de máximos anuales. Boya LP-II. Parámetros de las funciones de distribución de <i>Gumbel</i> , <i>Fréchet</i> y <i>Weibull</i> , obtenidos con los métodos MOM, MLE y LS. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05	85
3.10	Método de máximos anuales. Boya LP-II. Cuantiles asociados a periodos de retorno $T_r = 50, 100, 500$ y 1000 años.	87
3.11	Método de máximos anuales. Boya LP-II. Parámetros de las funciones de distribución GEV, K4P, GAM y WAK obtenidos con el método LMOM. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05	89
3.12	Método de máximos anuales. Boya LP-II. Cuantiles asociados a periodos de retorno $T_r = 50, 100, 500$ y 1000 años	89
3.13	Boya LP-I. Parámetros del modelo GPD estimados con el método LMOM e intervalos de confianza al 95 % según el método de bootstrapping para diferentes umbrales	97
3.14	Boya LP-II. Parámetros del modelo GPD estimados con el método LMOM con intervalos de confianza al 95 % según el método de bootstrapping para diferentes umbrales	97
3.15	Boya LP-I. Parámetros del modelo GPD, Error cuadrático medio (RMSE), estadístico de Kolmogorov-Smirnov para los datos observados, $\widehat{KS}$ , y valor teórico del test, $KS$ al 95 % de confianza	98
3.16	Boya LP-II. Parámetros del modelo GPD, Error cuadrático medio (RMSE) y estadístico de Kolmogorov-Smirnov para los datos observados, $\widehat{KS}$ y valor teórico del test, $KS$ al 95 % de confianza	98
3.17	Boya LP-I. Cuantiles asociados a periodos de retorno de 50, 100 y 500 años con intervalos de confianza al 95 %	100
3.18	Boya LP-II. Cuantiles asociados a periodos de retorno de 50, 100 y 500 años con intervalos de confianza al 95 %	100
3.19	Selección del valor umbral según diferentes métodos basados en el número de excedencias	103
3.20	Boya LP-I. Cuantiles de $H_s$ asociados a periodos de retorno de 100 y 500 años para diferentes ventanas temporales sobre un umbral de 2.7 metros	103

3.21	Boya LP-I. Parámetros de las distribuciones GPD y K4P, sobre diferentes umbrales, estimados a través del método LMOM, y test de Kolmogorov-Smirnov (KS). $KS_*$ representa el límite superior de aceptación con un nivel de significación 0.05. $m$ representa el número de datos sobre el umbral, $u$ . . . . .	106
3.22	Boya CO. Parámetros de las distribuciones GPD y K4P, sobre diferentes umbrales, estimados a través del método LMOM, y test de Kolmogorov-Smirnov (KS). $KS_*$ representa el límite superior de aceptación con un nivel de significación 0.05. $m$ representa el número de datos sobre el umbral, $u$ . . . . .	106
3.23	Boya LP-I. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años . . . . .	111
3.24	Boya CO. Cuantiles de $H_s$ asociados a periodos de retorno de 100, 500 y 1000 años . . . . .	111
3.25	Boya LP-I. Resultados del test de la Chi2 para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos. $T_c = 24.725$ para $\alpha=0,01$ .	113
3.26	Boya LP-II. Resultados del test de la Chi2 para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos. $T_c = 24.725$ para $\alpha=0.01$ .	113
3.27	Boya LP-I. Resultados del test de Rayleigh para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos. $\alpha=0.01$ . . . . .	114
3.28	Boya LP-II. Resultados del test de Rayleigh para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos. $\alpha=0.01$ . . . . .	114
3.29	Boya LP-I. Resultados del test modificado de Kolmogorov-Smirnov para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos. $\alpha = 0.01$ . Valor crítico=1.66 . . . . .	115
3.30	Boya LP-II. Resultados del test modificado de Kolmogorov-Smirnov para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos. $\alpha = 0.01$ . Valor crítico = 1.66 . . . . .	115
3.31	Boya LP-I. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para cada estación climática . . .	120
3.32	Boya LP-II. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para cada estación climática . . .	121

---

3.33	Boya LP-I. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para la serie temporal completa . . .	124
3.34	Boya LP-II. Cuantiles asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para la serie temporal completa . . . . .	124
3.35	Boya LP-I. Estación hidroclimática I. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %. . . . .	130
3.36	Boya LP-I. Estación hidroclimática II. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %. . . . .	130
3.37	Boya LP-II. Estación hidroclimática I. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %. . . . .	133
3.38	Boya LP-II. Estación hidroclimática II. Cuantiles de $H_s$ asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %. . . . .	133
4.1	Estimación de parámetros . . . . .	152



# 1

---

## Introducción

---

El ser humano ha intentado siempre entender las leyes que gobiernan el comportamiento de la naturaleza, con el fin de mejorar su capacidad para predecir eventos futuros. En el caso particular del oleaje, el enorme interés práctico implícito ha estimulado el desarrollo de una gran cantidad de estudios, los cuales, en conjunto con el impresionante desarrollo de los medios de computación automáticos, han permitido importantes avances en el conocimiento de dicho fenómeno.

La dinámica marina y los mecanismos involucrados en la generación del oleaje en su superficie obedecen a las leyes físicas de la Mecánica de Fluidos Geofísicos y consecuentemente, en principio, debería poder ser descrito en términos deterministas. Sin embargo, el oleaje es un proceso dinámico influenciado por un número muy elevado de factores que en muchas ocasiones interactúan mutuamente. Por consiguiente, en la práctica, debido a la complejidad del proceso, resulta extremadamente difícil lograr una descripción exacta del oleaje en un punto e instante de tiempo dados.

A pesar de la dificultad que entraña la caracterización del oleaje, los conocimientos adquiridos hasta el momento hacen que en la actualidad sea posible conocer con bastante detalle, y con unos pocos días de antelación, las condiciones de oleaje que se darán en un lugar dado del océano, en un instante concreto, mediante el uso de modelos numéricos (Janssen, 2008). No obstante, la fuerte demanda de resultados prácticos en el campo de la predicción del oleaje ha forzado la obtención de soluciones en las que la complejidad

del problema y las dificultades para obtener y analizar apropiadamente suficientes datos experimentales ha sido salvada empleando soluciones parcialmente empíricas. Generalmente, este tipo de solución fuerza la necesidad de ajustar algunos parámetros del modelo. El ajuste de dichos parámetros se realiza, en general, considerando el conjunto total de observaciones. En consecuencia, cuanto más raro es, o menor frecuencia de presentación posee un cierto valor de las observaciones, peor será representado por el modelo al realizar el ajuste de los parámetros (Cavaleri, 2009). Este hecho es especialmente significativo si en estos casos inusuales, o raros, la física del proceso en estudio cambia de modo sustancial. Esta situación se da claramente en el uso de modelos de oleaje para condiciones extremas, esto es, durante las condiciones de oleaje más severas. Por tanto, no resulta extraño comprobar que durante las tormentas más intensas los modelos muestran una marcada tendencia a infraestimar los valores de altura de ola, tanto más cuanto más severas son las condiciones de oleaje (Bidlot, 2002). Además, las secuencias de parámetros integrales derivados de los modelos presenta una estructura más suavizada que la correspondiente a las observaciones instrumentales, eliminando la variabilidad natural del fenómeno (Bertotti and Cavaleri, 2012). En definitiva, los modelos físico-matemáticos han sido desarrollados para caracterizar las condiciones medias del oleaje, no las situaciones más extremas, o menos frecuentes.

Además de la citada limitación de los modelos físico-matemáticos para reproducir las condiciones de oleaje durante condiciones extremas, existe otra deficiencia importante de los mismos, relacionada con la ventana de tiempo sobre la cual son capaces de proporcionar predicciones adecuadas. Según se ha comentado anteriormente, los conocimientos actuales permiten la implementación de modelos que ofrecen predicciones útiles con un horizonte temporal en el rango de unos pocos días. Naturalmente, la razón de estas limitaciones se encuentra en la complejidad y, por consiguiente, dificultad para predecir adecuadamente el fenómeno generador del oleaje (el viento) para periodos de tiempo de mayor extensión, así como la complejidad intrínseca del oleaje, resultado de las interacciones mutuas, a diferentes escalas espaciales y temporales, entre la atmósfera y la superficie del océano, además de otros factores que complican aún más el proceso.

Los valores más grandes y más pequeños adoptados por una variable aleatoria, en un número finito de medidas, son denominados valores extremos. Estos valores, así como otros valores en las partes más altas y bajas del rango de la variable, son generalmente de mayor interés que la mayoría de las observaciones, concentradas entorno al valor medio. Los valores extremos constituyen a su vez una variable aleatoria cuya predictibilidad resulta de enorme importancia práctica. Así, por ejemplo, el diseño óptimo de estructuras marítimas costeras, tales como diques rompeolas y muros de protección, así como las estructuras oceánicas,

---

tales como plataformas, aerogeneradores marinos o tecnologías para el aprovechamiento de la energía de las olas, requieren de la estimación precisa de las condiciones extremas del oleaje que la estructura deberá soportar durante su periodo de vida. Este periodo puede oscilar, dependiendo del tipo de estructura, desde unas pocas decenas de años hasta periodos superiores al siglo. En consecuencia, en la inmensa mayoría de los casos, resulta necesario conocer los valores extremos que pueden afectar a las estructuras durante periodos de tiempo sustancialmente superiores a aquellos de los que se dispone de observaciones.

Las deficiencias mostradas por los modelos fisico-matemáticos de predicción del oleaje para reproducir las condiciones más severas y de posibilitar predicciones a largo plazo de las condiciones de oleaje, eliminan la posibilidad de su uso para intentar conocer las condiciones de oleaje más severas que se espera encontrar en una zona dada para horizontes temporales muy grandes, tal como se requiere en campos como la Ingeniería de Costas y Oceánica. En este sentido, la descripción probabilística es la única herramienta viable para caracterizar un fenómeno cuando no se conocen todos los aspectos involucrados en el mismo, e incluso si se conoce todo, la caracterización probabilística continúa siendo un procedimiento conveniente si la complejidad del proceso es muy elevada (Sornette, 2004).

Con frecuencia, los daños causados por los fenómenos naturales están relacionados con una inadecuada seguridad estructural resultante de la infraestimación del riesgo, lo cual implica el conocimiento de la probabilidad de presentación de eventos que puedan provocar desperfectos en las estructuras. Por ello, la estimación precisa de los parámetros de diseño es de suma importancia para el éxito de las estructuras. No obstante, es necesario tener presente que los recursos pueden ser derrochados, no sólo cuando ocurren desperfectos, sino también cuando las estructuras son sobredimensionadas con el fin de evitarlos. La decisión se centra en un balance entre vivir con el riesgo y los costes de evitarlo. En este contexto, una aproximación científica que ha sido ampliamente utilizada para optimizar dicho balance es el denominado diseño probabilístico, cuyo fundamento reside en el análisis estadístico de las condiciones más severas, o extremas, de los fenómenos de interés.

Al aumentar la magnitud de un proceso peligroso, su frecuencia de ocurrencia disminuye. En consecuencia, las mayores catástrofes naturales son el resultado de un pequeño número de eventos de gran intensidad, o severidad, que raramente ocurren. Así, si se representa la magnitud de los eventos frente al logaritmo de su frecuencia de ocurrencia se observa un descenso aproximadamente lineal de la frecuencia de ocurrencia con el aumento de la intensidad. El intervalo de recurrencia, o periodo de retorno, es el intervalo de tiempo que, en promedio, transcurre entre eventos en los que la magnitud del proceso iguala o excede una intensidad específica. Una representación de los intervalos de recurrencia frente

a las magnitudes asociadas da lugar a un conjunto de puntos que se adaptan aproximadamente a una línea recta, en escala semilogarítmica, de forma que al aumentar el periodo de retorno aumenta la magnitud de la variable que es excedida.

De lo anterior se deduce que, el peligro asociado con la ocurrencia de un cierto fenómeno potencialmente catastrófico depende de su magnitud y de su frecuencia de presentación. Desde un punto de vista práctico resulta interesante conocer tanto la frecuencia con la que se superan determinados valores de una variable dada, como el intervalo de tiempo esperado entre eventos en los que se supera dicho valor. Naturalmente, la respuesta a las cuestiones anteriores incluyen un considerable nivel de incertidumbre. No obstante, el análisis estadístico de valores extremos resulta de gran utilidad práctica, permitiendo el uso de conceptos como el de evento de diseño, que representa la magnitud de un evento peligroso que una estructura debería ser capaz de soportar durante su periodo de vida.

Por definición, los eventos extremos ocurren con muy poca frecuencia y, por tanto, cualquier análisis estadístico relacionado con eventos extremos de una variable debe abordar el comportamiento de la cola de la distribución, o de los cuantiles extremos, haciendo uso de un conjunto reducido de datos históricos. Los métodos estadísticos tradicionales basados en la densidad de probabilidad de todos los valores de la variable no resultan adecuados para tal fin, dado que generalmente los mismos generan ajustes adecuados en el rango donde se encuentra la mayoría de las observaciones, pero a costa de un ajuste inadecuado en las colas. La teoría de valores extremos (EVT) constituye una rama de la estadística que se centra explícitamente en las observaciones extremas, es decir, su objeto de interés es la cola de la distribución y no la parte central. Esta teoría proporciona la posibilidad de realizar extrapolaciones fuera del rango de las observaciones, a partir de una cantidad reducida de registros de valores extremos, o tormentas significativamente intensas (Coles, 2001; Goda, 2000).

La teoría de los valores extremos, empleada para realizar extrapolaciones sobre horizontes temporales considerablemente superiores a los periodos que comprenden las observaciones disponibles, constituye una teoría matemática bien establecida, cuyos fundamentos se basan en la suposición de que los modelos (distribuciones probabilísticas) límite sugeridos por la teoría asintótica continúan siendo válidas para niveles extremos, pero finitos.

Los métodos de análisis basados en la teoría de valores extremos no predicen el futuro con exactitud, pero constituyen procedimientos adecuados para explicar y caracterizar el comportamiento de los eventos extremos observados. Estos modelos no son arbitrarios, sino que están basados en una teoría matemática rigurosa sobre el comportamiento de los

---

valores extremos. Es decir, la caracterización de los eventos extremos de un fenómeno dado no constituye un simple ejercicio de ajuste de distribuciones teóricas a distribuciones empíricas. No obstante, es importante tener en cuenta que en cualquier tipo de análisis de datos existen diferentes tipos de incertidumbre, y que además éstos pueden verse magnificados en el análisis de valores extremos.

En particular, la incertidumbre en la estimación de los parámetros resulta de gran importancia en este contexto. Así, incluso cuando se dispone de un número elevado de observaciones de calidad y un modelo adecuado, la estimación de los parámetros está sometida a un nivel de incertidumbre sustancial, y la inferencia es considerablemente sensible a pequeñas variaciones en los parámetros, especialmente en el parámetro de forma del modelo.

Por otra parte, también existe un cierto grado de incertidumbre asociado al modelo. Es posible disponer de un buen conjunto de datos experimentales pero un modelo no adecuado. Al hacer uso de la teoría de valores extremos trabajamos con una clase de modelos útiles y justificados teóricamente, aunque éstos son aplicables sobre valores umbrales altos, no existiendo un procedimiento robusto que permita elegir con certeza el valor del mismo. Si el umbral es establecido en un valor muy alto se tiene como resultado un número bajo de extremos, introduciendo mayor incertidumbre en los parámetros estimados. Si, por el contrario, se selecciona un umbral muy bajo la justificación teórica del modelo pierde consistencia. En tal sentido, Makkonen (2008) señala que un problema común en el análisis de valores extremos es que la creencia en la aplicabilidad de la teoría de valores extremos es tan fuerte que este tipo de análisis suele ser realizado incluso cuando no resulta razonable esperar un ajuste adecuado, debido a la naturaleza asintótica de la teoría.

Un tercer tipo de incertidumbre es la asociada a los datos. En este sentido, una muestra de miles de datos puede ser suficientemente amplia para realizar inferencias sobre la media de la distribución haciendo uso del teorema central del límite. Sin embargo, la inferencia en relación a la cola de la distribución está impregnada de una mayor incertidumbre, dado que sólo una cantidad reducida de puntos forman parte de esta zona de la distribución. En consecuencia, la introducción de nuevos datos a la serie de valores extremos puede tener un impacto sustancial en la inferencia. Otro aspecto de la incertidumbre asociada a los datos es la dependencia estadística. El análisis extremal está basado fundamentalmente en la hipótesis de valores independientes e idénticamente distribuidos. En la práctica, especialmente en procesos geofísicos, es frecuente encontrar efectos de agrupamiento de los valores extremos, tendencias, variaciones estacionales, así como otros tipos de dependencias.

La precisión de las técnicas empleadas para realizar extrapolaciones fuera del rango de las observaciones resulta considerablemente sensible a los diferentes tipos de errores anteriormente comentados, por lo que su minimización constituye una área de gran interés en el campo del análisis extremal y, lógicamente, los avances en este campo poseen una gran repercusión práctica.

Tal como se señaló anteriormente, la incertidumbre intrínseca a toda metodología estadística se acentúa notablemente en el análisis de valores extremos debido a la escasez de información. Dada la baja frecuencia de presentación de los eventos extremos, la cantidad de información es siempre un factor fuertemente limitante y, en consecuencia, se requiere de registros muy largos para poder reducir la incertidumbre asociada a las estimaciones y extrapolaciones realizadas, las cuales deben ir acompañadas de sus correspondientes intervalos de confianza.

Todo lo anterior continúa siendo válido, pero se complica sustancialmente, cuando los eventos extremos son el resultado de la acción conjunta de dos o más fenómenos naturales. Así, por ejemplo, las inundaciones costeras constituyen un grave peligro y suelen tener graves repercusiones socioeconómicas. Este fenómeno suele ser el resultado de la ocurrencia conjunta de estados de mar severos y niveles del mar elevados. En consecuencia, la caracterización de eventos extremos multivariados requiere conocer la distribución conjunta de los diferentes factores constituyentes. Es decir, la probabilidad de ocurrencia simultánea de dos o más variables que pueden estar parcialmente relacionadas. En el caso particular del oleaje y los niveles del mar, es de esperar que exista un cierto grado de dependencia entre las condiciones de oleaje y la componente de marea meteorológica, puesto que ambos fenómenos están relacionados con las condiciones meteorológicas locales (Hawkes et al, 2002), especialmente en situaciones en las que los sistemas meteorológicos capaces de generar fuertes sobreelevaciones del nivel medio y fuertes vientos soplando desde una dirección en la que el oleaje puede alcanzar un elevado grado de desarrollo energético. No obstante, la dependencia estadística entre ambos fenómenos puede resultar enmascarada por factores como la presencia de oleaje de fondo superpuesto al oleaje local (e.g., Rodríguez et al, 1998), haciendo que la caracterización conjunta de ambos fenómenos resulte más compleja aún.

La teoría de valores extremos multivariante es la herramienta utilizada para describir la distribución conjunta de eventos extremos generados por la combinación de dos o más variables y su desarrollo es mucho más reciente que la teoría univariante (Coles and Tawn, 1991; de Hann and de Ronde, 1998; Repko et al., 2004).

Con el objetivo de estudiar los aspectos mencionados en los párrafos anteriores, el presente trabajo se ha estructurado de la siguiente forma.

El Capítulo 2 ofrece una descripción de las condiciones climatológicas de la zona de estudio, así como de las series temporales de datos analizadas. Estos conjuntos de datos pertenecen a localizaciones geográficas con climatología marítima diferenciada.

El Capítulo 3 está dedicado al análisis univariante de eventos extremos de oleaje. En él se ofrece una introducción a la teoría básica y se presentan los resultados obtenidos en relación a varios de los aspectos antes citados.

En particular, se realiza una comparación entre las dos principales metodologías empleadas para el estudio de valores extremos, es decir, el método de máximos anuales y el de picos sobre un umbral (POT), ajustando diferentes funciones de distribución de probabilidad teóricas a los datos empíricos. Como indicador, se examinan las posibles diferencias existentes en los cuantiles asociados a periodos de retorno elevados.

Se explora la independencia estadística de los extremos seleccionados al establecer diferentes valores umbrales, así como la validez de la hipótesis de que su frecuencia de presentación se distribuye según una distribución de Poisson.

Se evalúan y comparan diferentes metodologías para la elección del valor umbral requerido en el método POT. En particular, se consideran dos tipos de métodos, unos basados en las propiedades teóricas del modelo y otros cuyo fundamento radica en el número de excedencias observadas.

Se introduce la distribución Kappa de cuatro parámetros (K4P) para la modelización de las excedencias sobre un umbral. Esta distribución contiene como caso particular a la distribución generalizada de Pareto (GPD) y añade mayor flexibilidad en el ajuste de los datos experimentales.

El estudio de los efectos de las covariables sobre el comportamiento de los eventos extremos es otro de los aspectos que se abordan en este capítulo. En concreto, se examina el efecto de la covariable estacionalidad, procediendo a su análisis a través de varias metodologías. En primer lugar, se hace uso de técnicas de análisis estadístico de variables circulares con el fin de evaluar la presencia o no de estacionalidad. Una vez verificada la existencia, o no, de dicho patrón se divide la serie temporal en diferentes periodos atendiendo a criterios climáticos e hidroclimáticos, y se examinan los efectos de estos procedimientos en las estimaciones resultantes.

En el Capítulo 4 se presenta una aplicación del análisis extremal multivariante al caso de la altura de ola y nivel del mar. En particular, se investiga la existencia de dependencia entre los valores extremos de ambas variables, haciendo uso de metodologías paramétricas y no paramétricas.

El Capítulo 5 recoge las conclusiones más relevantes del presente trabajo.



## 2

---

# Datos experimentales y descripción de la zona de estudio

---

En este capítulo se describen las características de los conjuntos de datos utilizados en el presente trabajo, las zonas de estudio y las diferentes fuentes de información para la medida del oleaje y nivel del mar.

## 2.1 Oleaje

Los registros de oleaje empleados en este estudio han sido obtenidos mediante boyas acelerométricas del tipo *Waverider* pertenecientes a la red REMRO (Red de Medida y Registro de Oleaje) del Ministerio de Fomento, gestionada por el Ente Público Puertos del Estado.

En concreto, estos datos corresponden a los registros obtenidos mediante boyas instaladas en la costa Norte y Este de la isla de Gran Canaria y en la costa de La Coruña. En la figura 2.1 se ilustra la localización de la boya situada en el norte de Gran Canaria, en adelante denotada como  $LP - I$ , y la boya situada en la costa Este, denotada como  $LP - II$ . La figura 2.2 muestra la boya situada en La Coruña, denotada como  $CO$ . Las características de las estaciones y su ubicación exacta se muestran en la tabla 2.1.

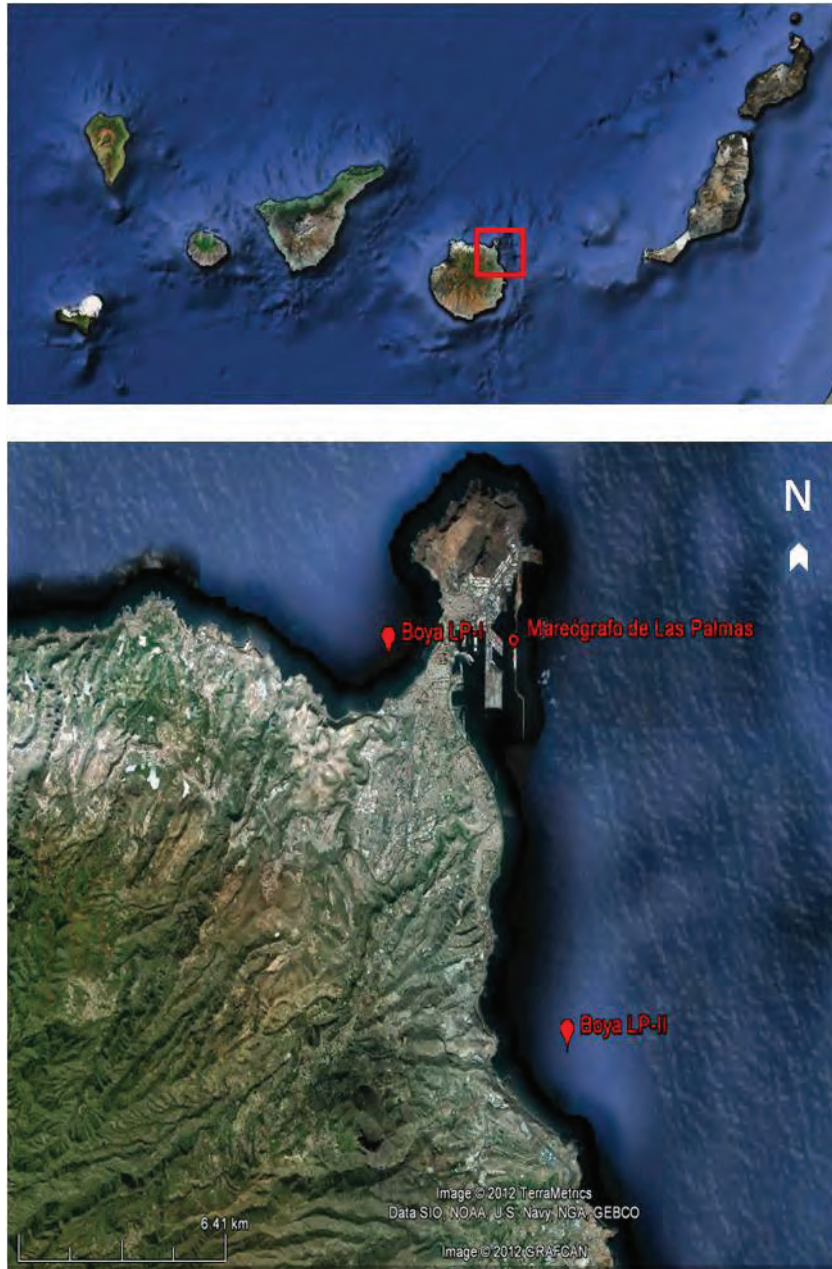


Figura 2.1: Localización de las boyas de oleaje LP-I, LP-II y mareógrafo del Puerto de Las Palmas



Figura 2.2: Localización de la boya de La Coruña (CO)

Las series de elevaciones de la superficie registradas en cada una de las boyas tienen una duración aproximada de 42 minutos y una cadencia de muestreo de una hora. La frecuencia de muestreo es de 2 Hz. Los periodos cubiertos por las medidas examinadas son 15 años (1987-2001) para la boya LP-I y 11 años (1991-2001) para la boya LP-II y CO.

La boya acelerométrica es el instrumento estándar que ofrece mayor precisión en la medida de la altura, periodo y dirección del oleaje, existiendo diferentes modelos comerciales. Dos de los modelos más destacados son las boyas WAVERIDER y TRIAXYS, fabricadas por las empresas DATAWELL (Holanda) y AXYS Technologies (Canadá), respectivamente.

El fundamento físico en el que se basa el funcionamiento de las boyas acelerométricas es sencillo. Un cuerpo esférico de dimensiones y densidad adecuadas flotando en la superficie del mar está sometido a movimientos ascendentes y descendentes, por efecto del oleaje, experimentando aceleraciones verticales que, si pudiesen ser medidas con suficiente precisión, permitirían obtener los desplazamientos verticales de la superficie libre del mar

Tabla 2.1: Características de las boyas Waverider

	LP-I	LP-II	CO
Latitud	28° 8.5' N	28° 4' N	43° 24.7' N
Longitud	15° 27.5' W	15° 23.8' W	8° 23.0' W
Profundidad	42 m	48 m	50 m
Sensor	Waverider	Waverider	Waverider
Longitud de la serie de datos	1987-2001	1991-2001	1991-2001

mediante una doble integración. Sin embargo, aunque la idea básica resulte elemental, su implementación práctica conlleva numerosos problemas.

La boya Waverider está constituida esencialmente por una esfera metálica de 1 metro aproximadamente de diámetro, en cuyo interior se encuentra un acelerómetro y la circuitería electrónica asociada para el tratamiento de las señales procedentes del transductor. En su parte superior se sitúa una antena que permite transmitir las señales medidas a una estación receptora, situada normalmente en la costa.

De esta forma, el dispositivo realiza la medida de las aceleraciones de forma continua, una doble integración de la señal para disponer de las sobreelevaciones sobre el nivel medio, y su transmisión hacia la estación receptora.

Prácticamente todos los pasos que constituyen el procedimiento de medida del oleaje empleando una boya Waverider pueden generar una serie de errores que deben ser minimizados si se desea disponer de registros con una calidad aceptable.

El primer problema de este tipo de sensor es el de mantener el acelerómetro en posición vertical. Para ello, la boya Waverider emplea una esfera, en la cual se suspende una plataforma pesada, en forma de disco, que sustenta el acelerómetro y que se encuentra inmersa en un líquido. El conjunto del disco, el acelerómetro y el fluido constituye un péndulo físico que se ajusta para que posea un periodo natural de oscilación críticamente atenuado, lográndose de esta forma que para frecuencias en el rango de las asociadas al oleaje (0.04 - 1 Hz) el acelerómetro mantenga efectivamente la verticalidad. El motivo de montar el acelerómetro en una plataforma estabilizada para mantener la verticalidad, no es otro que evitar el efecto de aceleraciones horizontales que puedan ser provocadas por los movimientos de balanceo y cabeceo de la boya, que puedan enmascarar los valores de la aceleración vertical, que es la significativa para el fenómeno de estudio.

Si el objetivo es medir las fluctuaciones de la superficie libre del mar en un punto determinado, resulta evidente que la boya debe ser fijada de alguna forma, de modo que su posición media no varíe con el tiempo. Para ello, se emplea un sistema de anclaje al fondo mediante una cuerda de material resistente y con un cierto grado de flexibilidad, que permita a la boya seguir la trayectoria de las partículas de fluido.

El sistema de anclaje de las boyas puede generar problemas importantes si su configuración no está correctamente estudiada, para adecuarla a las condiciones existentes en la zona de medida. Por ejemplo, en zonas donde las corrientes son muy intensas, como pueden ser las corrientes de marea en algunos lugares, el cabo del anclaje puede llegar a estar tenso, impidiendo el libre movimiento de la boya y afectando notablemente las medidas (e.g., Dillon y Kahn 1981). Este fenómeno también puede ocurrir ocasionalmente para olas esporádicas, de dimensiones no previstas durante el diseño del sistema de anclaje (e.g., James 1986).

En estas boyas, el acelerómetro está sometido al campo eléctrico creado por un fluido, cuya composición y características son secreto comercial. La conductividad del fluido se ajusta para que la boya y la circuitería que incorpora tengan la mayor sensibilidad posible. Puesto que la conductividad es función de la temperatura, el diseño del dispositivo se realiza de modo que las variaciones de conductividad, debidas a las variaciones de la temperatura, afecten lo mínimo posible a la sensibilidad de la boya. Por ello, Datawell proporciona curvas de compensación de la temperatura en condiciones de diseño. Pero los aumentos de la conductividad eléctrica del fluido modifican los efectos de la temperatura considerados bajo condiciones de diseño (e.g., Ribe 1981), haciendo que tales curvas de compensación no sean estrictas.

En ocasiones, no es posible de registros instrumentales suficientemente largos para abordar la caracterización de las condiciones a largo plazo y, en consecuencia, de las condiciones extremas del oleaje. Por ello, existen otras fuentes de datos que han sido o son utilizadas en tales condiciones. Entre éstas destacan las bases de datos visuales, las obtenidas mediante técnicas de retrosección y las derivadas de observaciones remotas.

En el caso de los *datos visuales*, los datos de oleaje y viento son recogidos por barcos meteorológicos situados en posiciones fijas sobre el océano, y por observadores voluntarios en barcos que navegan a lo largo de todos los mares y océanos. Sin embargo, los datos tomados de esta forma están distribuidos de forma irregular en el espacio y en el tiempo, además de poseer un elevado sesgo y una elevada incertidumbre. No obstante, la gran cantidad de información que se ha recogido de esta forma desde los años 50 hace que,

utilizada de forma adecuada, sea una fuente de información muy valiosa, si se carece de observaciones instrumentales.

Al contrario de lo que suele suceder con los datos meteorológicos, los cuales suelen tener una cobertura temporal amplia, la disponibilidad de datos oceanográficos, en este caso de oleaje, suele ser mucho más limitada. A partir de campos de vientos conocidos, y haciendo uso de modelos numéricos que simulan el comportamiento del oleaje, es posible reproducir las condiciones de este fenómeno ocurridas en una zona y periodo determinados. Esta técnica de estimación recibe el nombre de *retrospección* o *hindcasting*.

Tal como ya se ha comentado, en el contexto del análisis de los valores extremos del oleaje, la principal limitación que tiene esta metodología es que, en general, los modelos numéricos de oleaje no reproducen adecuadamente sus características en condiciones severas, por lo que, siendo adecuada la técnica para el estudio de las condiciones medias del oleaje, no es adecuado para el estudio de los valores extremos.

Otra posible alternativa frente a la no disponibilidad de medidas instrumentales in-situ es el uso de *técnicas de teledetección*. El estudio de los parámetros característicos del oleaje a través de observación remota se basa en el envío de ondas de radio a través de una fuente de condiciones conocidas para posteriormente estudiar la dispersión de dichas ondas tras su interacción con la superficie del mar. Los diferentes métodos existentes difieren entre sí por el rango de frecuencia de las ondas emitidas. Sin entrar en detalle, cabe destacar las técnicas de radar de alta frecuencia, radar de navegación en banda X y altimetría.

Su utilidad en el estudio de eventos extremos está aún bastante limitada. En el caso del altímetro se dispone de una buena cobertura espacial pero una frecuencia de muestreo muy baja. Los dispositivos de radar de alta frecuencia y radar de navegación, ubicados en costa, ofrecen un adecuado balance entre cobertura espacial y temporal. Sin embargo, los periodos durante los cuales se han obtenido registros, en unas pocas zonas del océano, no son aún suficientes para abordar estudios a largo plazo.

## 2.2 Nivel del mar

Con el objetivo de realizar el análisis conjunto del comportamiento extremal del oleaje y el nivel del mar se ha utilizado la información registrada por el mareógrafo situado en el Puerto de Las Palmas. Este mareógrafo está gestionado por el Ente Público Puertos del

Estado y pertenece a la red de medida REDMAR, que integra mareógrafos a lo largo de toda la costa española.

Concretamente se trata de un mareógrafo acústico modelo SONAR, fabricado por la empresa *SONAR Research and Development*, y está constituido por un emisor-receptor de ultrasonidos autocalibrado que se coloca a cierta distancia sobre la superficie del agua para registrar el nivel del mar, con su fecha y hora. Los datos son recogidos por una unidad que los transmite por radio o directamente a un PC, al que se puede acceder vía modem. La figura 2.3 muestra el mareógrafo situado en el Puerto de La Coruña.

La serie temporal comprende un periodo de 10 años, entre los años 1992 y 2001, con una frecuencia de muestreo horaria ofreciendo las siguientes variables:

- nivel del mar total (*Sea Water Level - SWL*): nivel del mar registrado por el mareógrafo
- marea astronómica: predicción del nivel del mar en base a la marea teórica
- marea meteorológica (*storm surge*): diferencia entre el nivel de mar medido por el mareógrafo y la marea astronómica



Figura 2.3: Mareógrafo SONAR ubicado en el Puerto de la Coruña. Fuente: *Puertos del Estado*

## 2.3 Características climatológicas de la zona de estudio

A excepción de las áreas que experimentan ciclones tropicales, en general, las mayores tormentas se producen en los denominados cinturones de tormentas, donde los vientos son de gran intensidad y notablemente variables. En estas zonas, las costas están sometidas en una elevada proporción a condiciones de oleaje de viento. En el hemisferio norte, el cinturón de tormentas cubre un área mucho más extensa que en el hemisferio sur, tal como se observa en la figura 2.4, y las tormentas ocurren fundamentalmente durante el invierno. Naturalmente, el oleaje generado en las zonas de tormentas se propaga hacia otras zonas del océano en forma de oleaje de fondo. De este modo, el oleaje generado en el cinturón de tormentas del norte del Atlántico Norte puede propagarse hacia el sur en forma de swell.

En particular, las tormentas localizadas en la región en la que se ubica el denominado sistema de bajas presiones, o de borrascas, de Islandia genera frecuentemente oleaje que se propaga en forma de swell hacia la zona de las Islas Canarias. En función de lo anteriormente expresado, el área de Canarias se encuentra entre las denominadas zonas dominadas por el swell, tal como se observa en la figura 2.4.

En la figura 2.5 (Titov, 1971) se observa como el swell que se aproxima a Canarias procede fundamentalmente del NNW, especialmente de la zona comprendida entre Canadá e Islandia, donde se sitúa la zona de bajas presiones de Islandia, dando lugar a fuertes tormentas. El oleaje generado en dicha zona se propaga hacia Canarias en forma de swell, que durante el invierno presenta un oleaje de fondo casi permanente y que durante aproximadamente el 20% de los casos supera los 3 metros de altura. Este oleaje de fondo persiste, aunque con menor intensidad, durante el verano, época en la que puede presentarse un oleaje de fondo generado al norte de las mismas.



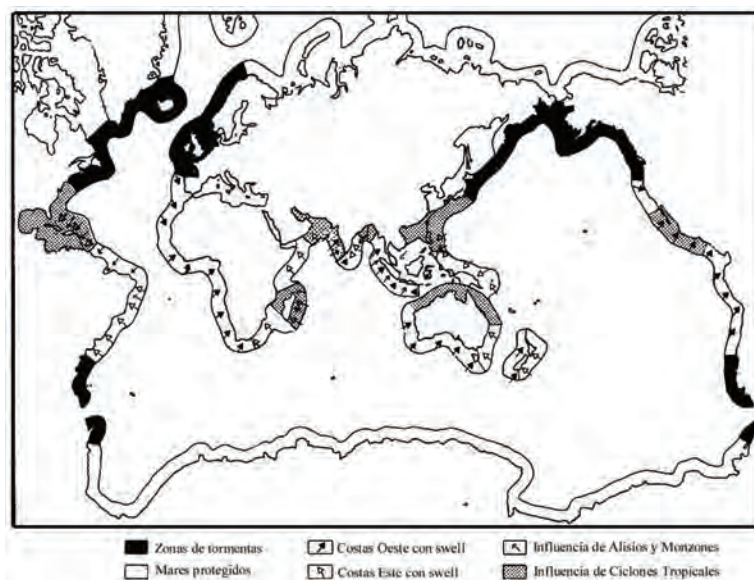


Figura 2.4: Ilustración de las zonas de tormentas y costas dominadas por el swell (Adaptada de Davies, 1972)

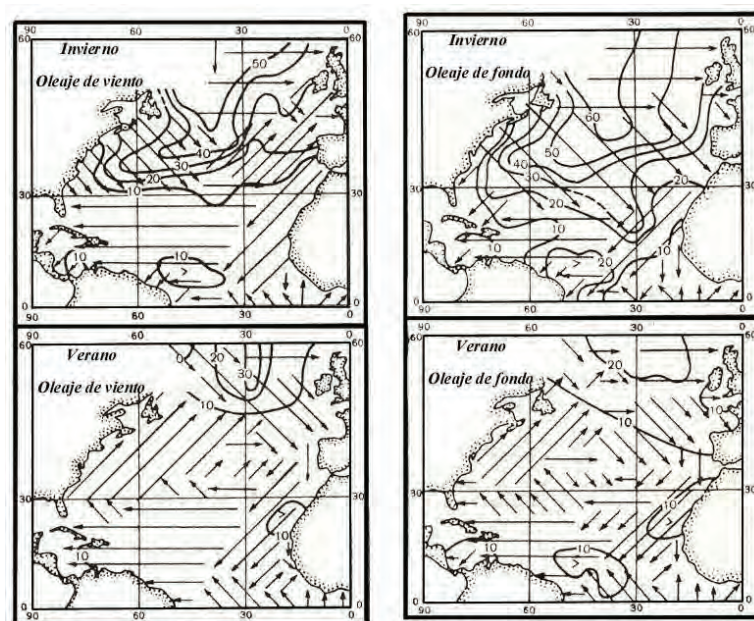


Figura 2.5: Condiciones de oleaje de viento y de fondo durante las épocas de invierno y verano en el Atlántico Norte (Adaptada de Titov, 1971)

Por otro lado, el régimen de vientos dominante en la zona es el de los vientos alisios, soplando con intensidad entre débil y moderada desde el N-NNE, especialmente en verano. De este modo, el oleaje generado por los alisios es en realidad un oleaje de viento que alcanza tanto la parte norte como la parte noroeste de la isla de Gran Canaria, tal como se observa en la figura 2.5.

Por el contrario, durante el invierno, cuando el anticiclón de las Azores se debilita y la actividad del ciclón de Islandia se intensifica, éste genera campos de oleaje de intensidad notable que se propagan en forma de swell hacia las islas, procedentes del NNW. En consecuencia, este tipo de oleaje afecta claramente al norte de Gran Canaria donde, con frecuencia, se combina con el oleaje de viento dando lugar a campos de oleaje mixtos, mientras la zona noreste, donde está ubicada la boya LP-II, se ve protegida por la propia isla, sintiendo mucho menos el efecto de este oleaje de fondo.

Por lo tanto, en general, el oleaje en el norte de las islas Canarias estará caracterizado por la presencia de un *swell* dominante procedente del NNW, especialmente durante el invierno, y un oleaje de viento procedente fundamentalmente del NNE, en especial durante el verano, pudiéndose detectar también en ocasiones un swell joven procedente de esta misma dirección. No obstante, también es necesario tener en cuenta la posibilidad de que, de forma esporádica, a las islas llegue oleaje de fondo generado por las tormentas tropicales que, tal como se observa en la figura 2.6, suelen seguir trayectorias que prácticamente rodean esta zona. Del mismo modo, de forma esporádica, las islas experimentan la presencia de tormentas que, generalmente, se aproximan desde el sector N-NW, y en menor medida desde el sur.

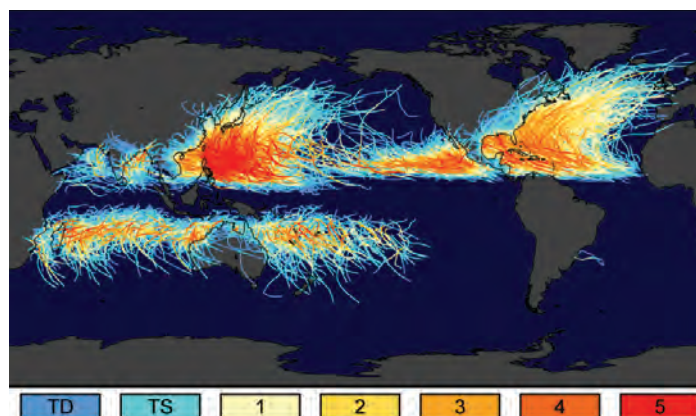


Figura 2.6: Trayectorias de tormentas tropicales y huracanes observadas durante el periodo 1851-2008. TD: Depresión tropical. TS: Tormenta tropical. 1-5: Categoría de huracán según escala de Saffir-Simpson. Fuente: National Oceanic and Atmospheric Administration (NOAA)

En función de lo anterior, resulta evidente que el oleaje en la zona norte de las islas presentará con bastante frecuencia una estructura frecuencial bimodal, debida a la combinación en este área del oleaje de fondo procedente del NNW y del oleaje de viento generado por los alisios, que soplan desde el NNE, tal y como se ilustra de forma esquemática en la figura 2.7.

En las figuras 2.8 y 2.9 se muestran los valores horarios medios durante el año para la altura de ola significativa, en el Norte de la isla (LP-I) y el Noreste (LP-II), respectivamente. Obsérvese que, mientras en la zona norte la altura de ola presenta un patrón estacional bastante claro, este rasgo no parece reflejado en la figura correspondiente a la zona noreste. En general, en la zona norte la altura de la ola es mayor durante el invierno y alcanza sus valores mínimos durante el verano, siendo ésta la única época en la que las alturas de ola son similares en ambas zonas.

Este patrón estacional se hace aún más evidente al examinar los valores medios horarios anuales de los periodos medios del oleaje en ambas zonas, mostrados en las figuras 2.10 y 2.11, para la boya LP-I y LP-II, respectivamente. En la figura correspondiente al norte de la isla se aprecia claramente el aumento del periodo medio durante los meses de invierno,

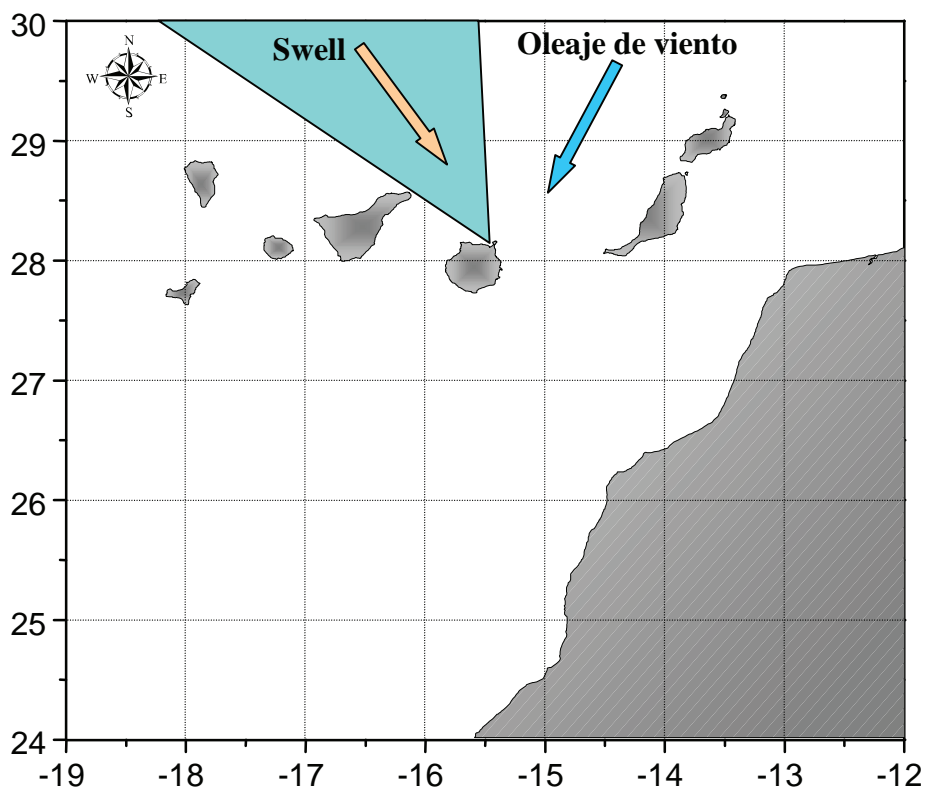


Figura 2.7: Oleaje incidente en la zona de Canarias

finales de otoño y comienzos de la primavera, causado por la presencia del swell procedente de la zona de tormentas del Atlántico Nor-noroeste, que no afecta a la zona noreste de la isla, al quedar ésta protegida por la propia isla de este tipo de oleaje. Durante el verano, debido a la intensificación del anticiclón de las Azores, y bajo el predominio de los alisios, en ambos puntos el periodo del oleaje alcanza valores similares.

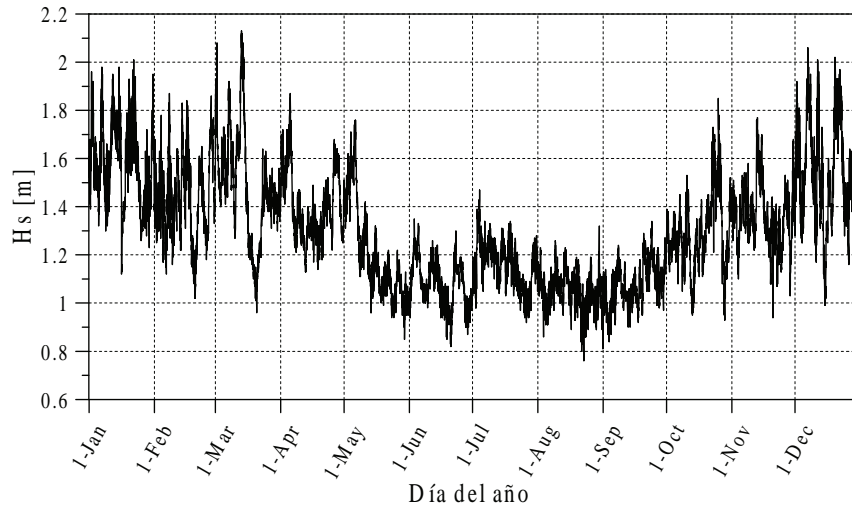


Figura 2.8: Valor medio horario anual de  $H_s$  en la boya LP-I (1987-2001)

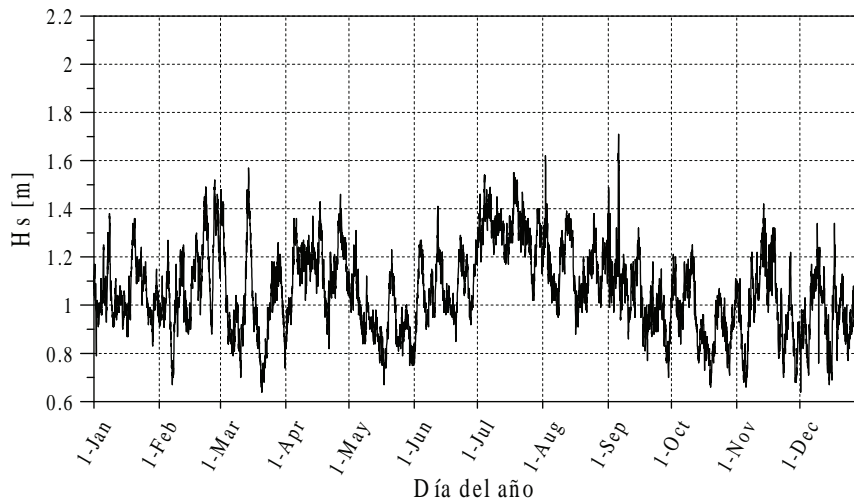


Figura 2.9: Valor medio horario anual de  $H_s$  en la boya LP-II (1991-2001)

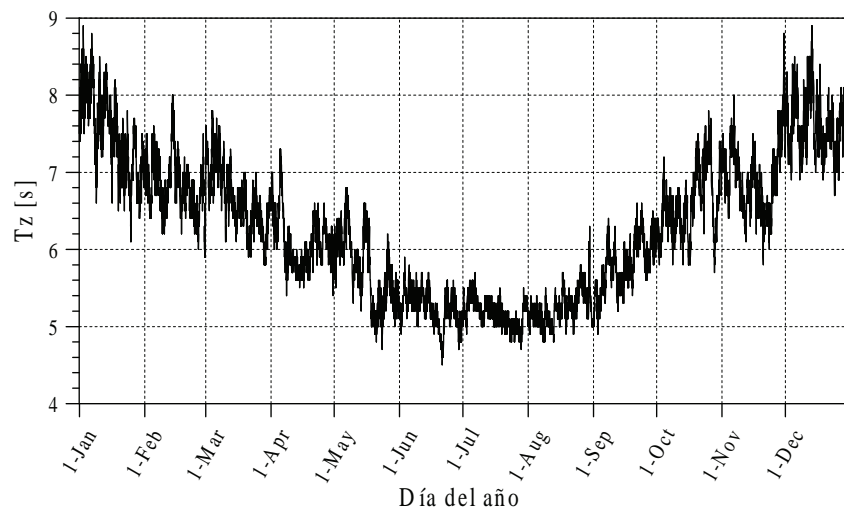


Figura 2.10: Valor medio horario anual del periodo medio en la boya LP-I (1987-2001)

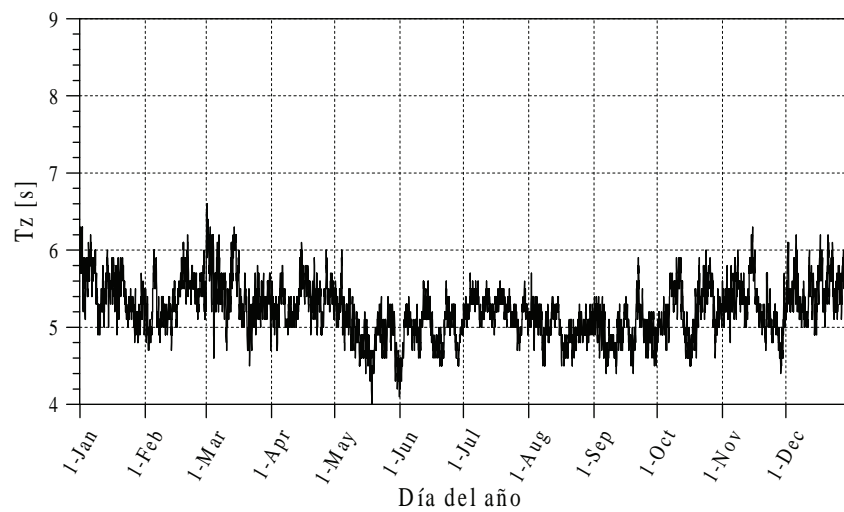


Figura 2.11: Valor medio horario anual del periodo medio en la boya LP-II (1991-2001)

La boya que registra las condiciones de oleaje de La Coruña se encuentra fondeada a una profundidad de 50 metros, donde incide un oleaje generado por el anticiclón de las Azores y por las borrascas originadas en altas latitudes del Atlántico Norte, según lo comentado anteriormente.

El clima marítimo en esta zona posee un carácter tormentoso dada su ubicación dentro de la zona de tormentas descrita en la figura 2.4, al contrario de lo que sucede con las islas Canarias que se encuentran en una zona dominada por la presencia de oleaje de fondo, observándose en esta zona una probabilidad mayor de observarse tormentas tropicales, según la figura 2.6.

La figura 2.12, extraída del informe de régimen medio elaborado por el Ente Público Puertos del Estado, muestra las condiciones de oleaje registradas por la boya CO. Puede observarse una mayor dispersión en la altura y periodo del oleaje durante el invierno, con periodos centrados entre 12-14 s, mientras que durante el verano la dispersión disminuye, centrandose los valores del periodo entre 8-10 s. Dicha figura también revela la existencia de un claro patrón estacional especialmente en términos de la altura de ola significativa que aumenta considerablemente en la estación de invierno y sufre una reducción notable en verano.

Si comparamos el clima marítimo de Galicia y Canarias, una de las diferencias más significativas a simple vista se pone de manifiesto en la altura de ola durante la estación de invierno, donde los temporales de mayor intensidad generan estados de mar mucho más severos, haciendo posible encontrar valores de altura de ola considerablemente mayores.

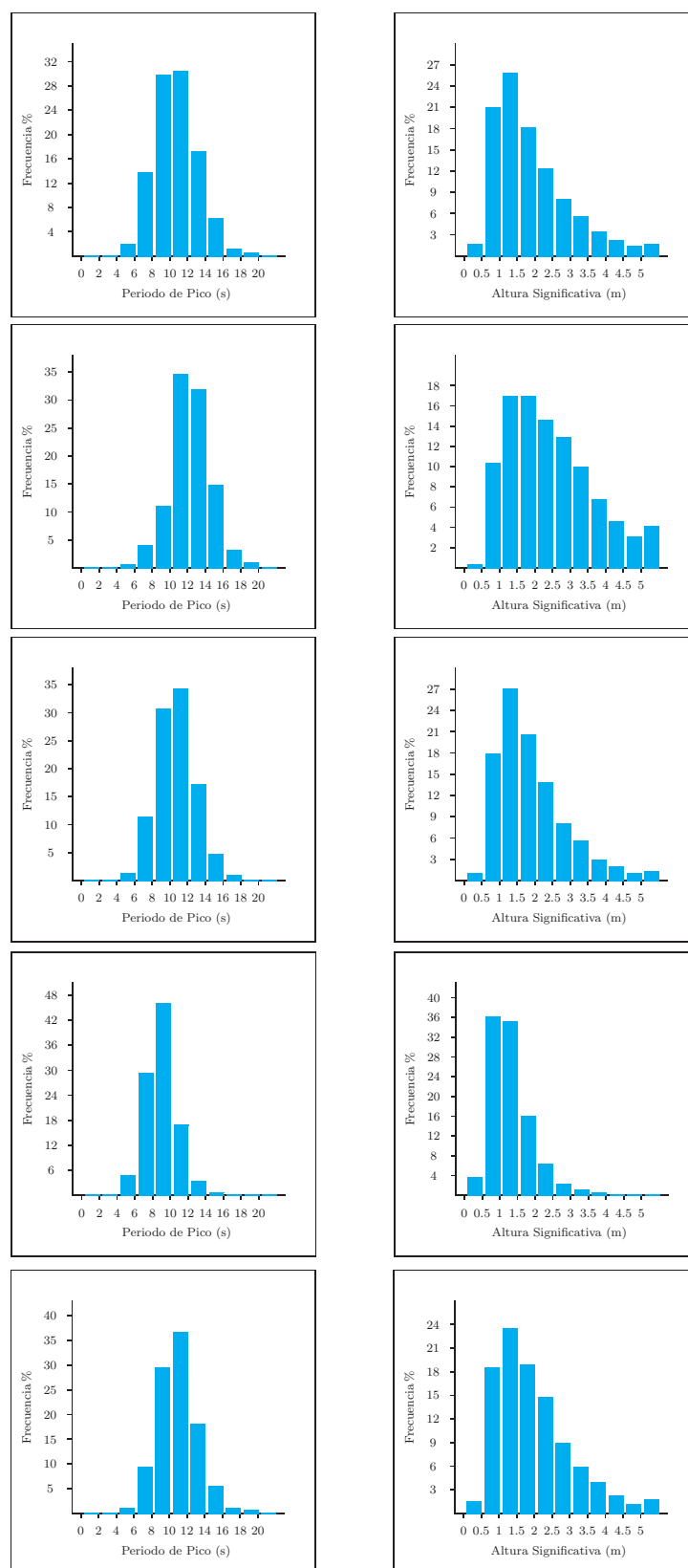


Figura 2.12: Buaya de La Coruña (CO). Estadísticas anuales de  $H_s$  y  $T_p$ . Serie temporal 1982-2011. Fuente: Puertos del Estado





# 3

---

## Análisis extremal univariante

---

### 3.1 Introducción

La caracterización estocástica del oleaje se puede abordar considerando diferentes escalas temporales. En la práctica, los estudios centran su atención sobre dos escalas temporales diferentes, aunque interrelacionadas, que reciben los nombres de estudios a corto plazo y a largo plazo. En el estudio a corto plazo la caracterización estocástica del oleaje se refiere fundamentalmente a la descripción del proceso aleatorio constituido por las fluctuaciones instantáneas de la superficie libre del mar,  $\eta(t)$ , donde  $t$  oscila, en general, desde unos pocos segundos hasta varias horas. La duración del proceso para su análisis a corto plazo debe ser tal que durante dicho periodo se satisfaga la condición de estacionaridad. Durante dicho periodo temporal, denominado estado de mar, se admite que los fenómenos físicos que gobiernan el comportamiento del oleaje permanecen estacionarios. En otras palabras, las características del oleaje durante un estado de mar permanecen estadísticamente invariantes. Sobre esta escala temporal las cantidades aleatorias de mayor interés son las propias fluctuaciones verticales instantáneas de la superficie libre, así como las alturas y los periodos de las olas individuales, definidas según determinados criterios. A largo plazo, la modelización del oleaje se refiere principalmente a la caracterización del comportamiento estocástico de las series temporales de parámetros espectrales, extraídos del análisis a corto plazo, sobre periodos temporales que oscilan desde uno a múltiples años. En este caso el

proceso es claramente no estacionario, incluyendo periodicidades, variabilidades interanuales, y a mayor escala temporal. Es decir, para escalas temporales superiores a la duración de un estado de mar no se impone la hipótesis de estacionariedad.

En conexión con lo anteriormente expuesto, el análisis de valores extremos se aplica también sobre ambas escalas temporales. En la escala temporal a corto plazo, el interés se centra en el estudio de los extremos de los parámetros característicos correspondientes a un estado de mar. Esto es, máximos o mínimos de  $\eta(t)$ , secuencias de alturas de ola individual, etc. En la escala a largo plazo, el objetivo es, generalmente, la caracterización estocástica de las excedencias de los parámetros espectrales, tales como la altura de ola significativa, que se espera ocurran durante periodos futuros sustancialmente elevados, normalmente 50 ó 100 años. Los contenidos del presente capítulo se refieren al estudio de los valores extremos de la altura de ola significativa a largo plazo.

El estudio de los valores extremos de la altura de ola significativa, como parámetro empleado para representar la severidad de un estado de mar, resulta de enorme interés en Ingeniería Oceánica, Ingeniería de Costas e Ingeniería Naval, así como en la gestión de las zonas costeras. En particular, su estudio resulta imprescindible en el diseño de estructuras como plataformas de prospección y extracción petrolífera, puertos, diques, rompeolas, el establecimiento de rutas de navegación óptimas, diseño de embarcaciones, establecimiento de normas de seguridad naval, protección de playas contra la erosión, y un largo etc.

Existen numerosos factores que afectan la precisión de las estimaciones de los valores de retorno, tales como la cantidad y calidad de los datos disponibles, los criterios empleados para definir las tormentas e identificar los máximos de tormentas independientes, la elección del valor umbral, las covariabilidades, etc. En consecuencia, resulta imprescindible utilizar metodologías que permitan valorar la bondad del modelo para caracterizar la estructura probabilística de los valores extremos, así como establecer intervalos de confianza que indiquen el grado de fiabilidad de las estimaciones resultantes del proceso de extrapolación hacia el rango de la variable en la que no existen observaciones.

Dos de las técnicas de análisis de valores extremos más utilizadas son la de máximos por bloque y la de excedencias sobre un umbral (POT). Una importante limitación y debilidad ambos métodos reside en que dichos procedimientos adoptan distribuciones asintóticas, cuya aplicabilidad y validez resultan difíciles de verificar. Recientemente, con la intención de aliviar las limitaciones implícitas en estas aproximaciones, Naess y Gaidai (2009) proponen una nueva y prometedora metodología, constituida por una combinación de un método estadístico para extraer la distribución de los valores extremos inherente a las observaciones

y un método para la extrapolación de la misma asumiendo un cierto comportamiento en la cola de la distribución. El procedimiento para extraer la distribución de valores extremos a partir de las observaciones recibe el nombre de *Average Conditional Exceedance Rate Method*, o método ACER de forma abreviada, el cual presenta ventajas sustanciales frente a los métodos clásicos cuando es adecuadamente implementado. De forma breve, comentar que este método proporciona una representación no paramétrica exacta de la distribución de los valores extremos de las observaciones. Además, resulta útil como una técnica de diagnóstico muy eficiente para revelar la estructura de dependencia de las series temporales analizadas. El procedimiento no requiere la independencia estadística de los datos. Es decir, el método hace uso de todas las observaciones y no resulta necesario un procedimiento inicial de desagrupamiento. Además, el método es aplicable a series temporales no estacionarias en aquellos casos en los que la cantidad de datos sea suficiente para reflejar apropiadamente la no estacionaridad del proceso. En consecuencia, en tales condiciones, no es necesaria la caracterización explícita de los efectos estacionales, los cuales son considerados automáticamente por el método. Algunas aplicaciones recientes de esta metodología pueden encontrarse en Karpa y Naess (2013), Skjong et al., (en prensa).

Naturalmente, la validación de esta metodología eliminaría en gran medida las limitaciones y problemas asociados a los métodos actualmente empleados en el análisis de valores extremos. No obstante, debido a su desarrollo tan reciente, no es considerada en el presente trabajo, aunque ya se está trabajando en tal sentido.

En los siguientes apartados se analizarán en detalle varios de estos aspectos, aplicando las diferentes técnicas de análisis a varios conjuntos de datos experimentales.

## 3.2 Métodos de selección de eventos extremos

La selección de los valores extremos de una variable aleatoria dada suele realizarse fundamentalmente a través de dos métodos, denominados método de los máximos anuales y método de picos sobre un umbral, los cuales se describirán de forma detallada a continuación. No obstante, existen otras opciones para abordar el estudio de los eventos extremos. Así, por ejemplo, algunos autores abordan dicho estudio considerando todos los valores registrados de altura de ola significativa, sin distinguir periodos de tormenta. Este procedimiento suele recibir el nombre de *método de la distribución inicial* (e.g., Goda, 1988; Berger, 1993; Mathiesen *et al.* 1994). Otra metodología utilizada para caracterizar los eventos extremos de oleaje consiste en seleccionar todos aquellos valores que exceden un cierto

valor umbral (e.g., Berger, 1993). Los procedimientos anteriores presentan el importante inconveniente de que los valores considerados presentan dependencia estadística entre sí, motivo por el cual no se recomienda su uso en la práctica y en consecuencia no son utilizados en el presente trabajo.

Otra de las técnicas existentes consiste en la selección de los  $r$  valores máximos en un año, en lugar de seleccionar simplemente el máximo anual, y ajustar posteriormente la distribución conjunta de esos  $r$  valores máximos a los datos (e.g., Smith, 1986). Ejemplos de aplicación de esta metodología al oleaje son las realizadas por Guedes Soares y Scotto (2004); Soukissian Y Arapi (2011). El problema principal de esta metodología radica en la selección del valor  $r$ . Esta metodología no es empleada en el presente trabajo por no ser común su uso en la práctica.

### 3.2.1 Método de muestreo basado en bloques

El modelo teórico está basado en el máximo  $M_n$  de un conjunto de datos con  $n \rightarrow \infty$ . Resulta evidente que en la práctica, la cantidad de información de la que se dispone es limitada, por lo que el procedimiento de trabajo consiste en dividir la serie temporal en bloques, escogiendo posteriormente el valor máximo dentro de cada uno de ellos. La forma tradicional de trabajar es tomar la longitud de cada bloque igual a un año, de forma que tendremos tantos valores extremos como años comprende la serie temporal. En este caso el método recibe el nombre de *método de máximos anuales*.

Una desventaja de este método es que presupone que el valor máximo anual de la variable es un valor extremo, hecho que no tiene por que ser cierto. Por otro lado, es necesario tener una serie temporal suficientemente larga, no menos de 15 años según algunos autores, para tener un número suficiente de extremos que garantice la fiabilidad estadística de los resultados. En aquellos casos en los que la información no es suficiente es posible reducir la longitud temporal de los bloques, de forma que, por ejemplo, se seleccione el valor máximo mensual. En este último caso se debe poner especial atención en la independencia estadística de los extremos, una de las hipótesis básicas en las que se fundamenta el modelo.

### 3.2.2 Método de muestreo basado en un umbral

La forma de obtener la población extremal en este caso se basa en lo que se conoce como *método de picos sobre un umbral* o *Peaks Over Threshold* (POT). La forma de proceder del

método POT permite seleccionar de una manera más racional aquellos eventos considerados como extremos, ya que se establece un valor umbral,  $u$ , a partir del cual se considerarán los eventos como sucesos extremos. De esta forma, se utiliza una mayor cantidad de información respecto al método de máximos anuales, tal como se ilustra en la figura 3.1, en la que se muestran los valores de  $H_s$  registrados cada tres horas durante un año, en la costa Noruega, identificándose el máximo anual (señalado en rojo) y los picos sobre el umbral de 8 metros (señalados en azul). Naturalmente, este incremento de información se traduce en una mayor fiabilidad estadística de los resultados finales.

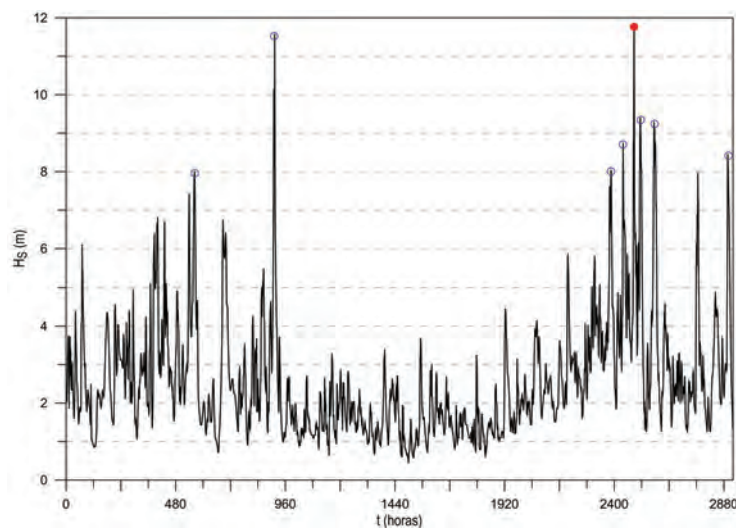


Figura 3.1: Extremos definidos según los métodos de máximos anuales (rojo) y POT (azul) para un registro anual de  $H_s$  con un intervalo de muestreo de 3 horas.

Sin embargo, es precisamente en la selección de este nivel umbral donde radica la mayor complicación del método. Un valor umbral demasiado bajo violaría una de las hipótesis básicas del modelo, la cual establece que  $u$  debe ser suficientemente alto, mientras que un valor umbral muy alto dejaría una población extremal muy pequeña, siendo imposible realizar estimaciones fiables.

Resulta importante destacar el hecho de que no existe un criterio unificado para definir estrictamente lo que se considera como un *pico* o *extremo*. Así, puede observarse en Coles (2001) como se emplea el método POT estableciendo un umbral  $u$  y considerando todos los datos que están por encima de este umbral, por lo que cada dato se consideraría como un pico o extremo. Otros trabajos como el de Valk (1994) definen periodos temporales durante el cual la variable permanece sobre el valor umbral de forma que el valor máximo de la variable en ese periodo es tomado como el valor extremo correspondiente. Por otro

lado, Simiu (1996), a partir de una serie temporal dada, calcula el valor máximo de la variable en periodos temporales que deben ser previamente definidos.

Una característica común que deberían cumplir todos estos métodos es la generación de una población extremal estadísticamente independiente, siendo fieles así a la hipótesis de independencia estadística en que se basa el modelo. En el presente trabajo se ha optado por trabajar con el método de Simiu (1996), utilizado también en González *et al.* (2005), y descrito a continuación.

El método comienza seleccionando el valor máximo de la variable dentro de un periodo temporal que recibe el nombre de *ventana temporal*, cuya longitud o duración, vendrá determinada por las características físicas del fenómeno bajo estudio. En el caso del oleaje, un valor típico para esta ventana temporal suele ser de 96 horas. Este valor temporal se justifica en el hecho de que este es el tiempo medio que suelen durar las perturbaciones atmosféricas causantes del oleaje, Simiu (1996) y Walton (2000). De esta forma se tiene cierta garantía de que cada pico o máximo seleccionado pertenezca a perturbaciones diferentes. Para asegurar la independencia estadística entre los picos de periodos sucesivos, si los máximos de dos periodos sucesivos están separados por menos de 48 horas (mitad del periodo inicial) el menor de los dos máximos es reemplazado por el siguiente menor valor en el periodo correspondiente, de forma que los picos estarán separados siempre, al menos, 48 horas, según se muestra en la figura 3.2. Como se demuestra posteriormente, este valor es coherente con el comportamiento observado en la función de autocorrelación de la serie temporal, aunque debe comentarse que otros valores para la ventana temporal podrían ser también válidos.

### 3.3 Frecuencia de presentación de los eventos extremos

Al estudiar el comportamiento estadístico de los eventos extremos de oleaje no solo resulta de interés caracterizar la intensidad de las tormentas sino también la frecuencia con la que éstas ocurren.

Así, se define la variable  $\theta$  como el tiempo transcurrido entre dos eventos consecutivos, y cuyas funciones de densidad,  $f(x)$ , y distribución,  $F(x)$ , de probabilidad toman la forma

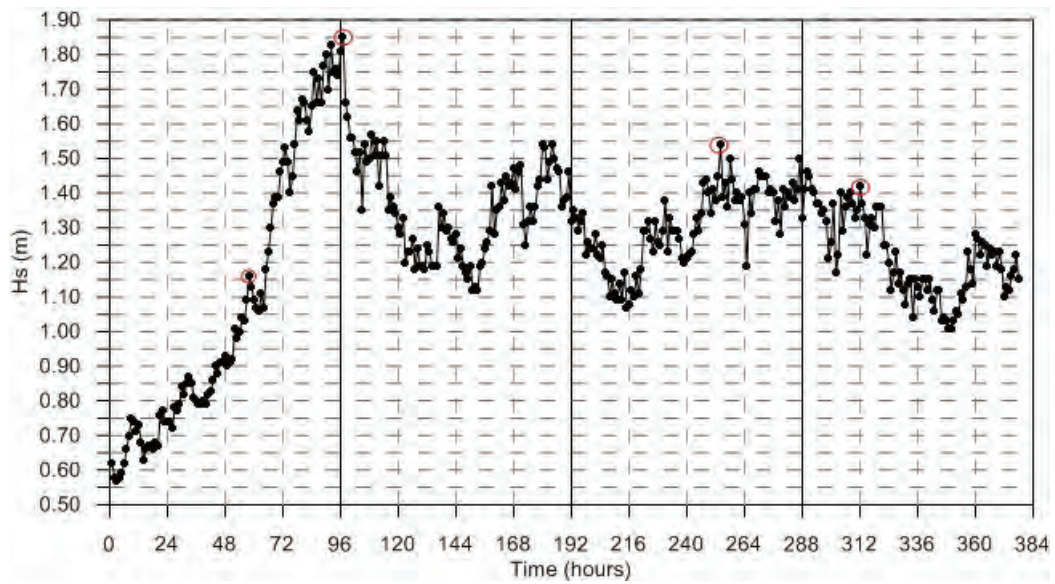


Figura 3.2: Método de Simiu (1996) para la selección de valores extremos

$$f(x)dx = \text{Prob} [x < \theta < x + dx] \quad (3.1)$$

$$F(d) = \text{Prob} [\theta < d] \quad (3.2)$$

El valor medio de la variable  $\theta$  queda definido como

$$T = E(\theta) = \int_0^{+\infty} \theta f(\theta) d\theta \quad (3.3)$$

Se definen además las siguientes variables:

- 1)  $N_t$ , número de eventos en el intervalo de tiempo  $[0, t]$ ,  $p_m[0, t] = \text{Prob}[N_t = m]$
- 2)  $M(t)$ , número medio de eventos en el intervalo de tiempo  $[0, t]$ ,  $M(t) = E(N_t)$
- 3)  $\lambda(t) = dM(t)/dt$ , intensidad del proceso
- 4)  $DI = \text{Var}(N_t)/E(N_t)$ , índice de dispersión

Normalmente el periodo temporal  $[0, t]$  corresponde a un año, es decir  $t = 1$ , por lo que a partir de ahora se usará la notación  $N$  en lugar de  $N_t$  para hacer referencia al número de eventos en un año.

La forma más sencilla y común de proceder a continuación es asumiendo que  $N$  puede modelizarse a través de una distribución de Poisson (Cunnane, 1979). En este sentido, el número  $N$  de eventos que tiene lugar en cualquier año constituye una variable de Poisson con probabilidad:

$$p_\lambda(m) = \text{Prob}(N = m) = e^{-\lambda} \frac{\lambda^m}{m!}, \quad m = 1, 2, \dots \quad (3.4)$$

El parámetro de frecuencia  $\lambda$ , que representa el número medio de eventos por año, puede ser estimado a partir de la media muestral, según  $\hat{\lambda} = \bar{m}$ . La asunción de que el número de eventos  $N$  puede ser caracterizado a través del modelo de Poisson implica que las ocurrencias deben ser estadísticamente independientes entre sí, hecho no siempre observable. Por otro lado, es posible demostrar, Lang *et al.* (1997), que el tiempo transcurrido entre dos eventos sucesivos puede caracterizarse a través de una distribución exponencial según

$$F(d) = \text{Prob}[\theta < d] = 1 - \exp(-\lambda d) \quad (3.5)$$

Así, el número de eventos  $N$  en el intervalo  $[0, t]$  queda caracterizado a través de una distribución de *Poisson* con valor medio  $E(N) = \lambda t$ , varianza  $\text{Var}(N) = \lambda t$  e índice de dispersión  $DI = 1$ . El tiempo transcurrido entre eventos sucesivos  $\theta$  está exponencialmente distribuido con valor medio  $T = 1/\lambda$ .

El índice de dispersión resulta útil en la medida que proporciona información a cerca de la idoneidad de la distribución de Poisson en la caracterización del número de eventos, de forma que cuanto más próximo esté al valor 1 mejor será la caracterización de  $N$ . Por el contrario, un valor del  $DI$  lejos de la unidad indicaría que la distribución de Poisson podría no ser la opción más adecuada, pudiendo entonces existir alguna otra distribución que caracterice mejor la variable  $N$ . Un estimador muestral para el  $DI$  queda definido según

$$\hat{DI} = \sigma^2 / \mu \quad (3.6)$$

donde  $\mu$  y  $\sigma^2$  representan la media y la varianza muestral respectivamente.

Los intervalos de confianza pueden ser calculados a través de la distribución  $\chi^2$  con  $M-1$  grados de libertad, donde  $M$  representa el número total de años que componen la



muestra. La hipótesis de que la distribución de Poisson caracteriza adecuadamente a la variable  $N$  no puede ser rechazada si el valor del  $DI$  está dentro del rango  $(\chi_{\alpha/2, M-1}^2/(M-1), \chi_{1-\alpha/2, M-1}^2/(M-1))$ , donde  $\alpha$  es el nivel de significación.

Otra forma de evaluar la hipótesis de *Poisson* es a través del uso del gráfico de Poisson (Hoaglin, 1980). Este gráfico se basa en la propiedad de linealidad que presenta la distribución de Poisson tras una serie de transformaciones. El valor medio del número de eventos  $N$  se define como

$$E(N) = Np_{\lambda}(m) = Ne^{-m} \frac{\lambda^m}{m!} \quad m = 0, 1, 2, \dots \quad (3.7)$$

Tomando logaritmos en ambos lados de (3.7) se obtiene

$$\log E(N) = \log(N) - \lambda + m \log(\lambda) - \log(m!) \quad (3.8)$$

Haciendo uso de las frecuencias observadas se define la serie  $x_0, x_1, x_2, \dots$ , donde cada término  $x_i$  representa el número de observaciones igual a  $i$  ( $x_0$  es el número de ocasiones en las que  $N = 0$ ,  $x_1$  el número de ocasiones en las que  $N = 1$ , etc). De esta forma, se asume que para un valor fijo de  $\lambda$  cada frecuencia observada  $x_m$  es igual a la frecuencia esperada  $E(N)$ . Obtenemos entonces la expresión

$$\log x_m = \log(N) - \lambda + m \log(\lambda) - \log(m!) \quad (3.9)$$

de forma que, representando  $\log(x_m) + \log(m!)$  frente a  $m$  obtendremos una línea recta de pendiente  $\log(\lambda)$  y corte con el eje de ordenadas  $\log(N) - \lambda$ .

Existen ocasiones en las que el modelo de *Poisson* no es adecuado para caracterizar el número de ocurrencias. Esto sucede en aquellos casos en los que las ocurrencias del fenómeno físico en cuestión presentan dependencia temporal, de forma que no pueden ser consideradas estadísticamente independientes.

Tales casos pueden presentarse durante el estudio de fenómenos físicos que presentan un marcado carácter estacional. En estos casos se ha comprobado que el uso de la distribución binomial negativa es una alternativa a la distribución de Poisson, mejorando significativamente el ajuste de los datos observados, *Cunnane (1979)*, *Ben-Zvi (1991)*, *Lang (1997, 1999)*, *Bhunya (2013)*. No obstante, cabe señalar que Onöz y Bayazit (2001) analizando

datos de inundaciones, observan que, incluso en aquellos casos donde la distribución binomial negativa ajusta mejor los datos observados que la distribución de Poisson, los periodos de retorno calculados en base a ambas distribuciones muestran resultados muy similares.

La distribución binomial negativa tiene por función de densidad de probabilidad

$$P(N = m) = \binom{r+m-1}{m} p^r (1-p)^m \quad m = 0, 1, \dots \quad (3.10)$$

donde  $r > 1$  y  $0 < p < 1$ .

Una particularidad de esta distribución es que su varianza es mayor que su valor medio, por lo que el índice de dispersión es mayor que 1. Se establece así un criterio teórico en base al cual queda justificada la utilización de la distribución de Poisson o binomial negativa en la caracterización del número medio de eventos por año. Cuando el valor del  $DI$  es mayor que uno la distribución binomial negativa debería caracterizar mejor a la variable  $N$ , mientras que para valores del  $DI$  próximos a la unidad la distribución de Poisson sería la más adecuada.

De forma similar a como sucede en el caso de la distribución de *Poisson*, tras una serie de transformaciones, es posible definir una relación de linealidad en la distribución binomial negativa. Siendo  $X$  el número de fracasos antes del  $r$ -ésimo éxito en una secuencia de pruebas independientes de *Bernouilli*, la función de densidad de probabilidad de  $X$  viene determinada por (3.10), donde tomando logaritmos se obtiene

$$\ln x_m = \ln N - \ln \binom{r+m-1}{m} + r \ln p + m \ln(1-p) \quad (3.11)$$

de forma que la representación de  $\ln x_m - \ln N + \ln \binom{r+m-1}{m}$  frente a  $m$  es lineal con pendiente  $\ln(1-p)$  y corte con el eje de ordenadas  $r \ln p$ . Por lo tanto, si  $N$  sigue una distribución binomial-negativa el gráfico debería presentar un comportamiento lineal.

### 3.4 Magnitud de los eventos extremos

La caracterización estadística de la magnitud de los eventos extremos se ha realizado tradicionalmente bajo el marco de la teoría clásica de valores extremos, a través de los resultados obtenidos por Fisher y Tippett (1928) quienes demostraron teóricamente las posibles distribuciones de probabilidad que pueden modelizar a tales eventos. Balkema

& de Haan (1974) y Pickands (1975) demuestran que las excedencias sobre un umbral suficientemente alto siguen una distribución generalizada de Pareto (GPD), sentando la base teórica para lo que se conoce actualmente como método de los *picos sobre un umbral* (POT), de sus siglas en inglés. Por esta razón, en esta sección se analizan por separado ambos modelos.

### 3.4.1 Distribuciones probabilísticas límite

#### Distribución generalizada de valores extremos

El modelo que representa la piedra angular de la teoría de valores extremos centra su atención en el comportamiento estadístico de

$$M_n = \max \{X_1, \dots, X_n\}$$

donde  $X_1, \dots, X_n$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas con función de distribución  $F$ . En la práctica,  $X_i$  normalmente representa valores de un proceso medido en una escala regular de tiempo, como niveles del mar horarios, o temperaturas medias diarias, por ejemplo, de forma que  $M_n$  representa el máximo del proceso sobre  $n$  unidades de tiempo.

Desde el punto de vista teórico, es posible obtener la distribución de probabilidad de  $M_n$  según la forma

$$\begin{aligned} \Pr \{M_n \leq z\} &= \Pr \{X_1 \leq z, \dots, X_n \leq z\} \\ &= \Pr \{X_1 \leq z\} \times \dots \times \Pr \{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned} \tag{3.12}$$

Sin embargo, en la práctica, es común el hecho de que la función  $F$  sea desconocida. Como solución a este desconocimiento de la forma de  $F$ , podría estimarse esta función  $F$  a partir de los datos muestrales existentes para ser utilizada posteriormente en (3.12). Desafortunadamente, esta forma de proceder no resulta adecuada, ya que, pequeñas variaciones en la estimación de  $F$ , producen desviaciones sustanciales en la estimación de  $F^n$ .

La alternativa a este procedimiento, aceptando que  $F$  es desconocida, intenta encontrar una familia de modelos teóricos para la variable  $F^n$ , la cual es estimada a partir de los valores extremos muestrales. Esto es similar a la aproximación de la media muestral por una distribución Normal, justificándose en el teorema central del límite. Por lo tanto, el objetivo sería entonces encontrar una distribución asintótica para la variable aleatoria  $M_n$ .

En primer lugar se estudia el comportamiento de  $F^n$  cuando  $n \rightarrow \infty$ . El primer problema que se presenta es que la función de distribución para el máximo no tiene en sí misma una distribución límite, ya que, a medida que aumenta  $n$ , la distribución se desplaza hacia la derecha (Buishand, 1989). Por lo tanto, no se produce la convergencia en distribución. Dicho de otra forma, para cualquier  $z < z_+$ , donde  $z_+$  es el valor más pequeño de  $z$  tal que  $F(z) = 1$ ,  $F^n(z) \rightarrow 0$  cuando  $n \rightarrow \infty$ , la distribución de  $M_n$  degenera en un punto de masa sobre  $z_+$ . Sin embargo, esta dificultad se resuelve definiendo una nueva variable en la forma

$$M_n^* = \frac{M_n - b_n}{a_n} \quad (3.13)$$

de forma que  $M_n^*$  posee una distribución límite para valores adecuados de  $\{a_n > 0\}$  y  $\{b_n\}$ . Valores apropiados de  $\{a_n > 0\}$  y  $\{b_n\}$  estabilizan la localización y escala de  $M_n^*$  evitando los problemas que surgen al tratar con  $M_n$ . De esta forma, se buscan distribuciones límite para  $M_n^*$ , con los adecuados valores de  $\{a_n > 0\}$  y  $\{b_n\}$ , en lugar de hacerlo para la variable  $M_n$ .

Las posibles distribuciones límite para  $M_n^*$  fueron determinadas por Fisher & Tippet (1928) a través del siguiente teorema:

**Teorema 3.1** *Si existe una secuencia de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que*

$$Pr\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z) \quad n \rightarrow \infty$$

*entonces  $G$  pertenece a una de las siguientes familias de distribuciones:*

*Tipo I o distribución de Gumbel*

$$G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty \quad (3.14)$$

*Tipo II o distribución de Fréchet*

$$G(z) = \begin{cases} 0, & z \leq b \\ \exp \left\{ - \left( \frac{z-b}{a} \right)^{-\beta} \right\}, & z > b \end{cases} \quad (3.15)$$

*Tipo III o distribución de Weibull*

$$G(z) = \begin{cases} \exp \left\{ - \left( \frac{b-z}{a} \right)^\beta \right\}, & z < b \\ 0, & z \geq b \end{cases} \quad (3.16)$$

siendo  $b$ ,  $a$  y  $\beta$  los parámetros de localización, escala y forma respectivamente, además de cumplir  $a > 0, b$  y en los casos II y III,  $\beta > 0$ .

El teorema 3.1 afirma que el máximo normalizado  $M_n^*$  tendrá por función de distribución una función perteneciente a una de las familias de distribuciones *Tipo I*, *Tipo II* o *Tipo III*. Las distribuciones de probabilidad definidas por el teorema 3.1 son conocidas como **distribuciones de valores extremos**, y suelen ser referenciadas también como *FT-I*, *FT-II* y *FT-III* respectivamente. En este punto debe comentarse que la distribución de Weibull propiamente dicha (Weibull, 1951) tiene la forma

$$G(z) = 1 - \exp \left[ - \left( \frac{z-b}{a} \right)^\beta \right] \quad z > b \quad \beta, a > 0 \quad (3.17)$$

por lo que la relación entre la distribuciones FT-III y Weibull es inmediata en la forma que la distribución FT-III de  $(-X)$  tiene por función de distribución de probabilidad una distribución de Weibull. A pesar de que en la práctica se hace referencia a la distribución de Weibull usando ambas distribuciones, debe quedar clara la diferencia entre ambas, ya que la distribución FT-III posee un límite superior fijado por  $b$ , mientras que la distribución de Weibull no se encuentra acotada superiormente.

El teorema 3.1 implica que, cuando  $M_n$  es estabilizado con los valores adecuados de  $a_n$  y  $b_n$ , el correspondiente máximo normalizado  $M_n^*$  tiene una distribución límite dentro de las tres distribuciones extremales. Un resultado importante es que los tres tipos de distribuciones extremales son las únicas distribuciones posibles para  $M_n^*$  a pesar de la distribución  $F$  de la población.

Notar que las distribuciones FT-II y FT-III están relacionadas a través de un simple cambio de signo. Además, estas dos distribuciones pueden ser transformadas en una distribución FT-I a través de los cambios

$$V = \log(Z - b) \quad y \quad V = -\log(b - Z)$$

respectivamente.

Los tres tipos de distribuciones límite que surgen como consecuencia del teorema 3.1 pueden combinarse en una única expresión cuya forma viene determinada por

$$G(z) = \exp \left\{ -(1 - kz)^{\frac{1}{k}} \right\} \quad (3.18)$$

la cual es conocida como forma de *Von Mises-Jenkinson* o *distribución Generalizada de Valores Extremos (GEV)*. En el caso de que  $k \rightarrow 0$  la distribución *GEV* se reduce a una distribución *Tipo I*. Los casos en los que  $k < 0$  y  $k > 0$  corresponden a distribuciones *Tipo II* ( $k = -1/\beta$ ) y *Tipo III* ( $k = 1/\beta$ ), respectivamente. Para  $k < 0$  la variable  $Z$  posee un límite inferior determinado por la cantidad  $1/k$ , mientras que para  $k > 0$  la misma cantidad marca el límite superior. Añadiendo a (3.18) los correspondientes parámetros de localización ( $\mu$ ) y escala ( $\sigma$ ) se obtiene

$$G(z) = \begin{cases} \exp \left\{ - \left[ 1 - k \left( \frac{z - \mu}{\sigma} \right) \right]^{\frac{1}{k}} \right\}, & k \neq 0 \\ \exp \left\{ - \exp \left[ - \left( \frac{z - \mu}{\sigma} \right) \right] \right\}, & k = 0 \end{cases}$$

con las siguientes condiciones:

$$\{z : 1 - k(z - \mu)/\sigma > 0\}$$

$$-\infty < \mu, k < \infty$$

$$\sigma > 0$$

La unificación de las tres familias de distribuciones extremales en una única distribución simplifica la implementación estadística evitando tener que ajustar las tres familias de distribuciones definidas por el teorema 3.1. A través de la inferencia de  $k$  son los propios datos los que determinan el comportamiento más adecuado de la cola o parte alta de la distribución y no es necesario suponer a priori un comportamiento u otro. Por este motivo podemos reescribir el teorema 3.1 de la siguiente forma:

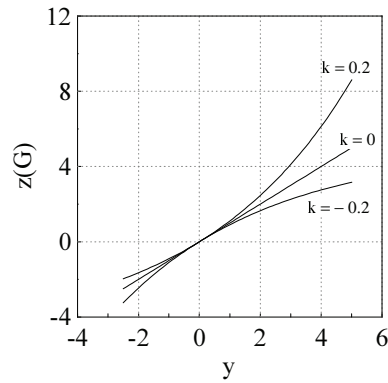


Figura 3.3: Función cuantil de la distribución GEV para diferentes valores del parámetro de forma  $k$  obtenida a través de (3.21)

**Teorema 3.2** Si existe una secuencia de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que

$$\Pr \{(M_n - b_n/a_n \leq z\} \rightarrow G(z) \quad n \rightarrow \infty \quad (3.19)$$

donde  $G$  es una función no degenerativa, entonces  $G$  pertenece a la familia definida por la distribución Generalizada de Valores Extremos (GEV):

$$G(z) = \begin{cases} \exp \left\{ - \left[ 1 - k \left( \frac{z-\mu}{\sigma} \right)^{\frac{1}{k}} \right] \right\}, & k \neq 0 \\ \exp \left\{ - \exp \left[ - \left( \frac{z-\mu}{\sigma} \right) \right] \right\}, & k = 0 \end{cases} \quad (3.20)$$

La función cuantil  $z(G)$  definida como la inversa de la función de distribución de probabilidad viene determinada por

$$z(G) = \begin{cases} \mu + \frac{\sigma}{k} (1 - \exp(-ky)) & k \neq 0 \\ \mu + \sigma y & k = 0 \end{cases} \quad (3.21)$$

donde  $y = -\ln \{-\ln G(z)\}$ . La forma que esta distribución toma, en función de los valores del parámetro  $k$ , se muestra en la figura 3.3.

Los parámetros de la distribución GEV pueden ser estimados a través de cualquiera de los métodos descritos en la sección correspondiente.

## Distribución Generalizada de Pareto

Este modelo centra su atención en el comportamiento de las excedencias de la variable sobre un determinado valor umbral,  $u$ . Siendo  $X_1, X_2, \dots$  una secuencia de variables aleatorias independientes e idénticamente distribuidas con función de distribución  $F$ , se suponen eventos extremos aquellos valores de  $X_i$  que superen un determinado valor umbral  $u$ . Denotando un término arbitrario de la secuencia  $X_i$  como  $X$ , la descripción del comportamiento estocástico de los eventos extremos vendrá dado por la probabilidad condicional

$$Pr \{X > u + y / X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0 \quad (3.22)$$

donde  $y$  es conocida como el valor de la excedencia, es decir, la diferencia entre el valor de la variable de estudio y el nivel umbral  $u$ .

Si la función de distribución  $F$  fuese conocida, la distribución de las excedencias sobre el umbral también serían conocidas según la ecuación (3.22). En la práctica esto no ocurre, de forma que se deben buscar distribuciones asintóticas para valores suficientemente altos de  $u$ . Este procedimiento es paralelo al uso de la distribución  $GEV$  como una aproximación a la distribución de los máximos para muestras suficientemente grandes.

Balkema & de Haan (1974) y Pickands (1975) demuestran que asintóticamente las excedencias sobre un nivel suficientemente alto siguen una distribución generalizada de Pareto (GPD) si y solo si la distribución de la variable está dentro del dominio de atracción de una distribución  $GEV$ . La expresión para la distribución GPD queda definida según

$$G(y) = \begin{cases} 1 - \left(1 - \frac{k(y-\xi)}{\alpha}\right)^{1/k} & k \neq 0 \\ 1 - \exp\left(\frac{-(y-\xi)}{\alpha}\right) & k = 0 \end{cases} \quad (3.23)$$

donde  $k$  es el parámetro de localización,  $\alpha > 0$  es el parámetro de escala y  $k$  ( $-\infty < k < \infty$ ) determina el parámetro de forma. Este resultado asintótico implica que (3.23) puede ser utilizada para caracterizar la función de distribución condicional de la variable  $Y = X - u$  sobre el umbral  $u$ , cuando  $X > u$  para valores suficientemente grandes de  $u$ . El rango de  $y$  viene determinado por  $0 \leq y < \infty$  para  $k \leq 0$  mientras que  $0 < y < \alpha/k$  para  $k > 0$ .

Aunando los resultados obtenidos hasta ahora es posible definir el siguiente teorema, conocido como teorema de Pickands-Balkema-de Haan:



**Teorema 3.3** *Siendo  $X_1, X_2, \dots$  una secuencia de variables aleatorias independientes e idénticamente distribuidas con función de distribución  $F$ , donde*

$$M_n = \max \{X_1, \dots, X_n\} \quad (3.24)$$

*Denotamos un valor arbitrario en la secuencia  $X_i$  por  $X$ , y suponemos que  $F$  satisface el teorema 3.1, de forma que para valores altos de  $n$ ,*

$$Pr \{M_n \leq z\} \approx G(z) \quad (3.25)$$

donde

$$G(z) = \exp \left\{ - \left[ 1 - k \left( \frac{z - \mu}{\sigma} \right) \right]^{\frac{1}{k}} \right\} \quad (3.26)$$

para diferentes valores de  $\mu, \sigma > 0$  y  $k$ . Entonces, para valores suficientemente altos de  $u$ , la función de distribución de la variable  $y = X - u$ , condicionada a  $X > u$ , es aproximadamente

$$G(y) = 1 - \left( 1 - \frac{k(y - \xi)}{\alpha} \right)^{1/k} \quad (3.27)$$

definida sobre  $0 \leq y < \infty$  para  $k \leq 0$  mientras que  $0 < y < \alpha/k$  para  $k > 0$ .

La ecuación (3.27) se justifica como una distribución límite a medida que  $u$  aumenta. La demostración de este teorema puede verse de forma esquemática en Coles (2001) y de una forma más rigurosa en Leadbetter *et al.* (1983).

El teorema 3.3 implica que si el máximo de la variable tiene por función de distribución de probabilidad  $G(z)$ , entonces las excedencias sobre el umbral  $u$ , tienen aproximadamente una función de distribución perteneciente a la familia Generalizada de Pareto. Además, los parámetros de la distribución GPD están determinados por aquellos parámetros asociados a la distribución GEV.

En particular, el parámetro  $k$  es igual en ambas distribuciones. Seleccionando diferentes, pero suficientemente altos, valores de  $n$  para los bloques, se verían afectados los valores

de los parámetros de la distribución GEV, pero no los correspondientes a la distribución GPD de las excedencias:  $k$  es invariante al tamaño del bloque.

Para los casos particulares en los que  $k = 0$  y  $k = 1$ , la distribución GPD se convierte en una distribución exponencial y uniforme, respectivamente.

## Distribuciones de probabilidad alternativas

Teóricamente puede demostrarse que los valores extremos de cualquier variable geofísica pueden ser caracterizados a través de una única función de distribución como es la generalizada de valores extremos o por alguna de los tres casos particulares que de ésta última se desprenden. Sin embargo, debido a que esta es una teoría asintótica que asume un número de datos tendiendo a infinito,  $n \rightarrow \infty$ , en la práctica es imposible cumplir esta hipótesis del modelo. Esto lleva a pensar en la existencia de otras funciones de distribución capaces de caracterizar los valores extremos de una variable aleatoria (Goda, 2000). Algunas funciones que han sido utilizadas son la distribución *Gamma Generalizada* (Ochi, 1992), la distribución *Kappa* de cuatro parámetros (Hosking, 1994) y la distribución de *Wakeby* (Park *et al.*, 2001).

### *Distribución Gamma Generalizada (GAM)*

La distribución *Gamma Generalizada* es una distribución triparamétrica con forma similar a la presentada por las distribuciones *Log-normal* y *Weibull*. Algunos trabajos, como el de Ochi (1992), evidencian que a pesar de que la distribución *Log-normal* caracteriza adecuadamente a la altura significativa del oleaje, los valores más altos o extremos de esta variable tienden a desviarse del comportamiento teórico de la distribución. La distribución *Gamma Generalizada* tiene como propiedad una mayor rapidez de convergencia a la unidad en comparación con la distribución *Log-normal*, por lo que, en aquellas ocasiones donde los datos empíricos muestren este comportamiento, esta distribución puede ser una alternativa interesante a tener en cuenta.

La función de distribución de probabilidad viene determinada por

$$G(z) = H(z/\beta, \alpha) \quad (3.28)$$

donde  $H(z, \alpha) = \{\Gamma(\alpha)\}^{-1} \int_0^z t^{\alpha-1} e^{-t} dt$  es la función gamma incompleta.

### *Distribución Kappa de 4 parámetros (K4P)*

Esta distribución fue introducida por Hosking (1994) como una generalización de la distribución *Kappa* de tres parámetros (Mielke, 1973), aplicándola en el campo de la meteorología para estudiar el comportamiento de las precipitaciones. Hasta el momento su uso ha estado restringido fundamentalmente al campo de la hidrología, no conociéndose referencias de su aplicación al análisis del oleaje. El hecho de disponer de cuatro parámetros permite un ajuste más flexible a los datos muestrales respecto de las distribuciones descritas anteriormente, conteniendo además como casos particulares a algunas de las distribuciones anteriormente descritas, tales como la GEV ( $h=0$ ) o la GPD ( $h=1$ ). Sin embargo, la adición de un parámetro se traduce en un aumento de la complejidad en el análisis y en la incertidumbre estadística de los resultados.

La función de densidad de probabilidad tiene la forma

$$f(x) = \alpha^{-1} [1 - k(x - \xi)/\alpha]^{(1/k)-1} [F(x)]^{1-h} \quad (3.29)$$

La función de distribución de probabilidad toma la forma

$$G(z) = \begin{cases} \left\{ 1 - h [1 - k(z - \xi)/\alpha]^{1/k} \right\}^{1/h} & \text{si } k \neq 0, h \neq 0 \\ \exp \left\{ - [1 - k(z - \xi)/\alpha]^{1/k} \right\} & \text{si } k \neq 0, h = 0 \\ \left\{ 1 - h \exp [-(z - \xi)/\alpha] \right\}^{1/h} & \text{si } k = 0, h \neq 0 \\ \exp \left\{ - \exp [-(z - \xi)/\alpha] \right\} & \text{si } k = 0, h = 0 \end{cases} \quad (3.30)$$

con función cuantil

$$z(G) = \xi + \frac{\alpha}{k} \left[ 1 - \left( \frac{1 - G^h}{h} \right)^k \right] \quad (3.31)$$

De los cuatro parámetros,  $\xi$  es el parámetro de localización,  $\alpha$  es el parámetro de escala y los parámetros  $k$  y  $h$  controlan la forma de la distribución. La única restricción en cuanto a los valores de estos parámetros es que  $\alpha > 0$ . Teniendo en cuenta que

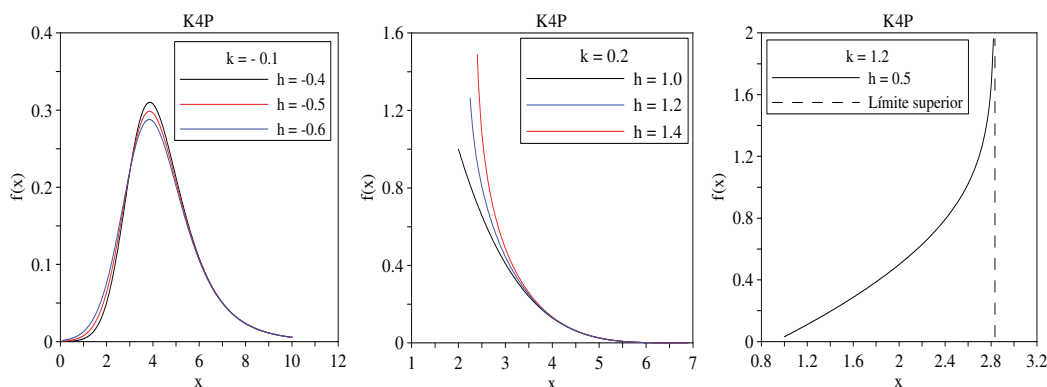


Figura 3.4: Diferentes formas de la distribución K4P en función de sus parámetros

$$\frac{1 - G^h}{h} \begin{cases} \rightarrow 0 & \text{cuando } G \rightarrow 1 \\ \rightarrow +\infty & \text{cuando } G \rightarrow 0 \text{ si } h < 0 \\ \rightarrow h^{-1} & \text{cuando } G \rightarrow 0 \text{ si } h > 0 \end{cases} \quad (3.32)$$

es posible calcular los límites de la distribución en la forma siguiente:

$$\begin{aligned} \xi + \alpha(1 - h^{-k})/k &\leq z \leq \xi + \alpha/k && \text{si } k > 0, h > 0 \\ \xi + \alpha \log h &\leq z < \infty && \text{si } k = 0, h > 0 \\ \xi + \alpha(1 - h^{-k})/k &\leq z < \infty && \text{si } k < 0, h > 0 \\ -\infty < z &\leq \xi + \alpha/k && \text{si } k > 0, h \leq 0 \\ -\infty < z &< \infty && \text{si } k = 0, h \leq 0 \\ \xi + \alpha/k &\leq z < \infty && \text{si } k < 0, h \leq 0 \end{aligned}$$

En este sentido, la función  $f(x)$  puede tomar diferentes formas dependiendo de la combinación de parámetros elegidas, tal y como se muestra en la figura 3.4.

El  $r$ -ésimo momento de una distribución de probabilidad se define como

$$E(X^r) = \int_0^1 [x(F)]^r dF \quad (3.33)$$

donde  $r$  es un número positivo entero. Para la distribución K4P esta integral existe bajo las siguientes condiciones

- 1) para todo  $r$ , si  $h \geq 0$  y  $k \geq 0$
- 2) para  $r < -1/kh$ , si  $h < 0$  y  $k \geq 0$
- 3) para  $r < -1/k$ , si  $k < 0$

Una particularidad de esta distribución es que, incluso conociendo sus primeros cuatro momentos, no es posible determinar de forma única los parámetros de la distribución. En general, esta distribución puede tener el mismo conjunto de primeros cuatro momentos para dos conjuntos de parámetros diferentes. Esto es debido a que algunas combinaciones de asimetría-curtosis  $(\tau_3, \tau_4)$  corresponden a distintos pares  $(h, k)$  (Hosking, 1994). Por esta razón las siguientes condiciones son impuestas sobre los parámetros:

- (a)  $k > -1$
- (b) si  $h < 0$ , entonces  $hk > -1$
- (c)  $h > -1$
- (d)  $k + 0.725h > -1$

Las condiciones (a) y (b) aseguran la existencia de los  $L$ -moments; (c) y (d) aseguran la unicidad de los parámetros, dados los primeros cuatro  $L$ -moments. Para más detalle puede verse Hosking (1994).

La estimación de los parámetros de la distribución K4P puede realizarse a través del método de los  $L$ -moments, Hosking (1994).

### ***Distribución de Wakeby (WAK)***

La distribución de Wakeby (WAK), introducida por Houghton (1978), posee 5 parámetros libres, por lo que puede considerarse como la distribución más flexible de todas las descritas en este trabajo. La función cuantil viene determinada por

$$z(G) = \xi + \frac{\alpha}{\beta} \left\{ 1 - (1 - G)^\beta \right\} - \frac{\gamma}{\delta} \left\{ 1 - (1 - G)^{-\delta} \right\} \quad (3.34)$$

no siendo posible obtener de forma explícita la función de distribución de probabilidad  $G(z)$ . Eligiendo adecuadamente el valor de los parámetros de esta distribución es posible obtener la distribución  $GEV$  y la distribución *Log-normal*. La estimación de estos parámetros puede realizarse a través del método de los  $L$ -moments, basado en el método de Landwehr *et al.*

(1979b). Básicamente ha sido utilizada en el contexto de la hidrología (Houghton, 1978; Park et al., 2001) no teniendo referencias hasta el momento de su aplicación a la altura significativa del oleaje.

### 3.4.2 Métodos de estimación de parámetros

Existen diferentes métodos para estimar los parámetros de una función  $f(x)$ . En el presente trabajo, y dependiendo de la función de distribución de probabilidad considerada, se han utilizado los siguientes métodos:

- mínimos cuadrados
- método de los momentos
- método de máxima verosimilitud
- método de los  $L$ -moments, propuesto por Hosking (1990)

#### Método de los mínimos cuadrados

El método de *mínimos cuadrados*,  $LS$ , encuentra las estimaciones para los parámetros en la ecuación eleccionada mediante la minimización de la suma de los cuadrados de las diferencias entre los valores observados de la variable respuesta y de aquellos proporcionados por la ecuación de predicción. Los estimadores por mínimos cuadrados poseen ciertas propiedades deseables bajo las siguientes suposiciones:

- se ha seleccionado la forma correcta de la ecuación de regresión. Esto implica que cualquier variabilidad en la variable respuesta que no pueda explicarse mediante el empleo de la ecuación de regresión se debe a un error aleatorio.
- los valores observados de la variable respuesta no se encuentran estadísticamente correlacionados. Se supone que cada valor observado está constituido por un valor real y una componente aleatoria no observable, de forma que la covarianza entre cualesquiera dos observaciones  $Y_i$  y  $Y_j$ , o entre los correspondientes errores aleatorios  $\epsilon_i$  y  $\epsilon_j$ , es cero para toda  $i \neq j$ .
- para toda  $i = 1, 2, \dots, n$  la media del error aleatorio  $\epsilon$  es cero y la varianza  $\sigma^2$ .

Tabla 3.1: Linealización de las distribuciones de probabilidad de Gumbel, Frechet y Weibull

	<b>Y</b>	<b>A</b>	<b>X</b>	<b>B</b>
Gumbel	$-\text{Ln}(-\text{Ln } G)$	$\frac{1}{a}$	$z-b$	-
Frechet	$\text{Ln}((\text{Ln } G)^{-1})$	$\beta$	$\text{Ln}(z-b)$	$-\beta \text{Ln}(a)$
Weibull	$\text{Ln}(-\text{Ln}(1-G))$	$\beta$	$\text{Ln}(z-b)$	$-\beta \text{Ln}(a)$

De forma concreta, para las distribuciones de probabilidad de Gumbel, Frechet y Weibull descritas en 3.1, su linealización en la forma  $Y = AX + B$ , donde  $A$  es la pendiente de la recta y  $B$  es la ordenada en el origen, permite obtener estimadores de sus parámetros, según se muestra en la tabla 3.1.

Para más detalles a cerca de la estima de los parámetros de estas tres distribuciones a través del método de mínimos cuadrados se recomienda consultar *Martínez, J., Martín, M.J. (1990)*.

### Método de los momentos

El método de los *momentos*, *MOM*, consiste en igualar los momentos apropiados de la distribución de la población con los correspondientes momentos muestrales para estimar un parámetro desconocido de la distribución. Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad de probabilidad  $f(x; \theta)$ , el *r-ésimo* momento alrededor del cero se define como

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r \quad (3.35)$$

Las expresiones matemáticas involucradas en la estima de los parámetros de las distribuciones de probabilidad de Gumbel, Frechet y Weibull a través del método de los momentos pueden verse ampliamente detalladas en *Martínez, J., Martín, M.J. (1990)*.

### Método de máxima verosimilitud

Para introducir el concepto de estimación de *máxima verosimilitud*, *MLE*, resulta muy ilustrativo el ejemplo propuesto en Canavos (1988) y que a continuación se presenta. El

desborde de ríos y lagos es un fenómeno natural que a veces tiene devastadoras consecuencias. Supóngase que en cierto año hubo dos serias inundaciones, por este fenómeno, en determinada región geográfica. Si se supone que el número de inundaciones por año en esta localidad es una variable aleatoria de Poisson con un valor del parámetro  $\lambda$ , desconocido, ¿cómo debe procederse para estimar el valor de  $\lambda$  con base a una sola observación  $x = 2$ ? Un posible método es seleccionar el valor de  $\lambda$  para el cual la probabilidad del valor observado es máxima. Es posible, para el valor observado, que  $\lambda$  sea cualquier número positivo. Para este ejemplo, supóngase que los posibles valores de  $\lambda$  son 1,  $3/2$ , 2,  $5/2$  y 3. Las probabilidades para el valor observado  $x = 2$  para cada uno de estos valores de  $\lambda$  son las siguientes:

$\lambda$	1	$3/2$	2	$5/2$	3
$p(2; \lambda)$	0.1389	0.2510	0.2707	0.2565	0.2240

Aparentemente  $p(2; \lambda)$  crece hasta un valor máximo de 0.2707 para  $\lambda = 2$ , y disminuye para  $\lambda > 2$ . El valor de 2 de  $\lambda$  es el que maximiza la probabilidad del valor observado. En otras palabras, la observación  $x = 2$  tiene una probabilidad mayor de ocurrencia para una distribución de *Poisson* con  $\lambda = 2$  que para cualquier otro valor del parámetro  $\lambda$ .

*Definición:* Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución con función (densidad) de probabilidad  $f(x; \theta)$ , y sea  $L(x_1, x_2, \dots, x_n; \theta)$  la verosimilitud de la muestra como función de  $\theta$ . Si  $t = u(x_1, x_2, \dots, x_n)$  es el valor de  $\theta$  para el cual el valor de la función de verosimilitud es máxima, entonces  $T = u(X_1, X_2, \dots, X_n)$  es el estimador de máxima verosimilitud de  $\theta$ , y  $T$  es el *estimador de máxima verosimilitud*.

Las expresiones matemáticas involucradas en la estima de los parámetros de las distribuciones de probabilidad de Gumbel, Frechet y Weibull a través del método de máxima verosimilitud pueden verse de forma detallada en *Martínez, J., Martín, M.J. (1990)*.

## Método de los *L*-moments

Los *L*-moments, *LMOM*, surgen como alternativa al método tradicional de los momentos, siguiendo la misma filosofía de igualación de los momentos muestrales y poblacionales pero con la diferencia en la forma de calcular dichos momentos.

La letra *L* hace referencia al hecho de que los momentos se obtienen a través de combinaciones lineales de los estadísticos ordenados. Dado que no se ha establecido hasta el



momento un nombre en castellano para este método, o al menos no se tiene conocimiento de ello, y para evitar confusiones terminológicas, se ha optado por mantener su nombre original a lo largo de todo este trabajo.

Si bien las bases teóricas de este método son relativamente antiguas, véase Gini (1912), Silitto (1951, 1969), Downton (1966), Chan (1967), Konheim (1971), Mallows (1973) y Greenwood et al. (1979), ha sido J.R.M. Hosking quien con su trabajo *L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics* publicado en 1990, pone de manifiesto las ventajas que este método ofrece respecto al método tradicional de los momentos. A partir de la publicación de dicho trabajo, numerosos artículos de investigación en el campo del análisis de valores extremos han sido publicados, véase Pearson (1993), Parida (1999), Park *et al.* (2001) y Pandey *et al.*, (2001) entre otros.

Uno de los aspectos que hacen interesante este método en el contexto del análisis extremal es su mayor robustez a la hora de trabajar con muestras pequeñas.

### **Definición**

$X$  representa una variable real aleatoria con función de distribución de probabilidad  $F(x)$  y función cuantil  $x(F)$ , mientras que  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  representan los estadísticos ordenados de una muestra aleatoria de tamaño  $n$  provenientes de la distribución de  $X$ .

Los  $L$ -moments de  $X$  se definen como

$$\lambda_r \equiv r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, \dots \quad (3.36)$$

La letra  $L$ , como se explicó anteriormente, enfatiza el hecho de que  $\lambda_r$  es una función lineal de los valores esperados de los estadísticos ordenados.

La esperanza matemática o valor esperado de un estadístico ordenado puede escribirse como (David, 1981)

$$EX_{j:r} = \frac{r!}{(j-1)!(r-j)!} \int x \{F(x)\}^{j-1} \{1-F(x)\}^{r-j} dF(x) \quad (3.37)$$

Sustituyendo esta expresión en 3.36, desarrollando los binomios en  $F(x)$  y sumando los coeficientes de cada potencia de  $F(x)$  se obtiene

$$\lambda_r = \int_0^1 x(F)P_{r-1}^*(F)dF, \quad r = 1, 2, \dots \quad (3.38)$$

donde

$$P_r^*(F) = \sum_{k=0}^r p_{r,k}^* F^k$$

y

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}$$

$P_r^*(F)$  es el  $r$ -ésimo polinomio de Legendre modificado, relacionado con el polinomio de Legendre común  $P_r(u)$  por la relación  $P_r^*(u) = P_r(2u - 1)$ .

Los primeros cuatro  $L$ -moments tienen la forma siguiente

$$\begin{aligned} \lambda_1 = EX &= \int_0^1 x(F)dF \\ \lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}) &= \int_0^1 x(F)(2F - 1)dF \\ \lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) &= \int_0^1 x(F)(6F^2 - 6F + 1)dF \\ \lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) &= \int_0^1 x(F)(20F^3 - 30F^2 + 12F - 1)dF \end{aligned} \quad (3.39)$$

El uso de los  $L$ -moments en la descripción de las distribuciones de probabilidad está justificado por el teorema descrito en Hosking (1990), según el cual:

- (a) Los  $L$ -moments  $\lambda_r, r = 1, 2, \dots$ , de una variable real aleatoria  $X$  existen si y sólo si  $X$  tiene valor medio finito.
- (b) Una distribución de probabilidad cuya media existe está caracterizada por sus  $L$ -moments  $\{\lambda_r : r = 1, 2, \dots\}$

Una de las implicaciones de este teorema es que el conjunto de  $L$ -moments que describen una distribución es único, hecho que no ocurre en el caso de los momentos convencionales.

$\lambda_2$  puede interpretarse como una medida de la escala o dispersión de la variable aleatoria  $X$ . Es conveniente en ocasiones estandarizar los momentos de orden superior,  $r \geq 3$  de

forma que sean independientes de las unidades de medida de  $X$ . Se definen entonces los  $L$ -moments *ratios* o razones de  $X$  como las cantidades

$$\tau \equiv \lambda_r / \lambda_2, \quad r = 3, 4, \dots$$

cumpléndose además  $|\tau_r| < 1$  para  $r \geq 3$ . En el caso de que  $X \geq 0$  entonces  $0 < \tau < 1$ .

De forma particular, los límites para  $\lambda_1$ ,  $\lambda_2$ ,  $\tau_3$  y  $\tau_4$  son

$$0 < \lambda_2 \quad -1 < \tau_3 < 1 \quad \frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1$$

Los  $L$ -moments están directamente relacionados con los *Probability Weighted Moments* (PWMs) o momentos ponderados definidos por Greenwood *et al.* (1979) según las relaciones

$$M_{p,r,s} = E [X^p \{F(X)\}^r \{1 - F(X)\}^s] \quad (3.40)$$

Estos PWMs pueden expresarse como combinaciones lineales de los LMOM, de forma que los procedimientos basados en los PWMs y LMOM son equivalentes. Sin embargo, los LMOM poseen mayor facilidad de interpretación como medidas de escala y forma de las distribuciones de probabilidad. En particular  $\lambda_1$ ,  $\lambda_2$ ,  $\tau_3$  y  $\tau_4$  pueden interpretarse como medidas de localización, escala, asimetría y curtosis respectivamente.

Para entender esto considérese la relación 3.39, donde se define  $\lambda_r$  como los valores esperados de las combinaciones lineales de los estadísticos ordenados.  $\lambda_1$ , el valor medio, es una medida de localización. Para interpretar  $\lambda_2$  considérese la configuración de una muestra de tamaño 2: si los dos valores tienden a estar juntos, como en la figura 3.5(a), entonces  $\lambda_2$  será menor que que en el caso donde los dos valores están más separados, como en la figura 3.5(b).  $\lambda_3$  es la segunda diferencia central respecto de la mediana. Muestras como la de la figura 3.5(c) que poseen una segunda diferencia central positiva surgen de distribuciones con asimetría positiva; la figura 3.5(d) es típica de distribuciones con asimetría negativa. Las distribuciones

simétricas tienen por tanto un valor de  $\lambda_3 = 0$ . Así,  $\lambda_3$  puede interpretarse como una medida de la asimetría de la distribución, aunque no es totalmente independiente de la escala.  $\lambda_4$ , es la tercera diferencia central de los valores esperados de los estadísticos ordenados, y mide lo mismo que el cuarto momento central convencional. La configuración mostrada

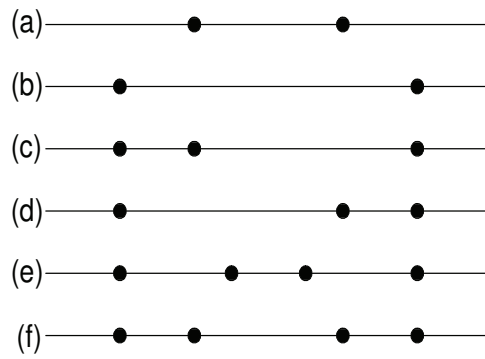


Figura 3.5: Configuraciones para muestras de tamaño 2,3 y 4

Tabla 3.2: Expresiones teóricas de  $\tau_3$  y  $\tau_4$  para diferentes distribuciones de probabilidad: Uniforme, Exponencial, Gumbel, generalizada de valores extremos (GEV) y generalizada de Pareto (GPD)

Distribución	$\tau_3$	$\tau_4$
Uniforme	0	0
Exponencial	1/3	1/6
Gumbel	0.1699	0.1504
GEV	$2(1-3^{-k})/(1-2^{-k})-3$	$(1-6 \cdot 2^{-k}+10 \cdot 3^{-k}-5 \cdot 4^{-k})/(1-2^{-k})$
GPD	$(1-k)/(3+k)$	$(1-k)(2-k)/(3+k)(4+k)$

por la figura 3.5(e) es típica de distribuciones de cola pesada o especialmente puntiagudas, cuya tercera diferencia central es positiva. Por el contrario, la figura 3.5(f) muestra una configuración típica de distribuciones planas o en forma de U, cuya tercera diferencia central es negativa. Por lo tanto,  $\lambda_4$  puede interpretarse como medida de la curtosis de la distribución. Las razones  $\tau_3$  y  $\tau_4$  son adimensionales y análogas a  $\lambda_3$  y  $\lambda_4$  respectivamente, y son así medidas de asimetría y curtosis.

Por lo tanto, las propiedades de una distribución de probabilidad deben quedar adecuadamente caracterizadas por los siguientes parámetros:  $\lambda_1$  o  $L$ -localización;  $\lambda_2$  o  $L$ -escala;  $\lambda_3$  o  $L$ -asimetría y  $\lambda_4$  o  $L$ -curtosis.

Especialmente interesante resulta la relación entre los parámetros  $\tau_3$  y  $\tau_4$ . Cada distribución de probabilidad teórica posee una relación característica entre estos dos parámetros, mostrada de forma analítica en la tabla 3.2 y de forma gráfica en la figura 3.6.

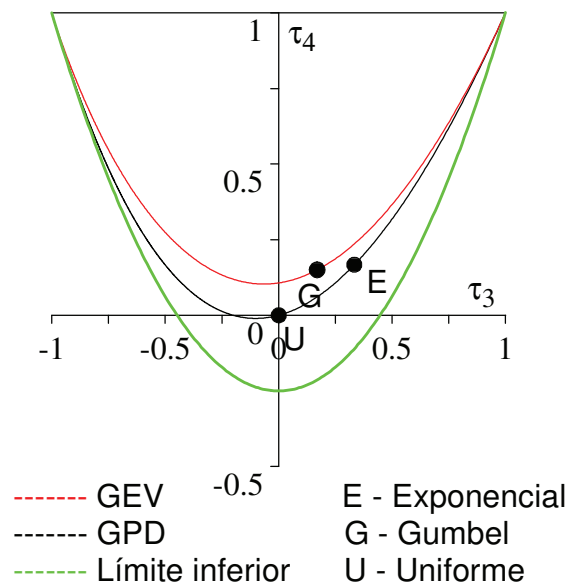


Figura 3.6: Relación entre  $\tau_3$  y  $\tau_4$  para diferentes distribuciones de probabilidad: Uniforme, Exponencial, Gumbel, generalizada de valores extremos (GEV) y generalizada de Pareto (GPD).

De esta forma, una representación de los valores de  $\tau_3$  y  $\tau_4$  obtenidos a partir de sus estimadores muestrales permitirían concluir, en función de la cercanía a las curvas teóricas, cual de las posibles distribuciones de probabilidad caracterizan adecuadamente la serie de valores extremos.

### *Estimación de los L-moments*

Las expresiones descritas hasta ahora para los  $L$ -moments están definidas para una distribución de probabilidad teórica. Sin embargo, en la práctica, lo normal es que haya que estimarlas a partir de un conjunto de datos de tamaño finito y medidos de forma experimental.

Se define una muestra de tamaño  $n$  según  $x_1, x_2, \dots, x_n$ , siendo  $x_{1:n} \leq x_{2:n} \leq \dots x_{n:n}$  la muestra ordenada. El  $r$ -ésimo  $L$ -moment muestral se define como

$$l_r = \sum_{k=0}^{r-1} p_{r-1,k}^* b_k \quad (3.41)$$

donde

$$b_r = n^{-1} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_{i:n} \quad (3.42)$$

Por lo tanto, un conjunto de datos queda caracterizado a través de sus  $L$ -moments, los cuales contienen información a cerca de la localización, escala, asimetría y curtosis. En este sentido  $l_r$  es el estimador natural de  $\lambda_r$ , por lo que a través de la igualación de los  $L$ -moments muestrales con sus homólogos poblacionales se podrán estimar los parámetros de cualquier distribución teórica. De igual forma, el estimador de  $\tau_r$  vendrá dado por  $t_r = l_r/l_2$ .

Algunas ventajas de los  $L$ -moments respecto a los momentos tradicionales son:

- para que los  $L$ -moments de una distribución de probabilidad existan, sólo es necesario que la distribución tenga media finita.
- para que los errores sean finitos sólo se requiere que la distribución tenga varianza finita.
- el ajuste de datos a distribuciones teóricas resulta más eficiente empleando los  $L$ -moments que los momentos ordinarios.
- son menos sensibles a la presencia de outlayers.

### *Distribución generalizada de valores extremos (GEV)*

Los  $L$ -moments de la distribución GEV descrita en el teorema 3.2 vienen dados por las expresiones

$$\lambda_1 = \mu + \frac{\sigma}{k} \{1 - \Gamma(1+k)\} \quad (3.43)$$

$$\lambda_2 = \frac{\sigma}{k} (1 - 2^{-k}) \Gamma(1+k) \quad (3.44)$$

$$\tau_3 = \frac{2(1 - 3^{-k})}{1 - 2^{-k}} - 3 \quad (3.45)$$

por lo que a través de las estimaciones correspondientes  $\hat{\lambda}_1$ ,  $\hat{\lambda}_2$  y  $\hat{\tau}_3$  es posible estimar los parámetros de la distribución.

### ***Distribución generalizada de Pareto (GPD)***

Los  $L$ -moments correspondientes a la distribución GPD vienen definidos a través de las relaciones

$$\lambda_1 = \xi + \frac{\alpha}{1+k} \quad (3.46)$$

$$\lambda_2 = \frac{\alpha}{(1+k)(2+k)} \quad (3.47)$$

$$\tau_3 = \frac{1-k}{3+k} \quad (3.48)$$

Así, usando los correspondientes  $L$ -moments muestrales, los parámetros de localización ( $\xi$ ), escala ( $\alpha$ ) y forma ( $k$ ) pueden ser estimados como

$$\begin{aligned} k &= \frac{1-3\tau_3}{1+\tau_3} \\ \alpha &= (1+k)(2+k)\lambda_2 \\ \xi &= \lambda_1 - (2+k)\lambda_2 \end{aligned} \quad (3.49)$$

### ***Distribución Kappa de 4 parámetros (K4P)***

La estimación de los parámetros de la distribución K4P se realiza en este trabajo a través del método de los  $L$ -moments descrito en Hosking (1994).

Recordando la definición de los momentos ponderados (PWMs) de una variable aleatoria  $X$  con función de distribución  $G$

$$M_{p,r,s} = E \{X^p [G(X)]^r [1-G(X)]^s\} \quad (3.50)$$

es posible determinar los  $L$ -moments como combinaciones lineales de éstos, según

$$\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \quad (3.51)$$

donde

$$\beta_r = M_{1,r,0} = E \{ X [G(X)]^r \} \quad (3.52)$$

Los PWMs de una distribución existen si la media de la distribución existe, lo que equivale en la distribución K4P a que  $h \geq 0$  y  $k > -1$ , ó  $h < 0$  y  $-1 < k < -1/h$ .

La cantidad  $\beta_r$  definida en 3.52 puede ser reescrita en la forma

$$\begin{aligned} \beta_r &= \int_0^1 x(G) G^r dG \\ &= (r+1)^{-1} \left( \xi + \frac{\alpha}{k} \right) - \frac{\alpha}{k} \int_0^1 \left[ h^{-1} (1 - G^h) \right]^k G^r dG \end{aligned} \quad (3.53)$$

En el caso de que  $k \neq 0$ ,  $\beta_r$  está dada por

$$r\beta_{r-1} = \begin{cases} \xi + \frac{\alpha}{k} \left[ 1 - \frac{r\Gamma(1+k)\Gamma(r/h)}{h^{1+k}\Gamma(1+k+r/h)} \right] & \text{si } k > -1, h > 0 \\ \xi + \frac{\alpha}{k} \left[ 1 - r^{-k}\Gamma(1+k) \right] & \text{si } k > -1, h = 0 \\ \xi + \frac{\alpha}{k} \left[ 1 - \frac{r\Gamma(1+k)\Gamma(-k-r/h)}{-h^{1+k}\Gamma(1-r/h)} \right] & \text{si } -1 < k < -1/h, h < 0 \end{cases} \quad (3.54)$$

De forma similar si  $k = 0$  se obtiene

$$r\beta_{r-1} = \begin{cases} \xi + \alpha [\gamma + \log h + \psi(1+r/h)] & \text{si } h > 0 \\ \xi + \alpha(\gamma + \log r) & \text{si } h = 0 \\ \xi + \alpha [\gamma + \log(-h) + \psi(-r/h)] & \text{si } h < 0 \end{cases} \quad (3.55)$$

donde  $\gamma = 0.5772\dots$  es la constante de Euler, y  $\psi$  es la función digamma,  $\psi(x) = d[\log \Gamma(x)]/dx$ . Los  $L$ -moments de la distribución K4P pueden ser entonces calculados a partir de las relaciones 3.51 y 3.54.

La estimación de los parámetros requieren la solución de (3.54) y (3.55) dados  $\beta_0, \beta_1, \beta_2$  y  $\beta_3$ . Sin embargo, no existe solución explícita, por lo que las ecuaciones deben ser resueltas de forma numérica. Para tal efecto se utiliza el programa *PELKAP* programado en FORTRAN 90 (Hosking, 2000) que utiliza el método de Newton-Raphson para la resolución de las ecuaciones correspondientes.



### *Distribución de Wakeby (WAK)*

Los  $L$ -moments de la distribución WAK vienen dados por las siguientes expresiones (Hosking, 1996)

$$\lambda_1 = \xi + \frac{\alpha}{(1+\beta)} + \frac{\gamma}{(1-\delta)} \quad (3.56)$$

$$\lambda_2 = \frac{\alpha}{(1+\beta)(2+\beta)} + \frac{\gamma}{(1-\delta)(2-\delta)} \quad (3.57)$$

$$\lambda_3 = \frac{\alpha(1-\beta)}{(1+\beta)(2+\beta)(3+\beta)} + \frac{\gamma(1+\delta)}{(1-\delta)(2-\delta)(3-\delta)} \quad (3.58)$$

$$\lambda_4 = \frac{\alpha(1-\beta)(2-\beta)}{(1+\beta)(2+\beta)(3+\beta)(4+\beta)} + \frac{\gamma(1+\delta)(2+\delta)}{(1-\delta)(2-\delta)(3-\delta)(4-\delta)} \quad (3.59)$$

$$\lambda_r = \frac{\alpha\Gamma(1+\beta)\Gamma(r-1-\beta)}{\Gamma(1-\beta)\Gamma(r+1+\beta)} + \frac{\gamma\Gamma(1-\delta)\Gamma(r-1+\delta)}{\Gamma(1+\delta)\Gamma(r+1-\delta)}, \quad r \geq 5 \quad (3.60)$$

Igualando los primeros  $p$   $L$ -moments de la población y de la muestra se obtiene el correspondiente sistema de ecuaciones de cuya solución se obtienen los parámetros estimados. Dado que no existe solución explícita para tal sistema de ecuaciones, se utiliza el método iterativo de Newton-Raphson para resolver el sistema. Landwehr *et al.* (1979) desarrollan un algoritmo para tal fin, el cual es implementado en el programa *PELWAK* (Hosking, 2000) utilizado en el presente trabajo. Este programa busca en primer lugar una solución de las ecuaciones para la cual todos los parámetros sean estimados. En caso de fallo en la convergencia del algoritmo, al parámetro de forma,  $\xi$  se le asigna el valor cero, buscando una nueva solución para la cual los restantes cuatro parámetros son estimados. Si esto fuese insuficiente, entonces se asume una distribución generalizada de Pareto (GPD) fijando los parámetros  $\alpha = 0$  ó  $\gamma = 0$ .

## 3.5 Modelos combinados Frecuencia-Magnitud

Según se ha comentado en secciones anteriores, la frecuencia de presentación de los eventos extremos puede ser modelizada a través de una distribución de Poisson, mientras que su magnitud puede modelizarse a través de alguna de las distribuciones de probabilidad anteriormente comentadas.

Basándose en las excedencias sobre un umbral, a continuación se proponen los modelos que combinan la caracterización de la frecuencia y magnitud de los eventos extremos, centrandó la atención en los modelos Poisson-GPD y Poisson-K4P.

### 3.5.1 Técnicas para la selección del umbral

Como ya se ha comentado, el cálculo del nivel umbral sobre el cual modelizar las excedencias no resulta una tarea sencilla. Actualmente existen diferentes metodologías propuestas para tal efecto, englobándose en dos grupos bien diferenciados.

Por un lado están las metodologías que buscan el cumplimiento de las hipótesis básicas del modelo, es decir, que el número de excedencias siga una distribución de Poisson y que la magnitud de las excedencias siga una distribución GPD. Para ello, se analizan por separado ambas hipótesis sobre un rango de umbrales suficientemente amplio, y finalmente se elige el umbral donde mejor ajuste se obtenga.

Por otro lado están los métodos que se basan en el número de excedencias sobre el umbral, asumiendo que en el contexto del análisis extremal, este número de excedencias debe ser una proporción determinada de la muestra total. Obviamente, el umbral seleccionado finalmente debería cumplir con las hipótesis básicas del modelo. Estos criterios suelen tener su justificación en la simulación numérica o en la propia experiencia de los autores.

Una comparación entre los resultados obtenidos a través de ambas metodologías pone de manifiesto la dificultad en la elección del valor umbral, observándose diferencias notables en los valores umbrales finales. Incluso las diferencias dentro de una misma metodología llegan a ser importantes.

#### Métodos basados en propiedades teóricas del modelo

Un primer método para seleccionar adecuadamente el nivel umbral  $u$  hace uso del valor medio de la distribución GPD. Si  $Y = X - u$  tiene una distribución generalizada de Pareto con parámetros  $\alpha$  y  $k$ , entonces

$$E(X - u | X > u) = \frac{\alpha - ku}{1 + k} \quad (3.61)$$

donde  $u > 0$  y  $\alpha - uk > 0$ . Así,  $E(X - u | X > u)$  será una función lineal de  $u$  con pendiente  $-k/1 + k$  y ordenada en el origen  $\alpha/1 + k$ . Además,  $E(X - u | X > u)$  representa el valor medio de las excedencias del nivel  $u$ , por lo que la media muestral de las excedencias de  $u$ , proporciona una estimación empírica. Esta función recibe el nombre de *función media de exceso* (MEF). De acuerdo con (3.61), estas estimaciones variarán linealmente con  $u$  para

aquellos niveles en los cuales el modelo GPD sea apropiado. Esto nos lleva al siguiente procedimiento:

$$\left\{ \left( u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_i - u) \right) : u < x_{max} \right\} \quad (3.62)$$

donde  $x_1, \dots, x_{n_u}$  es la serie de valores sobre el nivel  $u$  y  $x_{max}$  es el valor máximo de la serie. Este tipo de representación recibe el nombre de *mean residual life plot*, o gráfico de vida media residual. Por encima de un nivel  $u_0$  donde la distribución generalizada de Pareto caracterice adecuadamente las excedencias, el gráfico debería ser aproximadamente lineal con  $u$ .

Sin embargo, la interpretación de este tipo de gráfico no siempre es sencilla en la práctica, Davison y Smith (1992), Coles (2001), mostrando una gran variabilidad.

Otro procedimiento para seleccionar el valor umbral se basa en la estabilidad que el parámetro de forma,  $k$ , muestra respecto a  $u$ . En otras palabras, sobre un nivel  $u_0$  en el cual la distribución generalizada de Pareto es válida, las estimaciones del parámetro de forma  $k$  deberían ser aproximadamente constantes para valores de  $u > u_0$ . En ocasiones, la combinación de este método con el método descrito en el apartado anterior puede resultar de utilidad para decidir el valor final del umbral.

Ashkar y Rouselle (1987) proponen escoger el mínimo valor del umbral para el cual las excedencias de la variable sigan una distribución de Poisson. Este método tiene el inconveniente, en ocasiones, de seleccionar umbrales demasiado bajos.

Para el caso concreto de la distribución GPD, la idoneidad de la distribución sobre diferentes umbrales puede ser estudiada a través de los momentos lineales (L-moments) de tercer y cuarto orden. De esta forma, se construye un gráfico representando  $(\hat{\tau}_3, \hat{\tau}_4)$ , obtenidas a partir de los datos muestrales, para diferentes valores de  $u$ . En el mismo gráfico se representa la curva teórica correspondiente a la distribución GPD usando la aproximación (Hosking, 1990):

$$\tau_4 = \frac{\tau_3(1 + 5\tau_3)}{5 + \tau_3} \quad (3.63)$$

Las desviaciones de esta curva teórica pueden ayudar a decidir el rango de aplicabilidad del modelo.

## Métodos basados en el número de excedencias

En este apartado se consideran tres criterios propuestos por tres autores diferentes. De esta forma, Boos (1984), recomienda que la relación entre el número de excedencias,  $m$ , y el número total de datos que componen la muestra,  $n$ , debería ser

$$\frac{m}{n} = \begin{cases} 0.02 & \text{si } 50 < n < 500 \\ 0.1 & \text{si } 500 < n < 1000 \end{cases} \quad (3.64)$$

Hasofer (1996) sugiere usar  $m \cong 1.5\sqrt{n}$ , mientras que Gross *et al.* (1994) basado en simulaciones numéricas de velocidades del viento sugieren que el número de excedencias sobre el umbral debería ser del orden de 10 por año, en promedio.

### 3.5.2 Técnicas para estudiar la independencia estadística

Independientemente del método utilizado para la selección del umbral, es fundamental que la población extremal cumpla la condición de independencia estadística, ya que esta es una de las hipótesis básicas en las que se sustenta el modelo. Existe un gran número de test estadísticos propuestos en la literatura para analizar la independencia de una serie de datos. En este caso se hará uso del conocido como test de rachas o *Run test (RT)*. En este sentido una racha se define como una serie de valores sucesivos crecientes o decrecientes. El primer paso es calcular las diferencias  $(X_i - X_{i-1})$ . Un valor positivo de esta diferencia significa que la serie es creciente, mientras que un valor negativo se traduce en una serie decreciente. En otras palabras, si  $X_i > X_{i-1}$  se asigna un 1 para la observación, mientras que si  $X_i \leq X_{i-1}$  se asigna un 0.

Es posible calcular entonces la probabilidad de que una secuencia dada de rachas esté originada por un proceso aleatorio de dos estados (0 y 1), o de forma equivalente, que los datos sean independientes. Siendo  $n_0$  el número de observaciones por debajo del valor mediano y  $n_1$  el número de observaciones por encima, el número total de rachas en una secuencia es denotado por  $U$ .

Cuando  $N$  es relativamente grande ( $> 20$ ), la distribución de  $U$  puede ser aproximada por una distribución Normal (Bendat & Piersol, 1986). El valor esperado del número de rachas en una secuencia aleatoria es

$$\bar{U} = \frac{2n_0n_1}{n_0 + n_1} + 1 \quad (3.65)$$

La varianza en el número de rachas es

$$\sigma_{\bar{U}}^2 = \frac{2n_0n_1(2n_0n_1 - n_0 - n_1)}{(n_0 + n_1)^2(n_0 + n_1 - 1)} \quad (3.66)$$

En este sentido es posible crear un test ( $Z$ ) donde  $U$  es el número observado de rachas

$$Z = \frac{U - \bar{U}}{\sigma_{\bar{U}}} \quad (3.67)$$

La hipótesis nula es rechazada si el valor de  $Z$  es mayor que el valor crítico obtenido a partir de la distribución normal estándar. En otras palabras, la serie no puede ser considerada como aleatoria y por tanto la hipótesis de independencia es rechazada.

### 3.5.3 Modelo Poisson - GPD

De forma general, este modelo asume dos hipótesis básicas de partida:

1) el número de excedencias sobre el umbral sigue una distribución de Poisson (Cunnane, 1979).

2) la magnitud de las excedencias sobre un umbral suficientemente alto sigue una distribución generalizada de Pareto (GPD) (Pickands-Balkema-de Haan, 1974, 1975).

Como ya se ha comentado, uno de los objetivos del análisis extremal es obtener cuantiles asociados a periodos de retorno largos. Se define la variable de estudio  $Y = X - u$ , donde  $Y$  representa la excedencia de la variable  $X$  sobre el umbral. Teniendo en cuenta la definición de periodo de retorno dada a través de la expresión 3.80 y sustituyendo  $F$  por la distribución  $G$  definida por la expresión 3.23, el cuantil  $X_T$  excediendo un umbral  $u$  asociado a un periodo de retorno  $T$  viene dado por

$$X_T = \begin{cases} u + \xi + \frac{\alpha}{k}[1 - (\lambda T)^{-k}] & k \neq 0 \\ u + \xi - \alpha \ln(\lambda T) & k = 0 \end{cases} \quad (3.68)$$

El método elegido para la estimación de los parámetros de la distribución GPD es el método de los  $L$ -moments, descrito en la sección anterior.

### 3.5.4 Modelo de Poisson - $K4P$

Este modelo combinado de *Poisson-K4P* se plantea en este trabajo como una alternativa al modelo de *Poisson-GPD* descrito en la sección anterior. En este caso, se asume que el número de eventos sigue una distribución de Poisson, mientras que la magnitud de tales eventos es caracterizada a través de una distribución  $K4P$ .

No existe demostración teórica que justifique el uso de la distribución *Kappa* en la caracterización de eventos extremos, como puede suceder en el caso de la distribución generalizada de Pareto (Pickands, 1975). Sin embargo, esta última distribución surge como caso particular de la distribución  $K4P$ , por lo que es lógico pensar que su uso puede mejorar los resultados en aquellos casos en los que la distribución GPD no sea capaz de modelizar adecuadamente la magnitud de los eventos extremos.

Los trabajos que han sido publicados hasta el momento usando la distribución  $K4P$  se han centrado en el análisis de datos hidrológicos, tales como la precipitación atmosférica, Parida (1999) y Park (2002), no existiendo referencias de su uso en el contexto del estudio del oleaje. Además, la distribución ha sido utilizada en combinación con el método de muestreo de máximos anuales, no habiendo referencias de su uso en el contexto del análisis POT, tal y como se plantea en este trabajo.

Siguiendo el mismo procedimiento que en el caso del modelo Poisson-GPD, teniendo en cuenta ahora la distribución  $K4P$  dada por la expresión 3.30, la relación entre un cuantil  $X_T$  y su periodo de retorno asociado, toma la forma

$$X_T = u + \xi + \frac{\alpha}{k} \left[ 1 - \left( \frac{1 - (1 - (\lambda T)^{-1})^h}{h} \right)^k \right] \quad k, h \neq 0 \quad (3.69)$$

## 3.6 Pruebas de bondad de ajuste

Una prueba de bondad de ajuste se emplea para decidir cuando un conjunto de datos se ajusta a una distribución de probabilidad dada  $F_0(x)$ . Considérese una muestra aleatoria

de tamaño  $n$  de la distribución de una variable aleatoria  $X$  dividida en  $k$  clases exhaustivas y mutuamente excluyentes, y sea  $N_{ii} = 1, 2, \dots, k$ , el número de observaciones en la  $i$ -ésima clase.

Considérese la verificación de la hipótesis nula:

$$H_0 : F(x) = F_0(x) \quad (3.70)$$

donde el modelo de probabilidad propuesto  $F_0(x)$  se encuentra especificado, de manera completa, con respecto a todos los parámetros.

### 3.6.1 Estadística de Kolmogorov-Smirnov

Cuando  $F_0(x)$  es continua, la estadística de Kolmogorov-Smirnov (KS) se presenta como una opción adecuada para ser aplicada como test de bondad de ajuste. Esta prueba no necesita que los datos se encuentren agrupados y es aplicable a muestras de tamaño pequeño. Ésta se basa en la comparación entre las funciones de distribución acumulativa que se observan en la muestra ordenada y la distribución propuesta bajo la hipótesis nula. Si esta comparación revela una diferencia suficientemente grande entre las funciones de distribución muestral y propuesta, entonces la hipótesis nula de que la distribución es  $F_0(x)$ , se rechaza. Si la hipótesis nula es verdadera, entonces es lógico esperar que la diferencia sea relativamente pequeña.

Denótese por  $X_1, X_2, \dots, X_n$  a las observaciones ordenadas de una muestra aleatoria de tamaño  $n$  y defínase la función de distribución acumulativa muestral como

$$S_n(x) = 0, x < x_1 \quad S_n(x) = k/n, x_k \leq x < x_{k+1} \quad S_n(x) = 1, x \geq x_n$$

Dado que se especifica  $F_0(x)$  de manera completa, es posible evaluar a  $F_0(x)$  para algún valor deseado de  $x$ , y entonces comparar este último con el valor correspondiente de  $S_n(x)$ . Si la hipótesis nula es verdadera, entonces es lógico esperar que la diferencia sea relativamente pequeña.

La estadística  $KS$  se define como

$$D_n = \max |S_n(x) - F_0(x)| \quad (3.71)$$

El estadístico  $D_n$  tiene una distribución que es independiente del modelo propuesto bajo la hipótesis nula. Por esta razón, se dice que  $D_n$  es una estadística independiente de la distribución. Lo anterior da como resultado que la función de distribución de  $D_n$  pueda evaluarse en función del tamaño de la muestra y después usarse para cualquier  $F_0(x)$ . Los valores cuantiles superiores para  $D_n$  vienen tabulados en Cánavos (1988), para varios tamaños de la muestra. Para un tamaño del nivel de significación  $\alpha$  la región crítica es de la forma

$$P\left(D_n > \frac{c}{\sqrt{n}}\right) = \alpha \quad (3.72)$$

De acuerdo con lo anterior, la hipótesis nula  $H_0$  se rechaza si para algún valor  $x$  observado el valor de  $D_n$  se encuentra dentro de la región crítica de tamaño  $\alpha$ .

### 3.6.2 Raíz del error cuadrático medio

Sea  $T$  cualquier estimador de un parámetro desconocido  $\theta$ . Se define el error cuadrático medio de  $T$  como el valor esperado del cuadrado de la diferencia entre  $T$  y  $\theta$ .

Para cualquier estadística  $T$ , se denotará la raíz del error cuadrático medio por  $RMSE(T)$ , de sus siglas en inglés (*Root Mean Square Error*), de esta forma

$$RMSE(T) = \sqrt{E(T - \theta)^2} \quad (3.73)$$

### 3.6.3 Error cuadrático medio normalizado

La normalización del error cuadrático medio ( $NRMSE$ ) permite eliminar los efectos de escala, según la siguiente expresión:

$$NRMSE = \frac{RMSE}{y_{obs,max} - y_{obs,min}} \quad (3.74)$$

### 3.6.4 Coeficiente de determinación

El coeficiente de determinación se define según la siguiente expresión



$$CD = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (3.75)$$

donde  $0 \leq CD \leq 1$ . El coeficiente de determinación mide el grado de variabilidad total de la variable dependiente que es explicada por el modelo de regresión.

### 3.6.5 Gráficos cuantil-cuantil

Dada una muestra ordenada de observaciones independientes  $x_1 \leq x_2 \leq \dots \leq x_n$  de una población con función de distribución estimada  $\hat{F}$ , un gráfico cuantil-cuantil, definido aquí como *QQ plot*, consiste en

$$\left\{ \left( \hat{F}^{-1} \left( \frac{i}{n+1} \right), x_i \right) : i = 1, \dots, n \right\}$$

El nombre de *QQ plot* o *gráfico cuantil-cuantil*, deriva del hecho de que los valores  $x_i$  y  $\hat{F}^{-1}(i/n + 1)$  proporcionan estimaciones del cuantil  $i/n + 1$  de la distribución.

Si  $\hat{F}$  es un modelo razonable para la distribución de la población, al igual que antes, los puntos del *QQ plot* deberían situarse cerca de una línea de pendiente igual a la unidad.

## 3.7 Periodo de Retorno

El propósito fundamental del análisis extremal del oleaje es establecer una relación entre las alturas de ola elevadas y sus probabilidades de excedencia, las cuales pueden ser convertidas a periodos de retorno o intervalos de recurrencia media, siendo el periodo de retorno de una altura de ola dada,  $H_T$ , en años, la inversa de la probabilidad de que esa altura de ola específica sea excedida en cualquier año. Así,  $H_T$  representa la altura de ola que con un grado razonable de precisión se espera sea excedida, en promedio, una vez cada  $T$  años, motivo por el cual suele ser denominado *valor de retorno* o *valor de diseño asociado a un periodo de retorno  $T$* . En general, el nivel de retorno es denotado por  $X_T$ .

El periodo de retorno  $T$  puede considerarse una variable que contiene información conjunta sobre la frecuencia de presentación de un evento extremo y su magnitud. Como se ha descrito anteriormente, la situación más común es suponer que el número de eventos

extremos es caracterizado a través de una distribución de Poisson, mientras que la magnitud de tal evento puede modelizarse a través del método de máximos anuales o mediante el método alternativo de picos sobre umbral (POT).

### 3.7.1 Cálculo de $T$ a partir de la selección del máximo anual

Se define  $X$  como una variable aleatoria y  $X^*$  una variable aleatoria representando el máximo de  $X$  en un periodo dado (año, estación, mes). La longitud del periodo a partir del cual se extrae  $X^*$  es denotado por  $t$ . Definiendo la ocurrencia de un evento como la excedencia de un nivel dado  $x$  por  $X^*$ ,  $X^* > x$  se definen las siguientes relaciones

$T_X(x)$ , periodo de retorno

$X(T_x)$ , valor de  $X = x$  que tiene por periodo de retorno  $T_X(x)$

$F_X(x) = Prob[X^* < x]$ , función de distribución de  $X^*$

Asumiendo que los eventos están distribuidos según una distribución de Poisson, es posible obtener la relación entre  $F_X(x)$  y  $T_X(x)$ . Así,

$$Prob [m \text{ eventos en } N \text{ periodos}] = C_N^m [F_X(x)]^{N-m} [1 - F_X(x)]^m \quad (3.76)$$

la cual es una distribución binomial con media  $N[1 - F_X]$ . El número medio de eventos en el intervalo  $[0, t]$  es

$$M(x, t) = t[1 - F_X(x)] \quad (3.77)$$

Debido a que el periodo de retorno de un evento descrito por un proceso de Poisson es  $T = 1/\lambda$  donde  $\lambda = dM(x, t)/dt$  se obtiene

$$F_X(x) = 1 - \frac{t}{T_X(x)} \quad \text{ó} \quad T_X(x) = t \left( \frac{1}{1 - F_X(x)} \right) \quad (3.78)$$

En este sentido, el periodo de retorno  $T_X$  es el tiempo medio separando dos años sucesivos con valor máximo anual ( $t=1$ ) de la variable mayor que  $X(T_X)$ . Este tipo de periodo de retorno podría ser útil en casos en los que resulta indiferente conocer si existe más de una excedencia por periodo de muestreo.

### 3.7.2 Cálculo de $T$ a partir del método POT

Se define la variable aleatoria  $X$  y la variable  $X_u$  representando el máximo de  $X$  durante una excedencia del umbral  $u$ . La ocurrencia de un evento se define como la excedencia de un nivel  $x$  por el máximo  $X_u$ . Se definen entonces las siguientes cantidades

$T_u(X)$ , periodo de retorno

$X(T_u)$ , valor de  $X = x$  que tiene periodo de retorno  $T_u(x)$

$F_u(x) = Prob[X_u < x]$ , función de distribución de probabilidad de  $X_u$

Tenemos entonces dos tipos de eventos presentes: El evento  $X$  excediendo el umbral  $u$  con el que se asocia la distribución  $p_m(t)$  y el evento  $X$  excediendo el nivel  $x$  con distribución  $F_u(x)$ .

Asumiendo que los eventos están distribuidos según una distribución de Poisson, el número medio de eventos  $X_u > x$  toma la forma

$$M(x, t) = \lambda_u t [1 - F_u(x)] \quad (3.79)$$

Al igual que en el caso anterior  $T = 1/\lambda$  donde  $\lambda = dM(x, t)/dt$ , por lo que

$$F_u(x) = 1 - \frac{1}{\lambda_u T_u(x)} \quad \text{ó} \quad T_u(x) = \frac{1}{\lambda_u (1 - F_u(x))} \quad (3.80)$$

En este sentido, el periodo de retorno  $T_u$  es el tiempo medio que transcurre entre dos valores sucesivos mayores que  $u$  excediendo  $X(T_u)$ . Notar que esta interpretación es diferente de la obtenida en el caso anterior. Un estimador para  $\lambda_u$  es

$$\lambda = \frac{n_u}{N_y}$$

donde  $n_u$  es el número de datos sobre el umbral y  $n_y$  el número de años que forma la muestra.

## 3.8 Intervalos de confianza

Los intervalos de confianza asociados a cualquier estimación estadística desempeñan un papel clave en la interpretación de los resultados obtenidos al analizar una muestra de datos de tamaño limitado, correspondientes a un proceso aleatorio. Estos intervalos permiten valorar la incertidumbre asociada a una cierta estimación y la fiabilidad estadística de la misma.

Generalmente, los intervalos de confianza asociados a los parámetros de una distribución dada, o para otras estimaciones estadísticas, son obtenidos empleando una aproximación normal. Es decir, mediante la estimación de los errores estándar de los parámetros estimados y, por tanto, dependiendo del método de estimación particular empleado. Dichos intervalos de confianza poseen límites superiores e inferiores de la forma  $\hat{x} \pm c\hat{\sigma}_x$ , donde  $\hat{x}$  representa el parámetro estimado,  $c$  una constante que determina el grado de confianza, y  $\hat{\sigma}_x$  el error estándar de la estimación.

De forma intuitiva, en general, los intervalos de confianza para los valores de retorno deberían ser asimétricos, poniendo de manifiesto que al trabajar con valores extremos, la incertidumbre debe ser mayor hacia la zona de valores superiores al de las estimaciones que hacia la región de valores menores. En estas situaciones existe un claro contraste con aquellas en las que se desea determinar los intervalos de confianza de la media de una distribución simétrica.

Aunque existen metodologías que en determinados casos permiten obtener de forma analítica expresiones explícitas para los intervalos de confianza, con frecuencia, esta tarea resulta extraordinariamente complicada o imposible. En dichas situaciones, los métodos *bootstrap* (Efron y Tibshirani, 1993) representan una alternativa simple y útil para derivar los intervalos de confianza de las estimaciones.

Esta metodología, basada en técnicas de simulación computacional intensiva, ha experimentado un considerable aumento de popularidad en las últimas décadas y ha sido empleada en el campo de los valores extremos por diversos autores (e.g., Zucchini y Adamson, 1989; Naess and Clausen, 2001; Naess and Hungnes, 2001; Tajvidi, 2003; Coles y Simiu, 2003; etc.). En este tipo de metodología se puede hacer una clara subdivisión entre técnicas paramétricas y no paramétricas. El bootstrap no paramétrico se basa en el remuestreo con reemplazamiento a partir de la muestra dada y el cálculo de los parámetros necesarios a partir de un gran número de muestras repetidas, de donde es posible estimar los intervalos

de confianza correspondientes. Por su parte, las técnicas paramétricas se basan en generar muestras aleatorias a partir de un modelo, o distribución, paramétrica ajustada a los datos observados. Esta última versión debería ser elegida si se conoce un modelo que caracteriza adecuadamente la muestra observada, las muestras de datos son relativamente pequeñas y el interés se centra en la cola de la distribución (Davison y Hinkley, 1997), tal como ocurre en el estudio de valores extremos. Sin embargo, la versión no paramétrica parece ser más popular que la paramétrica en estudios en los que se aborda la estimación de parámetros de retorno y cuantiles asociados a eventos extremos. En tal sentido, Kyselý (2008) realiza un estudio comparativo de ambas metodologías y concluye que, para muestras de tamaño pequeño o moderado, los resultados del método no paramétrico deberían ser interpretados con cautela, puesto que conducen a la estimación de intervalos de confianza muy estrechos y tienden a infraestimar las incertidumbres reales. Aunque el método paramétrico produce intervalos de confianza ligeramente amplios, en general, mejora la estimación de las incertidumbres, incluso cuando el modelo adoptado para caracterizar la muestra no es adecuado. En el presente estudio se ha optado por el uso de la versión no paramétrica, por ser ésta la más empleada y contrastada en el estudio de los eventos extremos del oleaje.

En este método se asume que  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  es un vector formado por  $n$  observaciones independientes de una variable aleatoria  $V$ . Existen dos formas básicas de trabajar con esta técnica. La primera de ellas consiste en una aproximación no paramétrica donde la variable aleatoria  $V$  queda caracterizada a través de su función de distribución de probabilidad empírica (EDF). Por otro lado, está la aproximación paramétrica, asumiendo que la variable  $V$  tiene por función de distribución de probabilidad (CDF) una función conocida  $F_V(v, \hat{\theta}) = \text{Prob}(V \leq v)$  donde  $\theta$  denota un vector de parámetros desconocidos que determinan la CDF. Estos parámetros son estimados a partir de los datos observados  $\mathbf{v}$  mientras que  $F_V(v, \hat{\theta})$  es supuesta como la CDF de  $V$ . En el presente trabajo se utilizará la aproximación no paramétrica.

El objetivo es estimar un parámetro estadístico,  $R$ , a partir de los datos observados  $\mathbf{v}$  utilizando la EDF. De esta forma  $\hat{R}$  es la estimación de  $R$  basada en los datos observados. La técnica está basada en el remuestreo (con reemplazamiento) a partir de la EDF obtenida esta a partir de los datos observados.

El proceso se realiza según: a partir de la función de distribución de probabilidad empírica se obtienen nuevos valores empíricos de la variable  $V^*$ . Se genera un número  $l$  de muestras  $v_j^*, j=1, \dots, l$ , con  $n$  observaciones independientes de  $V^*$  en cada una de ellas. Cada muestra  $v_j^*$  genera una estimación de la cantidad  $R_j^*$  de  $R$ .

Un estimador para los intervalos de confianza de  $\hat{R}$  se obtiene a partir de la desviación estándar  $s_R^*$ :

$$s_R^* = \sqrt{\frac{1}{l-1} \sum_{j=1}^l (R_j^* - \bar{R}^*)^2} \quad (3.81)$$

donde  $\bar{R}^* = (1/l) \sum_{j=1}^l R_j^*$ . Para un nivel de confianza  $1 - q$  se obtiene entonces

$$(\hat{R} - w_{q/2} s_R^*, \hat{R} + w_{q/2} s_R^*) \quad (3.82)$$

donde  $w_{q/2}$  denota el cuantil de la distribución normal estándar  $100(1 - q/2)\%$ .

Con el objetivo de evitar la asunción de que las estimaciones realizadas están normalmente distribuidas la verdadera distribución puede aproximarse generando un número suficientemente grande de muestras donde  $l$  suele ser del orden de  $10^3$ . Entonces  $R_j^*$  es ordenado en sentido creciente de forma que el intervalo de confianza toma la forma  $(R_L^*, R_M^*)$  donde  $L = [ql/2]$  y  $M = [(1 - q/2)l]$  con un nivel de confianza  $100(1 - q)\%$ .

### 3.9 Estacionalidad de los eventos extremos

La homogeneidad de la serie temporal objeto de estudio es una de las hipótesis básicas en las que se sustenta el modelo combinado de Poisson-GPD. En otras palabras, los datos deben pertenecer a una misma población inicial, o lo que es equivalente, deben tener la misma distribución de probabilidad,  $F$ .

La mayoría de las variables geofísicas, ya sean atmosféricas u oceanográficas, presentan comportamientos marcadamente estacionales a lo largo de un año. A modo de ejemplo puede citarse la temperatura del mar, presentando sus valores más altos en verano y sus valores mínimos en invierno. La altura del oleaje suele presentar sus mayores valores en invierno, coincidiendo con las perturbaciones atmosféricas que lo generan. Este comportamiento estacional de la serie temporal se traduce en una no homogeneidad de la misma, debiendo por tanto ser cautelosos a la hora de aplicar modelos estacionarios que no tienen en cuenta este efecto (Athanasoulis et al., 1995 ; Guedes Soares et al., 1996).

Parece normal asumir que si la variable física bajo estudio presenta un comportamiento estacional, los valores extremos de dicha variable también presentarán un comportamiento estacional. Sin embargo, este aspecto no ha recibido demasiada atención en el contexto del análisis de valores extremos. El uso cada vez más común de la técnica POT ha estimulado

el interés sobre el aspecto estacional dentro de los valores extremos. Así, Morton *et al.* (1997), Méndez *et al.* (2006), Jonathan *et al.* (2011). Mackay *et al.* (2010) han evidenciado la importancia de tener en cuenta la estacionalidad a través de modelos que permitan modelizarla o eliminándola a través de alguna de las técnicas disponibles, como por ejemplo dividiendo la serie en periodos homogéneos.

La identificación de ciclos estacionales anuales en la altura significativa del oleaje resulta más o menos sencillo a simple vista en función de las condiciones climatológicas de la zona de estudio. Así, la figura 3.7 y 3.8 muestran el valor medio diario de  $H_s$  en las boyas LP-I y LP-II. En la primera figura puede observarse un claro patrón estacional, con valores mayores en invierno. Sin embargo, en la segunda figura el comportamiento estacional no resulta tan evidente.

Por otro lado, las figuras 3.9 y 3.10 muestran la distribución de los valores extremos de  $H_s$  durante el año para las boyas LP-I y LP-II respectivamente, para diferentes valores umbrales.

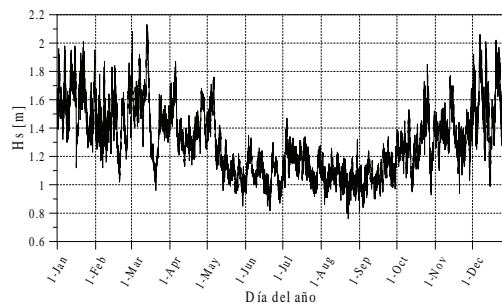


Figura 3.7: Boya Lp-I. Valor medio horario anual de la altura de ola significativa (1987-2001)

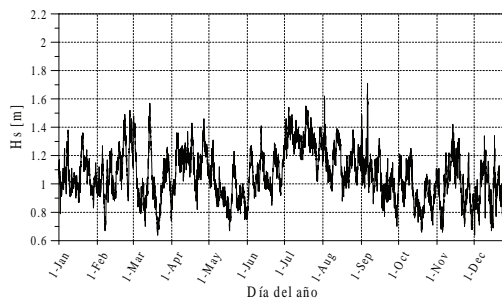


Figura 3.8: Boya LP-II. Valor medio horario anual de la altura de ola significativa (1991-2001)

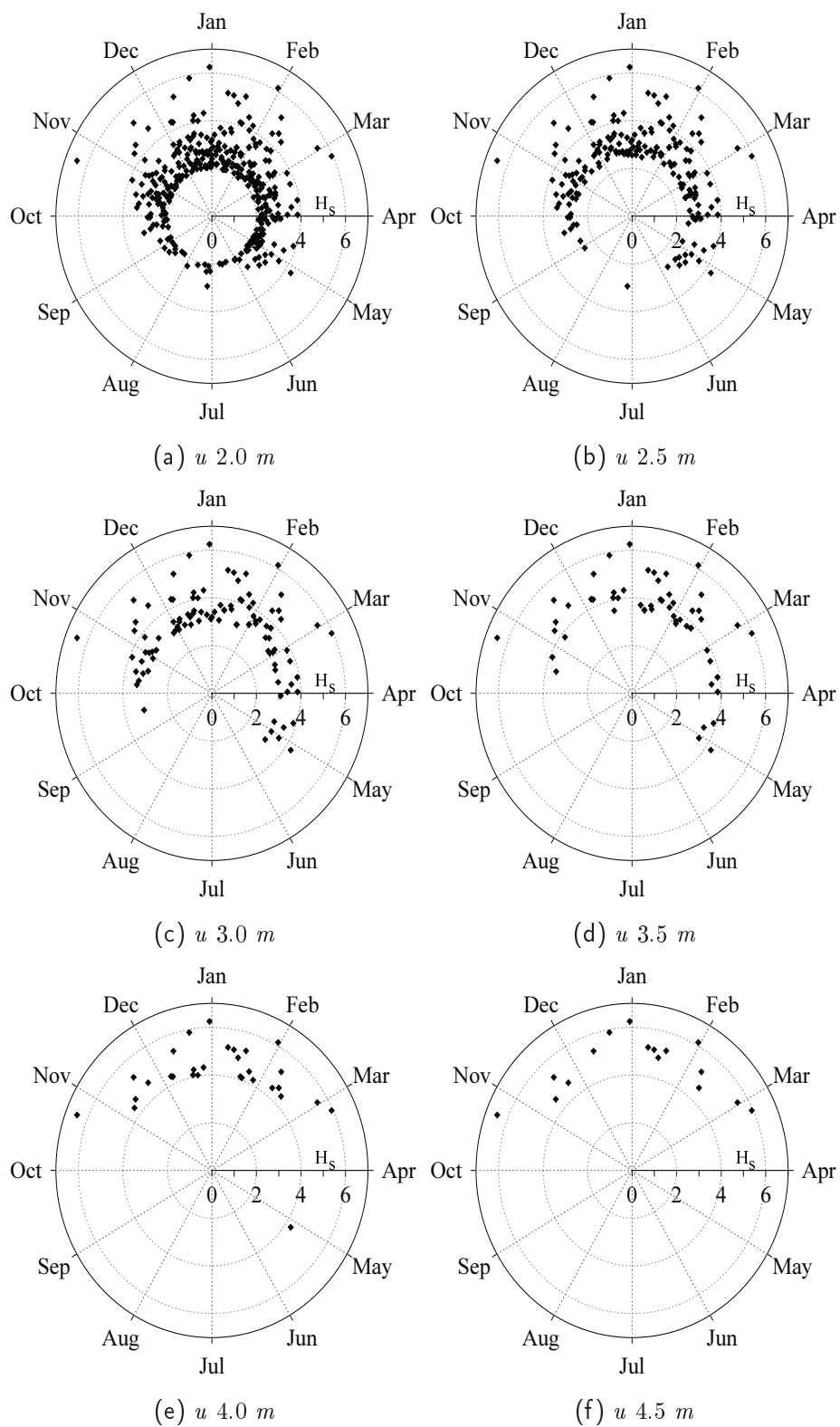


Figura 3.9: Boia LP-I. Distribución temporal de los valores extremos de la altura de ola significativa,  $H_s(m)$ , durante el año sobre diferentes umbrales



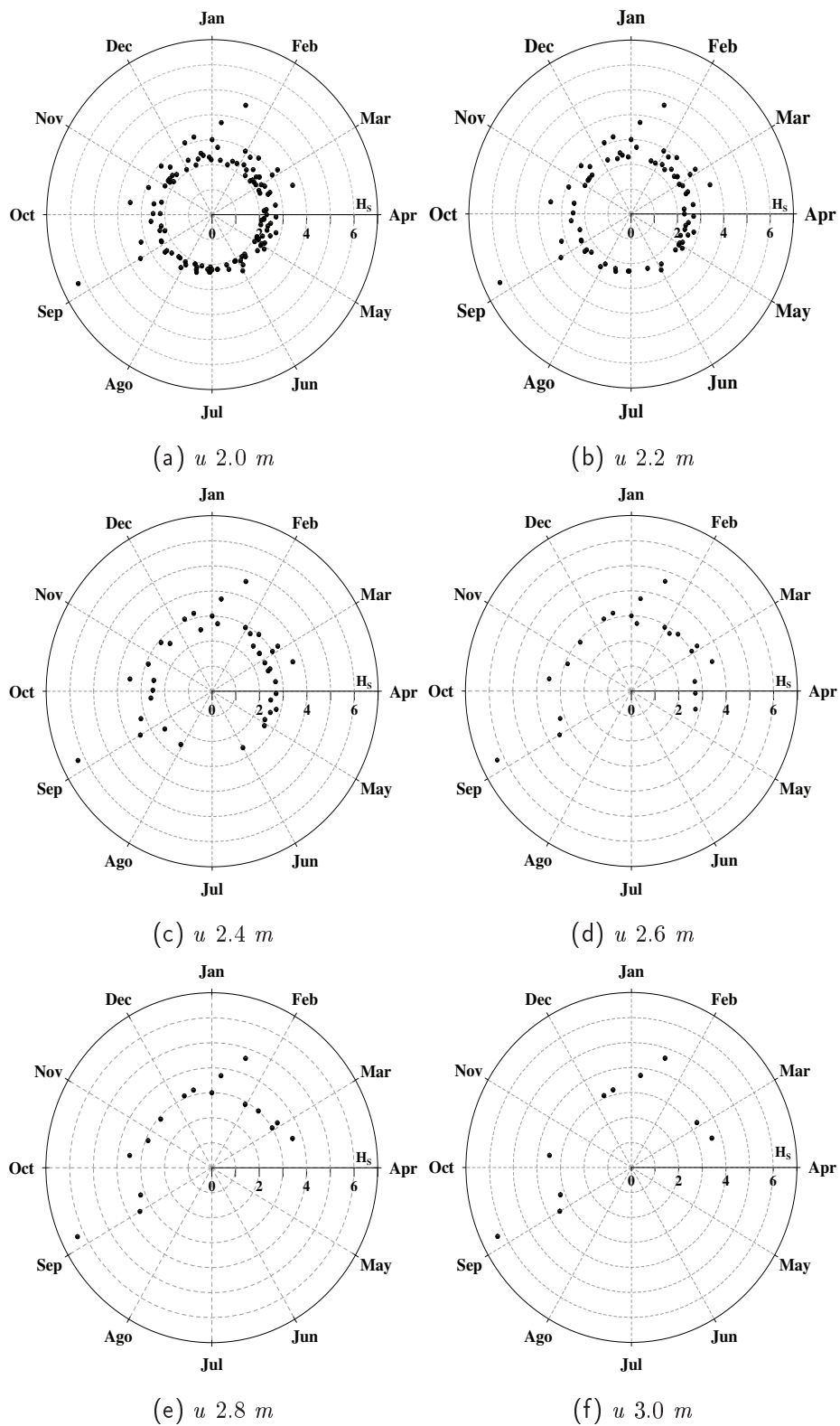


Figura 3.10: Boya LP-II. Distribución temporal de los valores extremos de la altura de ola significativa,  $H_s$ (m), durante el año sobre diferentes umbrales

Para el primer caso, con valores pequeños del umbral, se observan valores extremos distribuidos a lo largo de todo el año, aunque resulta evidente la existencia de una mayor concentración de puntos en los meses de invierno. A medida que el umbral aumenta, esta distribución de tormentas en los meses de invierno es más pronunciada, desapareciendo prácticamente la totalidad de los valores extremos en los meses de verano. Para el caso de LP-II, si bien es posible observar el mismo efecto, no menos cierto es que existe una mayor dispersión de los valores extremos a lo largo del año, indicando un menor carácter estacional de los valores extremos, en consonancia con lo mostrado en la figura 3.8.

En numerosas ocasiones en las que la variable de estudio presenta un comportamiento cíclico, la rama de la estadística conocida como *estadística circular* ofrece los métodos necesarios para su estudio (Mardia et al.). Así, por ejemplo, es posible considerar cada día del año ( $d$ ) donde sucede una tormenta como una variable aleatoria circular. La variable  $d$  debe entonces ser convertida en una variable angular  $\theta$  en radianes, según

$$\theta = \frac{2\pi}{365}d \quad (3.83)$$

Esto es útil para determinar si un conjunto de datos está o no distribuido uniformemente alrededor del círculo.

Entonces, para saber si los eventos extremos se distribuyen uniformemente a lo largo del año (hipótesis nula) o si por el contrario tienden a agruparse en periodos específicos (hipótesis alternativa), es necesario saber si el tiempo de ocurrencia de los eventos extremos sigue una distribución uniforme

$$f(\theta) = \frac{1}{2\pi} \quad (3.84)$$

Existen numerosos tests estadísticos para contrastar dichas hipótesis, y cuya descripción detallada puede verse en Fisher (1993). En el presente trabajo se utilizan tres estadísticos que se describen brevemente a continuación.

### ***Test de la Chi-cuadrado***

Este test es el más utilizado para evaluar el ajuste de un conjunto de datos a una determinada distribución teórica. Para su aplicación el círculo es dividido en 12 sectores,  $k = 12$ , de  $30^\circ$  de amplitud cada uno, calculando las frecuencias observadas  $o_i$  y esperadas  $e_i$  en cada mes  $i$ .

La expresión viene dada por:

$$T = \sum_k \frac{\sum_{i=1}^k (o_i - e_i)^2}{e_i} \quad (3.85)$$

La hipótesis nula (uniformidad) es rechazada si  $T$  excede el valor crítico  $T_c$  para  $k - 1$  grados de libertad. Cabe destacar que este test es robusto para muestras unimodales y multimodales.

### ***Test de Rayleigh***

Este test está basado en el valor estimado del vector resultante de longitud  $R$ , dado por

$$R = \sqrt{C^2 + S^2} \quad (3.86)$$

donde

$$C = \frac{1}{N} \sum_{i=1}^k n_i \cos \theta_i \quad (3.87)$$

$$S = \frac{1}{N} \sum_{i=1}^k n_i \sen \theta_i \quad (3.88)$$

siendo  $\theta_i$  el centro de cada intervalo mensual  $i$  expresado como un ángulo en el rango  $[0, 2\pi)$ . Este test es adecuado para detectar direcciones unimodales en una muestra de vectores.

El nivel de confianza  $P$  asociado con la longitud resultante media  $R$  en un test simple de Rayleigh viene dado por

$$P = e^{(-Z)} \left[ 1 + \frac{2Z - Z^2}{4N} - \frac{24Z - 132Z^2 + 76Z^3 - 9Z^4}{288N^2} \right] \quad (3.89)$$

donde

$$Z = N * R^2 \quad (3.90)$$

La hipótesis nula de uniformidad es rechazada si  $R$  excede un determinado valor crítico  $R_c$  para un nivel de significación  $\alpha$ , o dicho de otro modo, si el valor de la variable  $P$  es inferior al valor de significación  $\alpha$ .

Este test se considera adecuado sólo cuando es posible asumir que la distribución de la población tiene una única moda, asumiendo además que la muestra sigue una distribución de Von Mises.

Un valor de  $R = 1$  significa que todos los eventos ocurren el mismo día del año durante todos los años, mientras que un valor de  $R = 0$  implica que la probabilidad de un evento es la misma para cualquier día del año. Este último caso sólo ocurre cuando la distribución tiene una única moda.

### ***Test de Kolmogorov-Smirnov modificado***

Freedman (1979) sugirió una modificación del test clásico de Kolmogorov-Smirnov para examinar la estacionalidad de un conjunto de datos. Esta metodología no paramétrica elimina algunos inconvenientes derivados de la metodología convencional.

La hipotética (uniforme) distribución acumulada es una función definida por  $F(t) = t/12$ , donde  $t$  es el rango de cada mes del año. La distribución acumulada muestral es una función definida por  $F_N(t) = k/N$ , donde  $k$  es el número de eventos ocurridos durante todos los meses  $\leq t$ . El estadístico,  $V_n$ , viene dado por:

$$V_n = \max(F_N(t) - F(t)) + |\min(F_N(t) - F(t))|; 1 < t < 12 \quad (3.91)$$

La distribución de  $V_n$  no sigue ninguna distribución de probabilidad específica, aunque ha sido empíricamente evaluada por medio de simulaciones con el método Monte Carlo y tabulada en Freedman (1979). La hipótesis nula será rechazada si  $V_n * \sqrt{n}$  es mayor que el valor crítico 1.66 con un nivel de significación  $\alpha = 0.01$ .

Si existen indicios de una posible estacionalidad en la serie, una de las formas de abordar su estudio es a través del uso de modelos que tengan en cuenta este efecto, Coles 2001, en los cuales los parámetros libres del modelo se expresan como una función del tiempo. Por supuesto, estos modelos presentan una mayor complejidad de uso que los modelos descritos hasta el momento.

Como alternativa al método anterior, una de las estrategias más comunes es intentar convertir la serie temporal no homogénea en una serie temporal homogénea. Para ello

Tabla 3.3: Estaciones climáticas

	Comienzo	Final
Primavera	21 marzo	21 junio
Verano	21 junio	22 septiembre
Otoño	22 septiembre	21 diciembre
Invierno	21 diciembre	21 marzo

existen básicamente dos aproximaciones. La primera consiste en identificar la componente estacional de la serie y sustraerla posteriormente. Una desventaja de este método, en el contexto del análisis extremal, es que esta componente estacional es obtenida de forma promediada para toda la serie temporal, de forma que siendo válida para la parte central de la serie, puede no serlo para los valores más altos de la variable, que son en definitiva el objeto de estudio (Davison y Smith, 1990).

La segunda aproximación al estudio de la estacionalidad consiste en dividir la serie temporal en intervalos temporales que puedan ser considerados homogéneos, de forma que el modelo es utilizado en cada uno de esos intervalos. La cuestión aquí es determinar el número de intervalos temporales que deben ser considerados en el análisis. La forma más común de trabajar es considerando cada estación climática: *primavera*, *verano*, *otoño* e *invierno*, como un intervalo temporal homogéneo. Los límites de cada estación se muestran en la tabla 3.3. Sin embargo, algunos autores, ver Carter & Challenor (1981), proponen tomar como intervalo temporal homogéneo cada uno de los meses del año, generando así doce intervalos temporales. Davison & Smith (1990) proponen dividir el año en tan solo dos intervalos temporales basándose en el número de datos generados sobre un umbral determinado. De esta forma, el año es dividido en un primer intervalo abarcando los meses comprendidos entre Enero y Junio y un segundo intervalo que abarca los meses comprendidos entre Julio y Diciembre.

En esta misma línea de dividir el año en diferentes periodos homogéneos, Ouarda (1993) propone una metodología un tanto diferente para seleccionar periodos homogéneos dentro de la serie temporal original. El método plantea la representación del número total de excedencias  $n(t)$  durante un periodo  $[0, t]$  para cada año de la serie. Cada periodo temporal homogéneo es identificado entonces como aquel intervalo donde la cantidad  $dn/dt$  es constante. En otras palabras, el cambio de pendiente en el gráfico equivale a un cambio de estación. Esta metodología permite dividir el año de una forma más realista en intervalos temporales homogéneos.

Un aspecto importante a tener en cuenta es que lo que se pretende con este análisis estacional es determinar si existen diferencias entre los resultados obtenidos cuando el análisis es llevado a cabo sobre la serie temporal completa y cuando es llevado a cabo sobre los intervalos temporales homogéneos. Cuanto mayor sean las diferencias entre ambos mayor grado de estacionalidad presentará la serie. Para ello, dentro del contexto del análisis POT, la forma correcta de proceder sería determinando un valor umbral inicial para la serie completa, usando posteriormente ese mismo valor para cada intervalo temporal. El problema está entonces en el hecho de que si el número de intervalos temporales es demasiado grande se corre el riesgo de que no existan datos sobre el umbral en alguno de ellos, siendo imposible evaluar el efecto de la estacionalidad. Como referencia para evaluar los posibles efectos estacionales se puede escoger el cuantil asociado a un periodo de retorno determinado.

## 3.10 Resultados y discusión

### 3.10.1 Aplicación del método de máximos anuales

En este apartado se presentan los resultados obtenidos de la aplicación del método de máximos anuales a los datos registrados por la boya LP-I y LP-II. Dado que el método de muestreo consiste en seleccionar el valor máximo de la altura del oleaje dentro de cada año, se obtiene una población extremal de tamaño igual al número de años que componen la muestra, es decir, 17 datos en LP-I y 11 datos en LP-II.

Las distribuciones de probabilidad utilizadas en la caracterización de los valores máximos anuales se muestran en la tabla 3.4. Inicialmente se utilizan las tres distribuciones extremales clásicas de *Gumbel*, *Fréchet* y *Weibull*. Los métodos estadísticos de los momentos (MOM), máxima verosimilitud (MLE) y mínimos cuadrados (LS) son los elegidos en este caso para la estimación de los parámetros de las distribuciones.

Los parámetros estimados de las tres distribuciones extremales clásicas se recogen en la tabla 3.5 junto con los valores del error cuadrático medio (RMSE), error cuadrático medio normalizado (NRMSE), coeficiente de determinación (CD) y valor del test de Kolmogorov-Smirnov (KS).

Los tres métodos de estimación ofrecen resultados muy similares para cada una de las distribuciones utilizadas. Los valores de RMSE, NRMSE y CD son útiles para discernir

Tabla 3.4: Distribuciones de probabilidad utilizadas en la caracterización de los valores máximos anuales de la altura del oleaje

Distribución	Expresión
<i>Gumbel</i>	$G(z) = \exp \left\{ -\exp \left[ -\left( \frac{z-b}{a} \right) \right] \right\}$
<i>Fréchet</i>	$G(z) = \exp \left\{ -\left( \frac{z-b}{a} \right)^{-\beta} \right\}$
<i>Weibull</i>	$G(z) = \exp \left\{ -\left( \frac{b-z}{a} \right)^\beta \right\}$
<i>GEV</i>	$G(z) = \exp \left\{ -(1 - kz)^{\frac{1}{k}} \right\}$
<i>GAM</i>	$G(z) = H(z/\beta, \alpha)$ donde $H(z, \alpha) = \{\Gamma(\alpha)\}^{-1} \int_0^z t^{\alpha-1} e^{-t} dt$
<i>K4P</i>	$G(z) = \left\{ 1 - h [1 - k(z - \xi)/\alpha]^{1/k} \right\}^{1/h}$
<i>WAK</i>	$z(G) = \xi + \frac{\alpha}{\beta} \left\{ 1 - (1 - G)^\beta \right\} - \frac{\gamma}{\delta} \left\{ 1 - (1 - G)^{-\delta} \right\}$

Tabla 3.5: Método de máximos anuales. Boya LP-I. Parámetros de las funciones de distribución de *Gumbel*, *Fréchet* y *Weibull*, obtenidos con los métodos MOM, MLE y LS. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05

		$\alpha$	$\xi$	$\beta$	RMSE	NRMSE	CD	KS
<i>Gumbel</i>	MOM	0.794	4.617	0.000	0.079	0.129	0.958	0.188
	MLE	1.000	4.581	0.000	0.054	0.088	0.964	0.134
	LS	1.015	4.554	0.000	0.056	0.092	0.962	0.135
<i>Fréchet</i>	MOM	4.603	0.000	7.166	0.102	0.177	0.944	0.214
	MLE	4.460	0.000	4.370	0.066	0.115	0.945	0.115
	LS	4.442	0.000	4.546	0.070	0.124	0.943	0.129
<i>Weibull</i>	MOM	5.482	0.000	5.774	0.057	0.084	0.973	0.148
	MLE	5.476	0.000	6.198	0.067	0.108	0.969	0.160
	LS	5.523	0.000	4.880	0.041	0.057	0.977	0.114

entre los tres métodos de estimación propuestos, mientras que el test de Kolmogorov-Smirnov permite evaluar la idoneidad del modelo a través del contraste de hipótesis

$$H_0 : \text{ Serie distribuida según } G(z)$$
$$H_1 : \text{ Serie no distribuida según } G(z)$$

Para este último, dado que el valor del test es inferior al valor crítico de referencia de 0.318 para  $n = 17$ , no es posible rechazar la hipótesis nula, o lo que es lo mismo, debemos concluir que las tres distribuciones pueden caracterizar adecuadamente a la serie de datos extremales. Sin embargo, teniendo en cuenta los gráficos cuantil-cuantil mostrados en la figura 3.11, resulta evidente que la distribución de *Weibull* es la más adecuada para la caracterización de los valores máximos anuales. Tanto la distribución de *Gumbel* como la distribución de *Fréchet* presentan desviaciones respecto de la línea recta de pendiente unidad que deberían seguir los datos si el modelo fuese el adecuado. Dentro de la distribución de *Weibull*, el método de estimación que muestra los valores menores de RMSE, NRMSE y CD es el de mínimos cuadrados.

De esta forma, la distribución de *Weibull*, en conjunto con el método de mínimos cuadrados, parece ser la mejor opción para caracterizar los valores máximos anuales de la altura del oleaje. Sin embargo, el gráfico cuantil-cuantil correspondiente al método MLE parece ajustar mejor los valores más altos de la serie.



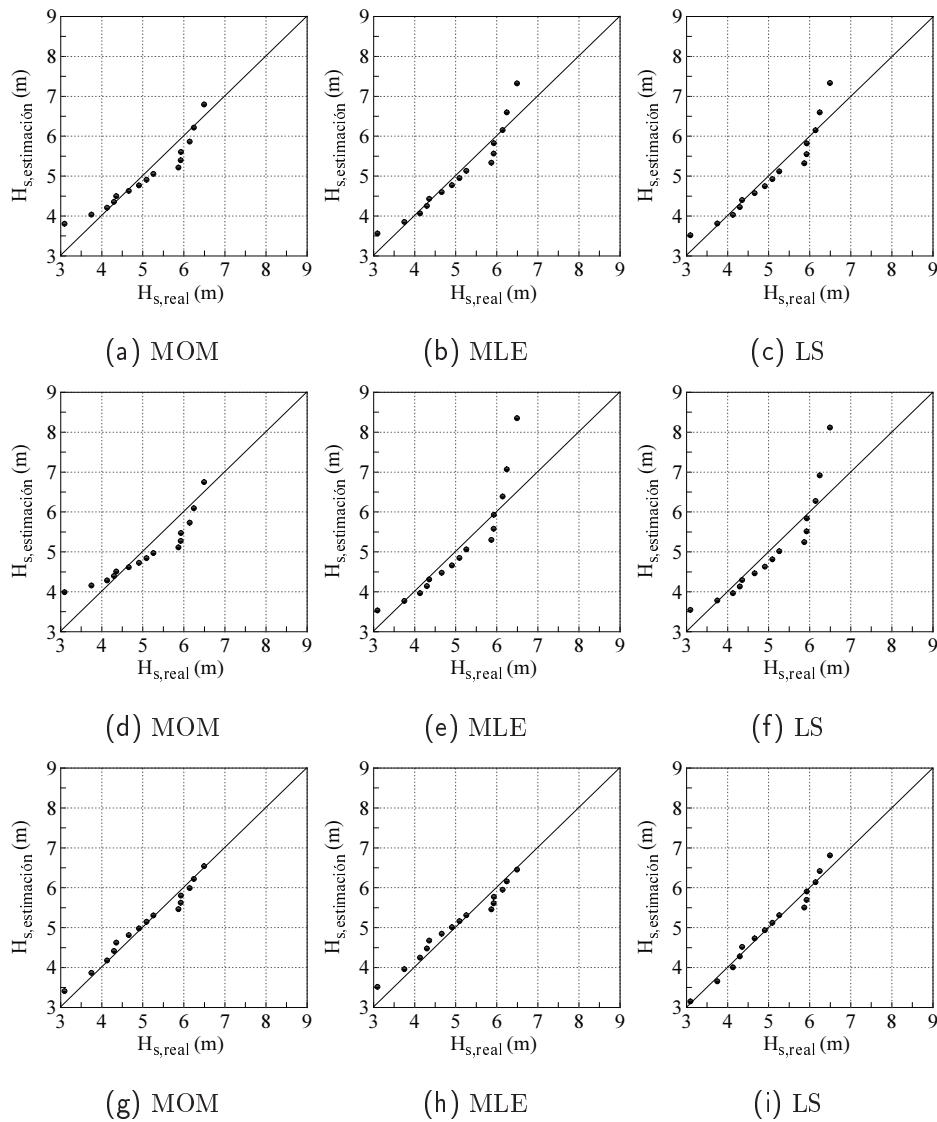


Figura 3.11: Método de máximos anuales para la boya LP-I. Gráficos cuantil-cuantil correspondientes a las funciones de distribución de *Gumbel* (superior), *Fréchet* (centro) y *Weibull* (inferior) utilizando los métodos MOM, MLE y LS en la estimación de parámetros

La tabla 3.6 recoge los cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100, 500 y 1000 años para las tres distribuciones de probabilidad y los tres métodos de estimación propuestos. De acuerdo con lo comentado anteriormente, los cuantiles más fiables serán los obtenidos con la distribución de *Weibull* y el método de mínimos cuadrados. Así, se pone de manifiesto la importancia en el método de estimación de los parámetros, ya que la diferencia entre los cuantiles estimados a través del método LS y los métodos MOM y MLE es, en promedio, un 5 % superior.

Tabla 3.6: Método de máximos anuales. Boya LP-I. Cuantiles asociados a periodos de retorno  $T_r = 50, 100, 500$  y 1000 años.

<i>Función</i>	Estimador	$H_{50}$	$H_{100}$	$H_{500}$	$H_{1000}$
<i>Gumbel</i>	MOM	7.715	8.269	9.551	10.101
	MLE	8.483	9.181	10.795	11.488
	LS	8.516	9.225	10.862	11.567
<i>Fréchet</i>	MOM	7.933	8.745	10.953	12.066
	MLE	10.892	12.779	18.486	21.666
	LS	10.478	12.218	17.423	20.294
<i>Weibull</i>	MOM	6.943	7.142	7.522	7.661
	MLE	6.823	7.005	7.352	7.479
	LS	7.305	7.553	8.031	8.207

Con el objetivo de evaluar la capacidad de las funciones de distribución de probabilidad alternativas para caracterizar la serie de extremos, se ajustan las distribuciones GEV, K4P, GAM y WAK a la serie de máximos anuales, utilizando el método de estimación de parámetros de los  $L$ -moments (LMOM).

Los parámetros estimados se recogen en la tabla 3.7 así como los valores de RMSE, NRMSE, CD y KS. Las distribuciones K4P y WAK ofrecen los mejores ajustes de entre todas las distribuciones utilizadas, atendiendo a los valores del RMSE. Este resultado es lógico si se tiene en cuenta que estas dos distribuciones poseen 4 y 5 parámetros respectivamente, permitiendo mayor flexibilidad en el ajuste. No obstante, los valores de RMSE, NRMSE y CD para ambas distribuciones son prácticamente iguales, indicando que la introducción de un quinto parámetro por parte de la distribución WAK no mejora significativamente el ajuste final. Este resultado también es verificable a través de los gráficos cuantil-cuantil mostrados la figura 3.12, donde puede observarse que los correspondientes a las distribuciones K4P y WAK son prácticamente idénticos.

Tabla 3.7: Método de máximos anuales. Boya LP-I. Parámetros de las funciones de distribución GEV, K4P, GAM y WAK obtenidos con el método LMOM. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05

<i>Función</i>	$\alpha$	$\xi$	$k$	h		RMSE	NRMSE	CD	KS
GEV	1.141	4.797	0.470	-		0.048	0.072	0.976	0.124
K4P	3.180	3.795	1.135	0.762		0.035	0.060	0.983	0.082
GAM	0.223	0.000	22.677	-		0.057	0.083	0.970	0.157
	$\alpha$	$\xi$	$\beta$	$\gamma$	$\delta$	RMSE	NRMSE	CD	KS
WAK	637.418	0.000	202.197	4.510	-1.327	0.034	0.061	0.984	0.077

Tabla 3.8: Método de máximos anuales. Boya LP-I. Cuantiles asociados a periodos de retorno  $T_r = 50, 100, 500$  y  $1000$  años.

<i>Función</i>	$H_{50}$	$H_{100}$	$H_{500}$	$H_{1000}$
GEV	6.838	6.947	7.096	7.132
K4P	6.562	6.580	6.592	6.594
GAM	7.496	7.878	8.688	9.013
WAK	6.533	6.544	6.550	6.551

De esta forma, se puede concluir que la distribución K4P sería la mejor opción para caracterizar la serie de máximos anuales, mejorando sesnsiblemente al ajuste proporcionado por la distribución de Weibull.

Los cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100, 500 y 1000 años se muestran en la tabla 3.8. Los cuantiles correspondientes a la distribución K4P deberían ser los más fiables, obsevándose una sobreestimación por parte de las distribuciones GEV y GAM.

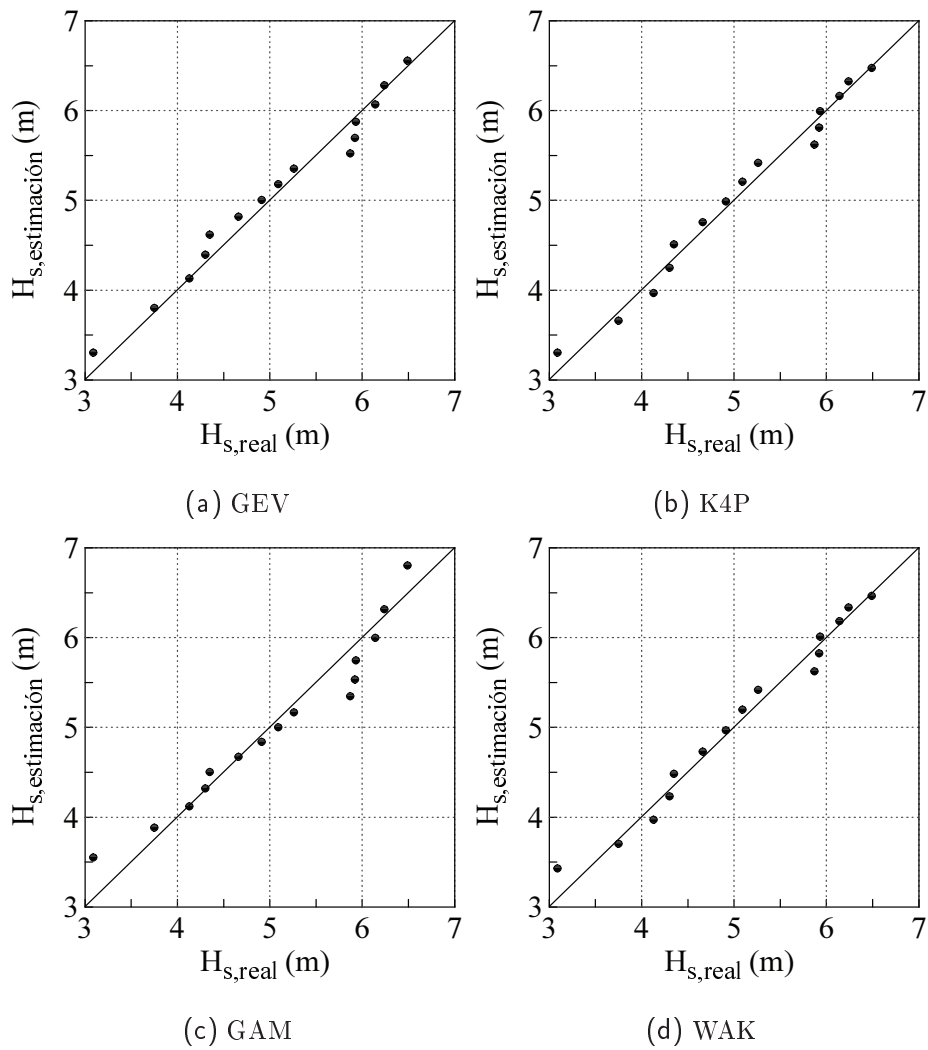


Figura 3.12: Método de máximos anuales para la boya LP-I. Gráficos cuantil-cuantil correspondientes a las funciones de distribución GEV, K4P, GAM y WAK

Tabla 3.9: Método de máximos anuales. Boya LP-II. Parámetros de las funciones de distribución de *Gumbel*, *Fréchet* y *Weibull*, obtenidos con los métodos MOM, MLE y LS. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05

		$\alpha$	$\xi$	$\beta$	RMSE	NRMSE	CD	KS
<i>Gumbel</i>	MOM	0.890	2.970	0.000	0.054	0.095	0.957	0.113
	MLE	0.700	3.036	0.000	0.056	0.102	0.973	0.083
	LS	1.189	2.890	0.000	0.081	0.141	0.927	0.152
<i>Fréchet</i>	MOM	2.971	0.000	4.824	0.044	0.080	0.985	0.058
	MLE	2.959	0.000	4.690	0.041	0.075	0.984	0.058
	LS	2.899	0.000	3.484	0.047	0.078	0.972	0.103
<i>Weibull</i>	MOM	3.880	0.000	3.367	0.094	0.151	0.891	0.183
	MLE	3.879	0.000	3.191	0.091	0.152	0.891	0.179
	LS	3.906	0.000	3.231	0.096	0.156	0.877	0.188

Idéntico análisis es realizado con los datos registrados por la boya LP-II, donde los parámetros de las distribuciones de *Gumbel*, *Fréchet* y *Weibull* se muestran en la tabla 3.9. La distribución de *Fréchet* ofrece los valores mínimos de RMSE y NRMSE, así como los valores máximos de CD, de forma que puede concluirse que esta es la distribución más adecuada para caracterizar los valores máximos anuales. Comparando los valores de KS mostrados en la tabla con el valor teórico correspondiente a  $n=11$  de 0.391, las tres distribuciones podrían ser aceptadas como candidatas a la modelización de los valores máximos anuales.

Atendiendo a los gráficos cuantil-cuantil mostrados en la figura 3.13 se confirma que la distribución de *Fréchet* es la más adecuada de entre las tres distribuciones extremales. La distribución es capaz de caracterizar adecuadamente a la mayoría de los datos, a excepción de los dos más altos, que parecen desviarse del comportamiento general de la serie. Los ajustes obtenidos, cuando se utilizan los métodos MOM y MLE en la estima de los parámetros, parecen caracterizar mejor la parte central de la serie, observándose una desviación importante en los dos datos más altos. Sin embargo, la distribución, cuando se utiliza el método LS, es capaz de modelizar mejor los dos datos más extremos, si bien esto es a costa de una pequeña desviación en la parte central de la serie. De esta forma, se observa como el método

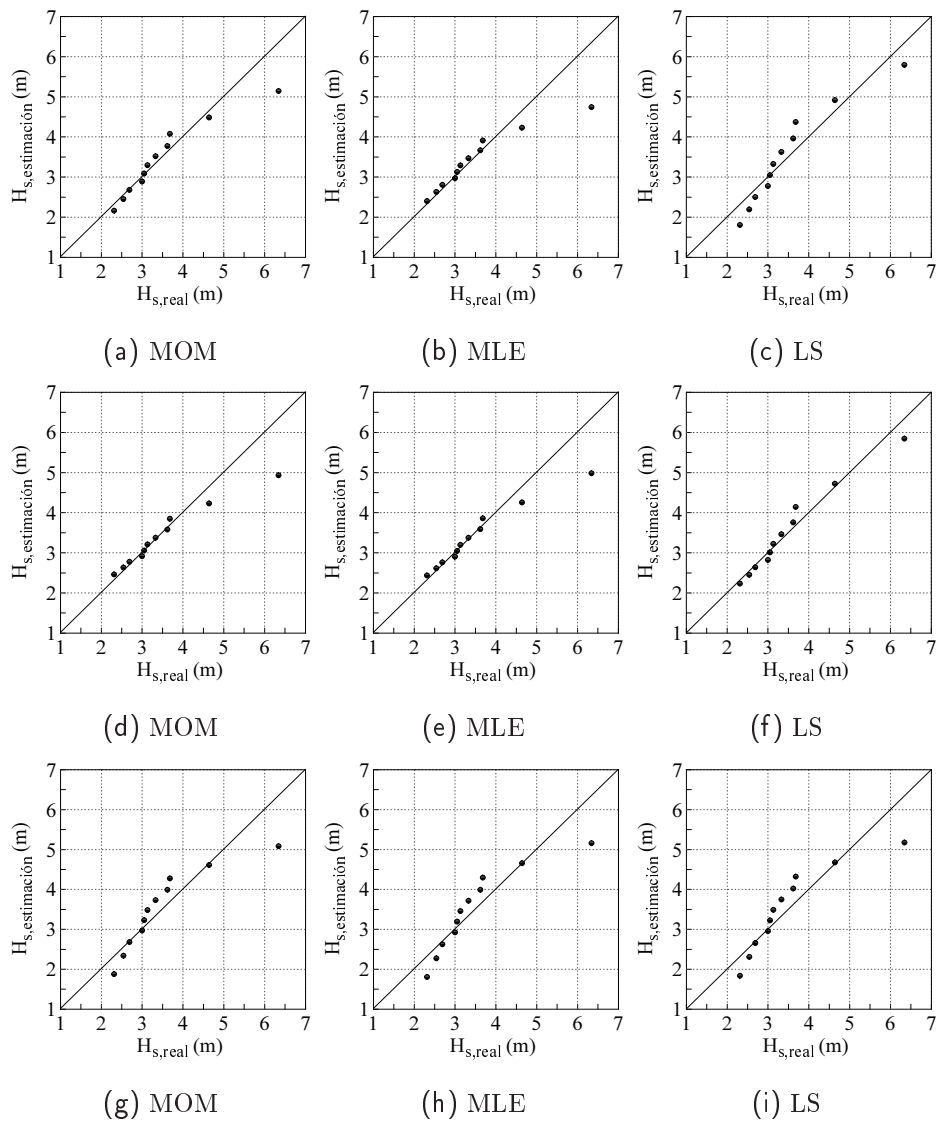


Figura 3.13: Método de máximos anuales para la boya LP-II. Gráficos cuantil-cuantil correspondientes a las funciones de distribución de *Gumbel*(superior), *Fréchet* (centro) y *Weibull* (inferior) utilizando los métodos MOM, MLE y LS.

Tabla 3.10: Método de máximos anuales. Boya LP-II. Cuantiles asociados a periodos de retorno  $T_r = 50, 100, 500$  y  $1000$  años.

<i>Función</i>	Estimador	$H_{50}$	$H_{100}$	$H_{500}$	$H_{1000}$
<i>Gumbel</i>	MOM	6.443	7.064	8.499	9.116
	MLE	5.767	6.256	7.385	7.871
	LS	7.531	8.362	10.281	11.107
<i>Fréchet</i>	MOM	6.672	7.711	10.773	12.438
	MLE	6.801	7.893	11.134	12.909
	LS	8.884	10.85	17.246	21.044
<i>Weibull</i>	MOM	5.818	6.107	6.675	6.888
	MLE	5.948	6.260	6.876	7.108
	LS	5.957	6.265	6.875	7.103

LS se encuentra influenciado por los datos más extremos en mayor grado que los métodos MOM y MLE.

Los cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100, 500 y 1000 años para las distribuciones de Gumbel, Fréchet y Weibull se muestran en la tabla 3.10. De acuerdo con los resultados obtenidos a través de los gráficos cuantil-cuantil, los métodos MOM y MLE subestiman los cuantiles asociados a largos periodos de retorno, debido a que no son capaces de predecir los dos valores más extremos de la serie.

Por otro lado, la serie de máximos anuales se modeliza a través de las distribuciones GEV, K4P, GAM y WAK utilizando el método LMOM en la estimación de los parámetros, mostrando los resultados en la tabla 3.11.

El primer comentario a destacar es que el método LMOM no converge en la estimación de los parámetros de la distribución K4P, de forma que los valores de  $\tau_3$  y  $\tau_4$  obtenidos son inconsistentes con una distribución K4P con valor de  $h > -1.0$ . Esto puede ser debido a la presencia de los dos extremos más altos de la serie, que presentan valores significativamente superiores al resto de extremos. Eliminando estos dos puntos y volviendo a realizar el ajuste para un valor de  $n=9$  el método converge adecuadamente. Los valores mínimos de RMSE y NRMSE se obtienen con la distribución WAK, mostrando los gráficos cuantil-cuantil de la figura 3.14 ajustes muy similares entre las distribuciones GEV y WAK.

En ambos casos, los dos extremos de mayor valor no son modelizados correctamente, por lo que puede concluirse que, en este caso, la introducción de dos parámetros adicionales no mejora significativamente el ajuste respecto a la distribución de Fréchet.

Los cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100, 500 y 1000 años se muestran en la tabla 3.12. Los cuantiles predichos por la distribución K4P no son significativos, al haber eliminado los dos extremos más altos de la serie, mientras que los correspondientes a las distribuciones GEV y WAK son muy similares.

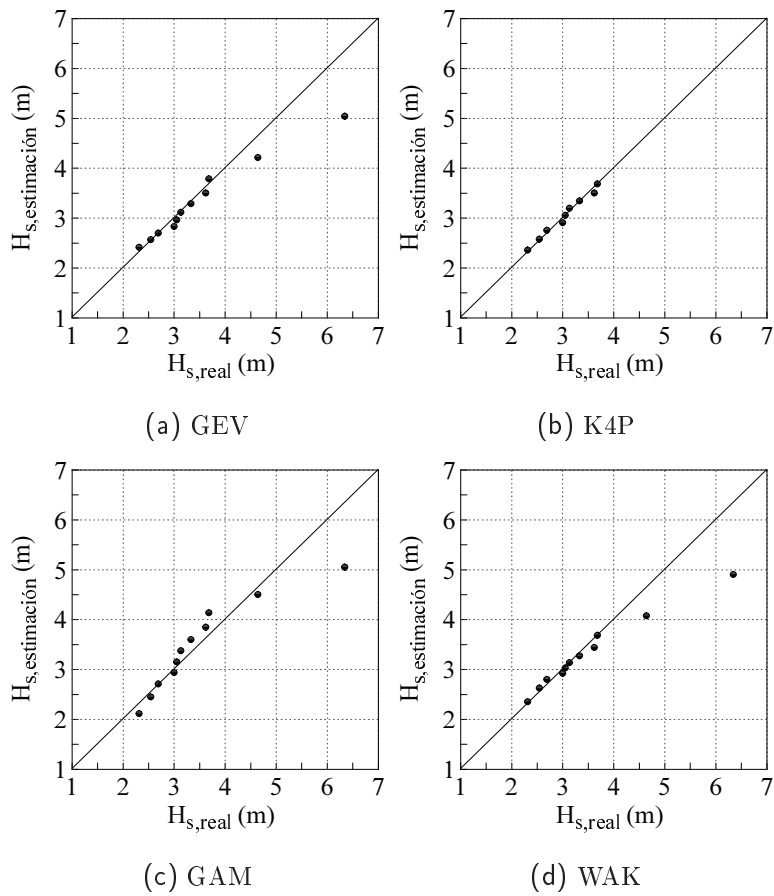


Figura 3.14: Método de máximos anuales para la boya LP-II. Gráficos cuantil-cuantil correspondientes a las funciones de distribución GEV, K4P, GAM y WAK utilizando el método LMOM para la estimación de los parámetros



Tabla 3.11: Método de máximos anuales. Boya LP-II. Parámetros de las funciones de distribución GEV, K4P, GAM y WAK obtenidos con el método LMOM. Valores de RMSE, NRMSE, CD y KS con un nivel de significación de 0.05

<i>Función</i>	$\alpha$	$\xi$	$k$	h		RMSE	NRMSE	CD	KS
GEV	0.593	2.886	-0.308	-		0.045	0.074	0.982	0.103
K4P	0.776	2.722	0.595	0.396		0.038	0.061	0.982	0.065
GAM	0.333	0.000	10.457	-		0.071	0.118	0.937	0.139
	$\alpha$	$\xi$	$\beta$	$\gamma$	$\delta$	RMSE	NRMSE	CD	KS
WAK	1.888	6.661	7.921	0.522	0.384	0.042	0.071	0.989	0.065

Tabla 3.12: Método de máximos anuales. Boya LP-II. Cuantiles asociados a periodos de retorno  $T_r = 50, 100, 500$  y  $1000$  años

<i>Función</i>	$H_{50}$	$H_{100}$	$H_{500}$	$H_{1000}$
GEV	7.364	8.899	14.003	17.106
K4P	3.898	3.942	3.994	4.005
GAM	6.036	6.467	7.397	7.776
WAK	7.486	9.354	16.190	20.714

Como conclusión se obtiene que en la boya LP-I la distribución de probabilidad que mejor ajusta la serie de máximos anuales es la distribución K4P, mientras que en la boya LP-II la es la distribución de *Fréchet* biparamétrica.

En función del método de estimación seleccionado pueden derivarse resultados significativamente diferentes, sobre todo si existen datos que presentan valores que se desvían del comportamiento general de la serie. La introducción de distribuciones de probabilidad con un mayor número de parámetros libres puede mejorar el ajuste, sobre todo en la parte alta de la serie, aunque esto se logra en ocasiones a costa de una pequeña desviación en la parte central de la serie.

Por otro lado, los cuantiles asociados a periodos de retorno elevados son significativamente menores en la boya LP-II, de acuerdo con las condiciones climatológicas de la zona.

### 3.10.2 Aplicación del método POT

En primer lugar se procede al análisis de la frecuencia de presentación de los eventos extremos a través del modelo de Poisson y de la distribución binomial negativa. Posteriormente se procede al análisis de la magnitud de los eventos extremos a través de los modelos GPD y K4P.

#### Selección de eventos extremos

Utilizando el método descrito en Simiu (1996) se obtiene una serie de 1191 valores extremos para la boya LP-I y 883 valores extremos para la boya LP-II. Este método tiene como objetivo obtener una serie de valores extremos independientes, separados entre sí por un espacio temporal que debe ser seleccionado inicialmente. Asumiendo que el tiempo medio de duración de las perturbaciones atmosféricas causantes del oleaje está próximo a 96 horas, se seleccionan los picos que estén separados como mínimo la mitad de esa duración media, es decir, 48 horas. En el caso de la boya LP-I la elección de este valor se ve en parte justificada por su función de autocorrelación, mostrada en la figura 3.15(a), la cual toma un valor próximo a 0.15 para un desfase temporal de 48 horas. Sin embargo, en el caso de la boya LP-II, este valor de 48 horas debe tomarse con cautela, ya que el valor de la función de autocorrelación para tal desfase temporal aumenta hasta 0.3, tal y como se ilustra en la figura 3.16(a).

La independencia estadística de los valores extremos dentro de cada serie temporal es analizada a través del *run test*,  $Z$ , mostrado en la figura 3.15(b), para LP-I y 3.16(b) para LP-II, sobre diferentes valores del umbral. Estos gráficos sugieren que a partir de un valor umbral de 2.0 metros en la boya LP-I y 2.5 metros en la boya LP-II, es posible aceptar la independencia estadística.

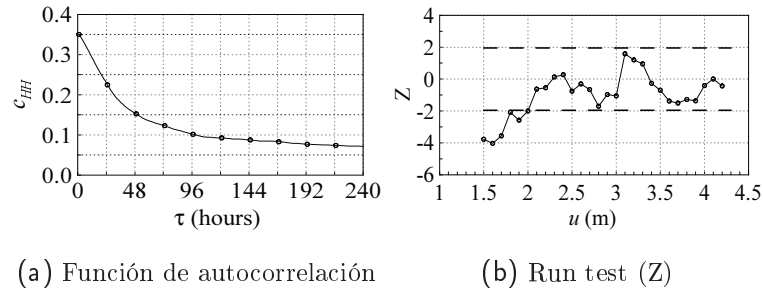


Figura 3.15: Boya LP-I. (a) Función de autocorrelación. (b) Run test ( $Z$ ) para la independencia sobre diferentes valores de  $u$  e intervalos de confianza al 95 % (línea discontinua).

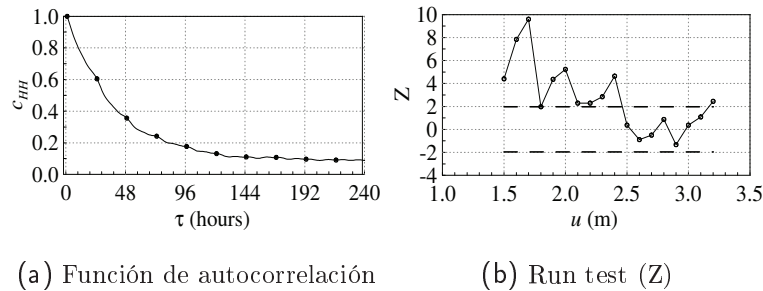


Figura 3.16: Boya LP-II. (a) Función de autocorrelación. (b) Run test ( $Z$ ) para la independencia sobre diferentes valores de  $u$  e intervalos de confianza al 95 % (línea discontinua).

### Análisis de la frecuencia de presentación de eventos extremos

La figura 3.17 muestra la evolución del índice de dispersión, DI, a medida que el valor umbral aumenta. Para la boya LP-I, entre los umbrales de 2.0 y 4.0 metros, el valor del DI se mantiene dentro de los límites de aceptación de la hipótesis nula, es decir, se acepta que el número medio de eventos extremos por año sigue una distribución de Poisson. No obstante, debe comentarse que el valor del DI parece estar más próximo a 1.5 que al valor teórico de 1.0, que debería tomar para la distribución de Poisson. Para la boya LP-II, el valor de DI se mantiene muy próximo al límite de aceptación, entre los umbrales de 1.5 y 2.4. A partir del umbral de 2.5 el DI disminuye y permanece dentro de la región de aceptación de la hipótesis nula. En base a este gráfico, podría concluirse que la distribución de Poisson es adecuada a partir de un umbral de 2.5 metros.

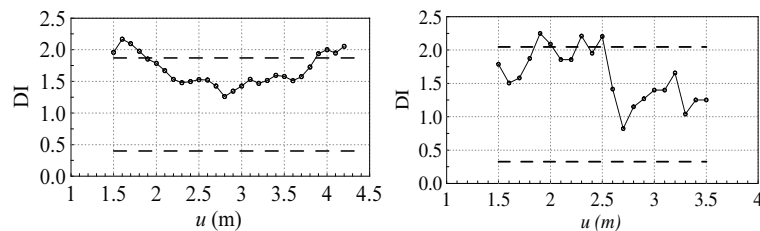


Figura 3.17: Índice de dispersión (DI) para diferentes valores de  $u$  e intervalos de confianza para  $DI=1$  al 95 % (línea discontinua). (a) Boya LP-I. (b) Boya LP-II.

Los gráficos de Poisson son utilizados de igual forma para comprobar tal hipótesis, mostrados en las figuras 3.18 y 3.19. El alto grado de linealidad mostrada en ambos casos viene a corroborar los resultados obtenidos a través del índice de dispersión. Además, se muestra el valor del parámetro  $\lambda$ , equivalente al valor medio de eventos extremos por año. Este valor disminuye a medida que el valor umbral aumenta debido al menor número de temporales contabilizados.

Por otro lado, dado que el índice de dispersión es superior al valor 1 sobre algunos umbrales, se hace uso también de la distribución binomial negativa para caracterizar la variable  $N$ . Las figuras 3.20 y 3.21 muestran el ajuste de la distribución binomial negativa, observándose cierto grado de linealidad, aunque en comparación con los gráficos de Poisson se desprende que no se mejora el ajuste a los datos observados. Por este motivo no resulta procedente utilizar un modelo de frecuencia-intensidad de eventos extremos basado en la distribución binomial negativa para el cálculo de los periodos de retorno. En este sentido, cabe señalar que Onoz y Bayazit (2001) analizando datos de inundaciones, observan que, incluso en aquellos casos donde la distribución binomial negativa ajusta mejor los

datos observados que la distribución de Poisson, los periodos de retorno basados en ambas distribuciones muestran resultados muy similares.

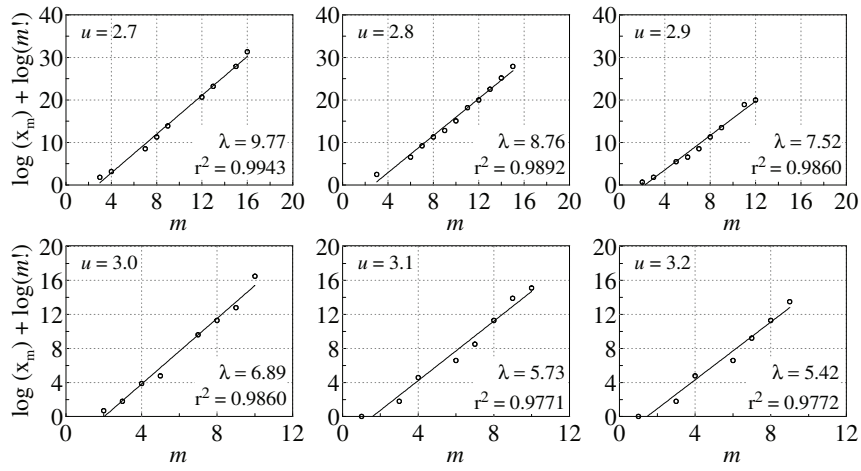


Figura 3.18: Boya LP-I. Gráficos de Poisson para diferentes valores del umbral  $u$

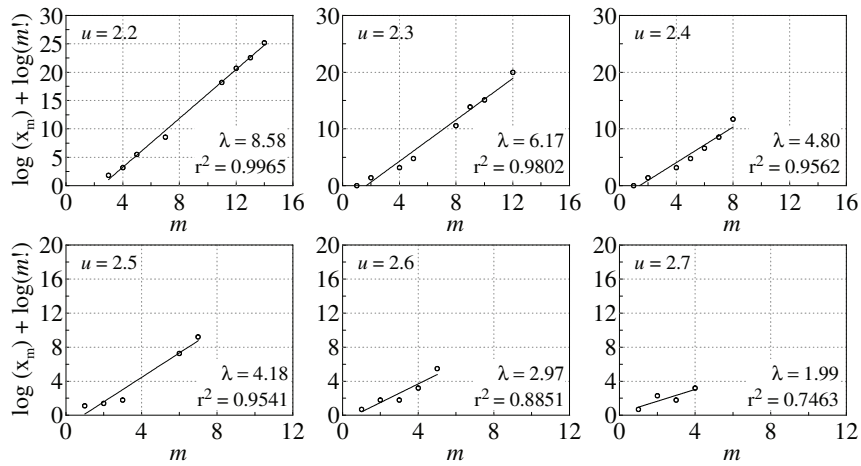


Figura 3.19: Boya LP-II. Gráficos de Poisson para diferentes valores del umbral  $u$

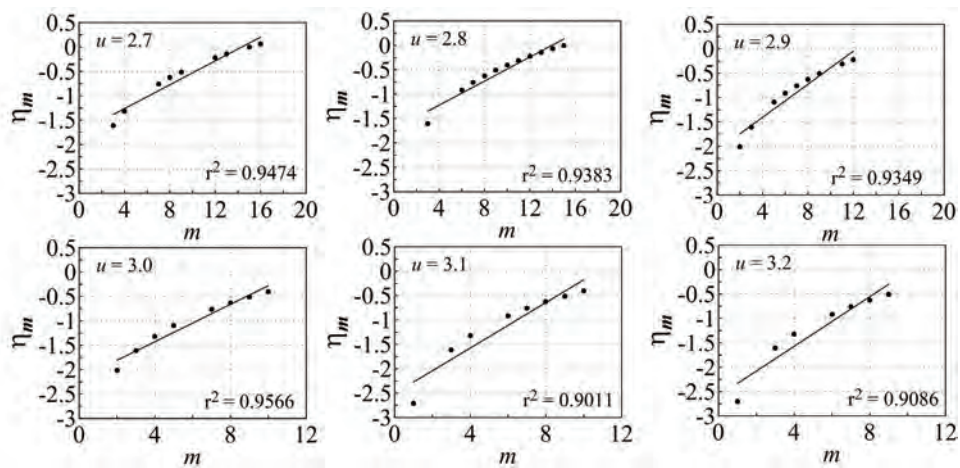


Figura 3.20: Boya LP-I. Ajuste de la distribución binomial negativa sobre diferentes valores del umbral  $u$

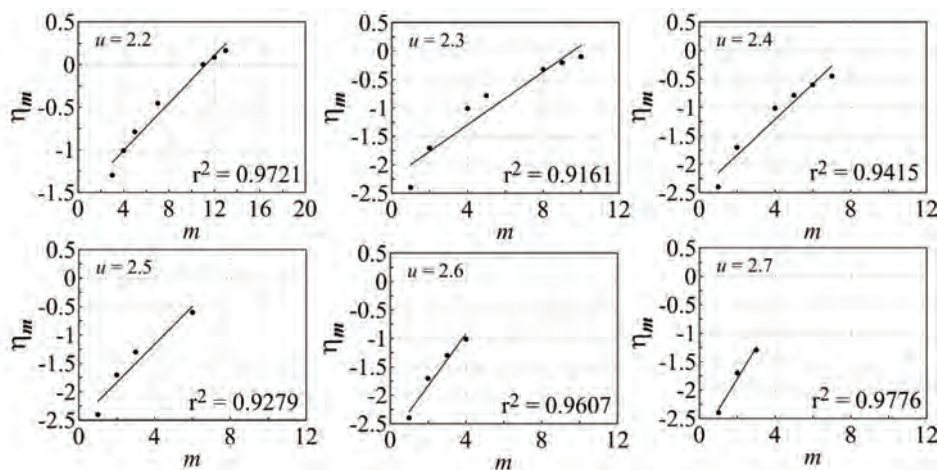


Figura 3.21: Boya LP-II. Ajuste de la distribución binomial negativa sobre diferentes valores del umbral  $u$

### Análisis de la magnitud de los eventos extremos. Modelo Poisson-GPD.

El primer paso será determinar los valores de los umbrales más adecuados, de forma que se garantice la condición de que éste debe ser suficientemente alto para satisfacer las hipótesis del modelo GPD.

La función MEF, dada por la expresión 3.61, se utiliza como primera aproximación para seleccionar el valor umbral, teniendo siempre presente que esta técnica no determina de forma definitiva el umbral más adecuado.

Dado que la función MEF correspondiente al modelo GPD varía linealmente con el valor umbral, se escogerá como valor umbral aquel a partir del cual la función MEF varíe linealmente.

Sin embargo, como ya se ha comentado, cuando se obtiene esta función a partir de los datos muestrales, no siempre es fácil realizar una interpretación adecuada. La figura 3.22 muestra la variación de la función MEF con diferentes valores umbrales para LP-I. En ella se puede apreciar la alta variabilidad de la función, pudiendo identificar hasta tres tramos de comportamiento lineal, con sucesivos cambios de pendiente. Por ello, resulta complicado utilizar este gráfico para la selección del umbral. Sin embargo, en el caso de la serie LP-II, tal y como muestra la figura 3.23, la MEF varía linealmente en la totalidad del rango de umbrales estudiados, por lo que el umbral puede escogerse tan alto como se desee, teniendo en cuenta que el número de excedencias sea suficiente para ofrecer resultados fiables. La variación del parámetro de forma  $k$  de la distribución GPD se muestra en las figuras 3.22(derecha) y 3.23(derecha) para las series LP-I y LP-II respectivamente.

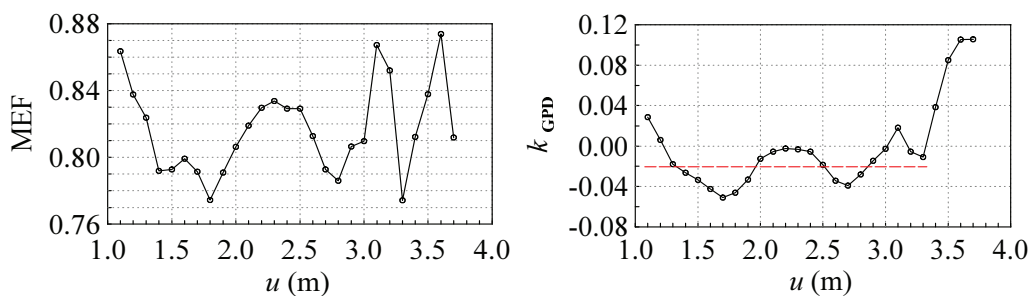


Figura 3.22: Boya LP-I. Función media de exceso para los datos observados (izquierda). Evolución del parámetro de forma,  $k$ , de la distribución GPD con el valor umbral (derecha)

Las tablas 3.13 y 3.14 muestran los valores de los tres parámetros, localización, escala y forma, con sus respectivos intervalos de confianza. En base a las propiedades teóricas

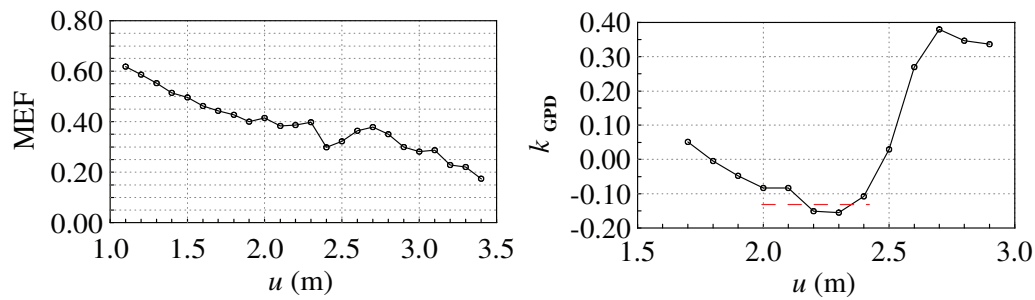


Figura 3.23: Boya LP-II. Función media de exceso para los datos observados (izquierda). Evolución del parámetro de forma,  $k$ , de la distribución GPD con el valor umbral (derecha)

del modelo GPD, el valor del parámetro  $k$  debería permanecer constante sobre aquellos umbrales en los que el modelo caracteriza adecuadamente las excedencias. Aunque parece observarse cierta variación o inestabilidad en las estimaciones del parámetro, es posible asumir cierta estabilidad entorno al valor de  $-0.02$  para LP-I y  $-0.12$  para LP-II, calculado entre los umbrales de  $1.5$  y  $3.2$  metros para LP-I y entre  $2.0$  y  $2.4$  para LP-II, mostrado por la línea discontinua en color rojo. A partir de valores umbrales de  $3.2$  metros para LP-I y  $2.4$  metros para LP-II, el parámetro de forma entra en una región de inestabilidad provocada por el bajo número de excedencias registradas sobre estos umbrales.

En el caso de LP-I, el hecho de que el parámetro de forma esté próximo al valor cero indica que las excedencias podrían ser caracterizadas a través de una distribución exponencial, ya que ésta surge como caso particular de la distribución GPD cuando  $k = 0$ .

Las tablas 3.15 y 3.16 muestran el valor RMSE y la estadística de Kolmogorov-Smirnov (KS) para los diferentes valores umbrales calculada sobre los datos observados  $\hat{KS}$ , y su valor teórico  $KS = c_{0.05}/\sqrt{n}$ . Para el caso de la boya LP-I, a partir de estos dos últimos valores, se puede concluir que el modelo Poisson-GPD no puede ser rechazado, ya que el valor de  $\hat{KS}$  es inferior al valor teórico. Esta conclusión está en consonancia con los gráficos cuantil-cuantil, mostrados en la figura 3.24, donde puede observarse como las alturas de ola predichas por el modelo están muy próximas a las observadas. En el caso de LP-II, sin embargo, solamente a partir del umbral de  $2.3$  podría aceptarse la validez del modelo, ya que para umbrales menores el valor de  $\hat{KS}$  es superior al valor de KS teórico. Los gráficos cuantil-cuantil en la figura 3.25 muestran una ligera desviación de las alturas de ola predichas por el modelo respecto a las observadas, hecho que se acentúa cuando la altura de ola es superior a  $3.0$  metros.



Tabla 3.13: Boya LP-I. Parámetros del modelo GPD estimados con el método LMOM e intervalos de confianza al 95 % según el método de bootstrapping para diferentes umbrales

$u$	$m$	$\xi$	$[\xi_{inf}, \xi_{sup}]$	$\alpha$	$[\alpha_{inf}, \alpha_{sup}]$	$k$	$[k_{inf}, k_{sup}]$
2.5	186	-0.024	[-0.013, 0.065]	0.798	[0.661, 0.980]	-0.020	[-0.115, 0.113]
2.6	168	-0.006	[-0.003, 0.045]	0.779	[0.635, 0.965]	-0.003	[-0.135, 0.100]
2.7	154	-0.013	[-0.050, 0.027]	0.774	[0.616, 0.968]	-0.039	[-0.144, 0.099]
2.8	137	-0.031	[-0.071, 0.011]	0.794	[0.626, 0.996]	-0.028	[-0.140, 0.116]
2.9	118	-0.022	[-0.064, 0.024]	0.817	[0.628, 1.051]	-0.014	[-0.138, 0.141]
3.0	104	-0.028	[-0.077, 0.020]	0.836	[0.635, 1.095]	-0.003	[-0.131, 0.155]
3.1	86	0.012	[-0.049, 0.086]	0.866	[0.629, 1.168]	0.018	[-0.133, 0.201]
3.2	78	0.018	[-0.040, 0.077]	0.829	[0.578, 1.158]	-0.006	[-0.169, 0.186]

Tabla 3.14: Boya LP-II. Parámetros del modelo GPD estimados con el método LMOM con intervalos de confianza al 95 % según el método de bootstrapping para diferentes umbrales

$u$	$m$	$\xi$	$[\xi_{inf}, \xi_{sup}]$	$\alpha$	$[\alpha_{inf}, \alpha_{sup}]$	$k$	$[k_{inf}, k_{sup}]$
2.1	118	0.085	[0.066, 0.105]	0.301	[0.238, 0.386]	-0.083	[-0.191, 0.089]
2.2	97	-0.033	[-0.052, -0.016]	0.268	[0.200, 0.367]	-0.151	[-0.261, 0.018]
2.3	71	-0.040	[-0.066, -0.019]	0.275	[0.188, 0.419]	-0.155	[-0.293, 0.048]
2.4	51	-0.052	[-0.084, -0.025]	0.312	[0.190, 0.547]	-0.107	[-0.285, 0.191]
2.5	36	-0.071	[-0.111, -0.037]	0.404	[0.214, 0.749]	0.029	[-0.237, 0.435]
2.6	25	-0.077	[-0.132, -0.016]	0.559	[0.284, 1.041]	0.269	[-0.063, 0.810]
2.7	19	-0.052	[-0.139, 0.047]	0.595	[0.278, 1.225]	0.380	[-0.015, 1.074]

Tabla 3.15: Boya LP-I. Parámetros del modelo GPD, Error cuadrático medio (RMSE), estadístico de Kolmogorov-Smirnov para los datos observados,  $\widehat{KS}$ , y valor teórico del test,  $KS$  al 95 % de confianza

$u$	$m$	$\xi$	$\alpha$	$k$	RMSE	$\widehat{KS}$	$KS_{teor}$
2.5	186	-0.024	0.798	-0.020	0.020	0.038	0.099
2.6	168	-0.006	0.779	-0.003	0.018	0.043	0.105
2.7	154	-0.013	0.774	-0.039	0.020	0.047	0.110
2.8	137	-0.031	0.794	-0.028	0.023	0.051	0.116
2.9	118	-0.022	0.817	-0.014	0.029	0.055	0.125
3.0	104	-0.028	0.836	-0.003	0.034	0.057	0.133
3.1	86	0.012	0.866	0.018	0.039	0.065	0.146
3.2	78	0.018	0.829	-0.006	0.039	0.056	0.154

Tabla 3.16: Boya LP-II. Parámetros del modelo GPD, Error cuadrático medio (RMSE) y estadístico de Kolmogorov-Smirnov para los datos observados,  $\widehat{KS}$  y valor teórico del test,  $KS$  al 95 % de confianza

$u$	$m$	$\xi$	$\alpha$	$k$	RMSE	$\widehat{KS}$	$KS_{teor}$
2.2	97	-0.033	0.268	-0.151	0.054	0.149	0.138
2.3	71	-0.040	0.275	-0.155	0.056	0.144	0.161
2.4	51	-0.052	0.312	-0.107	0.060	0.135	0.190
2.5	36	-0.071	0.404	0.029	0.060	0.135	0.226
2.6	25	-0.077	0.559	0.269	0.043	0.100	0.272

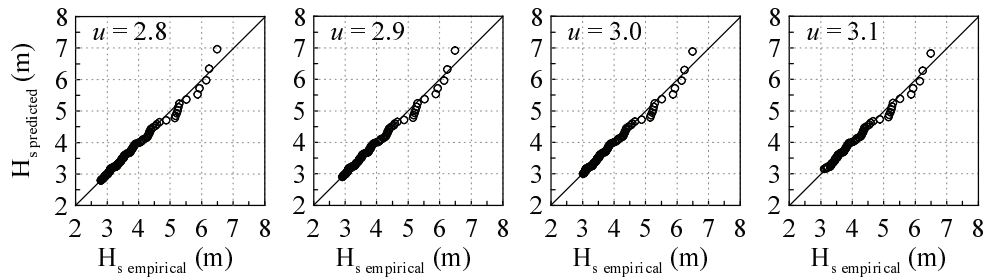


Figura 3.24: Boya LP-I. Gráficos cuantil-cuantil correspondientes a la función GPD sobre diferentes valores del umbral

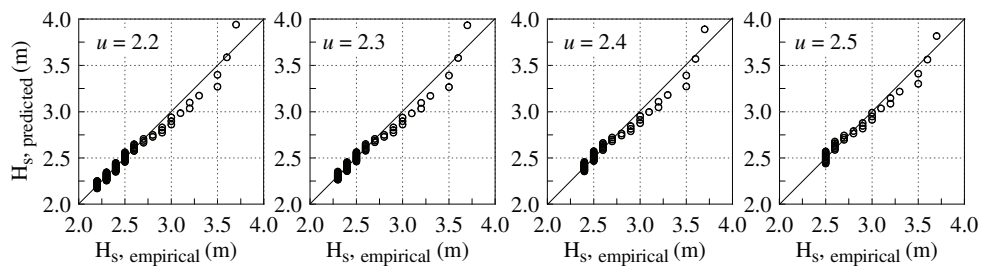


Figura 3.25: Boya LP-II. Gráficos cuantil-cuantil correspondientes a la función GPD sobre diferentes valores del umbral

En la tabla 3.17, para la boya LP-I, y en la tabla 3.18 para la boya LP-II se muestran las alturas de ola asociadas a periodos de retorno de 50, 100 y 500 años obtenidas a través de la relación 3.68, con sus respectivos intervalos de confianza. Esta información es además presentada de forma gráfica en las figuras 3.26 y 3.27. Para el caso de LP-I, estos resultados sugieren que las estimaciones de los cuantiles asociados a los periodos de retorno seleccionados se muestran poco sensibles respecto al umbral elegido.

Sin embargo, para el caso de LP-II, se evidencia una ligera tendencia de las alturas de ola a aumentar con el umbral, acentuándose este aumento entre los umbrales de 2.1 y 2.3 metros, para posteriormente volver a disminuir.

Tabla 3.17: Boya LP-I. Cuantiles asociados a periodos de retorno de 50, 100 y 500 años con intervalos de confianza al 95 %

$u$	$x_{50}$	$x_{100}$	$x_{500}$
2.5	9.72 [7.33, 12.45]	10.37 [7.58, 13.78]	11.94 [8.05, 17.23]
2.6	10.07 [7.50, 13.08]	10.80 [7.74, 14.58]	12.55 [8.28, 18.65]
2.7	10.27 [7.66, 13.44]	11.02 [7.93, 15.01]	12.85 [8.46, 19.27]
2.8	10.18 [7.60, 13.39]	10.88 [7.82, 15.03]	12.56 [8.26, 19.04]
2.9	10.07 [7.53, 13.37]	10.71 [7.72, 14.85]	12.22 [8.16, 18.86]
3.0	9.97 [7.47, 13.34]	10.57 [7.66, 14.85]	11.94 [8.03, 18.80]
3.1	9.78 [7.34, 13.36]	10.29 [7.48, 14.80]	11.46 [7.79, 18.80]
3.2	10.25 [7.54, 14.49]	10.86 [7.69, 16.27]	12.27 [7.98, 21.32]

Tabla 3.18: Boya LP-II. Cuantiles asociados a periodos de retorno de 50, 100 y 500 años con intervalos de confianza al 95 %

$u$	$x_{50}$	$x_{100}$	$x_{500}$
2.2	6.60 [4.63, 8.81]	7.29 [4.82, 10.28]	9.19 [5.26, 14.69]
2.3	6.91 [4.72, 9.33]	7.64 [4.91, 11.00]	9.68 [5.29, 16.58]
2.4	6.54 [4.32, 9.43]	7.09 [4.39, 11.02]	8.53 [4.57, 16.42]
2.5	5.41 [3.94, 8.37]	5.62 [3.96, 9.43]	6.11 [3.98, 13.11]
2.6	4.38 [3.64, 6.11]	4.42 [3.64, 6.47]	4.48 [3.64, 7.34]

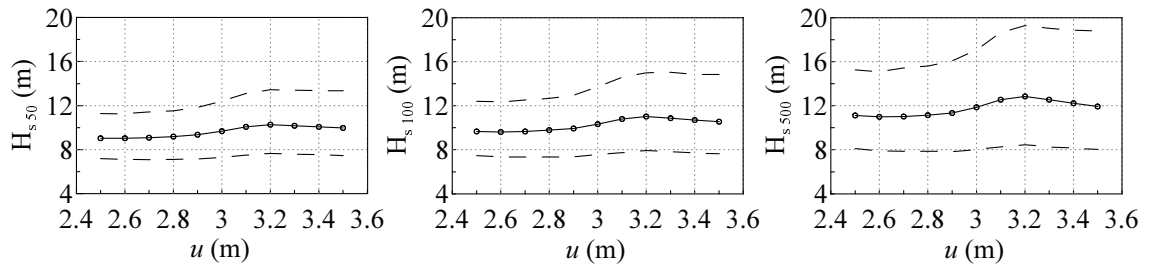


Figura 3.26: Boya LP-I. Variación de los cuantiles  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con diferentes valores del umbral. Intervalos de confianza al 95 % de confianza

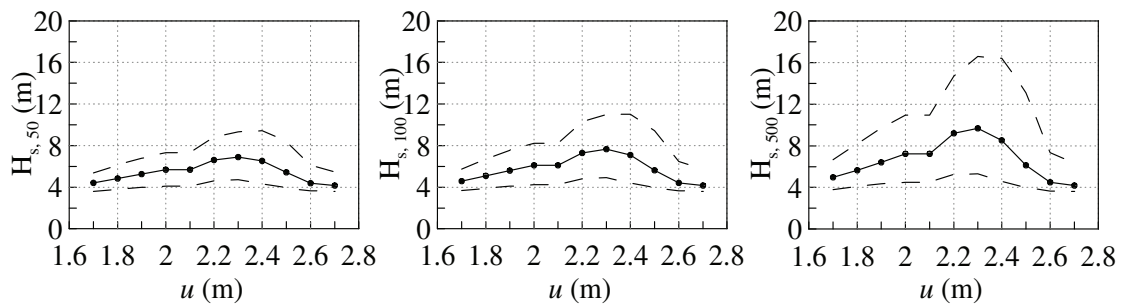


Figura 3.27: Boya LP-II. Variación de los cuantiles  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con diferentes valores del umbral. Intervalos de confianza al 95 % de confianza

La relación entre los  $L$ -moments de tercer y cuarto orden, para el modelo GPD y para los datos observados, se muestra en la figura 3.28. En el caso de la boya LP-I, el rango de umbrales más próximo a la relación teórica es el comprendido entre 2.0 y 2.6 metros, indicando este hecho que el modelo GPD podría ser adecuado para caracterizar las excedencias sobre este rango de umbrales. A partir del umbral 2.6 y hasta 3.1 metros, la diferencia entre los valores observados y los teóricos es algo mayor, aunque ésta no supera el valor de 0.01. Por ello, podría aceptarse que el modelo es adecuado sobre un umbral de 3.0 metros.

En el caso de la boya LP-II, los umbrales que se muestran más próximos a la curva teórica son los comprendidos entre 1.8 y 2.0 metros. A partir de un umbral de 2.1 metros, las diferencias entre los momentos observados y teóricos va en aumento, con la salvedad del umbral 2.6, donde la diferencia disminuye. Cabe destacar que las diferencias entre los valores observados y los teóricos a partir del umbral 2.1 son considerables, del orden de hasta 0.4 unidades. Por ello, se desprende de la figura que el modelo GPD podría aplicarse para modelizar las excedencias sobre un umbral de 2.0 metros, no siendo aconsejable utilizarlo con umbrales superiores.

Sin embargo, esto contradice la teoría, que supone que si el modelo es adecuado sobre un umbral determinado, lo será también para umbrales superiores. La explicación a esta contradicción radica en la escasa fiabilidad de las estimaciones realizadas sobre  $\tau_3$  y  $\tau_4$ , ya que han sido obtenidas con un número muy pequeño de excedencias.

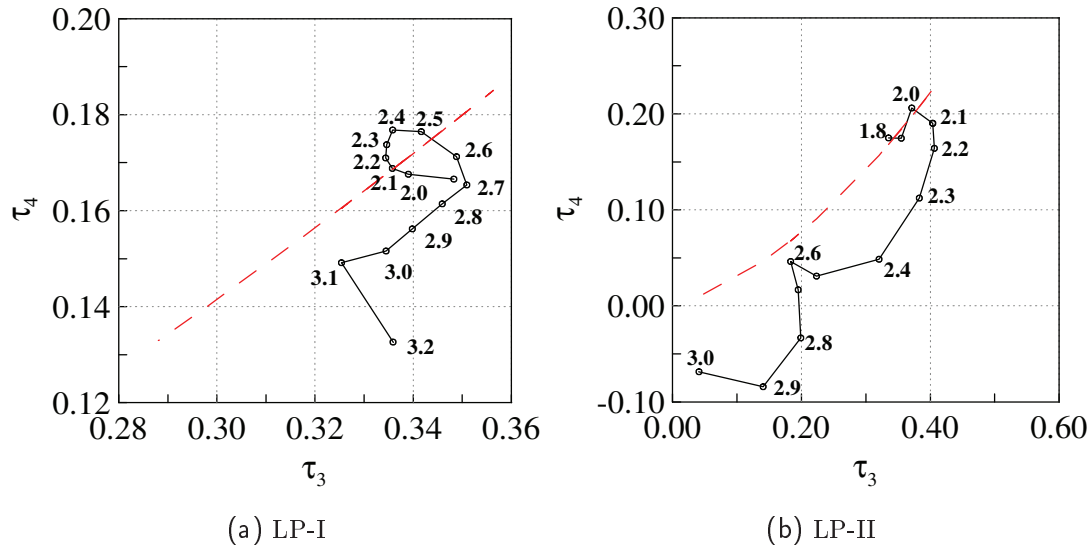


Figura 3.28: Relación entre los  $L$ -moments de tercer  $\tau_3$  y cuarto  $\tau_4$  orden de la distribución GPD. La línea continua representa la relación  $\tau_3 - \tau_4$  para los datos observados mientras que la línea discontinua representa idéntica relación teórica dada por la expresión (3.63)

La tabla 3.19 recoge los valores umbrales propuestos por los diferentes métodos basados en el número de excedencias, alternativa a los métodos basados en las propiedades teóricas del modelo. Se pone de manifiesto el alto grado de variabilidad existente, estando el rango de umbrales para la boya LP-I entre 2.2 y 3.5, y entre 2.0 y 2.4, para la boya LP-II.

Para el caso de los datos de la boya LP-I, un rango de umbrales aceptable podría estar entre 2.8 y 3.1 metros. En este rango, se cumple la hipótesis de independencia estadística, analizada a través de *run test*. La hipótesis de Poisson, analizada a través del índice de dispersión y los gráficos de Poisson, también se satisface. El parámetro de forma de la distribución GPD parece mostrar cierta estabilidad en este rango, mientras que el gráfico de  $L$ -moments, muestra una proximidad notable entre los valores observados y los teóricos. Finalmente, el número de excedencias generadas queda entre los números propuestos por Boos(1984) y Hasofer (1996).

Para el caso de LP-II, la elección del umbral resulta más complicado. Atendiendo a las hipótesis de independencia estadística y Poisson, el umbral debería ser elegido a partir de un valor de 2.4 metros. Sin embargo, a partir de este umbral, el parámetro

Tabla 3.19: Selección del valor umbral según diferentes métodos basados en el número de excedencias

	LP-I		LP-II	
	$m$	$u$	$m$	$u$
<b>Basado en propiedades teóricas</b>				
Ashkar y Rouselle (1987)	270	2.2	248	1.7
Función media de exceso-estabilidad de $k$	240	2.3	118 - 51	2.0 - 2.4
Relación $\tau_3 - \tau_4$	355 - 86	2.0 - 3.1	199 - 118	1.8 - 2.1
<b>Basado en el número de excedencias</b>				
Boos (1984)	118	2.8	88	2.2
Gross (1994)	154	2.7	110	2.1
Hasofer (1996)	55	3.5	45	2.4

Tabla 3.20: Boya LP-I. Cuantiles de  $H_s$  asociados a periodos de retorno de 100 y 500 años para diferentes ventanas temporales sobre un umbral de 2.7 metros

Ventana temporal (días)	$H_{100}$	$H_{500}$
3	8.08	9.22
4	8.45	9.90
5	8.13	9.25
6	7.88	8.77
7	8.05	9.05
8	7.99	8.97

de forma de la distribución GPD muestra una alta variabilidad, mientras que el gráfico de  $L$ -moments muestra una diferencia considerable entre los datos observados y teóricos. Además, el número de excedencias sobre el umbral de 2.5 es de tan solo 36, por lo que la fiabilidad estadística de las estimaciones se reduce considerablemente. Por todo esto, sería aconsejable reducir el valor umbral a un rango comprendido entre 2.0 y 2.3 metros, perdiendo algo de fiabilidad en la independencia estadística, a costa de ganar en el ajuste de la distribución GPD.

Como ya se ha comentado, el valor de 96 horas que determina la longitud de la ventana temporal, no debe interpretarse como un valor fijo o definitivo. Sin embargo, se comenta en Simiu (1996) que el tamaño de esta ventana no tiene efectos importantes sobre los resultados finales del modelo. Con el objetivo de evaluar este efecto de la longitud de la ventana temporal sobre los cuantiles asociados a determinados periodos de retorno, se realiza el análisis sobre series de valores extremos obtenidos con diferentes longitudes de ventana temporal para la boya LP-I.

La tabla 3.20 muestra los cuantiles asociados a periodos de retorno de 100 y 500 años sobre un umbral de 2.7 metros para diferentes ventanas temporales. El efecto de la variación de la longitud de esta ventana no parece tener efectos significativos sobre los cuantiles de  $H_s$ .

### **Modelo Poisson-K4P. Análisis de la magnitud de las excedencias.**

Con el objetivo de evaluar si la distribución K4P es capaz de mejorar el ajuste de los eventos extremos ofrecido por la distribución GPD, el modelo *Poisson-K4P* se aplica a los datos registrados por la boya LP-I y la boya CO, mostrando en la figura 3.29 la variación de los parámetros de forma  $k$  y  $h$ , sobre diferentes umbrales, de la distribución K4P. Estos valores se incluyen en la tabla 3.21 y 3.22 conjuntamente con los correspondientes valores del parámetro de forma  $k$  de la distribución GPD. Además, se muestra el valor del test de Kolmogorov-Smirnov para decidir sobre la idoneidad del modelo, concluyendo que para los dos registros analizados, ambos modelos, GPD y K4P, son aceptados para la modelización de la magnitud de los eventos extremos.

Para la boya LP-I, los gráficos cuantil-cuantil sobre diferentes valores umbrales se muestran en las figuras 3.30 y 3.31 para ambas distribuciones. La distribución GPD ajusta adecuadamente el conjunto de datos, con una pequeña desviación para alturas entre 4.5 y 5.5 metros. Esta desviación se aprecia en todos los umbrales seleccionados, salvo para



el umbral 3.1 donde prácticamente desaparece. Sin embargo, para los mismos umbrales la distribución K4P ofrece mejores ajustes reduciéndose de forma significativa las desviaciones existentes en el rango 4.5 - 5.5 metros. De este modo, sobre los umbrales 2.8, 2.9 y 3.0, la mejora en el ajuste resulta especialmente significativa. Sobre el umbral 3.1 los ajustes proporcionados por ambas distribuciones son muy similares, dado que el parámetro  $h$  se aproxima al valor 1.0, lo que significa que la distribución K4P tiende a una distribución GPD.

De igual forma, para la boya CO, puede observarse en las figuras 3.32 y 3.33 la mejora obtenida cuando se utiliza el modelo combinado *Poisson-K4P*. Especialmente la mejora se experimenta en el ajuste de los extremos comprendidos entre los umbrales de 5.5 y 5.9 metros. A partir de 6.0 metros los ajustes obtenidos por ambos modelos son prácticamente idénticos, debido una vez más a que el parámetro  $h$  se aproxima al valor 1.0, tendiendo la distribución K4P a una GPD. Cabe destacar el hecho de que el modelo presente en todos los casos una pequeña desviación en el ajuste de los extremos más altos.

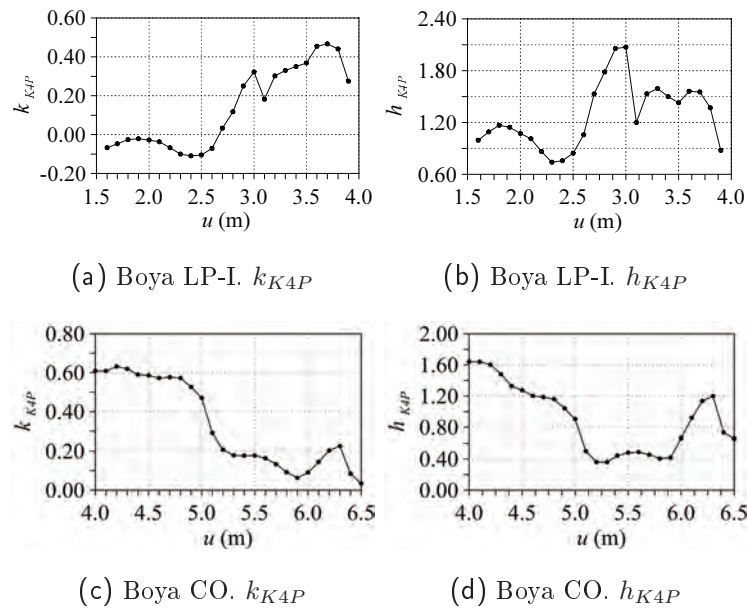


Figura 3.29: Variación de los parámetros de forma de las distribuciones GPD y K4P sobre diferentes valores umbrales

Tabla 3.21: Boya LP-I. Parámetros de las distribuciones GPD y K4P, sobre diferentes umbrales, estimados a través del método LMOM, y test de Kolmogorov-Smirnov (KS).  $KS_*$  representa el límite superior de aceptación con un nivel de significación 0.05.  $m$  representa el número de datos sobre el umbral,  $u$ .

		<i>GPD</i>				<i>K4P</i>					
$u$	$m$	$\xi$	$\alpha$	k	$KS$	$\xi$	$\alpha$	k	h	$KS$	$KS_*$
2.5	186	-0.024	0.798	-0.020	0.038	0.042	0.773	-0.032	0.954	0.038	0.099
2.6	168	-0.006	0.779	-0.003	0.043	-0.065	0.844	-0.001	1.108	0.039	0.105
2.7	154	-0.013	0.774	-0.039	0.047	-0.163	0.914	0.028	1.216	0.038	0.110
2.8	137	-0.031	0.794	-0.028	0.051	-0.182	0.936	0.039	1.214	0.041	0.116
2.9	118	-0.022	0.817	-0.014	0.055	-0.181	0.968	0.055	1.217	0.049	0.125
3.0	104	-0.028	0.836	-0.003	0.057	-0.194	0.996	0.070	1.220	0.058	0.133
3.1	86	0.012	0.866	0.018	0.065	-0.099	0.980	0.070	1.154	0.069	0.146
3.2	78	0.018	0.829	-0.006	0.056	-0.428	1.275	0.174	1.512	0.044	0.154

Tabla 3.22: Boya CO. Parámetros de las distribuciones GPD y K4P, sobre diferentes umbrales, estimados a través del método LMOM, y test de Kolmogorov-Smirnov (KS).  $KS_*$  representa el límite superior de aceptación con un nivel de significación 0.05.  $m$  representa el número de datos sobre el umbral,  $u$ .

		<i>GPD</i>				<i>K4P</i>					
$u$	$m$	$\xi$	$\alpha$	k	$KS$	$\xi$	$\alpha$	k	h	$KS$	$KS_*$
5.4	92	0.069	1.389	0.455	0.103	0.514	0.794	0.176	0.443	0.084	0.141
5.5	90	0.014	1.338	0.435	0.109	0.425	0.795	0.177	0.479	0.093	0.143
5.6	86	-0.008	1.267	0.410	0.114	0.378	0.766	0.163	0.487	0.087	0.146
5.7	81	-0.017	1.195	0.386	0.102	0.364	0.710	0.133	0.456	0.099	0.151
5.8	75	-0.006	1.110	0.354	0.111	0.373	0.641	0.091	0.404	0.092	0.157
5.9	69	0.014	1.005	0.306	0.101	0.355	0.599	0.063	0.418	0.086	0.163
6.0	63	0.051	0.867	0.225	0.116	0.247	0.643	0.089	0.663	0.099	0.171
6.1	60	0.021	0.795	0.178	0.118	0.071	0.739	0.145	0.918	0.101	0.175
6.2	57	-0.020	0.750	0.147	0.117	-0.111	0.851	0.203	1.137	0.100	0.180

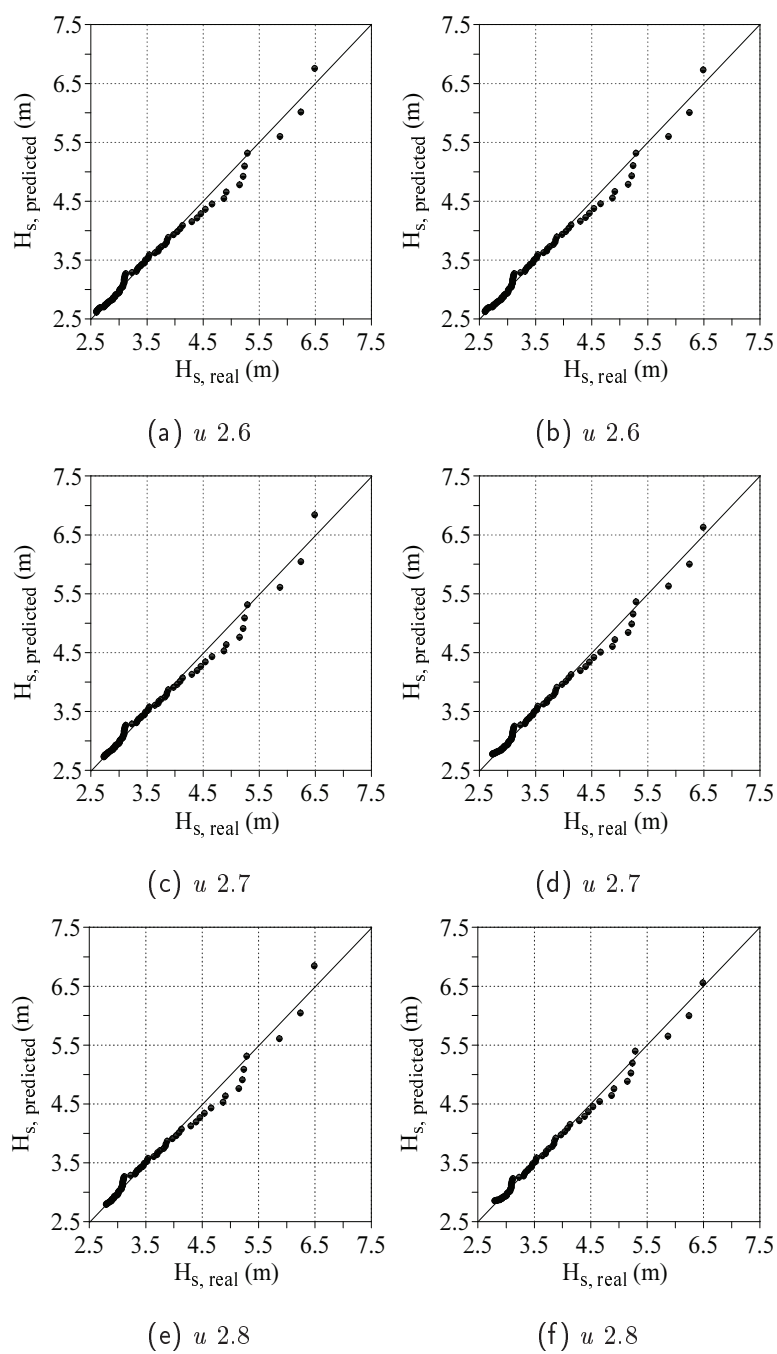


Figura 3.30: Boya LP-I. Gráficos cuantil-cuantil correspondientes a los modelos *Poisson-GPD* (izquierda) y *Poisson-K4P* (derecha) sobre el rango de valores umbrales 2.6 - 2.8 metros

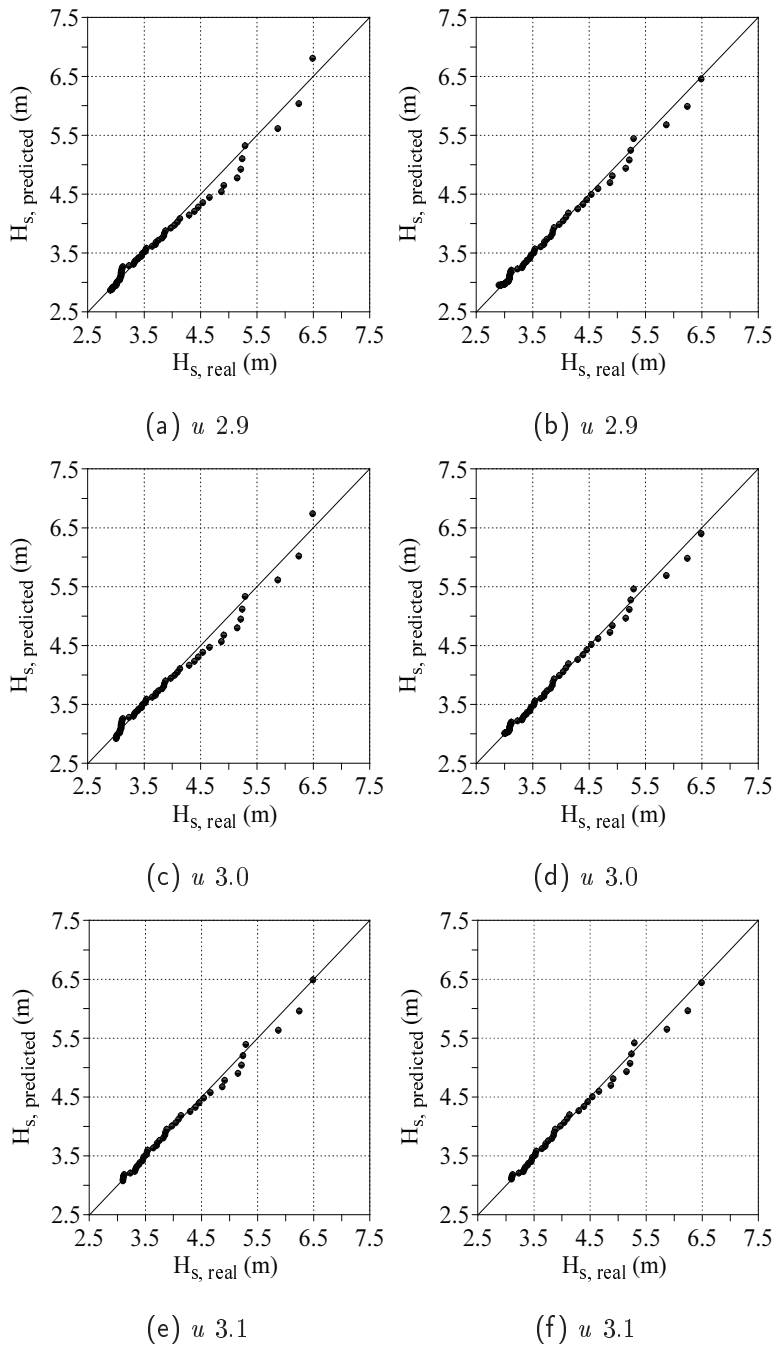


Figura 3.31: Boya LP-I. Gráficos cuantil-cuantil correspondientes a los modelos *Poisson-GPD* (izquierda) y *Poisson-K4P* (derecha) sobre diferentes valores umbrales

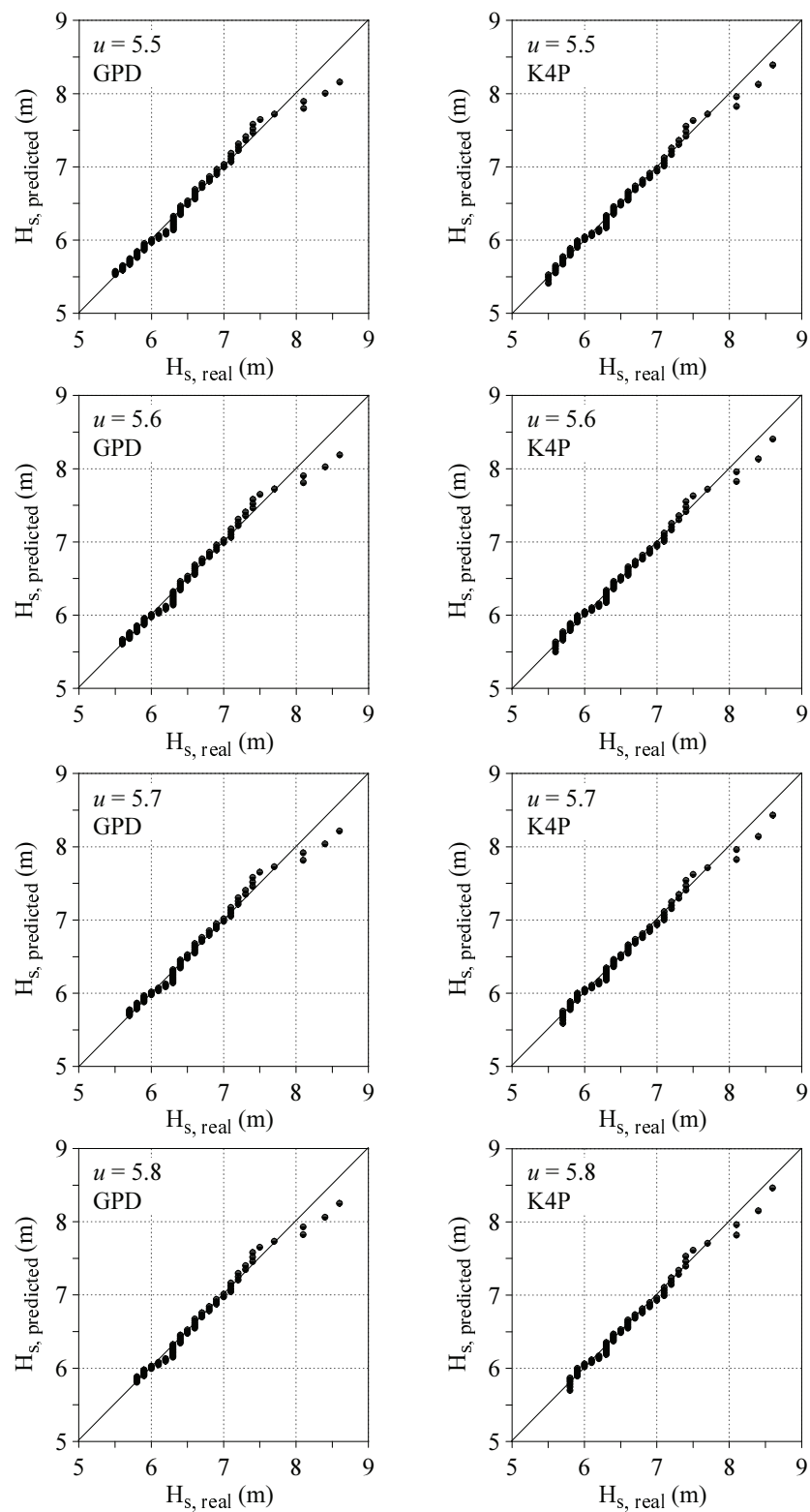


Figura 3.32: Boya CO. Gráficos cuantil-cuantil correspondientes a los modelos *Poisson-GPD* (izquierda) y *Poisson-K4P* (derecha) sobre diferentes valores umbrales

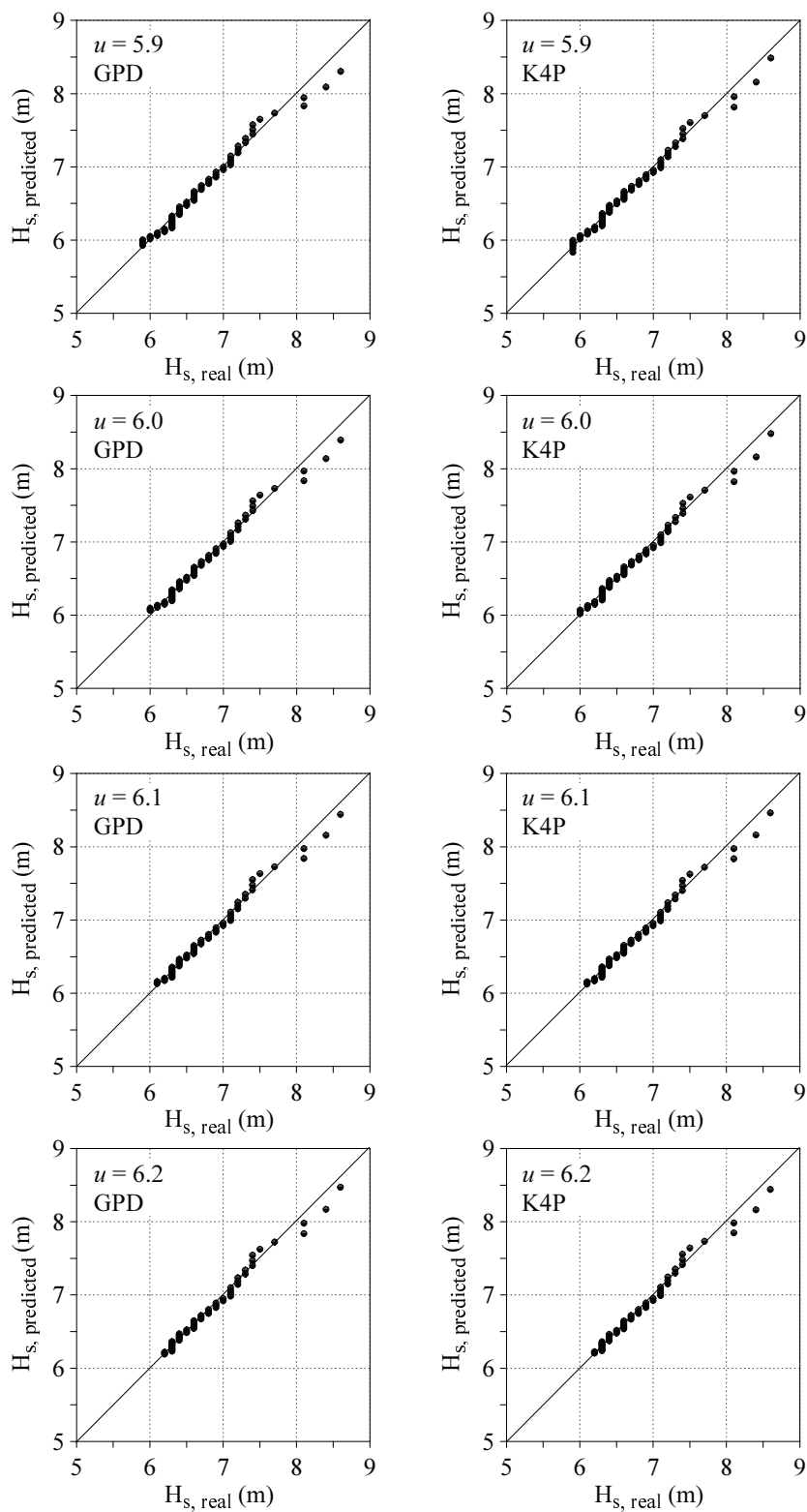


Figura 3.33: Boya CO. Gráficos cuantil-cuantil correspondientes a los modelos *Poisson-GPD* (izquierda) y *Poisson-K4P* (derecha) sobre diferentes valores umbrales

Tabla 3.23: Boya LP-I. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años

$u$	<i>GPD</i>			<i>K4P</i>		
	$H_{50}$	$H_{100}$	$H_{500}$	$H_{50}$	$H_{100}$	$H_{500}$
2.7	10.27	11.02	12.85	8.07	8.59	9.76
2.8	10.18	10.88	12.56	8.00	8.50	9.59
2.9	10.07	10.71	12.22	7.91	8.37	9.38
3.0	9.97	10.57	11.94	7.84	8.27	9.20
3.1	9.78	10.29	11.46	7.84	8.27	9.20

Tabla 3.24: Boya CO. Cuantiles de  $H_s$  asociados a periodos de retorno de 100, 500 y 1000 años

$u$	<i>GPD</i>			<i>K4P</i>		
	$H_{100}$	$H_{500}$	$H_{1000}$	$H_{100}$	$H_{500}$	$H_{1000}$
5.4	8.46	8.49	8.50	9.55	9.76	9.83
5.5	8.52	8.55	8.56	9.55	9.76	9.83
5.6	8.60	8.64	8.65	9.65	9.88	9.97
5.7	8.68	8.72	8.73	9.86	10.15	10.26
5.8	8.79	8.85	8.86	10.20	10.60	10.76
5.9	8.98	9.06	9.09	10.49	10.99	11.19
6.0	9.38	9.53	9.58	10.32	10.74	10.90
6.1	9.68	9.90	9.98	9.95	10.23	10.32
6.2	9.91	10.19	10.30	9.64	9.81	9.87

Para la boya LP-I, la tabla 3.23 muestra los cuantiles asociados a periodos de retorno de 50, 100 y 500, observándose como el modelo combinado *Poisson-GPD* ofrece estimaciones significativamente superiores a las del modelo *Poisson-K4P*. Para la boya CO, la tabla 3.24 muestra los cuantiles asociados a periodos de retorno de 100, 500 y 1000 años, observándose en este caso el efecto contrario, es decir, el modelo combinado *Poisson-GPD* ofrece estimaciones significativamente inferiores a las calculadas mediante el modelo *Poisson-K4P*.

### 3.10.3 Análisis de la estacionalidad

En primer lugar se procede a la aplicación del test estadístico de uniformidad que permita dilucidar si los extremos se distribuyen uniformemente a lo largo del año, en cuyo caso habría que cuestionar la existencia de estacionalidad, o si por el contrario puede rechazarse la hipótesis de uniformidad, pudiendo afirmar entonces la existencia de estacionalidad en los extremos.

Los resultados derivados de la aplicación de los test de la Chi<sup>2</sup>, Rayleigh y Kolmogorov-Smirnov modificado, mostrados en las tablas 3.25, 3.26, 3.27, 3.28, 3.29 y 3.30 revelan claramente la presencia de variaciones estacionales en la ocurrencia de eventos extremos en la boya LP-I, mientras que en la boya LP-II, en la mayoría de los casos, sucede lo contrario, no pudiendo rechazarse la hipótesis de uniformidad.

Este resultado concuerda con lo observado en las figuras 3.9 y 3.10, donde los extremos en la boya LP-I tienden a agruparse en los meses de invierno a medida que aumenta el umbral, mientras que en la boya LP-II la distribución de los extremos tiende a ser homogénea a lo largo del año.

Una vez determinada la estacionalidad desde el punto de vista estadístico, se procede a la separación de las series temporales en periodos homogéneos sobre los cuales se procederá a realizar el análisis de los valores extremos. Destacar el hecho de que, a pesar de que los tests estadísticos muestran una posible no estacionalidad en la boya LP-II, se procede a estudiar la serie en periodos temporales diferenciados, dado que en la figura 3.10 podría aventurarse la posibilidad de un ligero comportamiento diferenciado de los valores extremos a lo largo del año.



Tabla 3.25: Boya LP-I. Resultados del test de la Chi2 para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos.  $Tc = 24.725$  para  $\alpha=0,01$

Hs(m)	N	T	Uniformidad
> 2.4	210	125.657	rechazada
> 2.5	186	120.709	rechazada
> 2.6	168	116.714	rechazada
> 2.7	153	112.490	rechazada
> 2.8	132	100.909	rechazada
> 2.9	116	99.379	rechazada
> 3.0	104	96.076	rechazada
> 3.1	85	91.047	rechazada
> 3.2	77	88.350	rechazada
> 3.3	74	84.270	rechazada
> 3.4	64	75.875	rechazada
> 3.5	55	64.345	rechazada
> 3.6	47	62.021	rechazada

Tabla 3.26: Boya LP-II. Resultados del test de la Chi2 para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos.  $Tc = 24.725$  para  $\alpha=0.01$

Hs(m)	N	T	Uniformidad
1.8	204	21.882	no rechazada
1.9	167	22.053	no rechazada
2.0	122	12.950	no rechazada
2.1	100	9.440	no rechazada
2.2	74	10.648	no rechazada
2.3	54	15.777	no rechazada
2.4	39	14.846	no rechazada
2.5	27	12.555	no rechazada
2.6	21	17.285	no rechazada
2.7	18	16.666	no rechazada
2.8	16	17.000	no rechazada

Tabla 3.27: Boya LP-I. Resultados del test de Rayleigh para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos.  $\alpha=0.01$

Hs(m)	R	P	Uniformidad
> 2.4	0.5307	0.00	rechazada
> 2.5	0.5509	0.00	rechazada
> 2.6	0.5646	0.00	rechazada
> 2.7	0.5811	0.00	rechazada
> 2.8	0.5928	0.00	rechazada
> 2.9	0.6220	0.00	rechazada
> 3.0	0.6431	0.00	rechazada
> 3.1	0.6723	0.00	rechazada
> 3.2	0.6983	0.00	rechazada
> 3.3	0.6959	0.00	rechazada
> 3.4	0.7073	0.00	rechazada
> 3.5	0.7025	0.00	rechazada
> 3.6	0.7411	0.0	rechazada

Tabla 3.28: Boya LP-II. Resultados del test de Rayleigh para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos.  $\alpha=0.01$

Hs(m)	R	P	Uniformidad
1.8	0.1297	0.0643	no rechazada
1.9	0.1299	0.0563	no rechazada
2.0	0.1325	0.1218	no rechazada
2.1	0.1838	0.1663	no rechazada
2.2	0.2493	0.0772	no rechazada
2.3	0.3159	0.0315	no rechazada
2.4	0.3467	0.0177	no rechazada
2.5	0.4868	0.0346	no rechazada
2.6	0.5042	0.0050	rechazada
2.7	0.5524	0.0076	rechazada
2.8	0.5163	0.0051	rechazada

Tabla 3.29: Boya LP-I. Resultados del test modificado de Kolmogorov-Smirnov para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos.  $\alpha = 0.01$ . Valor crítico=1.66

Hs(m)	Max	Min	$V_n$	$V_n(n)^{0.5}$	Uniformidad
> 2.4	0.2071	-0.1357	0.3429	4.9685	rechazada
> 2.5	0.2070	-0.1478	0.3548	4.8394	rechazada
> 2.6	0.2024	-0.1607	0.3631	4.7063	rechazada
> 2.7	0.2141	-0.1552	0.3693	4.5678	rechazada
> 2.8	0.2348	-0.1439	0.3788	4.3519	rechazada
> 2.9	0.2414	-0.1638	0.4052	4.3638	rechazada
> 3.0	0.2308	-0.1827	0.4135	4.2165	rechazada
> 3.1	0.2451	-0.1863	0.4314	3.9771	rechazada
> 3.2	0.2825	-0.1710	0.4535	3.9791	rechazada
> 3.3	0.2905	-0.1577	0.4482	3.8555	rechazada
> 3.4	0.2969	-0.1615	0.4583	3.6667	rechazada
> 3.5	0.3500	-0.1061	0.4561	3.3822	rechazada
> 3.6	0.3670	-0.1312	0.4982	3.4157	rechazada

Tabla 3.30: Boya LP-II. Resultados del test modificado de Kolmogorov-Smirnov para evaluar la uniformidad a lo largo del tiempo de las ocurrencias de eventos extremos.  $\alpha = 0.01$ . Valor crítico = 1.66

Hs(m)	Max	Min	$V_n$	$V_n(n)^{0.5}$	Uniformidad
1.8	0.0882	-0.0049	0.0931	1.3303	no rechazada
1.9	0.1038	-0.0005	0.1043	1.3477	no rechazada
2.0	0.0929	-0.0096	0.1025	1.1317	no rechazada
2.1	0.1067	-0.0033	0.1100	1.1000	no rechazada
2.2	0.1261	-0.0225	0.1486	1.2787	no rechazada
2.3	0.1852	-0.0093	0.1944	1.4289	no rechazada
2.4	0.2308	0.0000	0.2308	1.4412	no rechazada
2.5	0.2222	-0.0370	0.2593	1.3472	no rechazada
2.6	0.2857	-0.0476	0.3333	1.5275	no rechazada
2.7	0.2778	-0.1111	0.3889	1.6499	no rechazada
2.8	0.2708	-0.1667	0.4375	1.7500	rechazada

En primer lugar son analizadas las cuatro estaciones climáticas tradicionales: *primavera*, *verano*, *otoño* e *invierno*, y cuyas fechas de comienzo y fin se muestran en la tabla 3.3.

Las figuras 3.34 y 3.35 muestran el comportamiento de los tres parámetros de la distribución GPD sobre diferentes valores del umbral para las boyas LP-I y LP-II, respectivamente.

En la boya LP-I, el parámetro de forma se muestra relativamente estable sobre diferentes valores del umbral en las estaciones de otoño e invierno, mientras que en las estaciones de primavera y verano se observa un aumento progresivo con el umbral. Este comportamiento se repite para el parámetro de escala, mientras que el parámetro de localización parece establecerse muy próximo a cero.

En la boya LP-II, el parámetro de forma en las estaciones de otoño e invierno se muestra muy estable con el umbral, además de presentar valores próximos entre sí. La estación de primavera, aún mostrando estabilidad respecto al umbral, presenta valores del parámetro de forma sensiblemente superiores a los del resto de estaciones. La estación de verano presenta un valor del parámetro de forma que decrece de forma continua a medida que el valor umbral aumenta. El parámetro de escala presenta un comportamiento similar al mostrado por el parámetro de forma, con las estaciones de primavera, otoño e invierno próximas entre sí y una estación de verano más alejada. El parámetro de localización permanece próximo a cero para la totalidad de estaciones.

Los variación de los valores de los diferentes parámetros de la distribución GPD a lo largo de las estaciones climáticas, ponen de manifiesto la presencia de estacionalidad en la presentación de los valores extremos. En la boya LP-I, esta estacionalidad se manifiesta claramente a partir de un valor umbral de 2.4 metros, existiendo una marcada diferencia entre las cuatro estaciones climáticas. Sin embargo, en la boya LP-II, la estación de primavera se diferencia del resto a lo largo de todos los valores del umbral seleccionados, así como la estación de verano se diferencia del resto a partir de un umbrales de 2.1 metros, aproximadamente. Las estaciones de otoño e invierno muestran comportamientos similares entre sí a lo largo de los umbrales estudiados.

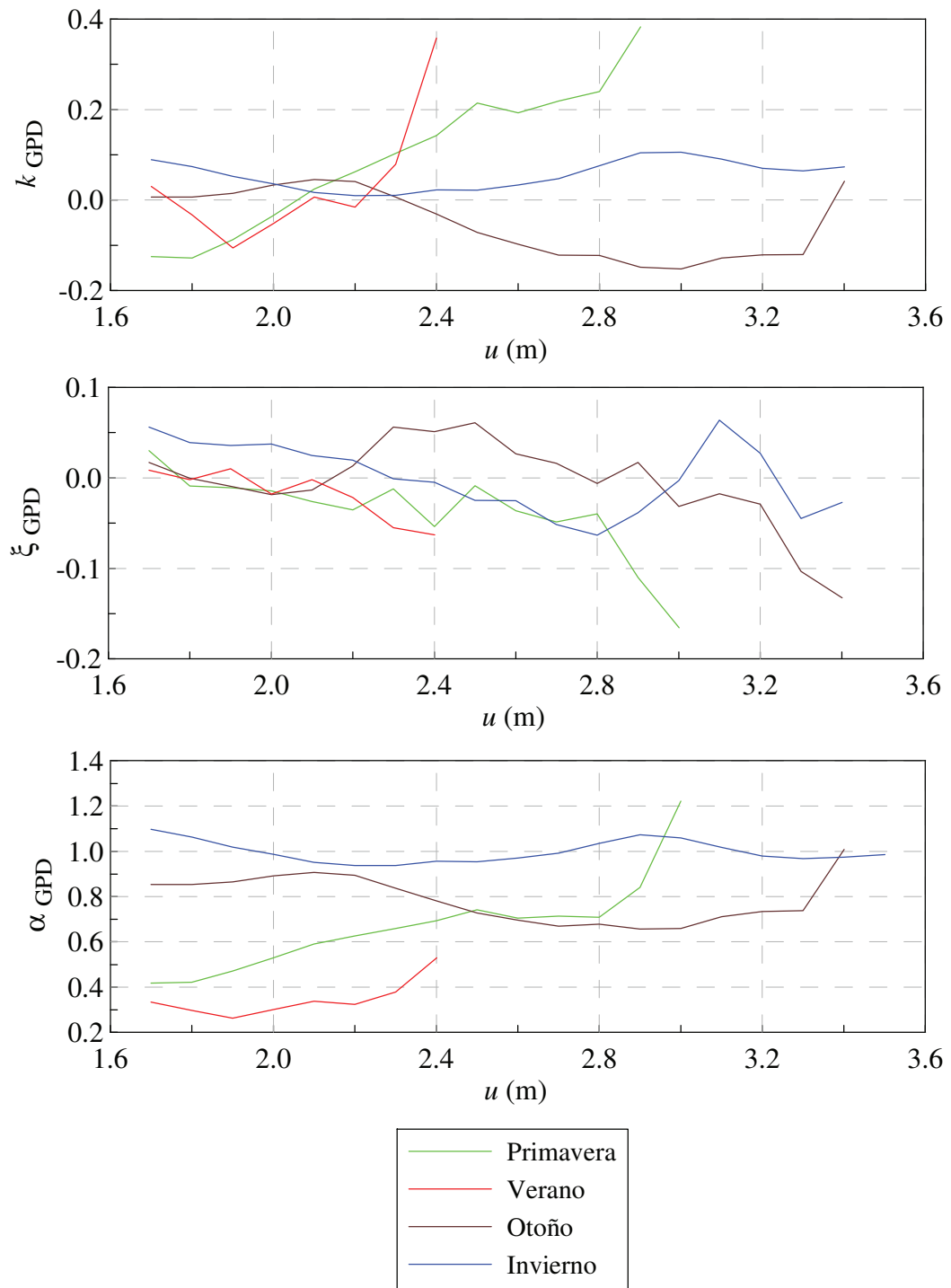


Figura 3.34: Boya LP-I. Variación de los parámetros de forma (superior), localización (medio), y escala (inferior) para las diferentes estaciones climáticas sobre diferentes valores umbrales

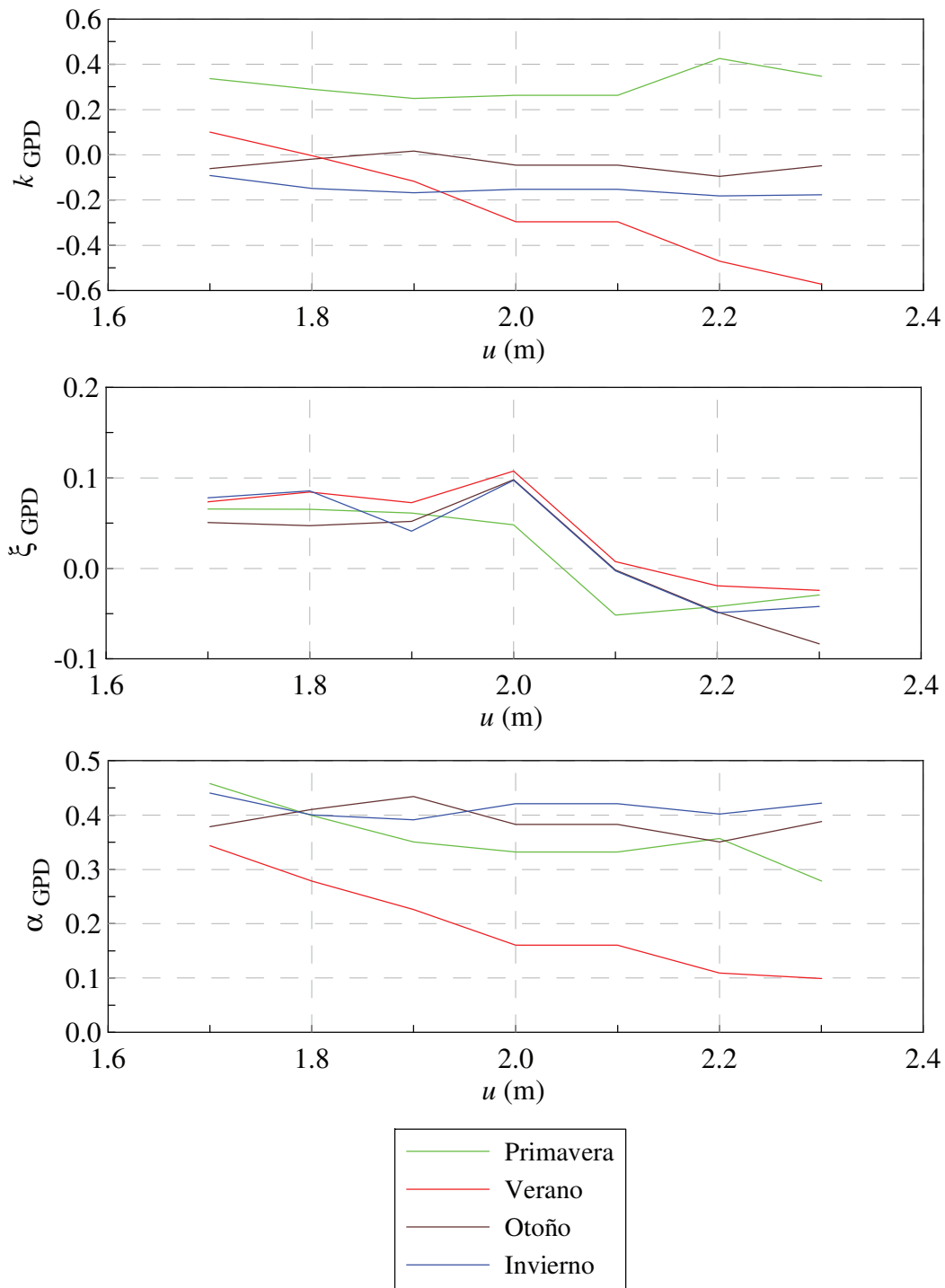


Figura 3.35: Boya LP-II. Variación de los parámetros de forma (superior), localización (medio) y escala (inferior) para las diferentes estaciones climáticas sobre diferentes valores umbrales

El siguiente paso es analizar el efecto de la estacionalidad sobre los cuantiles asociados a determinados periodos de retorno. Para ello, sobre cada estación climática se obtienen los cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años. Esta información se recoge en las tablas 3.31 y 3.32 y en las figuras 3.36 y 3.37 para las boyas LP-I y LP-II, respectivamente. Estos resultados se comparan con aquellos obtenidos a partir de la serie temporal completa, mostrados en las tablas 3.33 y 3.34.

Para el caso de LP-I, se desprende de la figura 3.36 la diferencia de comportamiento de los cuantiles según la estación del año. Para la estación de Primavera los cuantiles disminuyen progresivamente a medida que aumenta el umbral, mientras que el comportamiento en la estación de Otoño es exactamente la contraria. La mayor estabilidad se observa en la estación de Invierno. Esta inestabilidad de los cuantiles está directamente asociada a la inestabilidad de los parámetros de la distribución GPD, y más concretamente al parámetro de forma. Se observa una relación inversa si comparamos la figura 3.34 y 3.36 entre ambas cantidades, de manera que cuando el parámetro de forma aumenta, los cuantiles disminuyen. En cuanto a los valores absolutos de los cuantiles, también es posible observar diferencias en función de la estación, de forma que los mayores valores se obtienen para la estación de Otoño, mientras que los menores ocurren en la estación de Verano.

Para el caso de LP-II, figura 3.37, los cuantiles correspondientes a las estaciones de Primavera, Otoño e Invierno se muestran relativamente estables respecto al umbral. Sin embargo, los cuantiles correspondientes a la estación de Verano muestran un aumento progresivo con el umbral debido a la inestabilidad de los parámetros de la distribución, figura 3.35. Los mayores cuantiles se obtienen para la estación de Invierno, mientras que los valores menores se corresponden con la estación de Primavera.

Dado que lo que se pretende evaluar es el efecto de la estacionalidad en el contexto del análisis POT, las comparaciones entre estaciones climáticas deberían hacerse sobre un mismo valor umbral. Sin embargo, se debe poner especial atención en el número de excedencias, ya que, si éste es demasiado bajo, las estimaciones realizadas carecerían de la necesaria fiabilidad estadística. Como consecuencia, las diferencias en los resultados entre estaciones podrían ser más una consecuencia del bajo número de excedencias que del propio efecto estacional.

Tabla 3.31: Boya LP-I. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para cada estación climática

		<i>Primavera</i>		
$u$	m	$H_{50}$	$H_{100}$	$H_{500}$
1.9	85	7.65 [4.71, 11.11]	8.35 [4.85, 12.89]	10.14 [5.09, 18.49]
2.0	66	7.08 [4.46, 10.46]	7.57 [4.53, 11.99]	8.77 [4.64, 16.45]
2.1	53	6.56 [4.27, 10.03]	6.91 [4.32, 11.40]	7.69 [4.42, 15.34]
2.2	42	6.23 [4.14, 9.49]	6.49 [4.18, 10.73]	7.03 [4.21, 14.12]
2.3	35	5.97 [4.03, 9.50]	6.16 [4.04, 10.68]	6.54 [4.05, 14.46]
2.4	31	5.72 [3.96, 9.09]	5.86 [3.97, 10.16]	6.14 [3.97, 13.24]
		<i>Verano</i>		
1.5	102	3.98 [2.86, 5.79]	4.16 [2.89, 6.41]	4.55 [2.95, 8.12]
1.6	75	4.03 [2.79, 6.00]	4.20 [2.81, 6.61]	4.58 [2.84, 8.71]
1.7	55	4.16 [2.71, 6.02]	4.34 [2.72, 6.63]	4.74 [2.75, 8.44]
1.8	44	4.63 [2.79, 6.79]	4.90 [2.80, 7.80]	5.56 [2.83, 10.71]
1.9	32	5.40 [2.88, 9.02]	5.85 [2.89, 10.92]	7.05 [2.92, 16.99]
2.0	23	5.10 [2.77, 9.25]	5.42 [2.77, 11.17]	6.22 [2.80, 17.42]
		<i>Otoño</i>		
2.4	72	9.83 [5.63, 15.73]	10.54 [5.72, 18.11]	12.24 [5.87, 24.04]
2.5	65	10.78 [5.94, 17.33]	11.72 [6.04, 20.37]	14.08 [6.23, 29.49]
2.6	61	11.48 [6.19, 18.54]	12.60 [6.37, 21.89]	15.51 [6.67, 32.34]
2.7	55	12.29 [6.61, 19.98]	13.62 [6.78, 24.00]	17.18 [7.10, 36.38]
2.8	49	12.52 [6.51, 20.21]	13.88 [6.64, 24.27]	17.50 [6.94, 37.51]
2.9	42	13.61 [6.75, 22.66]	15.25 [6.88, 27.69]	19.78 [7.14, 45.77]
3.0	39	13.90 [6.78, 23.08]	15.60 [6.97, 28.67]	20.32 [7.20, 46.62]
		<i>Invierno</i>		
2.6	82	9.57 [6.49, 13.60]	10.08 [6.59, 15.18]	11.21 [6.79, 19.63]
2.7	75	9.41 [6.43, 13.40]	9.87 [6.52, 14.81]	10.87 [6.66, 18.89]
2.8	67	9.08 [6.24, 12.78]	9.45 [6.31, 13.95]	10.25 [6.41, 16.99]
2.9	58	8.79 [6.18, 12.34]	9.09 [6.22, 13.40]	9.71 [6.29, 16.51]
3.0	51	8.83 [6.31, 12.44]	9.13 [6.36, 13.58]	9.74 [6.43, 16.45]
3.1	44	9.08 [6.45, 13.32]	9.40 [6.52, 14.88]	10.08 [6.57, 18.87]
3.2	42	9.35 [6.56, 14.04]	9.72 [6.59, 15.80]	10.52 [6.64, 20.59]



Tabla 3.32: Boya LP-II. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para cada estación climática

<i>Primavera</i>							
$u$	m	$H_{50}$		$H_{100}$		$H_{500}$	
2.0	35	3.11	[2.74, 3.74]	3.14	[2.75, 3.92]	3.20	[2.76, 4.43]
2.1	35	3.11	[2.74, 3.74]	3.14	[2.75, 3.92]	3.20	[2.76, 4.43]
2.2	24	2.95	[2.70, 3.38]	2.96	[2.70, 3.46]	2.98	[2.70, 3.63]
2.3	18	3.00	[2.73, 3.55]	3.01	[2.73, 3.70]	3.04	[2.73, 4.10]
2.4	14	3.12	[2.72, 4.91]	3.15	[2.72, 5.91]	3.21	[2.72, 9.82]
<i>Verano</i>							
1.9	41	4.36	[2.54, 6.86]	4.72	[2.54, 8.18]	5.70	[2.55, 12.80]
2.0	28	5.71	[2.52, 8.51]	6.66	[2.53, 10.80]	9.77	[2.53, 19.93]
2.1	28	5.71	[2.52, 8.51]	6.66	[2.53, 10.80]	9.77	[2.53, 19.93]
2.2	23	7.83	[2.63, 10.46]	10.10	[2.63, 14.40]	19.32	[2.64, 33.37]
2.3	14	10.94	[2.59, 12.94]	15.24	[2.60, 19.35]	35.10	[2.60, 54.52]
<i>Otoño</i>							
1.9	29	4.75	[3.27, 6.18]	5.02	[3.30, 6.93]	5.63	[3.37, 9.17]
2.0	22	5.18	[3.44, 7.12]	5.55	[3.48, 8.45]	6.46	[3.54, 12.97]
2.1	22	5.18	[3.44, 7.12]	5.55	[3.48, 8.45]	6.46	[3.54, 12.97]
2.2	20	5.53	[3.53, 7.93]	6.02	[3.56, 9.89]	7.27	[3.59, 16.96]
2.3	16	5.36	[3.46, 8.95]	5.74	[3.47, 11.41]	6.68	[3.50, 21.96]
<i>Invierno</i>							
1.9	49	7.06	[4.03, 10.49]	7.99	[4.18, 12.90]	10.59	[4.39, 21.13]
2.0	34	7.26	[3.93, 11.14]	8.14	[4.00, 13.53]	10.59	[4.12, 22.53]
2.1	34	7.26	[3.93, 11.14]	8.14	[4.00, 13.53]	10.59	[4.12, 22.53]
2.2	31	7.72	[4.01, 11.83]	8.76	[4.10, 14.95]	11.76	[4.19, 25.81]
2.3	24	7.98	[3.92, 12.64]	9.04	[3.95, 16.14]	12.06	[4.03, 29.21]

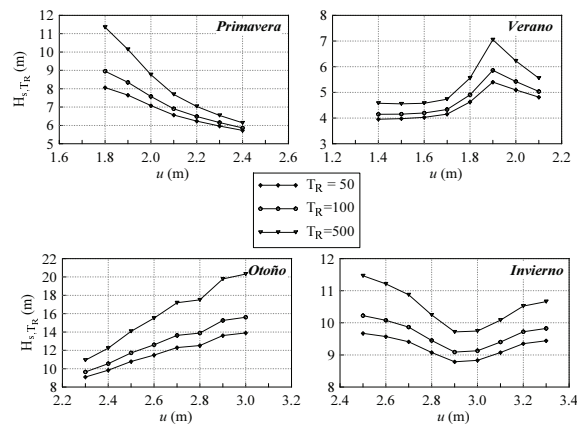


Figura 3.36: Boya LP-I. Variación de los cuantiles de  $H_s$  asociados a periodos de retorno,  $T_R$ , de 50, 100 y 500 años sobre diferentes valores umbrales

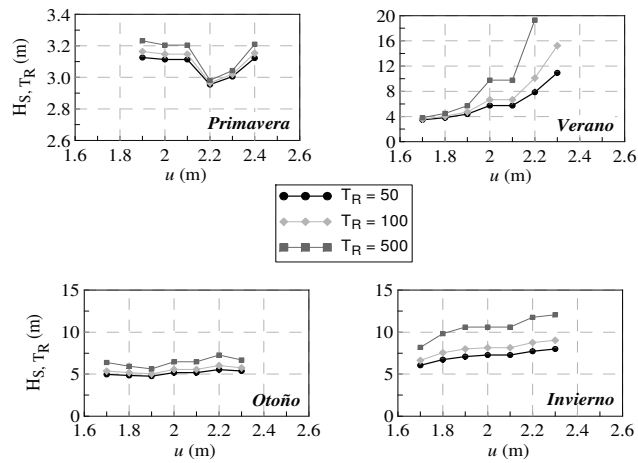


Figura 3.37: Boya LP-II. Variación de los cuantiles de  $H_s$  asociados a periodos de retorno,  $T_R$ , de 50, 100 y 500 años sobre diferentes valores umbrales

Para la boya LP-I, a modo de ejemplo, tomando el valor umbral de 2.7 metros, se obtiene a partir de la tabla 3.33 para la serie total original un valor para el cuantil  $H_{s,50}$  de 10.27 metros. Sin embargo, para el mismo valor umbral, se obtienen valores de  $H_{s,50}$  de 12.29 y 9.41 metros en las estaciones de otoño e invierno, respectivamente.

Para la boya LP-II, sobre un umbral de 2.3 metros, el cuantil de  $H_s$  asociado a un periodo de retorno de 50 años muestra un valor de 6.91 metros. Para el mismo parámetro y sobre el mismo valor umbral, las estaciones de invierno y otoño muestran valores de 7.98 y 5.43 metros, respectivamente.

En base a estos resultados, los cuales evidencian diferencias entre los parámetros de la distribución GPD y los cuantiles de  $H_s$  asociados a largos periodos de retorno, en función de la estación climática, se puede concluir que existe estacionalidad en las dos series temporales analizadas, si bien es cierto que en la boya LP-II el grado de estacionalidad es menor, ya que la distribución de eventos extremos es algo más errática que en la boya LP-I.

Las figuras 3.38 y 3.39 muestran los gráficos cuantil-cuantil estacionales para las series LP-I y LP-II, respectivamente, sobre diferentes valores umbrales. Para el caso de LP-I, en las cuatro estaciones climáticas y sobre todos los umbrales seleccionados, los valores de  $H_s$  predichos por el modelo son muy similares a los valores observados, concluyendo así que el modelo combinado de Poisson-GPD se muestra adecuado para caracterizar las excedencias sobre los umbrales seleccionados. Para el caso de LP-II, el modelo se muestra adecuado para la estación de primavera, mientras que para el resto de estaciones se observa una ligera desviación entre los valores predichos por el modelo y los observados para valores altos de  $H_s$ . Esta desviación es máxima en el caso de la estación de verano, donde los dos valores más altos de  $H_s$  se desvían considerablemente respecto de la línea  $Y = X$ .

Tabla 3.33: Boya LP-I. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para la serie temporal completa

$u$	$m$	$H_{s,50}$	$H_{s,100}$	$H_{s,500}$
2.5	186	9.72 [7.33, 12.45]	10.37 [7.58, 13.78]	11.94 [8.05, 17.23]
2.6	168	10.07 [7.50, 13.08]	10.80 [7.74, 14.58]	12.55 [8.28, 18.65]
2.7	154	10.27 [7.66, 13.44]	11.02 [7.93, 15.01]	12.85 [8.46, 19.27]
2.8	137	10.18 [7.60, 13.39]	10.88 [7.82, 15.03]	12.56 [8.26, 19.04]
2.9	118	10.07 [7.53, 13.37]	10.71 [7.72, 14.85]	12.22 [8.16, 18.86]
3.0	104	9.97 [7.47, 13.34]	10.57 [7.66, 14.85]	11.94 [8.03, 18.80]
3.1	86	9.78 [7.34, 13.36]	10.29 [7.48, 14.80]	11.46 [7.79, 18.80]
3.2	78	10.25 [7.54, 14.49]	10.86 [7.69, 16.27]	12.27 [7.98, 21.32]

Tabla 3.34: Boya LP-II. Cuantiles asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 % para la serie temporal completa

$u$	$m$	$x_{50}$	$x_{100}$	$x_{500}$
1.9	163	5.24 [3.98, 6.79]	5.58 [4.10, 7.55]	6.41 [4.34, 9.69]
2.0	118	5.68 [4.10, 7.33]	6.11 [4.22, 8.23]	7.21 [4.48, 10.95]
2.1	118	5.68 [4.10, 7.33]	6.11 [4.22, 8.23]	7.21 [4.48, 10.95]
2.2	97	6.60 [4.63, 8.81]	7.29 [4.82, 10.28]	9.19 [5.26, 14.69]
2.3	71	6.91 [4.72, 9.33]	7.64 [4.91, 11.00]	9.68 [5.29, 16.58]
2.4	51	6.54 [4.32, 9.43]	7.09 [4.39, 11.02]	8.53 [4.57, 16.42]
2.5	36	5.41 [3.94, 8.37]	5.62 [3.96, 9.43]	6.11 [3.98, 13.11]
2.6	25	4.38 [3.64, 6.11]	4.42 [3.64, 6.47]	4.48 [3.64, 7.34]

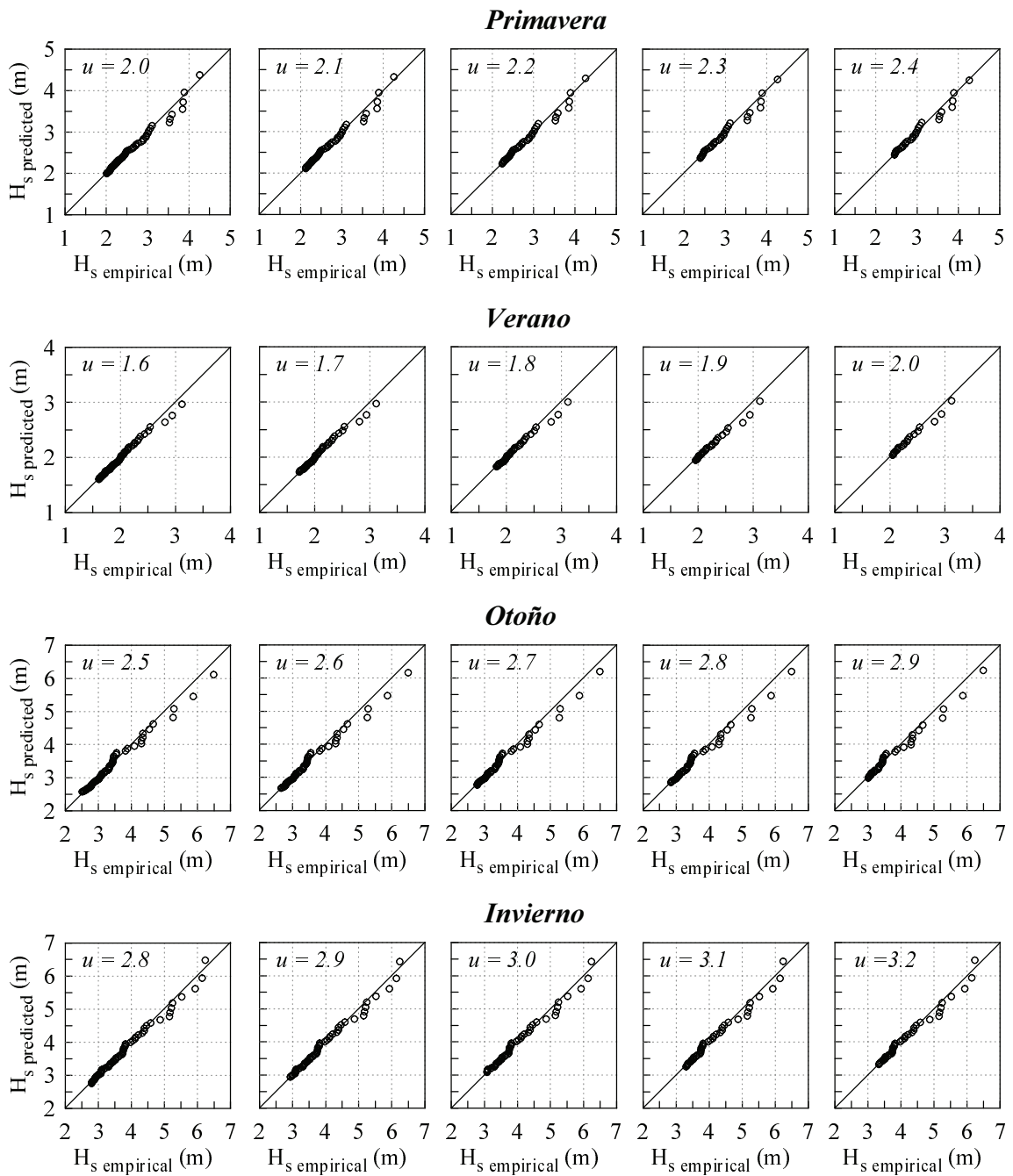


Figura 3.38: Boya LP-I. Gráficos cuantil-cuantil para cada estación climática sobre diferentes valores umbrales

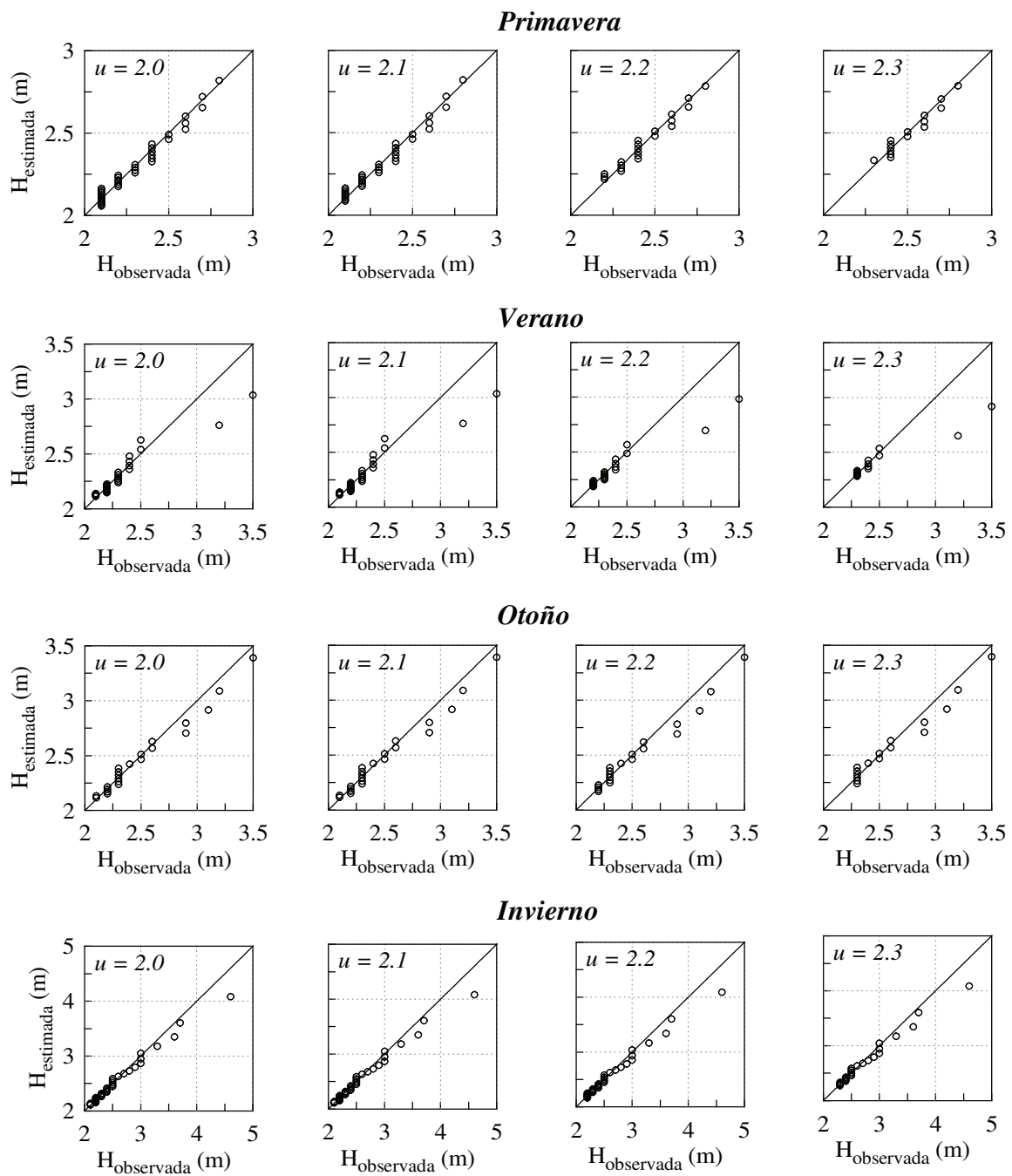


Figura 3.39: Boya LP-II. Gráficos cuantil-cuantil para cada estación climática sobre diferentes valores umbrales

Una vez analizadas las estaciones climáticas tradicionales se procede a la aplicación del método descrito en Ouarda (1993), que generará varios periodos temporales homogéneos que llamaremos *estaciones hidroclimáticas*. Este análisis es llevado a cabo sobre la serie temporal LP-I y LP-II y se muestra en las figuras 3.40 y 3.41 con un valor de  $t$  equivalente a 5 días. Cada una de las curvas corresponde a un valor umbral diferente. A cada día del año se le asigna un número de orden comprendido entre 1 y 365. De esta forma, al 1 de Enero le corresponde el número 1 mientras que al 31 de Diciembre le corresponde el 365.

Para el caso de LP-I, a partir de la figura es posible identificar dos estaciones independientes. La primera de ellas es denotada como estación I y comprende el periodo temporal que va desde el día 1 hasta el día 90 (30 de marzo) y desde el día 280 (7 de octubre) hasta el día 365 aproximadamente. Nótese que aunque en la gráfica ambos periodos aparezcan separados, uno es continuación del otro, formando una única estación, con igual valor de la pendiente. La segunda es denotada como estación II y comprende el periodo temporal que va desde el día 91 (1 de abril) hasta el día 279 (6 de octubre) aproximadamente. Esta segunda estación posee una pendiente prácticamente nula, o lo que es equivalente, no se acumulan datos en este periodo temporal. De la figura 3.40 se desprende que la duración de ambas estaciones depende ligeramente del valor umbral seleccionado. De esta forma, a medida que el valor umbral aumenta, la estación II se hace más larga, mientras que la estación I se acorta. Sin embargo, dado que este efecto no parece ser muy significativo se ha optado por mantener invariables los límites de cada estación respecto del valor umbral.

Para el caso de LP-II, la interpretación de la figura es algo más compleja. A priori podría interpretarse que existen dos estaciones bien diferenciadas, ya que existen dos zonas con pendientes diferentes. La primera estación, denominada estación I, vendría dada por la zona de pendiente pronunciada entre los días 1-90 (1 de enero - 31 de marzo) y 240-365 (28 de agosto - 31 de diciembre). La segunda estación, estación II, vendría dada por la zona de pendiente casi nula, entre los días 91-239 (1 de abril - 27 de agosto).

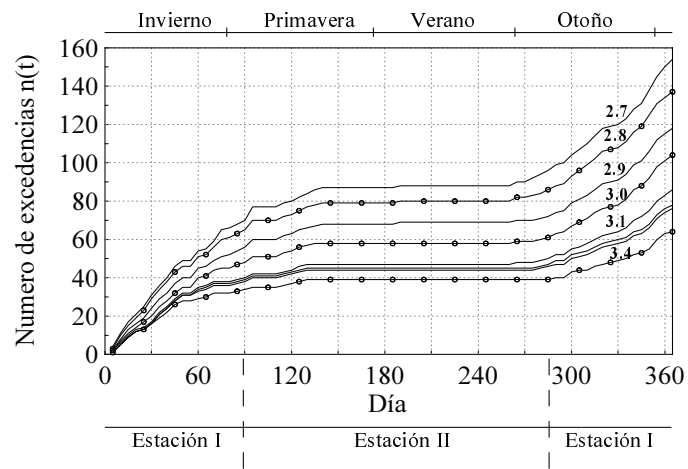


Figura 3.40: Boya LP-I. Identificación de estaciones hidroclimáticas homogéneas

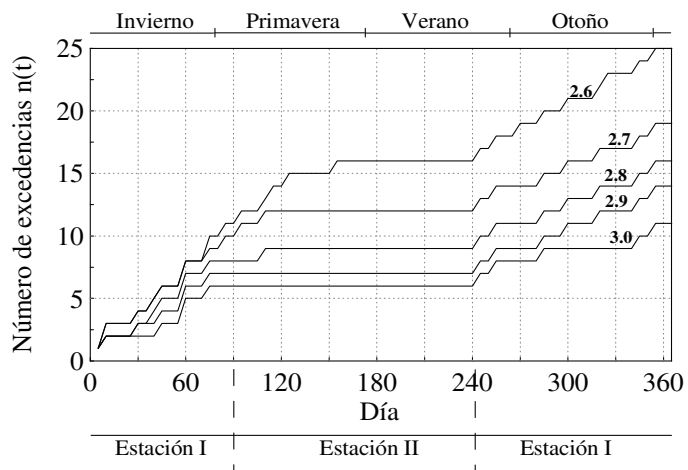


Figura 3.41: Boya LP-II. Identificación de estaciones hidroclimáticas homogéneas



Sin embargo, si se observa detalladamente, las pendientes correspondientes al primer tramo y al tercer tramo no son iguales. De hecho, la primera de ellas presenta un valor ligeramente superior. Esto podría interpretarse como la presencia de tres estaciones diferentes. Por otro lado, es evidente la fuerte relación existente entre el valor umbral y el comienzo de la estación II, de forma que éste se adelanta en cuanto el umbral aumenta. Tan estrecha es esta relación que el comienzo de la estación puede variar hasta 50 días dependiendo de si el umbral se fija en 2.6 o en 3.0 metros.

Esta relación entre la duración de las estaciones y el valor umbral se explica en base a que a medida que el umbral aumenta, los eventos extremos más severos tienden a concentrarse en determinados meses del año, normalmente en Otoño e Invierno. De esta forma, los periodos homogéneos tienden a ser más cortos.

Un aspecto importante que no debe pasarse por alto en la figura 3.41 es el bajo número de excedencias con las que se construyen las curvas  $n(t)$  para los diferentes umbrales. Surge entonces la cuestión de si los cambios de pendiente pueden interpretarse realmente como un cambio de estación o se debe a variaciones relacionadas con la incertidumbre estadística.

Por tanto, se asume la existencia de dos estaciones, mostradas en la figura 3.41, con el comienzo de la estación II, independiente del umbral, fijado en el día 91. Las tablas 3.35 y 3.36 muestran los cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años para las estaciones I y II identificadas en la serie LP-I. Una comparación entre ambas pone de manifiesto la diferencia entre los valores cuantiles, obteniéndose los mayores valores en la estación I. Este resultado es de esperar ya que la estación I agrupa la mayor parte de los meses de otoño e invierno, periodos donde se alcanzan los máximos valores de la altura del oleaje.

La comparación de la tabla 3.35 y la correspondiente a la serie temporal completa, tabla 3.33, pone de manifiesto la similitud entre los cuantiles de la estación I y los correspondientes a la serie completa. Este resultado es lógico si se tiene en cuenta que la estación I agrupa la mayor parte de los valores extremos de la serie, de forma que la información contenida en ella es prácticamente la misma que la contenida en la serie temporal completa.

Con el objetivo de comparar el efecto de la estacionalidad y los diferentes métodos utilizados, las figuras 3.42 y 3.43 muestran el valor del cuantil  $H_{50}$  obtenido con los diferentes métodos de análisis de la estacionalidad, además del valor obtenido a partir de la serie de datos completa.

Tabla 3.35: Boya LP-I. Estación hidroclimática I. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %.

$u$	$m$	$H_{s,50}$	$H_{s,100}$	$H_{s,500}$
2.5	168	9.27 [6.84, 12.24]	9.77 [7.01, 13.42]	10.89 [7.26, 16.57]
2.6	156	9.61 [6.98, 12.88]	10.16 [7.11, 14.18]	11.44 [7.40, 17.58]
2.7	143	9.84 [7.16, 13.32]	10.44 [7.34, 14.67]	11.82 [7.66, 18.71]
2.8	130	10.45 [7.37, 15.03]	11.80 [7.69, 19.16]	12.38 [7.81, 21.19]
2.9	113	9.83 [7.20, 13.44]	10.38 [7.34, 14.92]	11.63 [7.60, 18.92]
3.0	101	9.82 [7.19, 13.57]	10.34 [7.33, 15.01]	11.53 [7.52, 19.07]
3.1	84	9.76 [7.09, 13.56]	10.24 [7.22, 15.17]	11.32 [7.40, 19.01]
3.2	77	10.20 [7.31, 14.48]	10.77 [7.43, 16.33]	12.07 [7.64, 21.20]
3.3	75	10.32 [7.34, 14.96]	10.91 [7.46, 16.95]	12.26 [7.71, 22.51]

Tabla 3.36: Boya LP-I. Estación hidroclimática II. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %.

$u$	$m$	$H_{s,50}$	$H_{s,100}$	$H_{s,500}$
1.9	122	7.00 [4.59, 10.42]	7.66 [4.73, 12.18]	9.37 [5.04, 17.59]
2.0	96	6.71 [4.36, 10.59]	7.23 [4.46, 12.45]	8.54 [4.62, 17.90]
2.1	74	6.12 [4.03, 9.82]	6.46 [4.08, 11.13]	7.25 [4.17, 15.23]
2.2	57	5.70 [3.79, 9.14]	5.93 [3.81, 10.30]	6.42 [3.85, 13.21]
2.3	45	5.55 [3.71, 8.68]	5.73 [3.73, 9.81]	6.10 [3.75, 12.54]
2.4	37	5.48 [3.62, 8.52]	5.64 [3.63, 9.40]	5.96 [3.65, 11.60]
2.5	31	5.72 [3.65, 8.71]	5.91 [3.65, 9.73]	6.31 [3.66, 12.24]
2.6	26	6.73 [3.88, 11.74]	7.12 [3.90, 13.92]	8.08 [3.94, 21.41]

En dichas figuras se representa por tanto el valor de  $H_{50}$  calculado a partir de los siguientes escenarios:

- serie completa
- primavera
- verano
- otoño
- invierno
- estación hidroclimática I (HCI: Severa)
- estación hidroclimática II (HCII: Moderada)

Los cuantiles asociados a periodos de retorno de 50, 100 y 500 años para las estaciones I y II identificadas en la boya LP-II se muestran en las tablas 3.37 y 3.38, respectivamente. Una vez más, se observan diferencias significativas entre los cuantiles calculados sobre ambas estaciones, obteniendo los máximos valores para la estación I, la cual engloba las estaciones de invierno y otoño.

Para el caso de la boya LP-I, se observa que los mayores valores  $H_{50}$  se obtienen para la estación de otoño, mientras que la estación de invierno y la serie completa se asemejan a la estación hidroclimática I (severa). En otras palabras, el mayor grado de estacionalidad viene generado por las tormentas que se generan en el periodo de otoño. Por otro lado, la estación de primavera y la hidroclimática II ofrecen resultados similares, destacando la estación de verano con valores de  $H_{50}$  ligeramente inferiores.

En resumen, podrían identificarse tres estaciones o periodos diferenciados:

- el periodo de otoño, que presenta valores de  $H_{50}$  significativamente superiores al resto
- estación hidroclimática I, que coincidiría aproximadamente con la estación de invierno
- estación hidroclimática II que agruparía de forma aproximada a las estaciones de primavera y verano

Para el caso de la boya LP-II, se observa que los mayores valores de  $H_{50}$  se obtienen en general para la estación de invierno, mientras que la estación de verano presenta valores significativamente altos sobre los umbrales de 2.20 m y 2.30 m. La estación de otoño y la estación hidroclimática I presentan valores muy similares, y ligeramente inferiores a los valores obtenidos sobre la serie completa. Los valores obtenidos para la estación de primavera y la estación hidroclimática II son prácticamente iguales, y además significativamente inferiores a los obtenidos para la serie completa.

En este caso, al igual que sucede en el caso anterior para la boya LP-I, es posible identificar tres periodos diferenciados:

- estación de invierno, presentando los valores de  $H_{50}$  más altos
- estación hidroclimática I, que coincidiría aproximadamente con la estación de otoño y verano, esta última hasta valores umbrales de 2.20 m
- estación hidroclimática II que coincide prácticamente con la estación de primavera

Tabla 3.37: Boya LP-II. Estación hidroclimática I. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %.

$u$	$m$	$H_{s,50}$	$H_{s,100}$	$H_{s,500}$
1.8	109	5.65 [4.04, 7.60]	6.00 [4.13, 8.53]	6.81 [4.28, 10.99]
1.9	90	5.71 [4.07, 7.85]	6.05 [4.17, 8.84]	6.84 [4.31, 11.62]
2.0	65	5.79 [4.08, 8.03]	6.12 [4.13, 8.99]	6.90 [4.27, 11.90]
2.1	65	5.79 [4.08, 8.03]	6.12 [4.13, 8.99]	6.90 [4.27, 11.90]
2.2	59	6.17 [4.25, 8.99]	6.58 [4.31, 10.41]	7.58 [4.42, 14.60]
2.3	46	5.68 [4.00, 8.91]	5.95 [4.01, 10.30]	6.56 [4.05, 14.46]
2.4	34	4.80 [3.63, 7.98]	4.89 [3.66, 8.90]	5.06 [3.67, 11.72]
2.5	27	4.33 [3.50, 7.21]	4.36 [3.50, 8.01]	4.41 [3.50, 10.24]
2.6	20	3.97 [3.43, 5.21]	3.97 [3.43, 5.35]	3.98 [3.43, 5.64]

Tabla 3.38: Boya LP-II. Estación hidroclimática II. Cuantiles de  $H_s$  asociados a periodos de retorno de 50, 100 y 500 años con sus intervalos de confianza al 95 %.

$u$	$m$	$H_{s,50}$	$H_{s,100}$	$H_{s,500}$
1.7	109	2.83 [2.57, 3.19]	2.84 [2.57, 3.23]	2.85 [2.58, 3.31]
1.8	90	2.93 [2.62, 3.42]	2.94 [2.62, 3.49]	2.97 [2.63, 3.65]
1.9	73	3.03 [2.67, 3.73]	3.06 [2.67, 3.87]	3.10 [2.67, 4.24]
2.0	53	3.05 [2.66, 3.87]	3.08 [2.66, 4.05]	3.12 [2.66, 4.50]
2.1	53	3.05 [2.66, 3.87]	3.08 [2.66, 4.05]	3.12 [2.66, 4.50]
2.2	38	3.11 [2.64, 4.61]	3.14 [2.64, 5.00]	3.20 [2.64, 6.25]
2.3	25	3.16 [2.65, 4.65]	3.19 [2.65, 5.06]	3.26 [2.65, 6.32]
2.4	17	3.70 [2.69, 8.51]	3.85 [2.69, 11.51]	4.24 [2.69, 25.09]

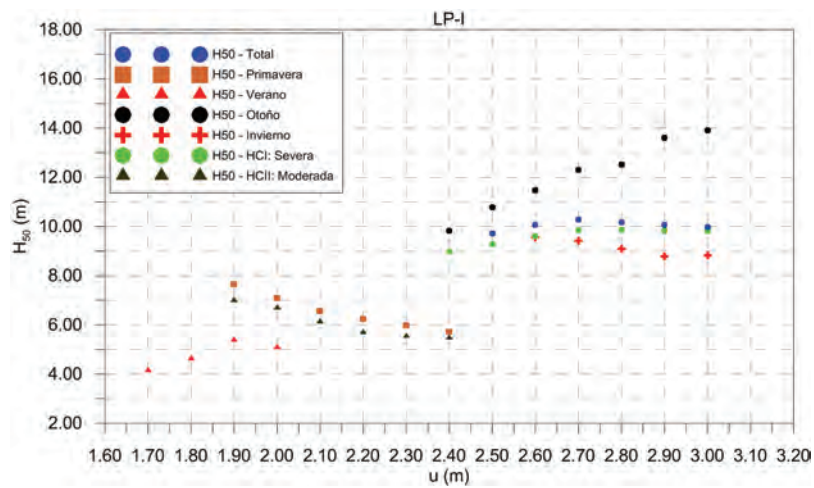


Figura 3.42: Boya LP-I. Comparación del cuantil  $H_{50}$  obtenido con los diferentes métodos de análisis de la estacionalidad

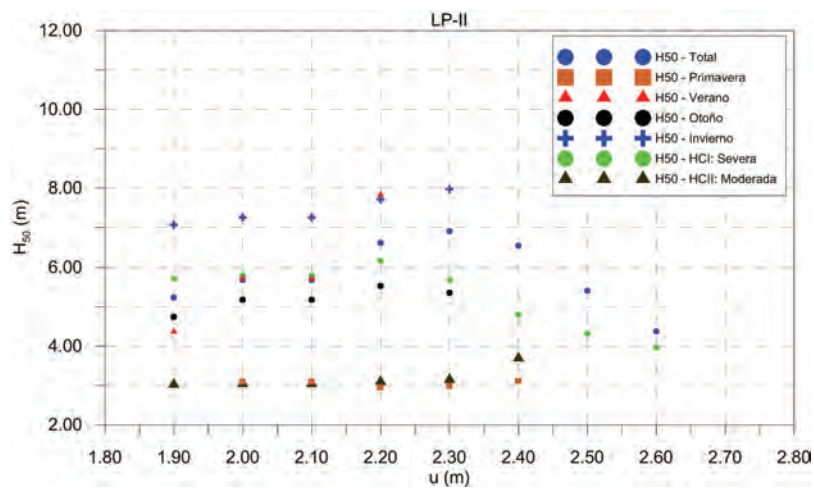


Figura 3.43: Boya LP-II. Comparación del cuantil  $H_{50}$  obtenido con los diferentes métodos de análisis de la estacionalidad

# 4

---

## Análisis extremal multivariante

---

### 4.1 Introducción

El análisis extremal multivariante analiza la presentación simultánea de valores extremos de  $n$  variables, de forma que, a través de la función de distribución de probabilidad conjunta, la probabilidad de que  $n$  variables presenten valores extremos en el mismo instante de tiempo es evaluada. La mayoría de los trabajos en este campo se han centrado hasta ahora en el estudio conjunto de dos variables, siendo posible una generalización de la teoría a  $n$  variables.

Las estructuras marítimas suelen estar construidas, bajo el marco de la teoría del análisis extremal univariante, en base a la ya mencionada altura de ola de diseño  $H_c$ . En determinadas situaciones es posible observar el fallo de estructuras marítimas en condiciones de altura de ola inferiores a la altura de referencia  $H_c$ . Este hecho suele ser debido al efecto combinado que sobre la estructura ejercen otras variables que se suman a la altura del oleaje y que no han sido tenidas en cuenta durante el proceso de diseño. Es por ello que en algunas ocasiones las estructuras son diseñadas teniendo en cuenta el efecto combinado de dos o más variables, como por ejemplo, la altura de ola y el nivel del mar. Una vez determinada la región de fallo, o dicho de otra forma, aquella combinación de valores de ambas variables que pueden hacer fallar la estructura, la teoría extremal multivariante es capaz de evaluar la probabilidad de presentación de tales situaciones. A modo de ejemplo,

se ilustra en la figura 4.1 la relación entre la altura de ola significativa  $H_s$  y el nivel del mar  $SWL$  en la costa norte de la isla de Gran Canaria. Suponiendo una estructura marítima ficticia cuya región de fallo viene determinada según se muestra en 4.1 todos los pares de valores  $(H_s, SWL)$  que estén dentro de esta zona serán potencialmente peligrosos para la estructura en cuestión.

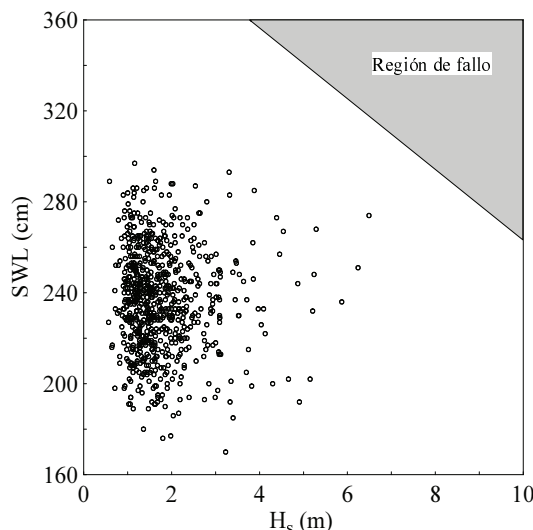


Figura 4.1: Altura de ola significativa ( $H_s$ ) frente a nivel del mar ( $SWL$ ) en la costa norte de Gran Canaria con su respectiva zona de fallo

Matemáticamente puede demostrarse que si dos variables aleatorias son estadísticamente independientes, entonces los máximos de ambas variables también serán independientes. Sin embargo, si ambas variables son dependientes, es posible que sus valores máximos estén o no relacionados, introduciéndose el concepto de *dependencia asintótica*. Así, a medida que aumenta el valor umbral que delimita la presencia del valor extremo, la dependencia entre ambas variables puede mantenerse o disminuir.

Un trabajo pionero en este campo es el desarrollado por Sibuya (1960), donde el autor extiende la teoría clásica unidimensional de valores extremos al caso bidimensional, considerando la distribución conjunta de los máximos de dos variables aleatorias. A diferencia del caso unidimensional, donde las posibles distribuciones límite para el máximo de la variable son sólo tres, en el caso bivariado se demuestra que existen infinitas distribuciones conjuntas límite para los dos máximos. Además, en este contexto bivariado se introduce el concepto de *función de dependencia* o *estructura de dependencia* que determinará el grado de dependencia entre los máximos de ambas variables. Esta función de dependencia recibe también el nombre de *función cópula*. La atención e interés sobre el estudio de valores



extremos multivariados a través de esta metodología ha crecido significativamente en los últimos años (Bonazzi, 2012 ; Genest *et al.* 2009, 2011 ; Salvadori, 2007 ; Martín, 2007).

Los modelos estadísticos propuestos para el análisis conjunto de valores extremos pueden separarse en dos grupos, los paramétricos y los no paramétricos. Los modelos paramétricos asumen un modelo específico para caracterizar la estructura de dependencia entre ambas variables, basando posteriormente las estimaciones de los parámetros en la teoría de máxima verosimilitud. Algunas referencias significativas al respecto son los trabajos de Tawn (1988), Coles & Tawn (1991, 1994) y de Haan *et al.* (2008), refiriéndose los dos primeros al estudio conjunto de la altura de ola y el nivel del mar.

Por otro lado, los modelos no paramétricos se muestran más flexibles a la hora de caracterizar la estructura de dependencia, sin asumir a priori una forma específica para tal estructura. Algunas referencias al respecto son de Haan & de Ronde (1998) y Draisma *et al.* (2004), donde al igual que en los trabajos citados anteriormente, se estudian los valores extremos conjuntos de la altura de ola y el nivel del mar.

Entre los años 1995 y 1997 se desarrolló el proyecto NEPTUNE entre diferentes Universidades y Centros de Investigación europeos cuyo objetivo final fue el análisis extremal conjunto de la altura del oleaje y el nivel del mar en las costas de Reino Unido y Holanda. Ambos tipos de modelos, paramétricos y no paramétricos, fueron utilizados ofreciendo resultados muy similares, mostrándose algo más efectivo el modelo no paramétrico.

Si bien es cierto que la teoría multivariante de valores extremos queda bien establecida desde un punto de vista teórico, su puesta en práctica no está exenta de complejidad. Un claro ejemplo de esta dificultad se encuentra precisamente cuando se observan los resultados derivados del proyecto NEPTUNE. Uno de los resultados principales del proyecto fue la existencia de una dependencia asintótica entre los valores extremos de la altura de ola y el nivel del mar (de Haan & de Ronde (1998)). Sobre este mismo conjunto de datos, Peng (1999), utilizando un coeficiente de dependencia modificado a partir del propuesto en Ledford y Tawn (1996), obtiene la misma conclusión sobre la existencia de dependencia asintótica. Sin embargo, Draisma *et al.* (2004) utilizando el estimador de Hill (Hill, 1975), el estimador de Peng (Peng, 1999) y el método de máxima verosimilitud, para evaluar el coeficiente de dependencia, obtiene que según los estimadores de Hill y máxima verosimilitud, debería rechazarse la hipótesis de dependencia asintótica entre ambas variables.

Si en el caso univariante la selección de la población extremal presenta cierta complejidad, en el caso multivariante este aspecto adquiere una complejidad de mayor magnitud, ya que deben seleccionarse los valores extremos conjuntos, para lo cual debería utilizarse,

además de los criterios establecidos por la teoría, el conocimiento del funcionamiento físico de los procesos bajo estudio.

El modelo utilizado en el presente trabajo es el modelo no paramétrico propuesto en de Haan, L. & de Ronde, J. (1998) y cuya descripción se ofrece en la siguiente sección.

## 4.2 Modelo teórico

El objetivo del análisis extremal es la caracterización estadística de aquellos valores de la variable que presentan escasa probabilidad de presentación, o dicho de otra forma, aquellos valores situados en la cola de la distribución de probabilidad. Este análisis, en el caso multivariante, puede reducirse al análisis individual de cada una de las variables que entran en juego y al análisis de la estructura de dependencia entre dichas variables.

El estudio de cada una de las variables por separado se reduce a un análisis extremal univariante común, ajustando las ya comentadas distribuciones de probabilidad extremales. Por otro lado, la forma más sencilla de estudiar la estructura de dependencia es a través de la transformación de las colas de las distribuciones marginales iniciales, de forma que pasen a ser distribuciones uniformes o distribuciones de Pareto, según convenga. De esta forma las distribuciones marginales límite serán distribuciones *exponenciales* o de *Fréchet*, respectivamente.

Después de esta transformación de las distribuciones marginales iniciales, la cola de la distribución bivariada posee una propiedad de homogeneidad muy útil que nos permite estimar probabilidades en el área donde no hay observaciones.

La estructura de dependencia de extremos multivariados aporta información a cerca de como están relacionadas dos variables en la cola alta de la distribución. Por ejemplo, queremos saber si una situación catastrófica es comunmente causada por altos niveles del mar, por altos valores de altura de ola o por ambos a la vez. Esta estructura de dependencia puede ser analizada a través de las *curvas cuantil* o a través de la *medida espectral*.

Para el primer método se dibujan líneas de igual probabilidad para cuadrantes extremos (curvas cuantil) usando los datos transformados. Estas curvas de nivel deberían ser convexas y deberían tener la misma forma para varios niveles (de esta forma se puede comprobar si el modelo es adecuado o no). Estas curvas revelan el grado de dependencia asintótica a través de su grado de curvatura, no siendo posible cuantificar de forma numé-

rica dicho grado de dependencia (en la teoría de valores extremos multivaridos no existe un análogo al coeficiente de correlación en la teoría clásica de distribuciones bivariadas).

Se dispone de un conjunto de variables aleatorias  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  independientes e igualmente distribuidas con función de distribución

$$F(x, y) = Prob(X < x, Y < y) \quad (4.1)$$

Las funciones de distribución marginales de  $X$  e  $Y$  son denotadas como  $F_1(x) = Prob(X \leq x)$  y  $F_2(y) = Prob(Y \leq y)$ . Respecto de la cola de  $F(x, y)$  se asume la condición de valores extremos bivariados, la cual asegura que *existen constantes*  $a_n > 0, b_n, c_n > 0$  y  $d_n$  tales que cuando  $n \rightarrow \infty$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left\{ \frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n} \leq x, \frac{\max(Y_1, Y_2, \dots, Y_n) - d_n}{c_n} \leq y \right\} \\ = \lim_{n \rightarrow \infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y) \end{aligned} \quad (4.2)$$

siendo  $G$  una función de distribución extremal bivariada.

La relación 4.2 implica la convergencia de las distribuciones marginales

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x, \infty) \quad (4.3)$$

$$\lim_{n \rightarrow \infty} F^n(c_n y + d_n) = G(\infty, y) \quad (4.4)$$

donde  $G(x, \infty)$  y  $G(\infty, y)$  son distribuciones generalizadas de valores extremos (GEV)

$$G(x, \infty) = \exp -(1 + \gamma_1 x)^{-1/\gamma_1} \quad (4.5)$$

$$G(\infty, y) = \exp -(1 + \gamma_2 y)^{-1/\gamma_2} \quad (4.6)$$

Tomando logaritmos en 4.2 y desarrollando en series de Taylor se obtiene

$$\lim_{n \rightarrow \infty} n \{1 - F(a_n x + b_n, c_n y + d_n)\} = -\log G(x, y) \quad (4.7)$$

de forma que para las distribuciones marginales se obtiene

$$\lim_{n \rightarrow \infty} n \{1 - F_1(a_n x + b_n)\} = -\log G(x, \infty) = (1 + \gamma_1 x)^{-1/\gamma_1} \quad (4.8)$$

$$\lim_{n \rightarrow \infty} n \{1 - F_2(c_n y + d_n)\} = -\log G(\infty, y) = (1 + \gamma_2 y)^{-1/\gamma_2} \quad (4.9)$$

### 4.2.1 Estructura de dependencia

La estructura de dependencia entre ambas variables puede ser analizada a partir de la *función de dependencia en la cola*,  $l(x, y)$ . Para ello se define la función cuantil  $Q_i(p) = (1 - F_i)^{\leftarrow}(p)$  como aquel cuantil que es excedido con probabilidad  $p$ . De esta forma es posible reescribir la relación 4.7 en términos de esta función cuantil, según

$$\lim_{p \rightarrow 0} p^{-1} \{1 - F(Q_1(px), Q_2(py))\} = -\log G\left(\frac{x^{-\gamma_1} - 1}{\gamma_1}, \frac{y^{-\gamma_2} - 1}{\gamma_2}\right) = l(x, y) \quad (4.10)$$

donde  $p = 1/n$ . La cuestión entonces es qué tipo de función es  $l$ . Esta función  $l$  tiene las siguientes propiedades:

1. Satisface  $l(x, \infty) = l(\infty, x) = x$ . Esto es una consecuencia de la normalización elegida. Para esta elección, la distribución límite extremal bivariada  $G$  tiene distribuciones marginales  $G(x, \infty) = \exp(-x)$  y  $G(\infty, y) = \exp(-y)$ .

2. Puede ser vista como una medida  $\nu$  del conjunto

$$\{(r, s) | r < x \text{ o } s < y\} = ([x, \infty) \times [y, \infty))^c$$

3. Es una función homogénea de orden 1, según

$$l(tx, ty) = tl(x, y)$$

4. Relaciona las probabilidades de excedencia bivariadas con las probabilidades de excedencia univariadas, de forma que para valores suficientemente grandes de  $x, y$

$$1 - F(x, y) \approx l(1 - F_1(x), 1 - F_2(y)) \quad (4.11)$$

A partir de esta expresión es posible ver que  $l(x, y)$  es una función que contiene toda la información a cerca de la dependencia entre  $X$  e  $Y$  en la cola de la distribución. Una vez que  $l$  es conocida la probabilidad conjunta de  $\{X > a_n x + b_n \text{ o } Y > c_n y + d_n\}$  depende solamente de las probabilidades marginales  $P(X > a_n x + b_n)$  y  $P(Y > c_n y + d_n)$ .

Una forma de visualizar la estructura de dependencia es a través de las *curvas cuantil*,  $Q_c$ , definidas según

$$Q_c = \{(x, y) | 1 - F(x, y) = c\}$$

Teniendo en cuenta la relación 4.11 se deduce que las curvas  $Q_c$  surgen como la solución asintótica de

$$l(1 - F_1(x), 1 - F_2(y)) = c \quad (4.12)$$

La curva  $Q_c$  tiene las siguientes propiedades:

1. Conecta los puntos  $(c, 0), (0, c)$
2. Es una función cóncava
3. La forma de  $Q_{c_1}$  es la misma que  $Q_{c_2}$  para  $c_1 \neq c_2$
4. En caso de independencia las curvas cuantil están dadas por  $\{(x, y) | x + y = c\}$  mientras que en el caso de dependencia total están dadas por  $\{(x, y) | \max(x, y) = c\}$ .

En este sentido, el grado de dependencia entre ambas variables viene determinado por el grado de curvatura que presenten las curvas cuantil.

Una forma diferente de estudiar la estructura de dependencia es a través de la *medida espectral*  $\Phi(\theta)$  de la distribución  $G$  (Einmahl *et al.* 1997).

A través de un cambio a coordenadas polares y considerando el conjunto definido por

$$B_{\theta,r} = \left\{ (u,v) \mid u^2 + v^2 > r^2, \arctan\left(\frac{u}{v}\right) \leq \theta \right\} \quad (4.13)$$

donde  $\theta \in [0, \pi/2]$  y  $r > 0$ . Dado que  $B_{\theta,r} = rB_{\theta,1}$  y utilizando la propiedad de homogeneidad de la función  $l$

$$\nu_0(B_{\theta,r}) = r\nu_0(B_{\theta,1}) \quad (4.14)$$

El conjunto  $B_{\theta,r}$  se representa en la figura 4.2. La función  $\nu_0(B_{\theta,1})$  es de hecho la función de distribución de una medida finita en el intervalo  $[0, \pi/2]$ . Esta medida es conocida como *medida espectral* de la distribución de valores extremos. De esta forma la medida  $\nu_0$  es determinada por su valor sobre todos los conjuntos de la forma  $B_{\theta,r}$ . Esto significa que la función de distribución

$$\Phi(\theta) = \nu_0(B_{\theta,1}) \quad (4.15)$$

de la medida espectral, determina  $\nu_0$  y, por lo tanto, también  $G$ . De esta forma  $G$  es caracterizada por tres cantidades:  $\gamma_1$ ,  $\gamma_2$  y la medida espectral  $\Phi$ .

Existen dos casos extremos para  $\Phi$ . Si  $\Phi$  está concentrada sobre los dos puntos extremos  $0$  y  $\pi/2$ , entonces  $G(x,y) = G(x,\infty)G(\infty,y)$  o lo que es lo mismo, las dos variables son independientes. Si  $\Phi$  está concentrada sobre algún punto en el intervalo  $(0, \pi/2)$  entonces existe dependencia total, o dicho de otra forma, una de las componentes queda perfectamente determinada por la otra de forma que

$$G(x,y) = \min \{G(x,\infty), G(\infty,y)\}$$

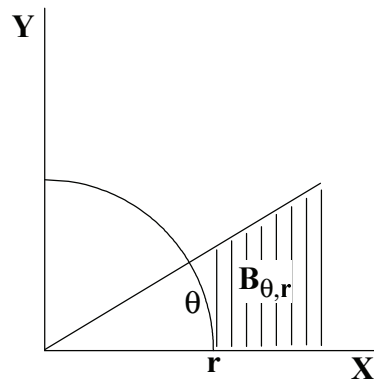


Figura 4.2: Representación en coordenadas polares del conjunto  $B_{\theta,r}$

### 4.3 Estimación

Una vez establecidas las funciones teóricas  $l$  y  $\Phi$ , es necesario desarrollar estimadores basados en las observaciones  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  con distribución inicial  $F$ . Es importante notar que  $-\log G$  es una aproximación sólo de la cola alta ( $1 - F(a_n x + b_n, c_n y + d_n)$ ) de la distribución original. En este sentido, solamente las observaciones más altas, o valores extremos, de  $F$  dan información a cerca de  $G$  o  $l$ , por lo que el número de extremos a utilizar,  $m$ , debería ser una pequeña fracción del número total de datos  $n$ . Entonces, la siguiente relación debería ser satisfecha

$$m = m(n) \rightarrow \infty, \quad \frac{m(n)}{n} \rightarrow 0 \quad (n \rightarrow \infty)$$

Partiendo de la relación 4.10 y reemplazando  $p$  por  $m/n$  es posible escribir

$$\lim_{n \rightarrow \infty} \frac{n}{m} Pr \left\{ 1 - F_1(X) < \frac{m}{n}x \text{ o } 1 - F_2(Y) < \frac{m}{n}y \right\} = l(x, y) \quad (4.16)$$

Se pretende convertir el primer término de la igualdad anterior en un estimador para  $l$ . Para ello se reemplaza  $Pr$  por su medida empírica  $P_n$  y  $F_1$  y  $F_2$  por sus respectivas medidas empíricas  $F_1^{(n)}$  y  $F_2^{(n)}$  dejando

$$\begin{aligned}
\hat{l}(x, y) &= \frac{n}{m} \frac{1}{n} \sum_{i=1}^n I \left\{ n(1 - F_1^{(n)}(X_i)) < mx \text{ o } n(1 - F_2^{(n)}(Y_i)) < my \right\} \\
&= \frac{1}{m} \sum_{i=1}^n I \left\{ X_i > X_{n-[mx],n} \text{ o } Y_i > Y_{n-[my],n} \right\} \\
&= \frac{1}{m} \sum_{i=1}^n I \left\{ R_i^X > n - mx \text{ o } R_i^Y > n - my \right\} \tag{4.17}
\end{aligned}$$

donde  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  son los estadísticos ordenados a partir de  $(X_1, X_2, \dots, X_n)$  y  $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$  los estadísticos ordenados a partir de  $(Y_1, Y_2, \dots, Y_n)$ ;  $R_i^X = \text{rango}(X_i)$  entre  $(X_1, X_2, \dots, X_n)$  y  $R_i^Y = \text{rango}(Y_i)$  entre  $(Y_1, Y_2, \dots, Y_n)$ . El estimador  $\hat{l}$  es un estimador consistente de  $l$ .

Bajo asunciones adicionales el proceso estocástico

$$\sqrt{m} \left\{ \hat{l}(x, y) - l(x, y) \right\}$$

converge en  $D([0, 1] \times [0, 1])$  a un proceso Gaussiano de media nula y estructura de covarianza conocida (Huang, 1992).

De forma similar es posible construir un estimador para la función  $\Phi$  de la medida espectral. Procediendo de forma análoga a la realizada en el caso de  $l$ , tomando  $r = m/n$

$$\hat{\Phi}(\theta) = \frac{1}{m} \sum_{i=1}^n I \left\{ n - R_i^X \vee R_i^Y < m, \frac{n - R_i^X}{n - R_i^Y} \leq \tan \theta \right\} \tag{4.18}$$

donde  $x \vee y = \max(x, y)$ .

El estimador  $\hat{\Phi}$  es un estimador consistente de  $\Phi$ .

Bajo asunciones adicionales (Einmahl et al. 1997) el proceso estocástico

$$\sqrt{m}(\hat{\Phi}(\theta) - \Phi(\theta))$$

converge a un proceso Gaussiano con media nula y estructura de covarianza conocida.



El proceso de estimación se lleva a cabo utilizando los rangos de los datos observados, sin hacer ningún tipo de presunción a cerca de la distribución de probabilidad a la que pertenecen dichos datos, por lo que se dice que el proceso de estimación es no paramétrico.

Una forma alternativa de estimar  $l$  y  $\Phi$  es a través de las relaciones

$$\hat{l}' = \frac{1}{m} \sum_{i=1}^n I \left\{ \hat{X}_i \left( \frac{n}{m} \right) > \frac{1}{x} \text{ o } \hat{Y}_i \left( \frac{n}{m} \right) > \frac{1}{y} \right\} \quad (4.19)$$

y

$$\hat{\Phi}'(\theta) = \frac{1}{m} \sum_{i=1}^n I \left\{ \hat{X}_i \left( \frac{n}{m} \right) \vee \hat{Y}_i \left( \frac{n}{m} \right) > 1, \arctan \left( \hat{Y}_i \left( \frac{n}{m} \right) / \hat{X}_i \left( \frac{n}{m} \right) \right) \leq \theta \right\} \quad (4.20)$$

donde

$$\begin{aligned} \hat{X}_i \left( \frac{n}{m} \right) &= \left( 1 + \hat{\gamma}_1 \frac{X_i - \hat{b}(n/m)}{\hat{a}(n/m)} \right)^{1/\hat{\gamma}_1} \\ \hat{Y}_i \left( \frac{n}{m} \right) &= \left( 1 + \hat{\gamma}_2 \frac{Y_i - \hat{d}(n/m)}{\hat{c}(n/m)} \right)^{1/\hat{\gamma}_2} \end{aligned} \quad (4.21)$$

Los estimadores  $\hat{l}'$  y  $\hat{\Phi}'$  son consistentes. Bajo asunciones adicionales

$$\sqrt{m} \left\{ \hat{l}'(x, y) - l(x, y) \right\} \quad y \quad \sqrt{m} \left\{ \hat{\Phi}'(\theta) - \Phi(\theta) \right\}$$

convergen a un proceso Gaussiano de media nula y estructura de covarianza conocida (de Haan & Resnick, 1993).

En este caso el proceso de estimación se lleva a cabo utilizando la función de distribución generalizada de valores extremos (GEV) para transformar los datos, por lo que el proceso de estimación se dice que es paramétrico.

Los estimadores  $\hat{\gamma}_1, \hat{\gamma}_2, \hat{a}(n/m), \hat{b}(n/m), \hat{c}(n/m)$  y  $\hat{d}(n/m)$  son los propuestos para el caso unidimensional.

Los estimadores propuestos en Dekkers *et al.* (1989) vienen dados de la siguiente forma:

Se definen las funciones

$$\bar{t} = t \wedge 0, \quad \rho_1 = \frac{1}{1 - \bar{t}}, \quad \rho_2 = \frac{2}{(1 - \bar{t})(1 - 2\bar{t})}$$

y para  $r = 1, 2$

$$H_{1,r} = m^{-1} \sum_{i=0}^{m-1} \{\log X_{n-i,n} - \log X_{n-m,n}\}^r$$

$$H_{2,r} = m^{-1} \sum_{i=0}^{m-1} \{\log Y_{n-i,n} - \log Y_{n-m,n}\}^r$$

Los estimadores toman la forma

$$\hat{\gamma}_j = H_{j,1} + 1 - \frac{1}{2} \left( 1 - \frac{H_{j,1}^2}{H_{j,2}} \right)^{-1}$$

$$\hat{b}(n/m) = X_{n-m,n}$$

$$\hat{a}(n/m) = \frac{X_{n-m,n} \sqrt{3H_{1,1}^2 - H_{1,2}}}{\sqrt{(\rho_1(\gamma_1))^2 - \rho_2(\gamma_1)}}$$

$$\hat{d} = Y_{n-m,n}$$

$$\hat{c}(n/m) = \frac{Y_{n-m,n} \sqrt{3H_{2,1}^2 - H_{2,2}}}{\sqrt{(\rho_1(\gamma_2))^2 - \rho_2(\gamma_2)}}$$

## 4.4 Coeficiente de independencia asintótica

El modelo bivariado de valores extremos descrito en la sección anterior asume la existencia de una dependencia asintótica entre las dos variables analizadas. Esto quiere decir que en el caso de que las dos variables presenten independencia asintótica, el modelo no será capaz de evaluar la probabilidad de un evento extremo. Esto puede verse retomando la relación 4.7 y reescribiéndola en la forma

$$\lim_{n \rightarrow \infty} nP \left\{ \frac{X - b_n}{a_n} > x \quad o \quad \frac{Y - d_n}{c_n} > y \right\} = -\log G(x, y) \quad (4.22)$$

y por lo tanto

$$\lim_{n \rightarrow \infty} nP \left\{ \frac{X - b_n}{a_n} > x, \frac{Y - d_n}{c_n} > y \right\} = \log G(x, y) - \log G_1(x) - \log G_2(y) \quad (4.23)$$

De esta forma, si las dos variables marginales son independientes, esto es  $G(x, y) = G_1(x)G_2(y)$ , la parte derecha de la ecuación anterior es exactamente cero.

Esto quiere decir que cuando las variables son independientes se asume una independencia exacta. Sin embargo, esta situación parece algo restrictiva, ya que en algunos casos podemos esperar que los extremos de las variables se aproximen a la independencia de forma asintótica más que mostrar una independencia exacta para valores grandes de las variables.

Con motivo de evitar esta situación, Ledford y Tawn (1996, 1997) introducen un submodelo en el cual la dependencia asintótica puede ser estudiada a través de un coeficiente  $\eta \in (0, 1]$ . De esta forma, la dependencia asintótica tiene lugar para  $\eta = 1$ , mientras que el caso de  $\eta < 1$  implica independencia asintótica. Ledford y Tawn sugieren también estimadores para este coeficiente de dependencia asintótica  $\eta$ , pero no establecen sus propiedades asintóticas.

Draisma *et al.* (2004) proponen una variación del estimador de  $\eta$  propuesto por Ledford y Tawn (1996) además de probar la normalidad asintótica de dicho estimador.

Por tanto, a través de este coeficiente  $\eta$  es posible realizar el siguiente contraste de hipótesis

$$H_0 : \quad \eta = 1 \quad (\text{dependencia asintotica})$$

$$H_1 : \quad \eta < 1 \quad (\text{independencia asintotica})$$

En el presente trabajo, para la estimación del coeficiente  $\eta$ , se utilizarán los estimadores propuestos en Draisma *et al.* (2004) y Peng (1999), los cuales son descritos en la siguiente sección.

#### 4.4.1 Estimador de Hill

El estimador de Hill fue introducido por el mismo autor en Hill (1975). Draisma *et al.* (2004) proponen una ligera variación en la forma de usar el estimador, definiéndolo según

$$\hat{\eta}_1 = \frac{1}{m} \sum_{i=1}^m \log \frac{T_{n,n-i+1}^{(n)}}{T_{n,n-m}^{(n)}} \quad (4.24)$$

donde  $m$  representa el número de estadísticos ordenados, o extremos, utilizado en el análisis, mientras que

$$T_i^{(n)} = \min \left( \frac{n+1}{n+1-R_i^X}, \frac{n+1}{n+1-R_i^Y} \right) \quad (4.25)$$

donde  $R_i^X$  representa el rango de  $X_i$  entre  $(X_1, X_2, \dots, X_n)$  y  $R_i^Y$  el rango entre de  $Y_i$  entre  $(Y_1, Y_2, \dots, Y_n)$ .

Bajo las asunciones adecuadas es posible demostrar que  $\sqrt{m}(\hat{\eta}_1 - \eta)$  es asintóticamente normal con media 0 y varianza

$$\sigma_1^2 = \eta_1^2(1-l)(1-2lc_x(1,1)c_y(1,1)) \quad (4.26)$$

Con motivo de construir intervalos de confianza para  $\eta_1$  o contrastar la hipótesis  $\eta_1 = 1$ , necesitamos estimadores consistentes para las cantidades presentes en la expresión 4.26 (Draisma *et al* 2004):

$$\hat{l} = \frac{m}{n} T_{n,n-m}^{(n)} \quad (4.27)$$

$$\hat{c}_x(1,1) = \frac{\hat{k}^{5/4}}{n} (T_{n,n-m}^{(n,\hat{k}^{-1/4})} - T_{n,n-m}^{(n)}) \quad (4.28)$$

con  $\hat{k} = m/\hat{l}$  y  $T_{n,i}^{(n,u)}$ ,  $i = 1, \dots, n$  los estadísticos ordenados de

$$T_i^{(n,u)} = \min \left( \frac{n+1}{n+1-R_i^X} (1+u), \frac{n+1}{n+1-R_i^Y} \right) \quad (4.29)$$

definiendo  $\hat{c}_y(1, 1)$  de forma análoga a  $\hat{c}_x(1, 1)$ .

Por tanto, como estimador de  $\sigma_1^2$  puede utilizarse

$$\hat{\sigma}_1^2 = (1 + \hat{\eta})^2(1 - \hat{l})(1 - 2\hat{l}\hat{c}_x(1, 1)\hat{c}_y(1, 1)) \quad (4.30)$$

#### 4.4.2 Estimador de Peng

Este estimador fue introducido por Peng (1999) en la forma

$$\hat{\eta}_2 = \log 2 / \log \left( \frac{S_n(2m)}{S_n(m)} \right) \quad (4.31)$$

con  $S_n(m) = \sum_{i=1}^n I\{X_i > X_{n,n-m} \text{ y } Y_i > Y_{n,n-m}\}$ .

Bajo asunciones adecuadas puede demostrarse que

$$2 \log(2)c(1, 1)\sqrt{m}(\hat{\eta}_2 - 1) \xrightarrow{d} N(0, \sigma^2) \quad (4.32)$$

donde

$$\begin{aligned} \sigma^2 = & 2c(1, 1)[1 - 4c_1(1, 1) - 4c_2(1, 1) + 6c_1(1, 1)c_2(1, 1)] \\ & + 4c(1, 2)c_1(1, 1)[1 - c_2(1, 1)] \\ & + 4c(2, 1)c_2(1, 1)[1 - c_1(1, 1)] + 2c_1(1, 1) + 2c_2(1, 1) \end{aligned} \quad (4.33)$$

Estimadores consistentes para  $c(x, y)$ ,  $c_1(1, 1)$  y  $c_2(1, 1)$  vienen dados por

$$\begin{aligned} \hat{\sigma}^2 = & 2\hat{c}(1, 1)[1 - 4\hat{c}_1(1, 1) - 4\hat{c}_2(1, 1) + 6\hat{c}_1(1, 1)\hat{c}_2(1, 1)] \\ & + 4\hat{c}(1, 2)\hat{c}_1(1, 1)[1 - \hat{c}_2(1, 1)] \\ & + 4\hat{c}(2, 1)\hat{c}_2(1, 1)[1 - \hat{c}_1(1, 1)] + 2\hat{c}_1(1, 1) + 2\hat{c}_2(1, 1) \end{aligned} \quad (4.34)$$

Las cantidades  $\hat{c}(x, y)$ ,  $\hat{c}_1(1, 1)$  y  $\hat{c}_2(1, 1)$  están definidas por

$$\begin{aligned}
\psi_1 &= \frac{1}{m} \sum_{i=1}^n I(X_i > X_{n,n-[m(1+m^{-1/4})]} \text{ and } Y_i > Y_{n,n-m}) \\
\psi_2 &= \frac{1}{m} \sum_{i=1}^n I(X_i > X_{n,n-m} \text{ and } Y_i > Y_{n,n-m}) \\
\psi_3 &= \frac{1}{m} \sum_{i=1}^n I(X_i > X_{n,n-m} \text{ and } Y_i > Y_{n,n-[m(1+m^{-1/4})]}) \\
\hat{c}(x, y) &= \frac{1}{m} \sum_{i=1}^n I(X_i > X_{n,n-mx} \text{ and } Y_i > Y_{n,n-my}) \\
\hat{c}_1(1, 1) &= m^{1/4} \{\psi_1 - \psi_2\} \\
\hat{c}_2(1, 1) &= m^{1/4} \{\psi_3 - \psi_2\}
\end{aligned}$$

Los intervalos de confianza pueden definirse a través de las cantidades

$$I_2(m) = -\frac{1.64\sigma}{2\hat{c}(1, 1)\sqrt{m} \log 2} + 1$$

con

$$\lim_{n \rightarrow \infty} P(\hat{\eta}_2 \leq I_2(m)) = 0.05$$

## 4.5 Método de selección de los eventos extremos

Las variables altura de ola (Hs) y nivel del mar (SWL) son analizadas de forma conjunta con el objetivo de poner en práctica la metodología descrita. El conjunto de datos de oleaje analizado corresponde a las boyas LP-I y LP-II, mientras que el conjunto de datos del nivel del mar pertenece al mareógrafo de Las Palmas, ubicado en el Puerto de Las Palmas.

La metodología utilizada para obtener los picos independientes de cada variable es la misma que la utilizada en el Capítulo 3 para el caso univariante, propuesta por Simiu (1996). De esta forma, se selecciona una ventana temporal de 4 días dentro de la cual se selecciona el valor máximo de cada variable, garantizando además que los picos adyacentes están separados como mínimo 2 días.

Como resultado, para el par de variables  $H_s$ -SWL, se obtienen 714 eventos o picos durante el periodo 1992-2001, mostrados en la Figura 4.3.

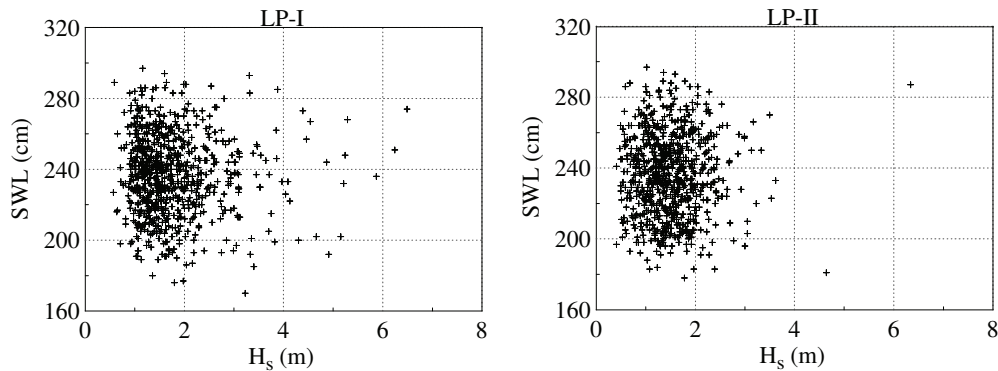


Figura 4.3: Valores conjuntos de altura de ola significativa ( $H_s$ ) y nivel del mar (SWL) registrados durante 714 tormentas.

## 4.6 Resultados y discusión

El análisis comienza con la estimación de los parámetros de las distribuciones extremales marginales,  $a(n/m)$ ,  $b(n/m)$ ,  $c(n/m)$ ,  $d(n/m)$ ,  $\gamma_1$  y  $\gamma_2$ .

La figura 4.4 muestra las estimaciones del parámetro de forma  $\hat{\gamma}$  para diferentes valores de  $m$ . En base al comportamiento del parámetro de forma a medida que aumenta el número de extremos se ha procedido a tomar un valor de  $m=50$  para los ajustes posteriores.

La tabla 4.1 muestra las estimaciones de los parámetros de la distribución extremal, mientras que la figura 4.5 muestra el ajuste de la distribución extremal a la cola de las distribuciones marginales empíricas.

El ajuste entre la distribución teórica y los datos empíricos en el caso de  $H_s$  en la boya LP-II y SWL parece aceptable. Sin embargo, para el caso de  $H_s$  en la boya LP-I parece haber una desviación de la distribución teórica respecto de los datos observados.

La dependencia asintótica es analizada a través de las curvas cuantil  $Q_c$  y de la medida espectral  $\Phi$ . La figura 4.6 muestra las curvas cuantil estimadas para ambas boyas. En la parte izquierda el conjunto  $\{(x, y) | \hat{l}(x, y) = c\}$  se obtiene de forma no paramétrica, usando los rangos de la muestra, para  $c = 0.140, 0.112, 0.098, 0.084, 0.070, 0.0560, 0.042$  y  $0.028$ . En la parte derecha el conjunto  $\{(x, y) | \tilde{l}(x, y) = c\}$  es calculado de forma paramétrica

Tabla 4.1: Estimación de parámetros

Hs						SWL		
LP			LP-II			LP		
$\gamma_{LP-I}$	$a_{LP-I}$	$b_{LP-I}$	$\gamma_{LP-II}$	$a_{LP-II}$	$b_{LP-II}$	$\gamma_{LP-II}$	c	d
0.100	0.691	3.070	0.255	0.421	2.390	-0.403	12.846	270.000

usando  $(\hat{X}(n/k), \hat{Y}(n/k))$  para los mismos valores de  $c$ . Los gráficos basados en los rangos parecen mostrar los resultados de forma más clara y precisa. Las propiedades de las curvas cuantil ya comentadas, linealidad e igualdad de forma en caso de independencia, parecen ser ciertas para las curvas estimadas. De este modo, debemos concluir que los valores extremos conjuntos de altura de ola significativa y nivel del mar son independientes en las dos localizaciones analizadas. Dicho de otro modo, la presentación de eventos extremos de altura de ola significativa no llevan asociados necesariamente valores extremos del nivel del mar, y viceversa.

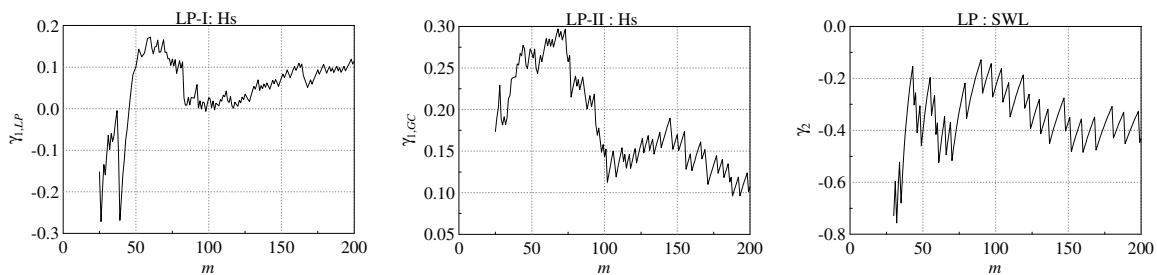


Figura 4.4: Estimaciones del parámetro de forma  $\gamma$  en función del número de extremos  $m$  para la altura de ola significativa  $H_s$  y nivel del mar SWL

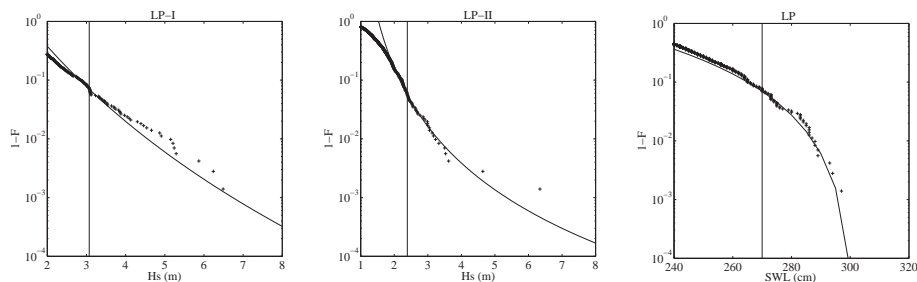


Figura 4.5: Probabilidades de excedencia marginales. La línea continua representa la estimación paramétrica. El símbolo (+) representa las estimaciones empíricas. La línea vertical representa el número de extremos  $(n - m)$  utilizados en el análisis, donde  $n = 714$  y  $m = 50$ .



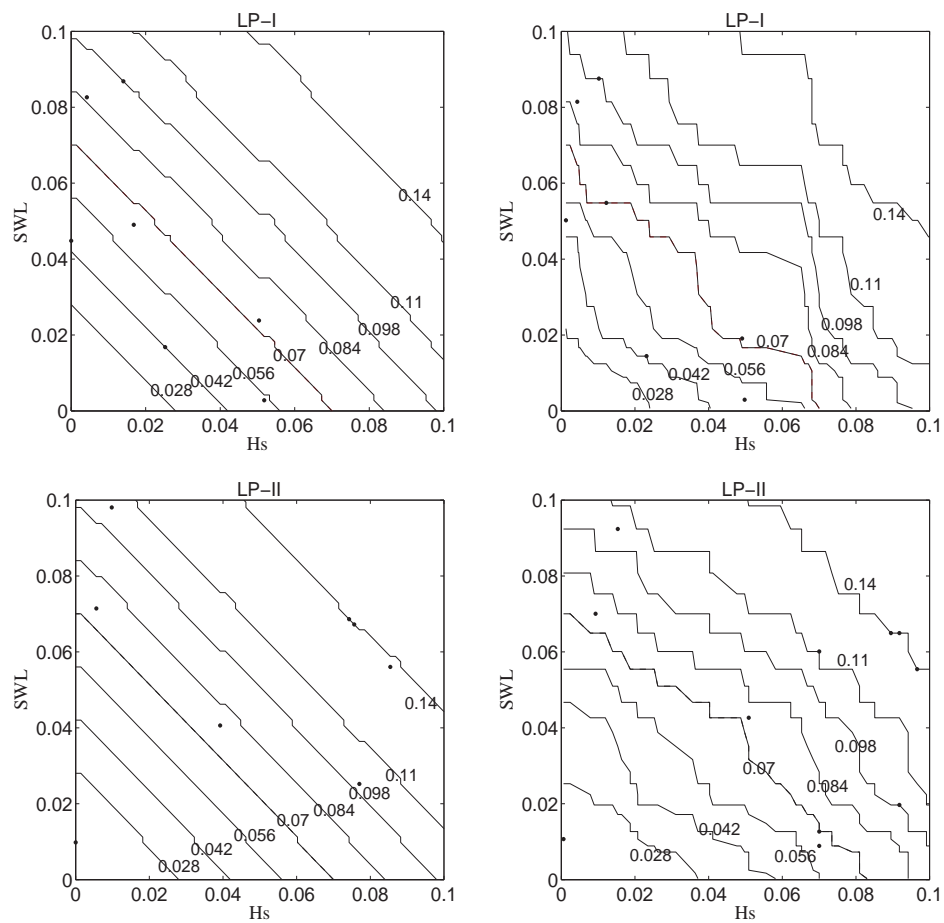


Figura 4.6: Curvas cuantil. Superior: Boya LP-I. Inferior: Boya LP-II. Niveles para la función  $\hat{l}$  (no-paramétrica)(izquierda) y  $\hat{l}'$  (paramétrica) (derecha). Las curvas cuantil corresponden a  $m = 20, 30, 40, 50, 60, 70, 80$  and  $100$

La figura 4.7 muestra las estimaciones de la medida espectral  $\Phi$ . En la parte superior se muestran los resultados correspondientes a la boya LP-I mientras que en la parte inferior los correspondientes a la boya LP-II. En la parte izquierda se presentan los gráficos basados en los rangos de la muestra y en la parte derecha los gráficos basados en las estimaciones paramétricas  $(\hat{X}_i(n/k), \hat{Y}_i(n/k))$ . Debido a que las observaciones están agrupadas en torno a los puntos 0 and  $\pi/2$ , debemos concluir que los valores extremos de ambas variables son independientes.

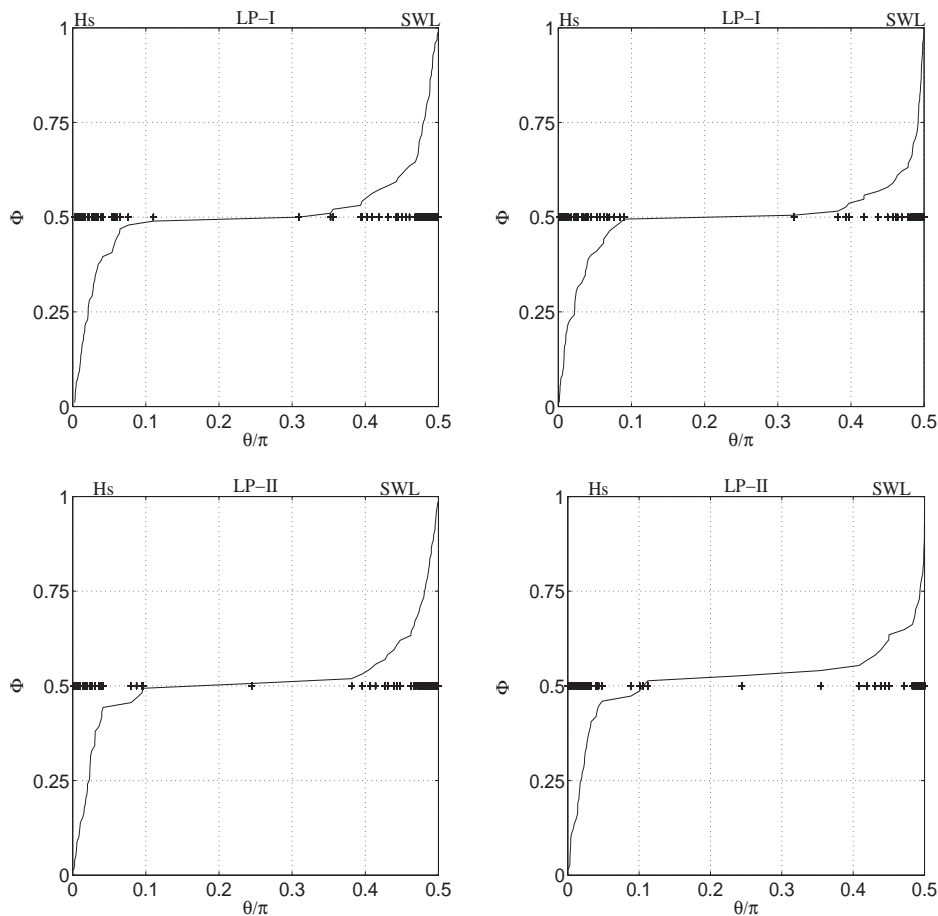


Figura 4.7: Medida espectral. Superior: Boya LP-I; Inferior: Boya LP-II. Valores individuales de  $\theta_i$  se muestran con el símbolo '+'. La línea continua representa la función de distribución  $\theta$ . En la parte derecha se muestran las estimaciones usando los rangos; en la parte izquierda se muestran las estimaciones usando  $(\hat{X}_i(n/m), \hat{Y}_i(n/m))$ . El eje horizontal se mueve en un rango de valores desde 0 (Hs) a  $\pi/2$  (SWL).

Las estimaciones del coeficiente de independencia, a través de los estimadores de Hill y Peng se muestran en las figuras 4.8 y 4.9, respectivamente, para las dos localizaciones analizadas. Las estimaciones obtenidas con el estimador de Hill nos llevan a rechazar la

hipótesis nula, es decir,  $\eta = 1$  (dependencia asintótica), debido a que las estimaciones están por debajo del límite de aceptación de la hipótesis nula. Por este motivo, se concluye que existe independencia asintótica entre los valores extremos de  $H_s$  y SWL, en los dos puntos analizados.

Por otro lado, los resultados obtenidos a través del estimador de Peng tienen difícil interpretación. Atendiendo solamente al valor del estimador  $\eta_2$  podría decirse que los valores son similares a los obtenidos con el estimador de Hill excepto para valores de  $m$  entre 30 y 60 aproximadamente en la boya LP-I.

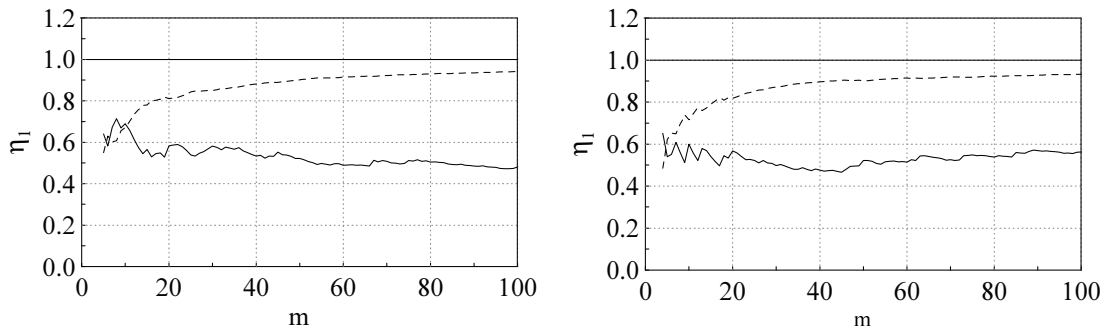


Figura 4.8: Estimador de Hill para el coeficiente de dependencia (línea continua) y límite de aceptación al 5 % para  $\eta = 1$  (línea discontinua). Izquierda: Boya LP-I. Derecha: Boya LP-II

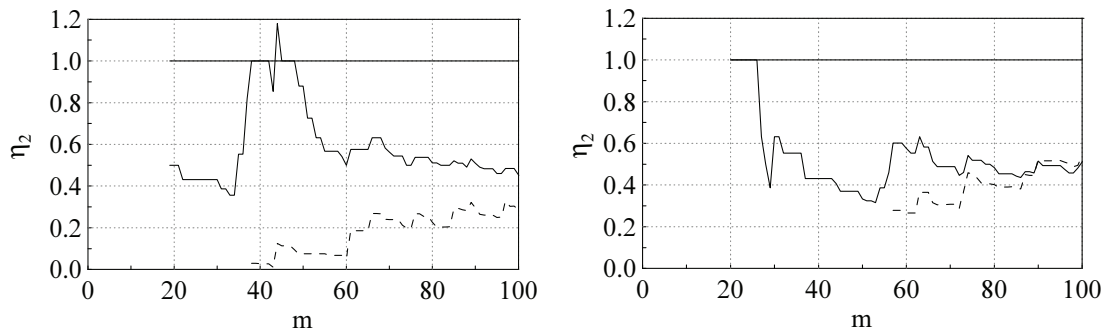


Figura 4.9: Estimador de Peng para el coeficiente de dependencia (línea continua) y límite de aceptación al 5 % para  $\eta = 1$  (línea discontinua). Izquierda: Boya LP-I. Derecha: Boya LP-II

El mayor problema se presenta en el cálculo del límite de aceptación de la hipótesis nula, que solamente puede ser obtenido a partir de un valor de  $m=40$  para la boya LP-I y  $m=60$  para la boya LP-II, dado que para valores inferiores de  $m$  la varianza resulta negativa. Teniendo en cuenta esto, los valores del estimador nos llevarían a concluir dependencia asintótica, ya que, los valores del coeficiente  $\eta$ , están dentro de la región de aceptación de  $\eta = 1$ . Sin embargo, teniendo en cuenta los resultados obtenidos a través del estimador de Hill, de las curvas cuantil y la medida espectral, no parece lógico aceptar este resultado de dependencia asintótica.

Este comportamiento inestable del estimador de Peng y de su varianza asociada puede ser explicado en base a la alta sensibilidad que muestran respecto de la estructura bivariada de los datos. De esta forma, una visión algo más detallada de la forma en que trabaja el estimador y su varianza, revelan una forma de trabajar basada en tan solo unos pocos valores extremos. Es decir, para un valor de  $m=20$ , el estimador no utiliza los 20 extremos, sino que de esos 20 datos utiliza aquellos que cumplen cierta condición. Sin embargo, el estimador de Hill, cuando  $m=20$ , utiliza los 20 extremos para el análisis. De ahí que, existiendo un valor estimado de  $\eta$  cuando  $m=20$  con el estimador de Hill, no exista esta estimación para el estimador de Peng. Algo similar ocurre en el caso del análisis extremal univariado donde el estimador de Pickands, que depende sólo de unos pocos extremos, se muestra menos preciso que el estimador de Hill (Drees, 1995).

Del análisis realizado sobre las series temporales de LP-I y LP-II el principal resultado obtenido es que no existe relación directa entre los valores extremos de la altura de ola significativa y el nivel del mar en la zona geográfica analizada. Este resultado está basado en las curvas cuantil, la medida espectral de la función extremal bivariada  $G$  y el coeficiente de independencia asintótica  $\eta$ .

Teniendo en cuenta los fenómenos físicos que entran en juego en la generación de los valores extremos del oleaje y del nivel del mar, podría pensarse que este resultado de independencia no resulta del todo convincente. De este modo, parecería lógico pensar que los valores extremos del oleaje están generados bajo condiciones atmosféricas adversas de baja presión, fenómeno que favorece a su vez un aumento en el nivel del mar total.

Varias son las razones que pueden llevar a concluir un resultado como éste. Por una lado, es posible que los fenómenos atmosféricos que favorecen aumentos sensibles del nivel del mar coincidan con la bajamar, de forma que el nivel total del mar no alcanza un valor extremo. En este sentido, un filtrado de la componente de marea podría eliminar este efecto, de forma que analizando la altura de ola y el residuo meteorológico (storm surge) podría llevar a una mayor correlación. Precisamente, en el trabajo publicado por Hawkes *et al.* (2002) se pone de manifiesto una mayor correlación entre los valores extremos del oleaje y el nivel del mar en aquellos lugares donde el rango de marea es pequeño, o lo que es lo mismo, donde no existe una componente de marea que enmascare la magnitud del residuo meteorológico. En Canarias, la altura del nivel del mar total suele presntear un valor medio de 2 metros, valor más que suficiente para enmascarar los efectos de los residuos meteorológicos.

Por otro lado, una breve descripción de la climatología en el área de estudio puede explicar este resultado de independencia. Las islas Canarias están situadas en una latitud próxima a los 28°N, dentro de la influencia de los vientos Alisios, que soplan durante todo el año generando un oleaje de viento, también conocido como *sea*. Por otro lado, el archipiélago canario está afectado por el oleaje de fondo, o tipo *swell*, generado por perturbaciones atmosféricas localizadas en la zona Noroeste del océano Atlántico. Los estados de mar resultantes suelen ser combinaciones de ambos tipos de oleaje, recibiendo el nombre de *mares mixtos*, con espectros frecuenciales característicos bimodales (Rodríguez, 1992). Este oleaje tipo *swell* no está relacionado con las condiciones atmosféricas locales, por lo que es posible tener situaciones donde el oleaje muestre valores extremos debido a un fuerte *swell* al mismo tiempo que las condiciones atmosféricas locales son de calma. De esta forma es posible observar valores extremos del oleaje con valores extremos y normales del nivel del mar.

Otro factor que puede enmascarar la dependencia entre ambas variables podría ser la existencia de un desfase temporal entre el valor máximo de cada variable.

A modo de ejemplo se ilustra a continuación un caso de dependencia asintótica entre las variables  $H_s$  y SWL en las costas holandesas. La figura 4.10 muestra una relación directa entre ambas variables de forma que los valores más altos de la altura de ola tienden a ocurrir en el mismo instante que los valores mas altos del nivel del mar. Esta relación directa se ve corroborada asintóticamente a través de las curvas cuantil en la figura 4.11(a) mostrando una curvatura importante y a través de la medida espectral 4.11(b) donde los puntos tienden a concentrarse en el centro de la gráfica.

Esta relación tan clara entre los valores extremos de ambas variables se pone de manifiesto en zonas en las que el oleaje generado por el viento en la zona (tipo *sea*), no está influenciado o mezclado con el oleaje de fondo (*swell*). Esta es la situación que se da en las costas holandesas debido a la localización geográfica de las mismas ofreciendo un escenario particularmente interesante para este tipo de estudios.

Sin embargo, cabe destacar que Peng (1999) analiza el mismo conjunto de datos haciendo uso del estimador propuesto por él mismo. El resultado de dicho análisis lleva a concluir que no se puede rechazar la independencia asintótica entre la altura de ola y el nivel del mar, poniendo de manifiesto la existencia de cierto grado de incertidumbre en estas técnicas de análisis extremal multivariante.

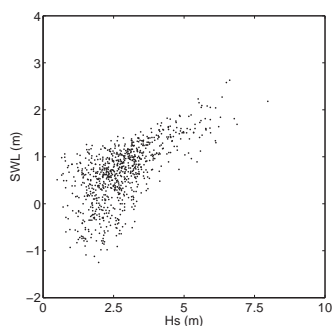
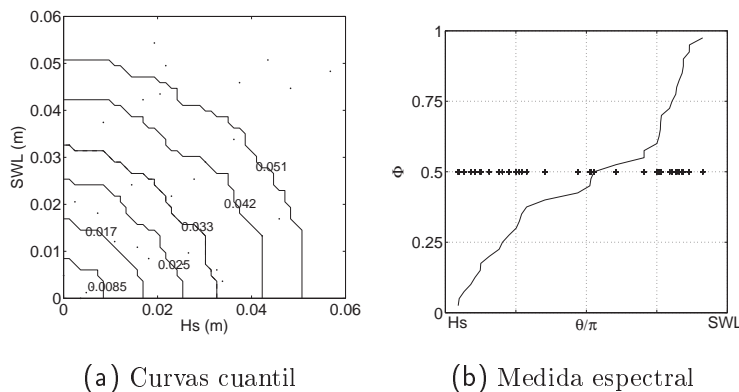


Figura 4.10: Relación entre  $H_s$  y SWL durante 718 tormentas registradas en las costas holandesas. Extraído del artículo de Ronde, J. de and de Haan, L. (1998) y reproducido con el permiso de los autores.



(a) Curvas cuantil

(b) Medida espectral

Figura 4.11: Curvas cuantil y medida espectral para las tormentas representadas en la figura 4.10

---

## Conclusiones

---

El análisis a través del método de máximos anuales sobre el conjunto de datos de la boya LP-I permite concluir que las distribuciones de valores extremos clásicas, es decir, Gumbel, Fréchet y Weibull pueden caracterizar adecuadamente la serie de extremos. Sin embargo, teniendo en cuenta los gráficos cuantil-cuantil, la distribución de Weibull es la que mejor ajuste presenta. Para dicha distribución, el método de estimación de parámetros que ofrece mejores resultados es el de mínimos cuadrados.

Los ajustes de las distribuciones GEV, K4P, GAM y WAK utilizando el método de estimación de parámetros LMOM ponen de manifiesto que la distribución K4P sería la mejor opción para caracterizar la serie de máximos anuales, mejorando sensiblemente el ajuste proporcionado por la distribución de Weibull.

En el conjunto de datos de la boya LP-II, es la distribución de Fréchet la que mejor ajuste ofrece, siendo capaz de caracterizar adecuadamente la mayoría de los extremos obtenidos a través del método de máximos anuales, a excepción de los dos más altos, que parecen desviarse del comportamiento general de la serie. Los ajustes obtenidos, cuando se utilizan los métodos MOM y MLE en la estima de los parámetros, parecen caracterizar mejor la parte central de la serie, observándose una desviación importante en los dos datos más altos. En este sentido, existe la posibilidad de que estos valores puedan ser medidas erróneas. No obstante, al utilizar el método LS, la distribución es capaz de modelizar mejor los dos datos más extremos, si bien esto es a costa de una pequeña desviación en la parte

central de la serie. De esta forma, se observa como el método LS se encuentra influenciado por los datos más extremos en mayor grado que los métodos MOM y MLE. Las distribuciones GEV, GAM y WAK no mejoran el ajuste proporcionado por la distribución de Fréchet, mientras que la distribución K4P no pudo ser ajustada por fallo en la convergencia del algoritmo numérico durante la estimación de los parámetros.

El análisis de la independencia estadística de los extremos en las series temporales LP-I y LP-II mediante el test de rachas sugiere que la independencia estadística puede aceptarse a partir de un valor umbral de 2.0 metros en LP-I y 2.5 metros en LP-II.

La hipótesis de que la frecuencia de presentación de los extremos sigue una distribución de Poisson se corrobora a través del cálculo del índice de dispersión (DI) y los gráficos de Poisson. La distribución binomial negativa no mejora el ajuste a los datos observados ofrecido por la distribución de Poisson.

La dificultad en la selección del umbral se hace patente a través de las diferentes metodologías utilizadas para su elección. Para el caso de los datos de la boya LP-I, los métodos basados en las propiedades teóricas del modelo sugieren un valor umbral próximo a 2.3 metros, mientras que los métodos basados en el número de excedencias sugieren un rango de umbrales entre 2.8 y 3.5 metros. Para el caso de la boya LP-II, el primer conjunto de técnicas sugieren un rango de umbrales comprendido entre 1.7 y 2.1 metros, mientras que el segundo conjunto de técnicas sugieren un rango entre 2.2 y 2.4 metros.

La caracterización de la magnitud de las excedencias sobre un umbral a través de la distribución K4P revela una mayor eficiencia que su modelización mediante la distribución GPD, sobreestimando esta última los cuantiles asociados a periodos de retorno elevados. Este efecto es aliviado parcialmente por la distribución K4P al incluir un parámetro adicional que le confiere mayor flexibilidad para ajustar los valores extremos.

Dentro del apartado del análisis estacional, los diferentes test estadísticos utilizados ponen de manifiesto la existencia de estacionalidad en el conjunto de datos de la boya LP-I, mientras que por el contrario, en los datos de la boya LP-II no puede rechazarse la uniformidad.

De igual forma, el efecto de la estacionalidad se observa en los diferentes valores que los parámetros de la distribución GPD presentan en función de la época del año, así como en los cuantiles asociados a los periodos de retorno más altos. Esto se observa también en la boya LP-II, a pesar de que el test estadístico no puede rechazar la uniformidad. Este hecho ilustra una mayor complejidad en el comportamiento de los extremos en este conjunto de



---

datos, pudiendo existir un pequeño grado de estacionalidad en los extremos que no llega a tener el suficiente peso para poder rechazar la hipótesis de uniformidad con fiabilidad estadística.

La metodología propuesta en Ouarda (1993) para dividir la serie en periodos homogéneos, denominadas estaciones hidroclimáticas, pone de manifiesto, en general, la existencia de dos periodos temporales bien diferenciados. Un periodo denominado moderado, que agrupa las estaciones de primavera y verano, y otro, denotado como periodo severo, que incluye las estaciones de otoño e invierno. No obstante, en el conjunto de datos de la boya LP-I, podría identificarse el periodo de otoño como un periodo homogéneo independiente, donde se alcanzan los mayores valores de altura de ola.

Respecto al análisis multivariado de los extremos de altura de ola y nivel del mar, se desprende del análisis realizado sobre los extremos de LP-I y LP-II que no existe dependencia asintótica entre los valores extremos de ambas variables. Este resultado está basado en las curvas cuantil, la medida espectral de la función extremal bivariada  $G$  y el coeficiente de independencia asintótica  $\eta$ .

Varias son las razones que pueden llevar a concluir un resultado como éste. Por un lado, es posible que los fenómenos atmosféricos que favorecen aumentos sensibles del nivel del mar coincidan con la bajamar, de forma que el nivel total del mar no alcanza un valor extremo. En este sentido, un filtrado de la componente de marea podría eliminar este efecto, de forma que analizando la altura de ola y el residuo meteorológico (storm surge) podría llevar a una mayor correlación. Precisamente, en Hawkes *et al.* (2002) se pone de manifiesto una mayor correlación entre los valores extremos del oleaje y el nivel del mar en aquellos lugares donde el rango de marea es pequeño, o lo que es lo mismo, donde no existe una componente de marea que enmascare la magnitud del residuo meteorológico. En Canarias, la altura del nivel del mar total suele presentar un valor medio de 2.0 metros, valor más que suficiente para enmascarar los efectos de los residuos meteorológicos. Otro posible factor que puede enmascarar la dependencia entre el oleaje y el nivel del mar es la presencia de oleaje de fondo, que contribuye al valor de la altura de ola total pero no está relacionado con el viento local que genera el residuo meteorológico.



---

## Bibliografía

---

- [1] Ashkar, F., Roussel, J. (1987)  
*Partial duration series modeling under the assumption of a Poissonian flood count*  
Journal of Hydrology, 90, 135-144
- [2] Athanassoulis, G.A., Stefanakos, A. (1995)  
*A non-stationary stochastic model for long-term time series of significant wave height*  
Journal of Geophysical Research, 100 (C8), 16149-16162
- [3] Balkema, A., de Haan, L. (1974)  
*Residual life time at great age*  
Annals of Probability, 2, 792-804
- [4] Beguería, S. (2005)  
*Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value*  
Journal of Hydrology, 303, 215-230
- [5] Bendat, J.S., Piersol, A.G. (1986)  
*Random data. Analysis and Measurement Procedures*  
John Wiley & Sons, 2nd Edition, 566 pp
- [6] Ben-Zvi, A. (1991)  
*Observed advantage for negative binomial over Poisson distribution in partial duration series*  
Stochastic Hydrology and Hydraulics, 5, 135-146
- [7] Berger, R. (1993)  
*Extreme value analysis of the sea state at 'Nordsee' research platform and off Hörnum/Syly*  
Dt. Hydrogr. Zeits, 45, 107-113

- [8] Bertotti, L., Cavaleri, L. (2002)  
*Modelling waves at Orkney coastal locations*  
J. Marine Systems, 96-97, 116-121
- [9] Bidlot J.-R., Holmes, D.J., Wittmann, P.A., Lalbeharry, R., Chen, H.S. (2002)  
*Intercomparison of the performance of operational ocean wave forecasting systems with buoy data*  
Weather and Forecasting, 17, 287-310.
- [10] Bonazzi, A., Cusack, S., Mitas, C., Jewson, S. (2012)  
*The spatial structure of European wind storms as characterized by bivariate extreme-value Copulas*  
Natural Hazards and Earth System Sciences, 12, 1769-1782
- [11] Boos, D.D. (1984)  
*Using extreme value theory to estimate large percentiles*  
Technometrics, 26, 33-39
- [12] Buishand, T.A. (1989)  
*Statistics of extremes in climatology*  
Statistica Neerlandica, 43
- [13] Bhunya, P.K., Berndtsson, R., Sharad, K. Jain,, Kumar, R. (2013)  
*Flood analysis using negative binomial and Generalized Pareto models in partial duration series (PDS)*  
Journal of Hydrology, 497, 121-132
- [14] Canavos, G. (1988)  
*Probabilidad y Estadística: Aplicaciones y Métodos*  
Ed. McGraw Hill
- [15] Castillo, E., Hadi, S. (1997)  
*Fitting the Generalized Pareto Distribution to Data*  
Journal of the American Statistical Association, 92(440), 1609-1620
- [16] Cavaleri, L. (2009)  
*Wave modeling - Missing the peaks*  
J. Physical Oceanography, 39, 2757-2778
- [17] Coles, S., Tawn, J. (1991)  
*Modelling extreme multivariate events*  
Journal of the Royal Statistical Society B, 53(2), 377-392

- [18] Coles, S., Tawn, J. (1994)  
*Statistical Methods for multivariate extremes: an application to structural design*  
Applied statistics, 43(1), 1-48
- [19] Coles, S. (2001)  
*An Introducing to Statistical Modeling of Extreme Value*  
Ed. Springer
- [20] Coles, S., Simiu, E. (2003)  
*Estimating uncertainty in the extreme value analysis of data generated by a hurricane simulation model*  
J. Engineering Mechanics, 129(11), 1288-1294
- [21] Cunnane, C. (1973)  
*A particular comparison of annual maxima and partial duration series methods of flood frequency prediction*  
Journal of Hydrology, 18, 257-271
- [22] Cunnane, C. (1979)  
*A note on the Poisson Assumption in Partial Duration Series Models*  
Water Resources Research, 15(2), 489-494
- [23] David, H.A. (1981)  
*Order Statistics*  
2nd Ed. New York:Wiley
- [24] Davies, J.L. (1972)  
*Geographical variation in coastal development*  
Oliver & Boyd, Edinburgh
- [25] Davison, A.C., Smith, R.L. (1990)  
*Models for Exceedances over High Thresholds*  
Journal Royal Statistical Society B, 52(3), 393-442
- [26] Davison, A.C., and D.V., Hinkley, D.V. (1997)  
*Bootstrap methods and their application*  
Cambridge University Press, 592 pp.
- [27] de Haan, L. (1994)  
*Extreme value statistics*  
In: Galambos J., Lechner J.A., Simiu E., editors. Extreme value theory and applications. Dordrecht: Kluwer Academic Publishers

- [28] de Haan, L., de Ronde, J. (1998)  
*Sea and Wind: Multivariate Extremes at Work*  
Extremes, 1(1), 7-45
- [29] de Haan, L., Neves, C., Peng, L. (2008)  
*Parametric tail copula estimation and model testing*  
Journal of Multivariate Analysis, 99, 1260-1275
- [30] Dekkers, A., Einmahl, J., and Haan, L. de (1989)  
*A moment estimator for the index of an extreme-value distribution*  
Annal of Statistics, 17, 1833-1855
- [31] Dillon, D.D., Kahn, L.A. (1981)  
*A wave meter mooring design for the Mid Atlantic Bight*  
EGG WASC report TR-E930-0001
- [32] Draisma, G., Drees, H., Ferreira, A., Haan, L. de (2004)  
*Bivariate tail estimation: dependence in asymptotic independence*  
Bernoulli, 10(2), 251-280
- [33] Drees, H. (1995)  
*Refined Pickands estimators of the extreme value index*  
Annal of Statistics, 23, 2059-2080
- [34] Dupuis, D.J. (1999)  
*Exceedances over High Thresholds: A guide to Threshold Selection*  
Extremes, 1(3), 251-261
- [35] Efron, B., Tibshirani, R. (1993)  
*An introduction to the bootstrap*  
New York, Chapman & Hall, 436 pp
- [36] Einmahl, J.H.J., Haan, L. de, Sinha, A.K. (1997)  
*Estimating the spectral measure of an extreme value distribution*  
Stochastic Process. Appl., 70, 143-171
- [37] Fisher, R.A., Tippett, L.H.C. (1928)  
*Limiting forms of the frequency distributions of the largest or smallest member of a sample*  
Proc. Camb. Phil. Soc., 24, 180-190

- [38] Fisher, N.I. (1993)  
*Statistical Analysis of Circular Data*  
Cambridge University Press
- [39] Ferreira, J.A, Guedes Soares, C. (1998)  
*An application of the peaks over threshold method to predict extremes of significant wave height*  
Journal of Offshore Mechanics and Arctic Engineering, 120, 165-176
- [40] Freedman, I.S. (1979)  
*The use of a Kolmogorov-Smirnov type statistic in testing hypotheses about seasonal variation*  
J. Epidemiology and Community Health, 33, 223-228
- [41] Genest, C., Segers, J.(2009)  
*Rank-based inference for bivariate extreme-value copulas*  
Ann. Stat., 37, 2990-3022
- [42] Genest, C., Kojadinovic, I., Neslehova J., Yan, J.(2011)  
*A goodness-of-fit test for bivariate extreme-value copulas*  
Bernoulli, 17, 253-275
- [43] Goda, Y. (1988)  
*On the methodology of selecting waves height*  
Proc. Int. Coastal Engineering Conference, 899-913
- [44] Goda, Y. et al. (1993)  
*Intercomparison of extremal wave analysis methods using numerically simulated data*  
Proc. Sec. Int. Symp. on Ocean Wave Measurement and Analysis, 963-977
- [45] Goda, Y. (2000)  
*Random Seas and Design of Maritime Structures*  
World Scientific Publishing, 2nd Ed, 443 pp
- [46] González, J., Vega, J., Rodríguez, G. (2005)  
*Extreme value analysis of sea waves in Canary islands by means of the POT method*  
Proc. 30th Bienal de Física RSEF, 388-389
- [47] Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979)  
*Probability Weighted Moments: Definition and relation to parameters of several distributions expressible in inverse form*  
Water Resources Research, 15, 1049-1054

- [48] Gross, J.L., et al.(1994)  
*Novel extreme value procedures: application to extreme wind data*  
Galambos, J., Lechner J.A., Simiu, E., editors. Extreme value theory and applications. Dordrecht: Kluwer Academic Publishers
- [49] Guedes Soares, C., Ferreira, A.M., Cunha, C., (1996)  
*Linear models of the time series of significant wave height in the Southwest Coast of Portugal*  
Coastal Engineering, 29, 149-167
- [50] Guedes Soares, C., Scotto, M.G. (2004)  
*Application of the r largest-order statistics for long-term predictions of significant wave height*  
Coastal Engineering, 51(5-6), 387-394
- [51] Gumbel, E.J. (1958)  
*Statistics of extremes*  
New York: Columbia University Press
- [52] Gumbel, E.J., Goldstein, N. (1964)  
*Analysis of empirical bivariate extremal distributions* Journal of the American Statistical Association, 59, 794-816
- [53] Hasofer, A.M. (1996)  
*Parametric estimation of failure probabilities*  
In: Casciati, F., Roberts, B., editors: Mathematical models for structural reliability analysis. Boca raton, FL: CRC Press
- [54] Hawkes, P.J., Gouldby, B.P., Tawn, J.A., Owen, M.W. (2002)  
*The joint probability of waves and water levels in coastal engineering design*  
Journal of Hydraulic Research, 40, 241-251
- [55] Hill, B. M. (1975)  
*A simple general approach to inference about the tail of a distribution*  
The Annals of Statistics, 3, 1163-1174
- [56] Houghton, J.C. (1978)  
*Birth of a parent: The Wakeby distribution for modelling flood flows*  
Water Resources Research, 14, 1105-1110



- [57] Hoaglin, D.C. (1980)  
*A Poissonnes Plot*  
The American Statistician, 34(3), 146-149
- [58] Hosking, J.R.M., Wallis, J.R. (1987)  
*Parameter and Quantile Estimation for the Generalized Pareto Distribution*  
Technometrics, 29, 339-349
- [59] Hosking, J.R.M. (1989)  
*Some theoretical results concerning L-moments*  
Research Report RC14492, IBM Research Division, Yorktown Heights, N.Y.
- [60] Hosking, J.R.M. (1990)  
*L-moments: Analysis and estimation of Distributions using Linear Combinations of Order Statistics*  
Journal Royal Statistical Society, B, 52(1), 105-124
- [61] Hosking, J.R.M. (1994)  
*The four parameter kappa distribution*  
IBM Journal Research Development, 38 (3), 251-258
- [62] Hosking, J.R.M. (1996)  
*Fortran routines for use with the method of L-moments*  
Research Report RC20525, IBM Research Division, Yorktown Heights, N.Y.
- [63] Hosking, J.R.M. (2000)  
*Fortran routines for use with the method of L-moments. Version 3.03*  
IBM Research Division, Yorktown Heights, N.Y., U.S.A.
- [64] Huang, X. (1992)  
*Statistics of bivariate extreme values*  
PhD Thesis, Erasmus University Rotterdam
- [65] James, I.D. (1986)  
*A note on the theoretical comparision of wave staff and wave rider buoys in steep gravity waves*  
Ocean Engineering, 13(2), 209-214
- [66] Jonathan, P., Ewans, K.C. (2011)  
*Modelling the seasonality of extreme waves in the Gulf of Mexico*  
Journal of Offshore and Arctic Engineering, 133, 021104-1-9

- [67] Janssen, P. (2008)  
*Progress in ocean wave forecasting*  
J. Computational Physics 227(7), 3572-3594
- [68] Karpa, O., Naess, A. (2013)  
*Extreme value statistics of wind speed data by the ACER method*  
J. Wind Engineering and Industrial Aerodynamics, 112, 1-10
- [69] Kyselý, J. (2008)  
*A cautionary note on the use of nonparametric Bootstrap for estimating uncertainties in extreme-value models*  
J. Applied Meteorology and Climatology, 47, 3236-3251
- [70] Lang, M., Rasmussen, P., Oberlin, G., Bobee, B. (1997)  
*Over-threshold sampling: Modeling of occurrences by renewal processes*  
Revue des Sciences de l'Eau, 3, 279-320
- [71] Lang, M. (1999)  
*Theoretical discussion and Monte-Carlo simulations for a Negative Binomial process paradox*  
Stochastic Environmental Research and Risk Assessment, 13, 183-200
- [72] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983)  
*Extremes and Related Properties of Random Sequences and Processes*  
Ed. Springer-Verlag
- [73] Ledford, A., Tawn, J.A. (1996)  
*Statistics for near independence in multivariate extreme values*  
Biometrika, 83(1), 169-187
- [74] Ledford, A., Tawn, J.A. (1997)  
*Modelling dependence within joint tail regions*  
Journal of the Royal Statistical Society, Serie B, 59, 475-499
- [75] Mackay, E., Challenor, P., Bahaj, A. (2010)  
*On the use of discrete seasonal and directional models for the estimation of extreme wave conditions*  
Ocean Engineering, 37(5-6), 425-442
- [76] Makkonen, L. (2008)  
*Problems in the extremes value analysis*  
Structural Safety, 30, 405-419

- [77] Mardia, K.V., Jupp, P.E.(2000)  
*Directional Statistics*  
John Wiley, Hoboken, N.J.
- [78] Martín Soldevilla, M.J. (2007)  
*Las cópulas en la caracterización de sucesos multivariantes. Aplicación al diseño de estructuras*  
Monografías CEDEX, ISSN 0211-8203 ; ISBN 978-84-7790-439-7; Centro de Estudios de Puertos y Costas, Centro de Estudios y Experimentación de Obras Públicas, Ministerio de Fomento
- [79] Martínez, J., Martín, M.J. (1990)  
*Descripción de las funciones de distribución más utilizadas en estudios de clima marítimo*  
Centro de Estudios y Experimentación de Obras Públicas, 103 pp
- [80] Mathiesen, M., Goda, Y., Hawkes, P., Mansard, E., Martín, M.J., Peltier, E., Thompson, E.F., Van Vledder, G. (1994)  
*Recommended practice for extreme wave analysis*  
Journal of Hydraulic Research, 32 (6), 803-814
- [81] Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J. (2006)  
*Estimation of the long-term variability of extreme significant wave height using a time-dependent POT model*  
Journal of geophysical Research, 111, C07024. doi:10.1029/2005JC003344
- [82] Mielke, P.W. (1973)  
*Another Family of distributions for describing and analyzing precipitation data*  
Journal Applied Meteorology, 12, 275-280
- [83] Morton, I.D., Bowers, J., Mould, G., (1997)  
*Estimating return period wave heights and wind speeds using a seasonal point process model*  
Coastal Engineering, 31, 305-326
- [84] Muir, L.R., El-Shaarawi, A.H. (1986)  
*On the calculation of extreme wave heights: A review*  
Ocean Engineering, 13(1): 93-118
- [85] Naess, A. (1998a)  
*Estimation of long return period design values for wind speeds*  
Journal of Engineering Mechanics, ASCE, 124(3), 252-259

- [86] Naess, A., Clausen, P.H. (1999)  
*Statistical extrapolation and the peaks over threshold method*  
Proceedings of OMAE99 On Offshore Mechanics and Arctic Engineering, Newfoundland, Canadá
- [87] Naess, A., Clausen, P.H. (2001)  
*Combination of the peaks-over-threshold and bootstrapping methods for extreme value prediction* Structural Safety, 23(4), 315-330
- [88] Naess, A., Hungnes, B. (2001)  
*Estimating confidence intervals of long return period design values by bootstrapping*  
J. Offshore Mechanics and Arctic Engineering, 124(1), 2-5
- [89] Naess, A., Gaidai, O. (2009)  
*Estimation of extreme values from sampled time series*  
Structural Safety, 31, 325-334
- [90] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983)  
*Extremes and Related Properties of Random Sequences and Processes*  
Ed. Springer-Verlag
- [91] Ochi, M. (1992)  
*New approach for estimating the severest sea state from statistical data*  
Coastal Engineering 1992: Proc. 23rd Int. Conf., Vol. 1, Venice, Italy, 512-525
- [92] Önöz, B., Bayazit, M. (2001)  
*Effect of the occurrence process of the peaks over threshold on the flood estimates*  
Journal of Hydrology, 244, 86-96
- [93] Ouarda, T.B.M.J., Ashkar, F., El-Jabi, N. (1993)  
*Peaks Over Threshold Model for Seasonal Flood Variations*  
Proceedings of the Symposium *Engineering Hydrology*, San Francisco, California, ASCE Pub., 341-346
- [94] Pacheco, M. (2003)  
*Análisis numérico-experimental del oleaje en mares mixtos*  
Tesis Doctoral, Universidad de las Palmas de Gran Canaria, 411 pp
- [95] Pandey, M.D., Van Gelder, P.H.A.J.M., Vrijling, J.K. (2001)  
*The estimation of extreme quantiles of wind velocity using L-moments in the peaks-*

- over-threshold approach*  
Structural Safety, 23, 179-192
- [96] Parida, B.P. (1999)  
*Modelling of Indian summer Monsoon rainfall using a four-parameter Kappa distribution*  
International Journal of Climatology, 19, 1389-1398
- [97] Park, J., Jung, H., Kim, R., Oh, J. (2001)  
*Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution*  
International Journal of Climatology, 21, 1371-1384
- [98] Pearson, C.P. (1993)  
*Application of L-moments to maximum river flows*  
The New Zealand Statistician, 1, 2-10
- [99] Peng, L. (1999)  
*Estimation of the coefficient of tail dependence in bivariate extremes*  
Statistics and Probability Letters, 43, 399-409
- [100] Pickands, J. (1975))  
*Statistical Inference Using Extreme Order Statistics*  
The Annals of Statistics, 3(1), 119-130
- [101] Rademakers, P.J. (1980)  
*Some details on the accelerometer stabilization as used in the Waverider*  
Datawell Inc., Holland
- [102] Repko, A., Van Gelder, P.H.A.J.M, Voortman, H.G., Vrijling, J.K (2004)  
*Bivariate description of offshore wave conditions with physics-based extreme value statistics*  
Applied Ocean Research, 26, 162-170
- [103] Ribe, R.L. (1981)  
*Calibrations and energy spectrum correction factors for Waverider buoys deployed under the ARSLOE programm*  
ARSLOE Informal Rep., 4, NOAA, Rockville, MD
- [104] Rodriguez, G. (1992)  
*Spectral and statistical characteristics of wind waves off Canary Islands*  
Proc 5th Int. Conference on Civil Engineering in the Oceans, 622-636, Texas, USA

- [105] Rodríguez, G. (1995)  
*Análisis de las ondas gravitatorias generadas por el viento en aguas profundas*  
Tesis Doctoral, Universidad de Las Palmas de Gran Canaria, 243 pp
- [106] Rosjberg, D. and Knudsen, J. (1983)  
*POT-estimation of extreme sea states and the benefit of using wind data*  
Proceedings of the NATO Advanced Study Institute on Statistical Extremes and Applications, Vimeiro, Portugal, pp. 613-622
- [107] Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R. (2007)  
*Extremes in nature: An Approach Using Copulas*  
Water Science and Technology Library, Vol. 56, XIV, ISBN 978-1-4020-4415-1, 292 p.
- [108] Sibuya, M. (1960)  
*Bivariate extreme statistics*  
Annals of the Institute of Statistical Mathematics, 11, 195-210
- [109] Simiu, E., Heckert, N.A. (1996)  
*Extreme Wind Distribution Tails: A Peaks Over Threshold Approach*  
Journal of Structural Engineering, ASCE, 122(5), 539-547
- [110] Skjong, M., Naess, A., Naess, O.E. (in press)  
*Statistics of extreme sea levels for locations along the Norwegian coast*  
Journal of Coastal Research
- [111] Smith, R.L. (1986)  
*Extreme value theory based on the  $r$  largest anual events*  
Journal of Hydrology, 86, 27-43
- [112] Sornette, D. (2004)  
*Critical Phenomena in Natural Sciences: Chaos, Fractals, Self organization, and Disorder : Concepts and Tools*  
Springer
- [113] Soukissian, T.H. and Arapi, P.M. (2011)  
*The effect of declustering in the  $r$ -largest maxima model for the estimation of  $H_s$ -design values*  
The Open Ocean Engineering Journal 4, 34-43
- [114] Tadviji, N. (2003)  
*Confidence intervals and accuracy estimation for heavy-tailed Generalized Pareto Dis-*

- tributions*  
Extremes, 6, 111-123
- [115] Tawn, J. (1988)  
*Bivariate extreme value theory: Models and estimation*  
Biometrika, 75, 397-415
- [116] Titov, L.F. (1971)  
*Wind-Driven Waves*  
Gidrometeorologicheskoe Izdatel'stro, Leningrad
- [117] Van Vleder, G. et al. (1993)  
*Case studies of extreme wave analysis: a comparative analysis*  
Proc. Sec. Int. Symp. on Ocean Wave Measurement and Analysis, 978-992
- [118] Walton, T.L., (2000)  
*Distributions for storm surge extremes*  
Ocean Engineering, 27, 1279-1293
- [119] Zucchini, W., Adamson, P.T. (1989)  
*Bootstrap confidence intervals for design storms from exceedance series*  
Hydrological Sciences J., 34(1), 41-48