

## **2. MATERIAL Y MÉTODOS**

En este capítulo se presenta el material necesario para la elaboración de la tesina y los métodos utilizados para el desarrollo de la misma. El material, en este caso, se compone de los datos de oleaje que proporcionan las boyas situadas en la costa catalana. En los métodos englobamos todo el desarrollo teórico en el que se basa la tesina y los programas que se han utilizado.

### **2.1. Fuentes de información de oleaje**

El estudio del oleaje (tanto a corto como a largo plazo) requiere de una información detallada del fenómeno. Existen distintas fuentes de datos sobre las que poder realizar el análisis y que pueden agruparse en tres categorías: datos visuales, datos instrumentales y datos procedentes de simulaciones numéricas. Aunque en la tesina solo se utilizan los datos instrumentales, es interesante conocer el resto de fuentes existentes.

#### **2.1.1. Datos visuales**

Los datos visuales provienen de observaciones realizadas por barcos en ruta, a lo largo de todas las costas del mundo, que colaboran con el Centro Meteorológico Mundial de Ashville (EEUU) y son obtenidos de forma sistemática desde aproximadamente 1950, de forma que un observador recoge (entre otras) información sobre la altura de ola, periodo y dirección de dos estados de mar, el mar de viento (SEA) y el mar de fondo (SWELL). Un estado de mar de viento se corresponde con el oleaje formado por la acción directa y continuada del viento existente y se caracteriza por presentar una gran irregularidad (aunque no siempre). Por el contrario, el oleaje de tipo SWELL es aquel que ya ha abandonado el área de generación (zona de desarrollo del mar de viento) y se propaga a través de la superficie; el resultado es un estado de mar más homogéneo y con menor dispersión direccional.

Una de las grandes ventajas de este tipo de datos es el gran número de observaciones existentes y la completa caracterización del suceso del oleaje (definido por altura, periodo y dirección). Así pues, el elevado número y la extensión temporal de las observaciones las dota de cierta fiabilidad estadística para algunos tipos de estudios.

Sin embargo, la desventaja de la información visual es la poca calidad, ya que es subjetiva porque depende de la experiencia y criterio de la persona que recoge los datos, y la gran heterogeneidad espacial, no quedando bien representados los episodios de mayor contenido energético puesto que los buques tienden a evitar la acción de grandes temporales por el consiguiente peligro para la navegación. Debido a esto no es aconsejable utilizar los datos visuales en el estudio de oleaje extremal, a menos que se efectúen correcciones oportunas a dichos datos, por ejemplo, comparándolos con los resultados de una boya.

---

### **2.1.2. Datos instrumentales**

Los datos instrumentales son adquiridos por sensores, principalmente boyas de oleaje. Son un instrumento moderno, capaz de recoger una gran cantidad de información de alta calidad. Son instrumentos que evolucionan continuamente ofreciendo cada día más prestaciones.

Pueden ser de dos tipos:

- direccionales
- escalares

Las boyas direccionales obtienen información de altura, periodo y dirección del oleaje, mientras que las escalares solo registran altura y periodo. En estos casos, para poder conocer la dirección se debe recurrir a información complementaria.

La ventaja de las boyas reside en la calidad de los datos que proporciona aunque desgraciadamente su instalación (fondeo) en nuestras costas es relativamente reciente. Además son un instrumento caro y por ello no existen tantas como sería deseable. Otro inconveniente de los datos instrumentales es la existencia de “huecos” en la serie de datos, factor preocupante, ya que se ha demostrado que la probabilidad de perder datos es mucho mayor en condiciones extremas.

En España, el Ministerio de Fomento, a través del programa REMRO (Red Española de Medida y Registro de Oleaje) gestiona un total de 17 boyas de oleaje a lo largo del litoral español. A su vez la Generalitat de Catalunya a través del programa XIOM (Xarxa de Instrumentació Oceanogràfica i Meteorològica) gestiona 4 boyas de oleaje a lo largo del litoral catalán.

### **2.1.3. Datos procedentes de simulaciones numéricas**

Cuando no es posible encontrar información ni visual ni instrumental, o bien alguna de ellas se encuentra incompleta (no existe información sobre la direccionalidad o existen grandes lagunas entre registros), es posible reproducir el oleaje a partir de información meteorológica mediante el uso de modelos numéricos de predicción de oleaje. Así pues, estos datos son útiles para rellenar huecos de información existentes en los datos registrados o para aquellas áreas marítimas donde sólo se dispone de cartas barométricas.

El problema de este tipo de datos es la necesidad de calibrar los modelos utilizados y el coste computacional que supone. Su fiabilidad depende de la calidad del campo de vientos utilizado y de la correcta calibración del modelo con datos registrados.

En la actualidad son cada vez más utilizados puesto que cubren de forma detallada la práctica totalidad de nuestras costas y empiezan a conformar una base de datos de más de tres años en algunos casos.

#### **2.1.4. Conclusiones**

Está claro que mezclar datos de dos fuentes diferentes no es recomendable, a menos que se conozca bien la relación entre los diferentes datos. Es decir, que los datos de la muestra pertenezcan a la misma población. Ya que los datos visuales son por naturaleza muy inexactos, la tendencia en los países modernos que pueden permitirse el tener una red de registros de oleaje es utilizar exclusivamente los datos instrumentales para estimar el oleaje extremal. Esto no significa que los datos de las boyas de registro no se puedan complementar con otros datos, por ejemplo meteorológicos; estos pueden servir, por ejemplo, para contrastar los datos instrumentales.

### **2.2. La xarxa d'instrumentació oceanogràfica meteorològica (XIOM)**

Los datos de esta tesina provienen de tres boyas de tipo escalar pertenecientes a la red XIOM. En este apartado se explica en que consiste la XIOM, como se formó y los instrumentos que la forman.

#### **2.2.1. Antecedentes**

En 1984 con el fin de obtener medidas para un mejor seguimiento de la evolución de las playas del litoral catalán, el Departament de Política Territorial i Obres Publiques de la Generalitat de Catalunya (DPTOP) impulsó la creación de un red de medida de oleaje distribuida a lo largo de la costa catalana. Dos boyas escalares se instalaron sobre fondo de 50m, en frente de las zonas más vulnerables al oleaje como son el Delta del Llobregat y el Delta de La Tordera.

Posteriormente, en el 1990, a la vista del creciente problema de regresión que sufría la línea de la costa en el Delta de l'Ebre, el DPTOP impulsó un proyecto para proponer soluciones estabilizadoras, el desarrollo del cual fue encargado al Laboratori d'Enginyeria Marítima de la Universitat Politècnica de Catalunya (LIM/UPC). En este proyecto se utilizaron dos boyas direccionales (Cap Tortosa y el Trabucador) y se incorporaron dos mareógrafos (L'Ampolla y Sant Carles de la Ràpita) y tres estaciones meteorológicas (L'Ampolla, Delta de l'Ebre y Sant Carles de la Ràpita) al paquete de instrumentos.

En Septiembre de 1992 se instaló una nueva boya escalar en el Golf de Roses.

Las boyas escalares del Golfo de Roses, Delta de La Tordera y Delta del Llobregat dejaron de ser operativas el año 1997, pero los instrumentos de Delta de l'Ebre siguieron funcionando, excepto la boya direccional del Trabucador que se perdió.

El conjunto de la red de boyas ha proporcionado las series temporales de oleajes más largas existentes, hasta hoy día, del litoral español.

En los últimos años, la necesidad de una mejor observación del medio ambiente y el desarrollo en el campo de la previsión y de los modelos numéricos, han ampliado los sectores interesados en la obtención de medidas de las variables costeras. En el 1999, el Servei de Meteorologia de Catalunya (SMC), Ports de la Generalitat (PG), el DPTOP y

LIM/UPC acordaron un nuevo impulso a esta red de medida consistente en volver a medir lo que había quedado interrumpido.

### 2.2.2. Estado actual

La XIOM está compuesta por un conjunto de equipos de medida de las variables costeras más significativas como pueden ser las condiciones meteorológicas y el estado del mar mediante boyas escalares y direccionales, estaciones meteorológicas y mareógrafos.

En la figura 1 se puede observar el estado actual de la red de medidas que, gracias a la colaboración de las instituciones citadas anteriormente, llegará a proporcionar datos para la caracterización exacta del clima marítimo de la costa catalana.

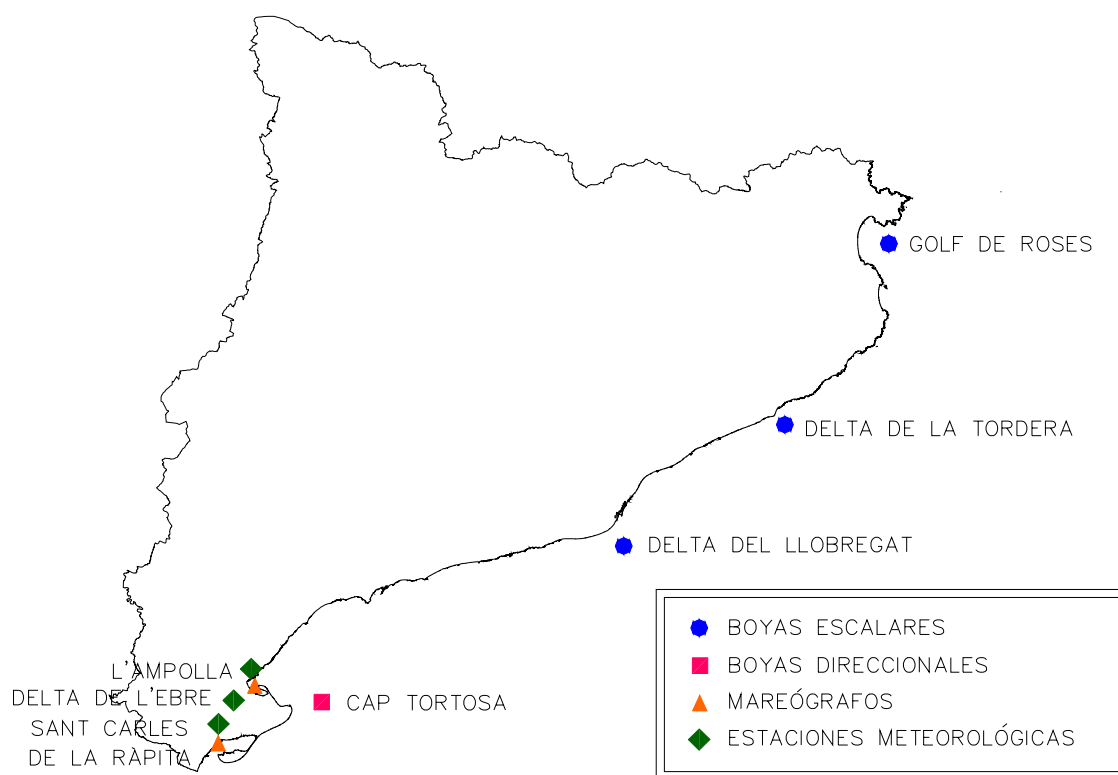


Figura 1: Situación de la red de medida en el año 2003

Boyas escalares: Golf de Roses  
Delta de La Tordera (Blanes)  
Delta del Llobregat\*

Boya direccional: Cap Tortosa

Mareógrafos: L'Ampolla  
Sant Carles de la Ràpita

Estaciones meteorológicas: L'Ampolla  
Delta de l'Ebre  
Sant Carles de la Ràpita

\* : Desde inicio del año 2004 la boya del Delta del Llobregat ha pasado a ser una boya direccional.

### **2.3. Clima de oleaje extremal**

El oleaje a largo término es un concepto que se aplica al estudio estadístico de una colección de datos cubriendo el máximo número de años posible. A diferencia del oleaje a corto término, proceso que se puede suponer estacionario, aleatorio y gaussiano, la estructura del oleaje a largo término es un reflejo de las características climatológicas, tanto locales como distantes. Esta descripción estadística del oleaje, denominada en términos generales “a largo término” suele plantearse para dos intervalos específicos de tiempo: el correspondiente al año climático medio (que estudia las condiciones climáticas más frecuentes, o medias) y el que abarca períodos de tiempo mucho mayores, de decenas de años (que estudia las condiciones más severas o extremas).

Estudiar el clima marítimo es imprescindible para el buen diseño de las obras marítimas. El clima medio es importante para el funcionamiento de las instalaciones. Es importante el número de horas al año en que la agitación del oleaje en una dársena de un puerto supera un cierto umbral. Es decir, con el clima medio se diseña para el estado límite de servicio. Pero a veces suceden episodios que pueden causar la destrucción, total o parcial, de una obra marítima. Prever estos sucesos es muy difícil, y como máximo se pueden hacer estudios estadísticos para hallar los parámetros de diseño de obras marítimas. En este caso, el diseño es para el estado límite último, y este concepto va unido a otro: el de vida útil de la obra.

El análisis extremal se centra en la caracterización de los eventos extremos o temporales para una zona de estudio, es decir, un subconjunto de datos dentro del total del registro.

El objetivo principal del análisis es poder determinar las acciones de diseño a soportar por una obra marítima determinada en función del riesgo asignado y su vida útil, siendo necesario conocer o estimar la altura de ola significativa asociada a una cierta probabilidad de excedencia dentro del periodo de vida del proyecto. Para determinar dicha altura se necesita modelar el comportamiento estadístico de aquellos valores de la serie de altura significativa, que por su magnitud pueden poner en riesgo la obra.

El clima extremal también es utilizado para determinar las acciones máximas a soportar durante las distintas fases de ejecución de una obra y que por supuesto son menores que las acciones de servicio de la obra. Es una práctica habitual el asegurar obras de gran valor económico durante la fase de ejecución puesto que resulta casi inevitable que ciertos tramos queden expuestos a la acción de temporales antes de poder ser finalizados produciéndose la destrucción del tramo y el retraso consiguiente en la finalización de la obra.

---

### 2.3.1. Métodos de selección de datos

En el análisis extremal los métodos más utilizados para modelar los valores extremos o máximos de una variable pueden clasificarse en dos grupos según la información que utilizan:

- Método de la muestra total o método de la distribución inicial

Fue propuesto por Draper (1963) y revisado por Goda (1979). La idea de este método es estimar el oleaje extremal a partir del clima de oleaje medio y se utiliza cuando el tiempo de registro es corto (inferior a 3 o 4 años).

El método parte de toda la información de oleaje (altura de ola) registrada en un periodo determinado y se apoya en la función de distribución de probabilidad de no excedencia de un año climático medio (régimen medio). Existen distintas aproximaciones en función del tipo de datos de partida (visual o instrumental). Así, si de lo que se dispone es de información instrumental, se hace uso de la siguiente expresión:

$$\Phi(Hs) = [F(Hs)]^{n(Hs)} \quad (1)$$

siendo  $\Phi$  la ecuación extremal,  $F(Hs)$  el régimen medio del oleaje y  $n(Hs)$  el número de pruebas estadísticas en un año medio y definido como:

$$n(Hs) = \frac{T \cdot n}{t(Hs)} \quad (2)$$

donde  $T$  es la duración real del año climático considerado,  $t(Hs)$  es el tiempo total en que se supera un cierto valor de  $Hs$  y  $n$  es el número de veces que se excede ese valor.

- Método de los valores de pico

Es el más utilizado aunque requiere un conjunto de datos mucho más extenso en el tiempo puesto que parte de valores máximos registrados en un periodo de tiempo determinado, que puede ser de un año o bien el máximo de una acción de temporal. Se diferencian por tanto dos grandes grupos:

*a. Método de los máximos anuales:*

Este método consiste en seleccionar tan sólo el valor máximo anual de la altura de ola en cierto punto de la costa a lo largo de muchos años. Es decir, para cada año seleccionar la máxima altura de ola que se produjo. Si se hiciese durante muchos años, se conocería aproximadamente la probabilidad de superar cierta altura de ola.

Si se hacen las hipótesis de años independientes y cambios climáticos inexistentes o muy lentos (y ninguna otra) es un método estadísticamente coherente.

El problema de este método es que las boyas son instrumentos modernos que llevan pocos años funcionando (la mayoría de las utilizadas en análisis llevan unos veinte años de registro, y es excepcional encontrar series de datos instrumentales de más de treinta años). Así pues, de la gran cantidad inicial de información sólo se seleccionan veinte o treinta datos. La incertidumbre estadística a causa de tener tan poca información es muy grande.

*b. Método de la serie de duraciones parciales o método de los valores de pico (POT, Peak Over Threshold):*

A diferencia del método anterior, con éste se aprovecha mejor toda la información disponible y se pueden hacer estimaciones más exactas. Además es un método más fiable que el de máximos anuales cuando la información disponible corresponde a un periodo inferior a 20 años, como sucede en este caso.

El método POT se basa en suponer que los máximos que superan un cierto umbral y que están separados por un cierto intervalo de tiempo (normalmente de dos a cuatro días) son independientes entre sí. Esto implica que si en un año ha habido temporales muy importantes no es a causa de que los sucesos extremos tiendan a concentrarse en ciertos años, sino que esta concentración ha sido por azar. Si los temporales son independientes, el hecho de que en un año haya habido un gran temporal no significa que la probabilidad de que ese mismo año existieran otros temporales importantes sea mayor (o menor). Aceptada la hipótesis, la muestra a analizar estará constituida, pues, por los máximos locales o bien por los picos de temporales que superen un determinado umbral de altura ( $H_0$ ). En cualquier caso, los máximos seleccionados deben cumplir la condición de independencia. Para este método se define el número medio de temporales al año ( $\lambda$ ) como la relación entre el número total de temporales registrados ( $N_T$ ) y el tiempo efectivo de medida ( $T_{ef}$ ):

$$\lambda = \frac{N_T}{T_{ef}} \quad (3)$$

En el método POT será importante seleccionar el umbral de manera que todos los picos pertenezcan a la misma población. Si se escoge un umbral muy bajo algunos máximos pertenecerán a temporales secundarios, que nada tienen que ver con el tipo de temporales que dan los mayores valores (esto nunca puede asegurarse, aunque puede ser cierto en algunos lugares). En cambio, si ese umbral es muy alto se obtienen muy pocos datos y se pierde información. Así pues, la elección del umbral,  $H_0$ , es un tema de discusión. Por lo general, se intenta que el

umbral de altura de ola se establezca según las características climáticas del área de estudio.

Una vez resuelta la elección del umbral  $H_0$ , se establece un segundo umbral  $H_1$  mayor que  $H_0$ , y se contabiliza el número de temporales que lo superan ( $N_I$ ). La relación entre el número de datos  $N_I$  y el total de temporales registradas ( $N_T$ ) se conoce como “censoring parameter” y se denomina  $\nu$ , definiéndose mediante la siguiente expresión:

$$\nu = \frac{N_I}{N_T} \quad (4)$$

Por último también se ha establecido un tercer umbral  $H_2$  mayor que  $H_1$ .

El hecho de aumentar el nivel umbral o el “threshold level” hace que el ajuste se realice sobre un número menor de datos y que corresponden a los máximos del conjunto. El ajuste sobre los diferentes grupos de datos que resultan al variar el nivel umbral afecta a las estimas a largo término de una manera variable, es decir, en un caso las estimas que resultan son más altas si se censuran más datos y en otro caso son más bajas. Este efecto se puede observar en las siguientes figuras, que quieren representar el ajuste a unos datos con selección de la muestra por el método POT y dos umbrales distintos,  $H'$  y  $H''$ . En la figura 2 se ve como el hecho de aumentar el umbral hace que las estimas también aumenten, y en la figura 3 al aumentar el umbral pasa el efecto contrario, las estimas disminuyen:

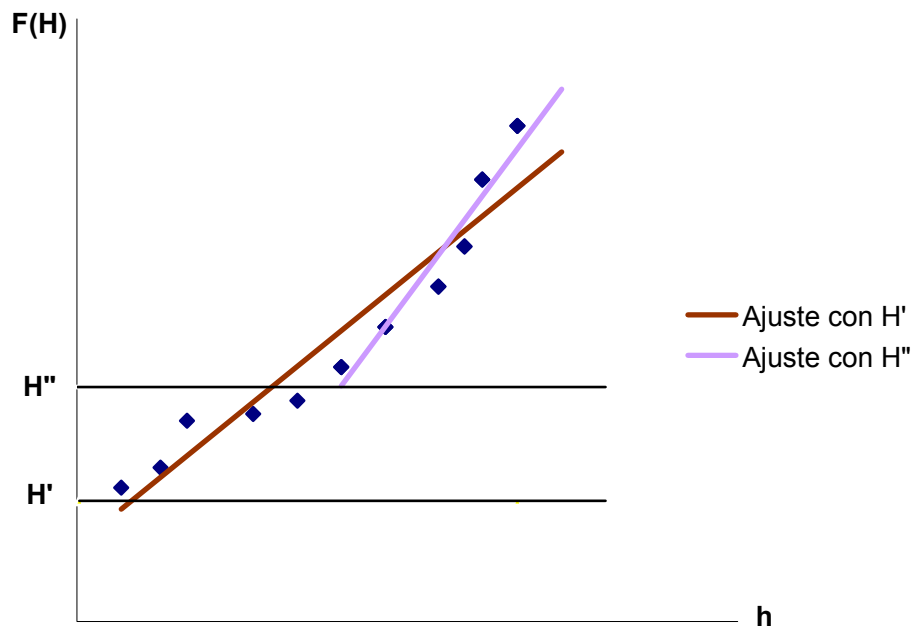


Figura 2



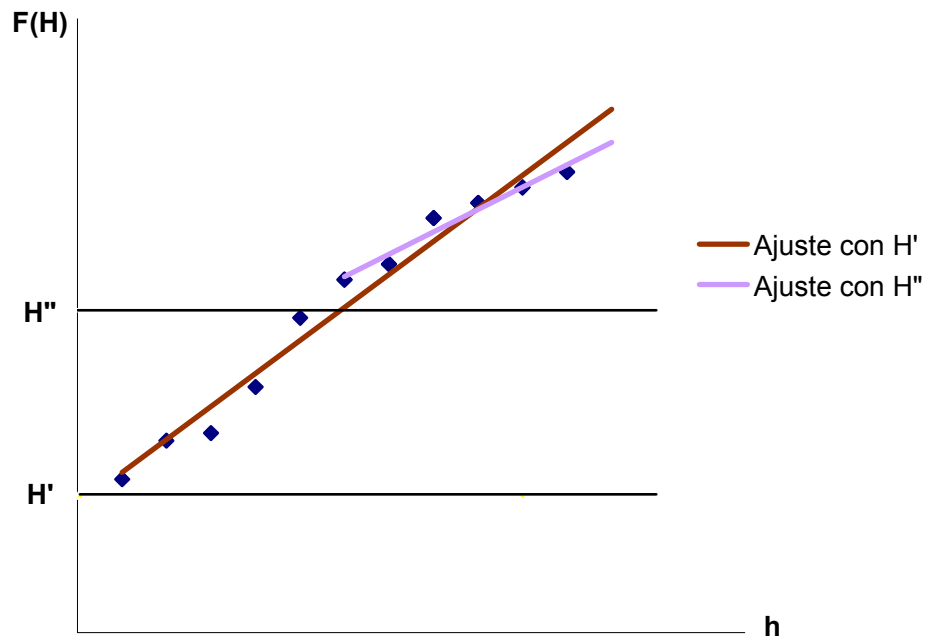


Figura 3

En ambos ejemplos el método de ajuste ha sido el método gráfico, aunque seguramente por cualquier otro método los resultados hubiesen sido parecidos. Para ambos casos el ajuste realizado con el nivel umbral mayor se ajusta mejor a la cola, y se podría considerar el ajuste más idóneo, pero el hecho de aumentar excesivamente el nivel umbral puede suponer una pérdida de información y llevar a resultados erróneos.

El método POT será con el que se realizará el análisis en esta tesina.

### 2.3.2. Funciones de distribución de probabilidad

Existe una amplia colección de distribuciones de probabilidad teóricas que pueden utilizarse para describir sucesos, pero no todas son adecuadas para describir ciertos fenómenos. Uno de los objetivos de esta tesina es conocer cuál de ellas se ajusta mejor a los datos.

Es frecuente en diversos campos de previsión de riesgos naturales calcular los periodos de retorno a partir de funciones de distribución de probabilidad de máximos o mínimos.

Las funciones de distribución se utilizan para asignar un determinado periodo de retorno a una altura de ola dada. El periodo de retorno se define como el número de años que en término medio transcurre entre dos sucesos extremos de determinada magnitud. Igual de interesante, o incluso más, es conocer cuál es el valor de altura de ola asociada a un cierto periodo de retorno.

Lo que se pretende con esta tesina es conseguir una función que asigne a cada altura de ola dada un periodo de retorno y viceversa. El problema se centra en seleccionar aquella función que mejor se ajuste a los datos de que se disponen.

Las funciones de distribución utilizadas en el análisis son las que se muestran a continuación en la tabla 1:

Distribución	Función de distribución	Campo de definición		
		x	A	B
Weibull	$1 - \exp\left[-\left(\frac{x-A}{B-A}\right)^c\right]$	$\geq A$	$> X_{min}$	$> A$
Gumbel (A.I)	$\exp\left\{-\exp\left[-\left(\frac{x-A}{B}\right)\right]\right\}$	-	-	$> 0$
Frechet (A.II)	$\exp\left[-\left(\frac{x}{B}\right)^c\right]$	$\geq 0$	-	$> 0$

Tabla 1: Funciones de distribución de probabilidad para el análisis extremal

donde:

X representa el evento altura de ola máxima significativa de los datos de origen

x representa el valor del suceso

F equivale a la función de probabilidad de no excedencia, siendo A, B y C los parámetros de posición, escala y forma que deben modificarse para que la función se ajuste lo máximo posible a los datos reales.

Una de estas tres funciones será la que mejor se ajuste a los datos y con la que asignaremos un periodo de retorno a cada altura de ola y viceversa. Esto será válido si aceptamos la hipótesis de que los máximos de oleaje se distribuyen según la distribución seleccionada. En realidad es muy improbable que las alturas de ola en la naturaleza se distribuyan según una distribución conocida. Por lo tanto, estimar la altura de ola para periodos de retorno superiores al tiempo de registro será siempre una extrapolación, tanto más inexacta cuanto más grande sea el periodo de retorno. Así que hay que ser conscientes que a mayor periodo de retorno, más inexactitud en la estima de la altura de ola.

### 2.3.3. Métodos de ajuste de las funciones de distribución de probabilidad

Una vez definidas las funciones hay que ajustarlas lo máximo posible a los datos. El ajuste consiste en estimar los parámetros de las distribuciones. Existen tres métodos clásicos que se explican a continuación:

- Método de los momentos muestrales

Es un método sencillo que se basa en que los parámetros de la función de distribución pueden determinarse a partir de los momentos muestrales (media, varianza o momentos superiores). Así pues, tan sólo se deben calcular los momentos de la población muestral, y suponer que los valores obtenidos se corresponden con los momentos de la distribución a ajustar. De igualar los momentos muestrales con los de la distribución se deben despejar todos los parámetros buscados. Es decir, si la función de distribución tiene  $n$  parámetros o incógnitas se deben calcular  $n$  momentos para poder plantear tantas ecuaciones como incógnitas.

- Método de la máxima verosimilitud

Es un método directo para la deducción de los estimadores de los diferentes parámetros. Consiste en determinar el valor de los parámetros de la distribución de forma que se haga máxima la probabilidad o verosimilitud de obtener el resultado muestral observado.

Si se supone que las muestras son independientes entre sí, se puede definir la probabilidad de obtener una muestra dada  $(x_1, x_2, \dots, x_n)$ , o verosimilitud  $L$  de esa muestra como

$$L(\theta_1, \dots, \theta_m; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_m) \quad (5)$$

donde los símbolos  $\theta$  representan los parámetros a determinar y  $f$  es la función de densidad de probabilidad que lógicamente depende de estos parámetros. Ahora el problema será encontrar los valores de estos  $\theta$  tales que maximicen la función de verosimilitud  $L$ .

Para simplificar el problema se suele maximizar el logaritmo de la función de verosimilitud. Suponiendo que el máximo está situado en un punto diferenciable de la función de verosimilitud, la derivada de esta función respecto a cada uno de los parámetros  $\theta$  debe ser nula.

$$\frac{\partial \ln L}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \left[ \sum_{i=1}^n \ln f(x_1, \dots, x_n; \theta_1, \dots, \theta_m) \right] = 0 \quad j=1, \dots, m \quad (6)$$

De estas  $m$  ecuaciones se pueden despejar los parámetros  $\theta$ .

- Método gráfico

Consiste en ajustar una recta en papel probabilístico a los puntos

$$(h_i, \widehat{F}_{est}(h_i)) \quad i=1, \dots, n \quad (7)$$

donde  $\widehat{F}_{est}(h_i)$  es una estima de  $F(h_i)$ . Se han deducido muchas fórmulas para efectuar estas estimas que reciben el nombre de *plotting position*. La función de distribución elegida representada en papel probabilístico tiene la propiedad de ser una recta. Se trabaja con variables reducidas. El problema del método es elegir la *plotting position* adecuada.

Una sofisticación del método gráfico consiste en utilizar la regresión lineal simple, entonces se denomina método de los mínimos cuadrados. Para aplicar este método se minimiza la siguiente expresión:

$$\sum_{i=1}^n [(h_i)_{observada} - (h_i)_{estimada}]^2 \quad (8)$$

donde  $(h_i)_{estimada}$  es la estima de la altura de ola correspondiente al periodo de retorno de la  $i$ -ésima ola observada,  $(h_i)_{observada}$ , calculado a partir de la *plotting position* elegida. Esta propiedad parece muy óptima pero se debe recordar que la *plotting position* elegida es sólo una aproximación, así que si la aproximación no es buena, tampoco lo serán las estimas, por mucho que se minimice la suma de cuadrados.

El problema de elegir la *plotting position* adecuada para cada función de distribución, es decir, la probabilidad acumulada que debe ser asignada a cada dato de la muestra se soluciona con estas recomendaciones según el tipo de función de distribución:

- Fórmula de Weibull:  $F(H_s) = 1 - \frac{i}{N+1}$ , recomendada para todas las distribuciones (9)

- Fórmula de Gringorten:  $F(H_s) = 1 - \frac{i-0,44}{N+0,12}$ , recomendada para Gumbel (10)

- Fórmula de Petruskas y Aagard:  $F(H_s) = 1 - \frac{i-0,3-\frac{0,18}{c}}{N+0,21+\frac{0,32}{c}}$ , recomendada para Weibull (11)

- Fórmula de Goda:  $F(H_s) = 1 - \frac{i-0,2-\frac{0,27}{\sqrt{c}}}{N+0,20+\frac{0,23}{\sqrt{c}}}$ , recomendada para Weibull (12)

• Fórmula de Goda y Onozawa:  $F(H_s) = 1 - \frac{i - 0,2 - \frac{0,27}{\sqrt{c}}}{N + 0,20 + \frac{0,23}{\sqrt{c}}}$ , recomendada para Frechet (13)

En la tabla 2 se observa de forma resumida todas las combinaciones de funciones de distribución y métodos de ajuste utilizados:

FUNCIÓN DE DISTRIBUCIÓN	MÉTODO DE AJUSTE	
Gumbel	Método de los momentos	
	Máxima verosimilitud	
	Mínimos cuadrados	Fórmula de Weibull
		Fórmula de Gringorten
Frechet	Método de los momentos	
	Mínimos cuadrados	Fórmula de Weibull
		Fórmula de Goda y Onozawa
Weibull	Método de los momentos	
	Máxima verosimilitud	
	Mínimos cuadrados	Fórmula de Weibull
		Fórmula de Petruskas y Aagard
		Fórmula de Goda

**Tabla 2: Combinaciones estudiadas de funciones de distribución y métodos de ajuste**

### 2.3.4. Intervalos de confianza

El oleaje en condiciones extremas está sometido a una gran incertidumbre debido, fundamentalmente, al tamaño limitado de las muestras frente a los periodos de retorno a los que se extrapola y al desconocimiento de la función de distribución que realmente siguen los máximos de oleaje. La obtención solamente de la estima puntual no es muy correcta cuando se trabaja con técnicas estadísticas, sino que es conveniente dar medidas de la exactitud de esa estima. Una manera adecuada de expresar la inexactitud de las estimas es mediante intervalos de confianza.

Obtener una estima puntual de la altura de ola correspondiente a un periodo de retorno no es obtener mucha información. En realidad, la probabilidad de que la altura sea exactamente un valor estimado es cero, y en general se desconoce la probabilidad de que ese valor sea superado. Una estima puntual tampoco da idea alguna de la exactitud de la estima. La solución a estos problemas es dar una estima por intervalo además de la estima puntual, aunque para periodos de retorno muy altos la estima puntual apenas tiene ningún valor por su inexactitud. Un intervalo de confianza es sencillamente un intervalo, estimado de alguna manera a partir de la muestra, que tienen una confianza determinada (normalmente el 80%, 90% o el 95%) de contener el valor real de la altura de ola que se desea estimar. Hay muchos métodos para hallar intervalos de confianza,

---

pero siempre hay que aceptar alguna hipótesis para poder obtener un intervalo. Normalmente la hipótesis que debe hacerse es que los datos se distribuyen según un modelo de distribución concreto.

En esta tesina se plantean 5 métodos para obtener el intervalo de confianza: método de simulación estadística con experimentos tipo bootstrap, fórmulas indicadas por el PIANC (*Permanent International Association of Navigation Congresses*), fórmulas empíricas propuestas por Goda, fórmula de Cramer y simulación estadística. En los dos casos de simulaciones estadísticas se realizan 2000 iteraciones, número recomendado por Vergés (1995).

### 2.3.5. *Bondad de ajuste*

De todas las funciones de distribución y los métodos de ajuste presentados, hay que decidir cual es la combinación que mejor se ajusta a los datos.

Primero de todo se ha de comprobar si los ajustes obtenidos son aceptables, y posteriormente quedarse con el ajuste adecuado. Estas tareas no son nada fáciles, y lo máximo que se puede conseguir es tomar la decisión con más probabilidad de ser correcta (si se aceptan algunas hipótesis) o conocer la probabilidad de hacer un cierto tipo de error.

La decisión no es única; por ejemplo, si se consideran todos los ajustes aceptables, se puede escoger aquel ajuste que tenga mínima incertidumbre. En cambio, se puede decidir escoger la distribución que tenga la menor probabilidad de ser incorrecta, independientemente de cuál sea la estima.

Para tomar todas estas decisiones se han desarrollado una serie de tests, llamados de bondad de ajuste, que en general se basan en considerar si es probable o improbable que la muestra obedezca a la distribución ajustada. Así pues, se debe efectuar una hipótesis inicial que se suele denominar  $H_0$ : que los datos obedecen a la distribución hipotética (o ajustada). Se define la significación de un test de bondad de ajuste  $\alpha$ , como la probabilidad de rechazar la hipótesis  $H_0$  condicionada a que la hipótesis es cierta (este es un error tipo I).

$$\text{significación } (\alpha) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) \quad (14)$$

Está claro que interesa un valor de  $\alpha$  pequeño, pero entonces la posibilidad de aceptar hipótesis falsas (error tipo II) será muy alta. Lo normal es aceptar un valor suficientemente pequeño de  $\alpha$  (por convenio se suele tomar 0,1 o 0,05).

Se han desarrollado tests de bondad de ajuste para distribuciones concretas, pero son los tests generales los más utilizados. Estos son el test  $\chi^2$  y el de Kolmogorov-Smirnov. Los tests, en general, se basan en calcular un estadístico función de la muestra y de la función de distribución hipotética. Si este estadístico se encuentra dentro de un intervalo que es función de la significación del test se acepta la hipótesis y, si no, se rechaza.

El test utilizado en esta tesina es el de Kolmogorov-Smirnov. El contraste o test de Kolmogorov-Smirnov (KS), compara la función de distribución teórica con la empírica.

Es un test muy sencillo y gráfico, válido para cualquier tamaño muestral, es decir, no se requieren grandes muestras, como sucede en este caso. Se ha considerado un nivel de significación de 0,1, es decir, si se cumple el test, tan sólo un 10% de los datos no siguen la distribución teórica elegida.

Otro estadístico de bondad de ajuste, que normalmente se utiliza en los ajustes por mínimos cuadrados, es el coeficiente de correlación. Se basa en suponer que si los datos obedecen a una cierta distribución, se distribuirán alrededor de una línea recta en el papel de probabilidad correspondiente a esa distribución. El estadístico utilizado en el test es el coeficiente de correlación,  $r$ , que para un conjunto de  $n$  puntos  $(x, y)$  se define como:

$$r^2 = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2} \quad (15)$$

El coeficiente de correlación es una buena medida de lo alineados que están los puntos. Si está por debajo de cierto valor se considera que los puntos no están suficientemente alineados y se rechaza la hipótesis y el ajuste.

#### **2.4. Programa Stones**

Una vez seleccionados los máximos a partir del método POT, los siguientes pasos son ajustar estos máximos a las distribuciones teóricas escogidas mediante los diferentes métodos de ajuste y escoger la distribución que mejor ajuste a partir de un método de bondad de ajuste, con el objetivo de poder calcular con esta función la altura de ola correspondiente a un periodo de retorno dado y su intervalo de confianza.

Una vez explicadas las funciones de distribución, los métodos de ajuste, el intervalo de confianza y el método de bondad de ajuste que se utilizan, solo queda por explicar el programa con el que se realizan estos cálculos: el programa Stones.

El programa de fortran (Stones.for) fue desarrollado por Vergés (1995) para el estudio del comportamiento de las funciones de distribución utilizadas normalmente en el estudio extremal del oleaje, es decir, la estimación de la altura de ola significativa correspondiente a cada periodo de retorno. También se pretende estudiar con él los diferentes métodos de ajuste que se pueden utilizar, así como elegir la mejor distribución gracias al criterio de bondad de ajuste.

Como es habitual en el lenguaje fortran, a parte de con el programa, se acostumbra a trabajar con un archivo de entrada de datos y varios de salida con los resultados. Después de que el programa localice el archivo de entrada de datos, se prosigue con la lectura de los datos en él almacenados. En este archivo es donde, previamente a la ejecución del programa, se ha de escribir toda la información necesaria para que el programa pueda realizar todos los cálculos programados. En posteriores apartados se explicará con más detalle la información en él registrada, pero a grandes rasgos se

---

puede decir que en el archivo de entrada se escriben las características de la serie de máximos (número de datos, duración del registro, el valor de los máximos,...), la distribución y el ajuste a utilizar, el método para hallar el intervalo de confianza, la información para el test de bondad de ajuste y los periodos y alturas para los que se quiere buscar las alturas y periodos respectivamente. Después de la ejecución del programa, se muestran los resultados en tres archivos de resultados que, igual que para el de entrada, se explicará su contenido en apartados posteriores.