

UNIVERSIDAD DE CANTABRIA

DEPARTAMENTO DE CIENCIAS Y TÉCNICAS  
DEL AGUA Y DEL MEDIO AMBIENTE

---

TESIS DOCTORAL

ESTUDIO DE LA VARIABILIDAD CLIMÁTICA DE VALORES  
EXTREMOS DE OLEAJE

---

Presentada por: CRISTINA IZAGUIRRE LASA

Dirigida por: FERNANDO J. MÉNDEZ INCERA  
IÑIGO J. LOSADA RODRIGUEZ

Octubre 2010

# CAPITULO I

## INTRODUCCIÓN

### 1.1 Marco de la tesis

El estudio de esta tesis se centra en el análisis de la variabilidad climática del oleaje extremal a distintas escalas espaciales y temporales. El mar y la costa son elementos clave en los procesos naturales y han jugado siempre un papel importante en el desarrollo de la humanidad.

El océano es uno de los componentes principales del sistema climático que ejerce, en éste, un papel fundamental de regulación. Por otro lado, su importancia, desde el punto de vista ambiental, es muy elevada al dar cabida a un gran número de ecosistemas.

Desde el punto de vista socioeconómico, a lo largo de la historia, el hombre ha sido capaz de encontrar en el mar un gran número de usos que le han permitido desarrollarse y evolucionar. Desde la antigüedad los usos como fuente de alimentación y vía de comunicación hicieron de los asentamientos costeros centros neurálgicos para las actividades comerciales, dando lugar a ciudades tan importantes como Constantinopla o Babilonia. Hoy en día, el desarrollo de estos intereses económicos junto con otros como el turismo, recreo o recursos energéticos han hecho que el 70 % de la población mundial se concentre en la franja costera (IPCC). Ejemplos de las ciudades más importantes del mundo son Nueva York, Sydney o Tokio, todas ellas situadas en la costa.

### **1.1.1 Circulación general del sistema climático**

El sistema climático es un sistema dinámico gobernado por la recepción, transformación y redistribución de energía, mediante la interacción de la atmósfera, el océano y las masas de tierra. Los fenómenos derivados del sistema climático se pueden clasificar en función de su escala espacio-temporal: los fenómenos de macroescala, asociada a miles de kilómetros y períodos en torno a la semana; los fenómenos mesoescalares, de cientos de kilómetros y períodos de días y los fenómenos de microescala, con longitudes de metros y períodos de horas.

La radiación solar entrante es la mayor fuente de energía que propulsa el sistema climático pero ésta no incide homogéneamente en toda la superficie de la tierra, siendo mayor en el ecuador y los trópicos y menor en las regiones polares. Esta diferencia en el calor incidente es equilibrada mediante un transporte meridional hacia los polos, el cual se lleva a cabo por la atmósfera y por los océanos. La atmósfera domina este transporte en latitudes medias mientras que la mayor parte del transporte en latitudes bajas corre a cargo del océano.

En la atmósfera, a escala global, la transferencia de calor hacia los polos se realiza a través de tres grandes células (Figura 1.1), siendo su estructura conocida como Circulación General de la Atmósfera.

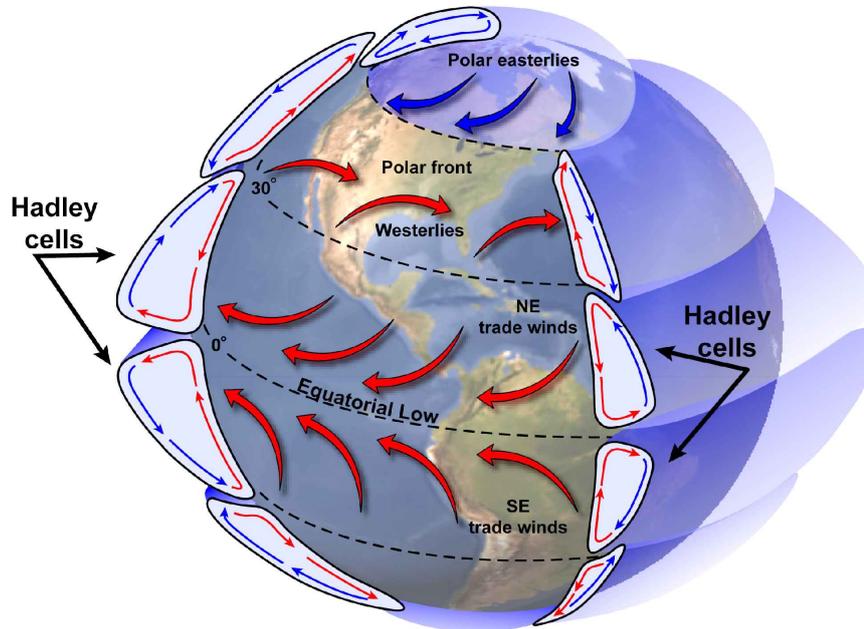


Figura 1.1 Circulación general de la atmósfera. Fuente: NASA

El aire caliente de las regiones ecuatoriales inicialmente asciende y se mueve hacia los polos, a latitudes mayores. Durante su propagación hacia los polos el aire se enfría y desciende hacia la superficie. Esto ocurre, en ambos hemisferios, cerca de la latitud 30° donde se forman los sistemas de altas presiones. Cuando el aire que desciende alcanza la superficie terrestre, una parte es recirculada hacia el ecuador formando los denominados vientos alisios (*trade winds*) y la otra se propaga hacia los polos. De manera similar, el aire frío desciende sobre los polos y se propaga hacia el ecuador. Alrededor de la latitud 60° los dos flujos de aire se encuentran, de forma que la convergencia fuerza al aire a ascender. La convergencia de estos dos flujos, uno muy frío y otro templado da lugar a un fuerte gradiente meridional de temperatura y presión que como resultado genera los fuertes vientos del oeste (*westerly winds*). El resultado de la interacción de estos flujos de aire da lugar a tres grandes células meridionales: las células tropicales son conocidas como “células de Hadley”, las células de latitudes medias como “células de Ferrel” y las de latitudes polares como “células polares”.

En los océanos el transporte de calor se produce mediante dos tipos de circulación. Una parte forzada por el viento, que domina en las capas superficiales del océano, y otra forzada por las diferencias de densidad, llamada circulación termohalina y dominante en profundidad. Sin embargo, la circulación del océano no es tan simple como la suma de ambas partes, pues además ocurren una serie de complejas interacciones. En la Figura 1.2 se puede ver un esquema de las principales corrientes oceánicas forzadas por el viento.

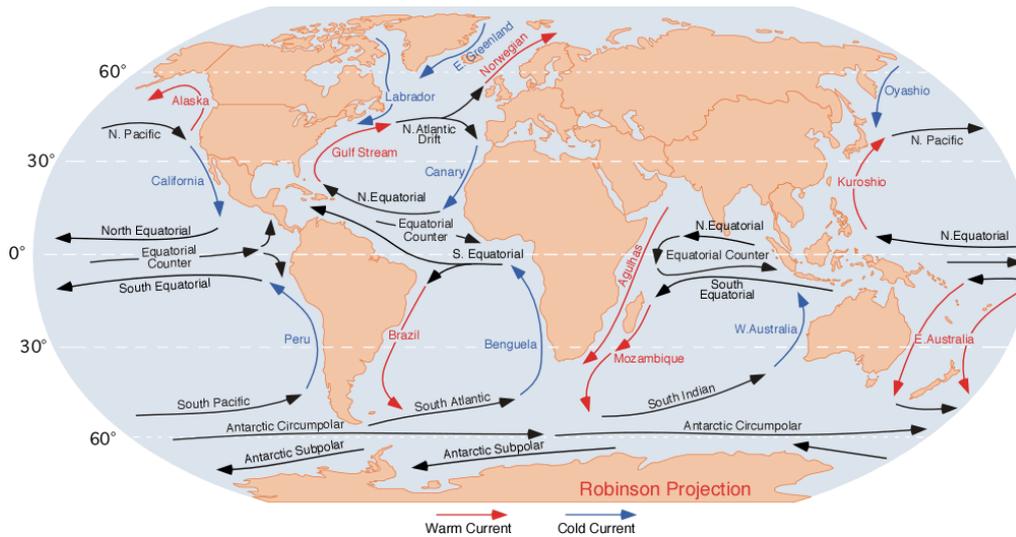


Figura 1.2 Esquema de las principales corrientes marinas. Fuente: [http://es.wikipedia.org/wiki/Corriente\\_marina](http://es.wikipedia.org/wiki/Corriente_marina)

La circulación generada por el viento es forzada por la transferencia de momento de la atmósfera a la superficie del océano, con patrones de circulación en sentido horario en el Atlántico Norte y Pacífico Norte, así como patrones de circulación antihorarios en el hemisferio sur. De forma general se pueden distinguir dos tipos de corrientes. En las cuencas occidentales de los océanos discurren las corrientes en dirección a los polos que transportan agua caliente de los trópicos a latitudes medias. En el Hemisferio Norte destacan la corriente del Golfo en el Atlántico Norte y la corriente de Kuroshio en el Pacífico Norte. La corriente de Brasil y Agulhas representan, de manera menos pronunciada, su equivalente en el Hemisferio Sur. Por otro lado, en las cuencas orientales de los océanos discurren las corrientes asociadas al transporte de agua fría hacia el ecuador. Éstas son las de California y Perú (o corriente de Humbolt) en el Pacífico, las corrientes de Canarias y Benguela en el Atlántico y la corriente de Leeuwin en el oeste de Australia. En general las corrientes del Oeste son más intensas, estrechas, rápidas y profundas, mientras que las del Este son más anchas, someras y lentas.

Este balance de energía llevado a cabo por la atmósfera y el océano tiene como objetivo mantener unas condiciones climáticas cuasi-estacionarias, caracterizadas por procesos atmosféricos como el viento, la temperatura del aire o la precipitación y oceánicos como el oleaje, las corrientes o la temperatura superficial del agua. Sin embargo, el estado del sistema climático no es constante sino que varía como resultado de forzamientos externos y variaciones internas.

### **1.1.2 Variabilidad del sistema climático**

El sistema climático sufre una serie de fluctuaciones temporales alrededor de su estado medio, al conjunto de las cuales se le denomina variabilidad climática. Esta variabilidad, que proporciona la estabilidad al sistema es, por tanto, una característica natural.

La variabilidad del sistema climático ocurre a distintas escalas dependiendo del forzamiento que induzca la fluctuación. Este rango de escalas va desde el muy largo plazo (escala geológica) donde se habla de cambios en miles de años, a la escala diurna, donde se producen cambios climáticos de la noche al día. Entre estos dos extremos se puede hablar de cambios a escala estacional, interanual, decadal y de largo plazo, donde estaría incluido el cambio climático. Históricamente, la percepción de la variación del clima a distintas escalas ha ido cambiando. Hoy en día, el período de tiempo considerado para determinar una descripción estadística del clima en términos de media y variabilidad es de 30 años.

La regularidad de las fluctuaciones diurnas y estacionales es un proceso bien conocido pero no se puede decir lo mismo de las variaciones de tipo interanual y decadal inducidas por fenómenos combinados de océano-atmósfera y que tienen manifestaciones cuasi-regulares. El estudio de estas variaciones, y las de largo plazo se ha convertido en las últimas décadas en un área de gran interés científico.

En lo que se refiere a la escala espacial el sistema climático también presenta una variabilidad a lo largo del globo. Atendiendo a los fenómenos macroescalares de tipo atmosférico se pueden distinguir las principales zonas climáticas del globo determinadas por las tres grandes células meridionales. Desde el punto de vista mesoescalar, la configuración de una determinada zona y condiciones locales de

orografía serán claves para determinar sus características climáticas. Por último, también el estudio de la escala local también refleja variaciones en el clima.

### 1.1.3 Eventos extremos

Además de la variabilidad a distintas escalas, los fenómenos extremos son una parte natural de un sistema climático estable. En general, cuando se habla de eventos extremos se piensa en episodios intensos de corto plazo, como por ejemplo un huracán, lluvias torrenciales o una ola de calor. Sin embargo, más que tratarlos aisladamente, únicamente en términos de duración o intensidad, es más útil estudiarlos desde el punto de vista estadístico. Para ello, es necesario conocer el registro temporal de la variable de estudio. Si se representan los valores de una variable en particular, por ejemplo la temperatura o precipitación, frente a su probabilidad de ocurrencia, se observará que los valores medios son los más probables, mientras que los valores extremos se encontrarán en la cola de la distribución (Figura 1.3).

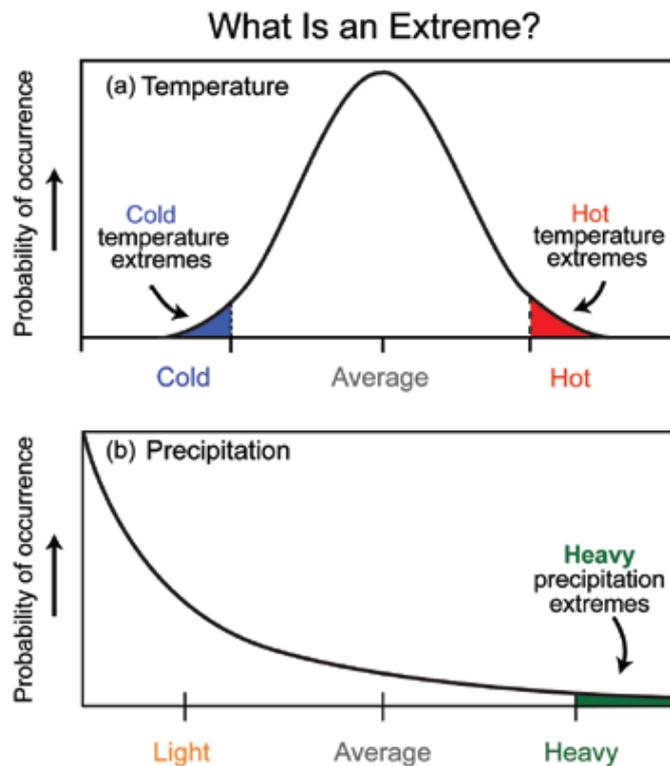


Figura 1.3 Funciones de densidad de temperatura y precipitación diarias. Fuente: Weather and Climate Extremes in a Changing Climate (U.S. Climate Change Science Program, Synthesis and Assessment Product 3.3, 2008)

De acuerdo con el glosario del Cuarto Informe de Evaluación (IPCC, 2007) del Panel Intergubernamental de Cambio Climático (IPCC) un evento extremo climático es un evento raro en un lugar e instante determinado. De esta manera, se pueden definir los eventos raros como aquellos con una probabilidad de excedencia menor del 10 % (áreas señaladas en las curvas de densidad de probabilidad de la Figura 1.3).

El estudio estadístico de los eventos extremos permite observar una variabilidad temporal a distintas escalas, como puede ser la estacional, interanual o el largo plazo. Los sistemas humanos y naturales se han ido adaptando a esta variabilidad de forma histórica. Sin embargo, debido a su excepcionalidad, estos eventos extremos llevan asociados unos costes y beneficios, tanto de tipo socioeconómico como ambiental. Por ejemplo, los huracanes proporcionan lluvias necesarias en ciertas áreas donde algunas comunidades de plantas tropicales dependen de que los vientos huracanados derriben los árboles grandes para así poder recibir más luz solar y desarrollarse. Sin embargo, para el ser humano muchas veces provocan muertes y daños irreparables, como ocurrió con el huracán Katrina en 2005. La adaptación de los sistemas (naturales y humano) al rango histórico de fenómenos extremos hace que la mayoría de los eventos fuera de este rango tengan impactos negativos. Hay que tener en cuenta que hay evidencias científicas de que un calentamiento global va acompañado de cambios en la intensidad, duración, frecuencia y extensión de los extremos climáticos (IPCC, 2007). La regulación del clima global debido al continuo incremento de la concentración de gases de efecto invernadero en la atmósfera hace que muchos aspectos de la naturaleza de los extremos se vean potencialmente modificados (Easterling et al., 2000a,b). Los aspectos más comunes son la intensidad y la frecuencia. En la Figura 1.4 se puede ver como la variación en la media de la distribución de la variable (por ejemplo temperatura) da lugar a un cambio importante en los extremos. En el caso de la temperatura el incremento de la media produce valores extremos mayores de los ocurridos históricamente y un aumento de la probabilidad de los extremos históricos.

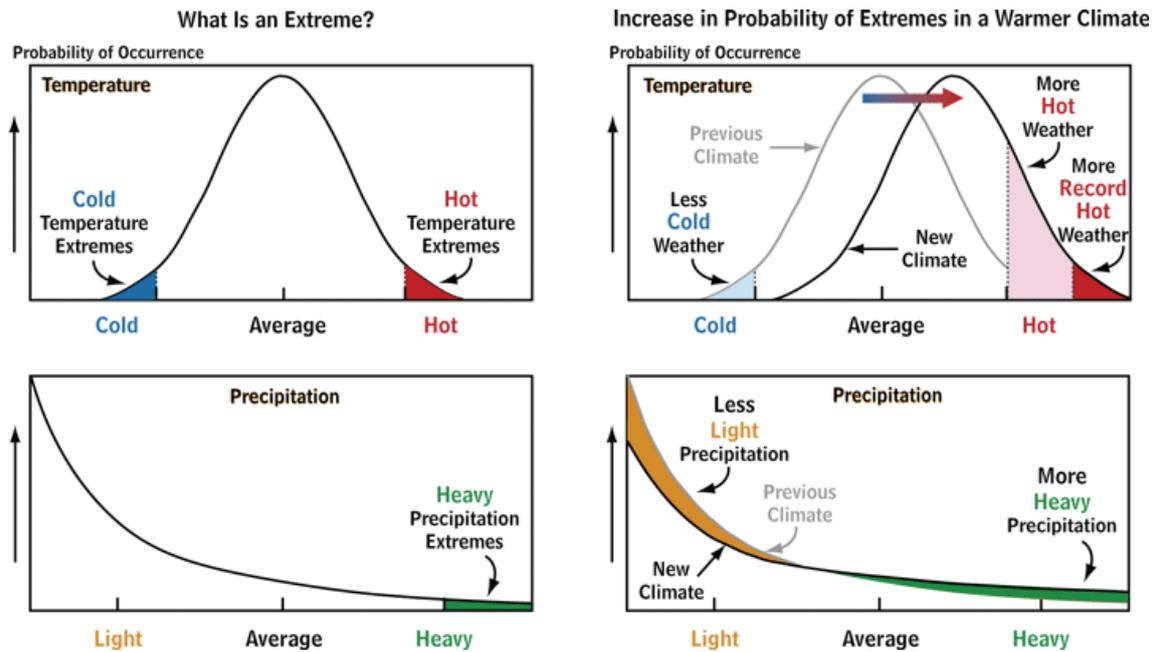


Figura 1.4 Variaciones en la función de densidad de las variables temperatura diaria y precipitación.  
Fuente: Weather and Climate Extremes in a Changing Climate (U.S. Climate Change Science Program, Synthesis and Assessment Product 3.3, 2008)

Debido a la vulnerabilidad de los sistemas frente a los eventos extremos, es imprescindible un buen conocimiento de la variabilidad temporal y espacial de las variables características de procesos como el oleaje, en el caso de esta tesis. Para ello, la estadística de extremos permite investigar estas variaciones a partir de los registros históricos de la variable y proyectar a situaciones futuras bajo escenarios de cambio climático.

Hay que tener en cuenta también que en el caso del sistema humano los impactos debidos a los cambios en los extremos no solo dependen de los cambios en el clima, sino también de la vulnerabilidad del sistema social. La vulnerabilidad está determinada por factores tales como la población o el estatus económico y otros como las medidas de adaptación reflejadas en las normas de construcción de edificios, uso eficiente del agua, etc. Los desastres ocurridos como consecuencia del tsunami de Indonesia en 2004, las inundaciones de Perú en 2010 o los terremotos de Haití y Chile en 2010 son ejemplos de las consecuencias de fenómenos extremos sobre comunidades altamente vulnerables.

### 1.1.4 Oleaje

El oleaje es el fenómeno que mejor representa el estado del mar. Este fenómeno es la respuesta de la superficie del mar a la variación en las condiciones atmosféricas. Sin embargo, su estudio es complejo y no ha sido hasta la segunda mitad del siglo XX cuando los primeros modelos analíticos han empezado a ser capaces de ofrecer una descripción completa del oleaje.

El oleaje se analiza como un proceso estocástico, lo que requiere que el proceso cumpla una serie de propiedades, entre ellas que sea estacionario, o al menos débilmente estacionario. Para garantizar estas propiedades, es necesario trabajar con períodos del proceso de una duración tal que, durante ese tiempo la variación de los estadísticos (por ejemplo media y varianza) del proceso sea inapreciable. Estos intervalos de tiempo en que el proceso se puede considerar débilmente estacionario se denominan “estados de mar”. El análisis estadístico y espectral del oleaje en estados de mar se denomina análisis de corto plazo. Su resultado da lugar al conocimiento de ciertos parámetros de estado de mar de tipo estadístico, como altura de ola significativa ( $H_s$ ) o período medio ( $T_m$ ) o de tipo espectral, como altura de ola del momento de orden cero espectral ( $H_{m0}$ ) o período de pico espectral ( $T_p$ ), los cuales proporcionan una completa definición del oleaje.

Los parámetros de estado de mar varían en cada estado de mar, constituyendo, a lo largo del tiempo, una serie temporal de estados de mar. Las características estadísticas de esta serie describen el clima marítimo a largo plazo en una determinada zona.

El estudio de los eventos extremos del oleaje y su variabilidad climática es solo posible mediante el análisis estadístico del clima marítimo a largo plazo.

### 1.1.5 Motivación

El oleaje es uno de los procesos derivados del sistema climático y, como tal, lleva asociado una ocurrencia de eventos extremos y una variabilidad de éstos. En cuanto a la variabilidad temporal se pueden distinguir distintas escalas: hay meses del año en los que los extremos de oleaje son mayores y más frecuentes (estacional); no todos los años

tienen el mismo número e intensidad de extremos de oleaje (interanual) y a lo largo de los años los extremos pueden ir cambiando (largo plazo). De la misma forma, también se observa una variabilidad a escala espacial en este proceso. Atendiendo a la escala global se pueden distinguir zonas de oleaje ligadas a fenómenos atmosféricos de macroescala como las borrascas o huracanes. En la escala regional se pueden distinguir oleajes generados por tormentas o vientos condicionados por la orografía y condiciones atmosféricas de la zona como, por ejemplo, el oleaje del Mar Mediterráneo. Por último, a escala local, en su propagación, el oleaje se ve afectado por procesos físicos tales como refracción, difracción, asomeramiento o rotura, procesos condicionados por las características del fondo.

Examinando los distintos intereses socioeconómicos y ambientales se pueden distinguir, a la hora de su diseño y gestión, distintas escalas temporales y espaciales de interés. Por ejemplo, el uso del océano como vía de comunicación ha requerido el diseño de rutas marítimas. El diseño adecuado de rutas comerciales se debe llevar a cabo mediante el conocimiento de la variabilidad espacial del oleaje a escala global y la variabilidad temporal del oleaje dentro del año (estacional). Otro ejemplo es el diseño de infraestructuras marítimas como, por ejemplo, los puertos. El diseño estará condicionado por la escala temporal de largo plazo de los extremos de oleaje en la zona (escala espacial local). Sin embargo, una vez iniciada la construcción del puerto, los trabajos estarán condicionados por la variabilidad estacional de los extremos de oleaje en la zona. En el estudio de las pesquerías y zonas de marisqueo interesará conocer la variabilidad estacional e interanual de los extremos de oleaje en la zona ya que condicionará el desarrollo y captura de las comunidades. En lo que respecta a las zonas costeras, su alta vulnerabilidad, debido a la cantidad de población concentrada en esta zona, hace que los eventos extremos del oleaje tengan a menudo consecuencias catastróficas. Las inundaciones, el daño en infraestructuras o la degradación de ciertos ecosistemas son algunos ejemplos. Para una gestión adecuada de la costa será necesario un buen conocimiento de los extremos de oleaje a largo plazo en la zona de interés. A este respecto, el cambio climático toma un papel importante. Las tendencias observadas en los extremos del oleaje en los últimos años pueden variar ante el calentamiento global, por lo que habrá que tener en cuenta los posibles escenarios de cambio climático propuestos por el IPCC para el análisis adecuado del largo plazo.

Por tanto, el conocimiento de la variabilidad espacial y temporal de los extremos del oleaje es necesario, desde el punto de vista del buen conocimiento del proceso físico, y clave para el diseño y gestión de los intereses socioeconómicos y ambientales que el hombre tiene en la costa y el océano. En esta tesis se va a estudiar la variabilidad temporal de los extremos de oleaje a escala estacional, interanual y de largo plazo y la variabilidad espacial a escala global, regional y local. El estudio a cada una de las escalas implicará una serie de condicionantes derivados de las bases de datos disponibles y los procesos físicos dominantes en cada escala.

El estudio de la variabilidad de los extremos de oleaje se va a llevar a cabo, principalmente, a través de la estadística de extremos. Sin embargo, el estudio de ciertas escalas de interés como el largo plazo implicará el uso de técnicas de *downscaling* estadístico y otras como la escala local requerirá de la combinación de técnicas de *downscaling* dinámico y *downscaling* estadístico que permitan transferir las dinámicas de oleaje de aguas profundas a aguas someras.

## 1.2 Estado del arte

El punto de partida de cualquier trabajo de investigación es el análisis y revisión de los trabajos y aportaciones hechas en la línea de estudio. En este apartado se resume el estado del conocimiento en el estudio de valores extremos de oleaje y su variabilidad temporal y espacial. En primer lugar, se describen los modelos estadísticos desarrollados en el estudio de los valores extremos de oleaje y las técnicas estadísticas utilizadas en el estudio de su variabilidad climática. A continuación, se describe el estado del conocimiento en lo que respecta a la variabilidad espacial del clima marítimo, distinguiendo entre la escala global y regional y la escala local. Por último, se describen los estudios hechos sobre la variabilidad a largo plazo de los extremos de oleaje en el siglo XXI mediante proyecciones.

## 1.2.1 Métodos estadísticos

### 1.2.1.1 Estadística de extremos

El desarrollo de la ingeniería moderna a lo largo de los últimos dos siglos ha dado lugar también al desarrollo y estudio con gran éxito de la estadística de valores extremos. El diseño de una infraestructura siempre se hace con la premisa de que ésta sea capaz de soportar las máximas solicitaciones en un período de tiempo determinado. En muchos ámbitos de la ingeniería estas solicitaciones son fenómenos ambientales, por ejemplo, una obra de abrigo de un puerto se diseñará para que sea capaz de soportar los temporales más grandes durante su vida útil. Es decir, el diseño viene condicionado por los valores extremos de altura de ola. Esta necesidad de conocer el comportamiento de los valores extremos dio lugar al desarrollo de la Teoría de Valores Extremos (TVE) y a su aplicación en distintos campos de la ingeniería, la climatología y la hidrología principalmente.

La TVE es la especialidad estadística que trata de estudiar los valores extremos (máximos o mínimos) de una muestra de datos, es decir, los valores que se sitúan en la cola de la distribución probabilística de la variable de estudio. Para ello, considera la distribución probabilística de estos valores. Ya a finales del siglo XIX empiezan a surgir los primeros estudios sobre extremos y en la primera mitad del siglo XX cabe destacar el trabajo de Bortkiewicz (1922), quien introduce el concepto de distribución de probabilidad de valores extremos y los de Fréchet (1927), Fisher y Tippett (1928), cuya demostración de que sólo son posibles tres familias paramétricas de distribuciones límite para máximos y sus equivalentes para mínimos ha sido uno de los resultados que más influencia ha tenido en el desarrollo de la TVE. Von Mises (1936) y Rice (1939) también abordan el problema de la distribución de los extremos y más tarde Gnedenko (1943) define el teorema de los tipos de extremos. En 1958 Gumbel presenta el libro “Statistics of Extremes”, que se convierte en una de las principales referencias de la TVE. Ya en la segunda mitad del siglo XX destacan los trabajos de Leadbetter et al. (1983) o Galambos (1987) sobre los desarrollos matemáticos que caracterizan las propiedades de las distribuciones de extremos. Un libro de referencia sobre la TVE y sus aplicaciones en ingeniería es “Extreme Value and Related Models with Applications in Engineering and Science” de Castillo et al. (2005).

Dado un problema en el que se requiere el estudio de los valores extremos, si consideramos que partimos de una secuencia de observaciones aleatorias e independientes  $X_1, \dots, X_n$  con una función de distribución común asociada  $F$ , tal que  $F(x) = \Pr\{X_i \leq x\}$  para cada  $x$  dentro del rango  $0 < F(x) < 1$ , el objetivo será modelar el comportamiento estadístico de la función de distribución de valores extremos, a los que se denotará de la forma  $M_n = \max\{X_1, \dots, X_n\}$ . La función de distribución de extremos puede derivarse teóricamente a partir de todos los valores, considerando la  $n$ -ésima potencia de  $F$ :

$$\Phi_n(x) = \Pr\{M_n \leq x\} = \Pr\{X_1 \leq x, \dots, X_n \leq x\} = \prod_{i=1}^n \Pr\{X_i \leq x\} = \{F(x)\}^n \quad (1.1)$$

En primer lugar, habrá que tener en cuenta que para poder aplicar la TVE hay que asumir tres hipótesis principales de partida:

1. El proceso es estocástico, es decir, se trabaja con la secuencia ordenada de la variable aleatoria y su probabilidad asociada.
2. Las observaciones de partida son independientes e idénticamente distribuidas, *iid*. Lo que quiere decir que constituyen una variable aleatoria donde cada observación tiene la misma función de distribución que las otras y todas son mutuamente independientes.
3. La variable de estudio es estacionaria. Esto implica que los valores muestrales pueden ser mutuamente independientes, pero sus propiedades estocásticas son homogéneas a lo largo del tiempo.

En segundo lugar, hay que elegir la muestra de valores extremos a la que se va a aplicar la TVE.

## Selección de los valores extremos

Hay distintos métodos de selección de la muestra de extremos a estudiar. Según las características de la serie temporal y el estudio que se quiera llevar a cabo serán más apropiados unos métodos u otros.

- *Método de máximos en un bloque de tiempo*

El método de máximos en un bloque de tiempo es el método más sencillo de selección de extremos. Se trata de definir una unidad de tiempo y escoger el valor máximo de la variable de estudio ocurrido en ella. La duración del bloque de tiempo deberá ser lo suficientemente grande como para representar los sucesos extremos que se quiere modelar, pero también será necesario tener en cuenta la longitud total de la serie temporal de la variable de estudio, puesto que habrá que garantizar un número mínimo de datos en la muestra para poder aplicar el análisis estadístico.

El método más clásico derivado de máximos en un bloque de tiempo es el Método del Máximo Anual (MMA) (Coles, 2001). Con este método la serie temporal se divide en unidades de tiempo de 1 año de duración y se escogen los máximos de la variable en esa unidad. De esta forma se tiene una serie continua de valores extremos a lo largo del tiempo. El principal inconveniente de este método es su dependencia con la longitud de la serie temporal de la variable. En el caso de variables oceanográficas las series de las que se dispone tienen registros cortos, lo que hace que la muestra de extremos obtenida a partir de este método tenga un escaso número de datos. Otras variaciones de este método son el Método de Máximos Mensuales (MMM) o el Método de Máximos Semanales (MMS), los cuales permiten solventar el problema de la falta de datos en la muestra y además obtener muestras más apropiadas para estudiar distintos patrones temporales. Sin embargo, hay que estudiar la dependencia entre temporales, garantizando que los eventos elegidos son independientes entre sí.

Hay que tener en cuenta también la falta de datos de la serie temporal de la variable de estudio en el bloque de tiempo. Si el porcentaje de huecos (datos no disponibles respecto al número de datos esperable en el bloque temporal) es superior al 40 % no es aconsejable utilizar esa unidad temporal.

- *Método de  $r$ -máximos en un bloque de tiempo*

El método de los  $r$ -máximos en un bloque de tiempo es una variación del método anterior. En este caso aparte de definir la unidad temporal también se define el número de eventos a elegir ( $r$ ), lo cual permite aumentar el tamaño de la muestra. El valor de  $r$  deberá ser el máximo posible siempre que se cumpla la aproximación asintótica del máximo dominio de atracción (MDA), ya que de esta manera aumentará la precisión del modelo. Sin embargo, si el valor de  $r$  es demasiado grande se estarán incluyendo en la muestra valores que no pertenecen a la cola de la distribución, por lo que no se ajustarán bien al modelo de máximos. Se aconseja no tomar valores de  $r$  mayores a 10-15 puesto que no aportan buenos resultados (Coles, 2001, Guedes Soares y Scotto, 2004).

- *Método de excedencias sobre un umbral*

El método de excedencias sobre un umbral (*Peak Over Threshold*, POT) selecciona los valores que superen un valor umbral de corte ( $u$ ). Con este método hay que fijar el valor  $u$ , pero también hay que garantizar que los eventos seleccionados sean independientes, lo cual queda determinado con el valor mínimo de tiempo entre eventos ( $\Delta t$ ).

La buena selección del umbral de corte y la independencia entre temporales serán determinantes en la pericia del modelo ajustado. Ambos parámetros deben elegirse en función de la física de la variable y según las características de la población de datos.

La selección del umbral de corte deberá ser lo suficientemente alto como para garantizar que todos los eventos elegidos pertenecen a la cola de la distribución, puesto que, en caso de no ser así se producirá sesgo en los resultados. Si por el contrario el umbral elegido es demasiado alto, la muestra contará con pocos datos y esto provocará una varianza muy grande del modelo. Según el estudio de extremos que se quiera llevar a cabo el tipo de umbral elegido podrá ser determinante para capturar la información necesaria de la población de datos. Un umbral constante a lo largo del tiempo proporcionará una muestra de eventos extremos absolutos para toda la serie temporal, mientras que un umbral variable con el tiempo ( $u(t)$ ) permitirá incluir, en la muestra de

datos, extremos más locales que permitirán estudiar variaciones inter o intra anuales de los extremos.

En cuanto a la independencia entre temporales  $\Delta t$ , el valor debe cumplir, por un lado, las bases matemáticas de partida, y por otro, debe ser un período de tiempo que explique razonablemente la independencia basándose en la física de la variable y el tipo de evento extremo (Luceño et al., 2006). En el caso de variables geofísicas, y más en concreto de oleaje, se han llevado a cabo numerosos estudios para medir la duración de las tormentas que generan un evento extremo, proponiéndose distintos  $\Delta t$ . Así Morton et al. (1997) proponen  $\Delta t = 1.25$  días, Guedes Soares y Scotto (2004)  $\Delta t = 20$  días, Méndez et al. (2006) utilizan  $\Delta t = 3$  días y Caires et al. (2006)  $\Delta t = 2$  días. Alves (2006) presenta un cuadro con la duración de las tormentas en las distintas partes de los océanos. Si bien, hay que tener en cuenta las características del oleaje en la zona de estudio, puesto que será determinante a la hora de elegir la dependencia entre temporales.

- *Método de máximos condicionado a una determinada duración del evento extremo*

Hasta ahora, los métodos de selección de extremos presentados se centran en la intensidad de la variable. Sin embargo, otro dato de relevancia para el estudio de extremos en el ámbito de la ingeniería oceanográfica y de costas es la duración de la excedencia ( $D$ ) o persistencia. El diseño de estructuras, como por ejemplo diques en talud, requiere del análisis estadístico de persistencias, ya que la máxima altura de ola de un estado de mar en una distribución Rayleigh está relacionada con la duración de la tormenta (Vidal et al., 2006). Por otro lado, en el estudio de la morfodinámica de playas también es necesario el análisis estadístico de las persistencias de oleaje (Callaghan et al., 2008).

En este caso, la selección de extremos se llevará a cabo fijando una duración de excedencia,  $D$ , y eligiendo los valores extremos de la variable de estudio que exceden o igualan esa duración.

## Modelos estacionarios de extremos

La TVE estudia el comportamiento de los valores extremos a partir de modelos basados en distribuciones de probabilidad. Existen dos tipos de aproximaciones a los modelos de extremos, los modelos exactos y los modelos asintóticos. En esta tesis los estudios realizados se han llevado a cabo en base a modelos asintóticos, los cuales se describen a continuación.

Los modelos asintóticos consideran que, bajo ciertas hipótesis de partida, es posible definir una expresión asintótica para  $\Phi_n(x)$  y de esta obtener  $F(x)$ ; de esta manera el modelo permite extrapolar a valores que no se encuentran en los datos de la muestra. Los dos métodos actuales de resolución de modelos asintóticos de extremos se basan en los siguientes teoremas de la TVE:

- La aproximación clásica de valores extremos basada en el Teorema I o Teorema de Fisher-Tippet (Galambos, 1987).
- La aproximación basada sólo en el ajuste de la cola de la distribución. Se define a partir del Teorema II o Teorema de Pickands (Pickands, 1975).

El Teorema I considera que los datos provienen de la población completa de valores (método de máximos o de  $r$ -máximos en un bloque de tiempo), mientras que el Teorema II sólo tiene en cuenta la rama alta de la población de datos, es decir, aquellos que superan un determinado umbral (método de excedencias sobre un umbral).

- *Modelo basado en la Distribución Generalizada de Extremos (GEV)*

El problema de modelar el comportamiento estadístico de la función de distribución  $\Phi_n(x)$  de los valores extremos  $M_n$  pasa por obtener su expresión teórica, que se puede derivar de la función de distribución  $F$  asociada a cada secuencia de observaciones  $X_n$ . Sin embargo, en la práctica esto no es del todo útil, ya que la función  $F$  es, en la mayoría de las veces, desconocida.

Una solución a este problema pasa por la aproximación alternativa de aceptar que  $F$  es desconocida y buscar una familia de modelos para  $F^n$ , que sólo podrán ser estimados en base a la TVE. Esto es similar a la práctica usual que se lleva a cabo en la estadística de valores medios, en la que la distribución del valor medio se aproxima a la distribución normal mediante el teorema del límite central.

La variable  $M_n$  presenta ciertos inconvenientes. Para cualquier  $x < x_+$ , donde  $x_+$  es el valor más grande de  $F$  (o el valor más pequeños de  $x$  tal que  $F(x)=1$ ), entonces  $F^n(x) \rightarrow 0$  cuando  $n \rightarrow \infty$ , de forma que la distribución de  $M_n$  degenera en una masa puntual en  $x_+$  (solo tomará valores 0 y 1). Esta dificultad se puede evitar mediante una normalización lineal de la variable  $M_n$ :

$$Z_n = \frac{M_n - b_n}{a_n} \tag{1.2}$$

para una secuencia de números reales  $\{a_n > 0\}$  y  $\{b_n\}$ . Valores apropiados de  $\{a_n\}$  y  $\{b_n\}$  estabilizan los parámetros de localización y escala de  $Z_n$  a medida que  $n$  aumenta, evitando así las dificultades de la variable  $M_n$ .

El Teorema de Fisher-Tippet establece que, si existe una secuencia de valores  $\{a_n > 0\}$  y  $\{b_n\}$  tales que

$$\Pr\{Z_n \leq x\}^{n \rightarrow \infty} = \Pr\{(M_n - b_n)/a_n \leq x\}^{n \rightarrow \infty} \rightarrow G(x) \tag{1.3}$$

donde  $G$  es una función de distribución no degenerada, entonces  $G$  pertenece a alguna de las familias de distribuciones que engloba la Función Generalizada de Extremos (*Generalized Extreme Value*, GEV) ampliamente conocidas como familias de Gumbel, Fréchet y Weibull. Cada familia tiene un parámetro de localización y escala respectivamente y además, las familias de Fréchet y Weibull tienen un parámetro de forma.

La función de distribución de la GEV de máximos tiene la expresión:

$$G(x; \mu, \psi, \xi) = \begin{cases} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\psi} \right) \right]_+^{-1/\xi} \right\} & \text{si } \xi \neq 0 \\ \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\psi} \right) \right] \right\} & \text{si } \xi = 0 \end{cases} \quad (1.4)$$

Y la función de densidad se define mediante:

$$g(x; \mu, \psi, \xi) = \begin{cases} \frac{1}{\psi} \left[ 1 + \xi \left( \frac{x - \mu}{\psi} \right) \right]_+^{-(1+1/\xi)} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\psi} \right) \right]_+^{-1/\xi} \right\} & \text{si } \xi \neq 0 \\ \frac{1}{\psi} \exp \left( - \frac{x - \mu}{\psi} \right) \exp \left[ - \exp \left( - \frac{x - \mu}{\psi} \right) \right] & \text{si } \xi = 0 \end{cases} \quad (1.5)$$

donde  $[a]_+$  implica el  $\max(a, 0)$ ,  $-\infty < \mu < \infty$  es el parámetro de localización,  $\psi > 0$  es el parámetro de escala y  $\xi$  es el parámetro de forma.

El parámetro de localización representa los valores medios de la variable aleatoria  $x$  y define el valor de la variable con probabilidad de no excedencia  $\exp(-1)$ . El parámetro de escala representa la dispersión de la muestra. Por último, el parámetro de forma es el que define la familia de la distribución en función del comportamiento de la cola. En el caso de que exista  $\xi$  la expresión de  $G(x)$  engloba dos familias de distribuciones.

- Cuando  $\xi > 0$  (y  $x^+ = \infty$ ) la cola de la distribución decae como una función de potencias (colas pesadas) y hablamos de familia tipo Fréchet.
- Cuando  $\xi < 0$  (y  $x^+ < \infty$ ) la cola está acotada y hablamos de tipo Weibull.
- Por último, si el parámetro de forma es 0 (y  $x^+ = \infty$ ),  $G(x)$  toma otra expresión que define la familia de distribuciones tipo Gumbel, en este caso la cola de la distribución decae exponencialmente dando lugar a colas ligeras.

Según la naturaleza de la variable de estudio la GEV derivará a una de estas familias. En el caso de oleaje, es común encontrar distribuciones de tipo Gumbel en aguas profundas, mientras que las limitaciones físicas del fondo en aguas someras dan lugar a distribuciones tipo Weibull.

La línea argumental expuesta está basada en el modelado extremal de datos agrupados en secuencias de observaciones de longitud  $n$  ( $M_n = \max\{X_1, \dots, X_n\}$ ), de forma que la muestra de extremos a la que se podrá ajustar la distribución GEV será  $\{M_{n,1}, \dots, M_{n,m}\}$ , obtenida mediante el método de máximos en un bloque de tiempo. Uno de los métodos más usuales es el método del máximo anual (MMA), en el que  $n$  es el número de observaciones en un año y  $m$  el número total de años.

- *Modelo de Pareto basado en el método POT*

El modelo GEV presentado en la sección anterior está basado en el método de máximos en un bloque de tiempo para seleccionar la muestra de datos a modelar. Este método presenta ciertos inconvenientes cuando la serie de datos no es muy extensa. Una alternativa a este método es el método POT que proporciona las excedencias sobre un umbral dado y en el que está basado el modelo Pareto que permite modelar las intensidades de la excedencia.

Sea  $X_1, X_2, \dots$  una secuencia de *iid* variables aleatorias con función de distribución marginal  $F$ . Con el método POT se consideran eventos extremos aquellos  $X_i$  que exceden un umbral  $u$ . Sea la variable aleatoria  $X$ , cuya función de distribución es  $F(x) = \Pr(X_i \leq x)$ . Las excedencias sobre  $u$  serán  $Y = X - u$ , condicionado a que  $X > u$ . La probabilidad condicionada permite describir el comportamiento estocástico de los eventos extremos definiendo la función de distribución de las excedencias sobre un umbral  $u$  de la forma:

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0 \tag{1.6}$$

Si la distribución  $F$  fuera conocida, la distribución de las excedencias sobre  $u$  también sería conocida, pero al igual que para el caso de máximos en un bloque de tiempo, en las aplicaciones reales esto no es así.

Sea  $M_n = \max\{X_1, \dots, X_n\}$  y asumiendo la normalización lineal de  $M_n$ , para valores grandes de  $n$ ,

$$\Pr\{M_n \leq x\} \approx G(x) \tag{1.7}$$

donde  $G(x)$  tiene la expresión (1.4) para  $\mu, \psi > 0$ . Por lo tanto, para un umbral  $u$  suficientemente grande, la función de distribución de  $Y = X - u$  condicionada a  $X > u$  se puede aproximar por la función:

$$H(y; \sigma, \xi) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{si } \xi = 0 \end{cases} \tag{1.8}$$

definido en  $y > 0$  y  $(1 + \xi y / \sigma) > 0$ , donde

$$\sigma = \psi + \xi(u - \mu) \tag{1.9}$$

La familia de distribuciones definida con la ecuación (1.8) se denomina distribución generalizada de Pareto (*Generalized Pareto Distribution, GPD*). Esto implica que, si los máximos en un bloque de tiempo tienen una distribución aproximada  $G$ , entonces, las excedencias sobre un umbral tienen su correspondiente distribución aproximada en la familia generalizada de Pareto. Los parámetros de la distribución generalizada de Pareto de excedencias sobre un umbral están determinados por los parámetros de la distribución GEV asociada de máximos en un bloque de tiempo. El parámetro de forma  $\xi$  es el mismo para ambas distribuciones (no depende del valor del umbral ni del tamaño de bloque de tiempo) y, al igual que para la GEV, este parámetro define la cola de la distribución, dando lugar a tres posibles familias.

- Cuando  $\xi > 0$  la distribución no está acotada ( $0 < x < \infty$ ) y estamos ante el caso de Pareto equivalente a Fréchet en la GEV.
  - En el caso de  $\xi < 0$  la distribución presenta un límite superior en  $u - \sigma/\xi$  y equivale a la familia de Weibull.
  - Por último si  $\xi = 0$  la distribución tampoco está acotada y tomando el límite cuando  $\xi \rightarrow 0$  la expresión de la distribución corresponde a una exponencial de parámetro  $1/\sigma$  (expresión inferior (3.8)). Este caso se corresponde con la familia Gumbel en la GEV.
- 
- *Modelo combinado Pareto-Poisson*

La distribución de Pareto modela las intensidades de las excedencias sobre el umbral. Sin embargo, el estudio de la frecuencia con que suceden los eventos extremos es igualmente interesante y complementario del modelado estadístico de extremos.

La muestra de eventos extremos que ocurren en instantes aleatorios en el tiempo se puede considerar un proceso de Poisson y su estudio puede llevarse a cabo mediante la variable aleatoria de Poisson. Para que un proceso pueda considerarse de Poisson debe satisfacer una serie de hipótesis:

1. La probabilidad de que ocurra un evento en un intervalo de tiempo de longitud pequeña es proporcional a la longitud del intervalo.
2. La probabilidad de que ocurra más de un evento en un intervalo suficientemente pequeño es despreciable (no se producirán sucesos simultáneos).
3. El número de eventos que ocurren en dos intervalos disjuntos son variables aleatorias independientes.

La función de distribución de Poisson se define como:

$$F(x; \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \tag{1.10}$$

para  $x=0,1,\dots,\lambda>0$ , donde  $\lambda$  es un parámetro adimensional que equivale al número esperado de ocurrencias durante un intervalo de tiempo  $(0, T]$  y se obtiene de la forma:

$$\lambda = \int_0^T \nu(t) dt \quad (1.11)$$

donde  $\nu(t)$  es la tasa de ocurrencia de eventos. Si  $\nu(t)$  es constante a lo largo del tiempo se dice que el proceso de Poisson es homogéneo y su parámetro viene dado por  $\lambda = \nu T$ .

La aleatoriedad e *iid* de las variables de estudio permite combinar el modelo de Pareto para intensidades y el de Poisson para frecuencias, obteniéndose el modelo Pareto-Poisson. Para ello, según las propiedades del límite de la distribución de Poisson, se demuestra que el número  $N$ , de excedencias sobre un umbral  $u$  para cualquier año sigue una distribución de Poisson con media  $\lambda$ . Además, condicionado a que  $N \geq 1$ , los valores de excedencia  $Y_1, \dots, Y_N$  serán *iid* en la distribución de Pareto.

La probabilidad de que el máximo anual sea menor que  $x$ , para  $x > u$  viene dada por la función de distribución del modelo Pareto-Poisson:

$$G(x; \sigma, \xi, \lambda) = \exp \left\{ -\lambda \left( 1 + \xi \frac{x-u}{\sigma} \right)^{-1/\xi} \right\} \quad (1.12)$$

Los parámetros de la distribución de Pareto-Poisson están unívocamente relacionados con los de la GEV mediante las expresiones:

$$\begin{aligned} \sigma &= \psi + \xi(u - \mu) \\ \lambda &= \left( 1 + \frac{\xi(u - \mu)}{\psi} \right)^{-1/\xi} \end{aligned} \quad (1.13)$$

Como se puede ver, ambos parámetros varían en función del umbral  $u$ , aspecto que se debe tener en cuenta a la hora de elegir este modelo, junto con la premisa de que las excedencias son *iid*.

### Método de ajuste. Diagnóstico. Obtención de cuantiles y bandas de confianza

- *Método de ajuste*

Una vez definidos los posibles modelos estadísticos con los que caracterizar la muestra de extremos, el objetivo será estimar, a partir de los valores de la muestra, los valores de los parámetros del modelo. De esta forma se podrá definir la función de distribución del modelo e inferir nuevos valores asociados a distinta probabilidad. Los valores propuestos para cada parámetro se denominarán estimadores puntuales y se representarán con un acento circunflejo sobre la letra que representa al parámetro, por ejemplo  $\hat{\theta}$ . De esta manera se está indicando que esos valores no son los valores reales de los parámetros, sino los estimados y por lo tanto tienen una función de probabilidad de ocurrencia.

Hay distintas técnicas para obtener los estimadores del modelo, siendo los más comunes en estadística de extremos el método de los papeles probabilísticos, el método de los momentos y el método de máxima verosimilitud. El método utilizado en este estudio es el de máxima verosimilitud (*Maximum Likelihood Estimation*, MLE) puesto que, analizados los pros y contras de cada método, se concluye que el MLE proporciona estimadores menos sesgados, más eficientes y con buena bondad de ajuste (Lettenmaier y Burges, 1982). Además cabe destacar como propiedades de estos estimadores que son consistentes, asintóticamente normales, asintóticamente eficientes e invariantes.

Dada la variable aleatoria  $X$  con función de densidad  $f(x;\theta)$  y función de distribución  $F(x;\theta)$  definidas por el vector de parámetros  $\theta = (\theta_1, \dots, \theta_p)$ , la función que proporciona la probabilidad de los valores observados es la función de verosimilitud  $L(x;\theta)$  y tendrá la forma:

$$L(x; \theta) = f(x_1, \dots, x_N; \theta) = \prod_{i=1}^N f(x_i; \theta) \quad (1.14)$$

El objetivo del método MLE es encontrar el modelo, de entre todos los considerados, que asigne la mayor probabilidad a los valores de la muestra. Esto se traduce en encontrar los estimadores puntuales de los parámetros de la función  $L$  que hagan máxima la probabilidad de observar los datos de la muestra  $\{x_1, \dots, x_n\}$ .

La función de verosimilitud  $L(x; \theta)$  es la función de densidad conjunta asociada a todos los extremos de la muestra, a partir de una familia de distribuciones elegida previamente. Puesto que los cálculos matemáticos son complejos, se utiliza el artificio matemático de tomar logaritmos neperianos a la ecuación (1.14) y trabajar con la función logarítmica de verosimilitud  $l(x; \theta)$ , tal que  $l(x; \theta) = \log L(x; \theta)$ . De esta forma la función a maximizar será:

$$l(x; \theta) = \sum_{i=1}^N \log L(x_i; \theta) \quad (1.15)$$

Los estimadores máximo-verosímiles  $\hat{\theta}$  son los valores de los parámetros que maximizan la función y se obtienen de la siguiente forma:

$$\frac{\partial l(x; \theta)}{\partial \theta_j} = 0 \quad (1.16)$$

para  $j=1, \dots, p$  parámetros.

A continuación se muestra la expresión que toma la función logarítmica de máxima verosimilitud para la familia de distribuciones GEV (1.17) y Pareto-Poisson (1.18).

Sean  $\{x_1, \dots, x_N\}$  las  $N$  observaciones de máximos en un bloque de tiempo, la función logaritmo de verosimilitud para la familia de distribuciones GEV se define como:

$$l(x; \mu, \psi, \xi) = -N \log(\psi) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^N \log\left(1 + \xi \frac{x_i - \mu}{\psi}\right) - \sum_{i=1}^N \left(1 + \xi \frac{x_i - \mu}{\psi}\right)^{-1/\xi} \quad (1.17)$$

Sean  $\{y_1, \dots, y_N\}$  los  $N$  valores extremos que exceden un umbral  $u$  y caracterizan cada pico de muestra observada. La función logarítmica de verosimilitud para la familia de distribuciones Pareto-Poisson se define:

$$l(y; \lambda, \sigma, \xi) = N \log \lambda - \lambda T - N \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^N \log\left(1 + \xi \frac{y_i}{\sigma}\right) \quad (1.18)$$

donde  $T$  es el intervalo de tiempo (años) del período de estudio. Los dos primeros términos de la función representan el proceso de Poisson, con media  $\lambda T$  y valor máximo para  $\hat{\lambda}$  en  $N/T$ . El resto de términos hacen referencia al modelado de las intensidades mediante los parámetros de Pareto,  $\sigma$  y  $\xi$ .

- *Diagnóstico*

Dado que el ajuste de un modelo probabilístico se hace con el objetivo de obtener más información de la población de datos estudiada, será necesario comprobar la precisión del modelo elegido.

Para comprobar que el modelo se ajusta bien a los datos de la muestra comúnmente se usan dos técnicas gráficas, la técnica de papeles probabilísticos (PP *plot*) y la técnica de cuantiles (QQ *plot*).

Dada la muestra de datos  $x_1, \dots, x_n$  con función de distribución desconocida  $F$  se quiere saber si el modelo obtenido por máxima verosimilitud con función de distribución  $\hat{F}$  ajusta bien a los datos de la muestra.

La función de distribución empírica  $\tilde{F}$  de la población de datos con función de distribución  $F$  se puede obtener a partir de los datos ordenados de la muestra

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Para cada  $x_{(i)}$ ,  $i$  de las  $n$  observaciones tienen un valor menor o igual que  $x_{(i)}$  por lo que la función de distribución empírica se define:

$$\tilde{F}(x) = \frac{i}{n+1} \quad (1.19)$$

para  $x_{(i)} \leq x \leq x_{(i+1)}$ .

Puesto que  $\tilde{F}$  es un estimador de la función de distribución  $F$ , y  $\hat{F}$  también lo es (pero proviene de un modelo de ajuste),  $\tilde{F}$  y  $\hat{F}$  deberían ser similares. Así, la técnica de papeles probabilísticas consiste en representar los puntos correspondientes a

$$\left\{ \hat{F}(x_{(i)}), \frac{i}{n+1} \right\} \quad (1.20)$$

para  $i = 1, \dots, n$ .

Si  $\hat{F}$  corresponde a un modelo que ajusta adecuadamente la población de datos estudiada, los puntos del gráfico PP deberán alinearse cerca de la diagonal. Si los puntos se separan considerablemente, esto evidenciará que el modelo no es bueno para los datos estudiados.

La técnica de cuantiles consiste en representar con puntos los valores de la variable correspondientes a distintas probabilidades:

$$\left\{ \hat{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)} \right\} \quad (1.21)$$

para  $i = 1, \dots, n$ .

Si  $\hat{F}$  es un buen estimador de  $F$ , el *QQ plot* consistirá en puntos cercanos a la diagonal. Ambas técnicas de diagnóstico contienen la misma información pero a diferentes escalas, de forma que se complementan y permiten llevar a cabo un diagnóstico más completo.

- *Obtención de cuantiles y bandas de confianza*

Como se dijo anteriormente los eventos extremos de oleaje son cruciales en diferentes procesos naturales y su conocimiento será necesario para poder diseñar y gestionar las actuaciones humanas en el medio marino. Por tanto, el modelado estadístico de los valores extremos de oleaje se hace con el objetivo de poder inferir valores extremos de gran magnitud, asociados a bajas probabilidades de presentación, y por lo tanto, muchas veces no presentes en la muestra de datos.

Para hacer referencia a este tipo de valores se introduce el concepto de período de retorno,  $R$ , que es el período de tiempo (expresado normalmente en años) en que, por término medio, se iguala o supera un suceso de una determinada magnitud. Este concepto es fundamental en el diseño ingenieril.

Dada una variable aleatoria  $X$  con función de distribución  $F(x)$ , la relación entre período de retorno y probabilidad de ocurrencia viene definida por la expresión:

$$R = \frac{1}{1 - F(x)} \tag{1.22}$$

Para el cálculo de valores asociados a períodos de retorno pequeños, es mejor utilizar la expresión:

$$R = -\frac{1}{\log(F(x))} \tag{1.23}$$

Esta expresión es equivalente a la anterior, aplicando en ella la aproximación matemática de límites (Beran y Nozdryn-Plotnicki, 1977). Para el caso de períodos de retorno elevados ambas expresiones conducen al mismo valor de la variable.

El cálculo del cuantil asociado a un período de retorno para el modelo basado en la familia de distribuciones GEV se obtendrá a partir de la ecuación explícita de la variable  $X$  en la función de distribución (1.4). Con la expresión (1.23) del valor del período de retorno el valor del cuantil se obtiene a partir de:

$$X_R = \begin{cases} \mu - \frac{\psi}{\xi} \left[ 1 - \left\{ -\log \left( 1 - \frac{1}{R} \right) \right\}^{-\xi} \right] & \text{para } \xi \neq 0 \\ \mu - \psi \log \left\{ -\log \left( 1 - \frac{1}{R} \right) \right\} & \text{para } \xi = 0 \end{cases} \quad (1.24)$$

Para el caso de los modelos basados en la familia de distribuciones Pareto-Poisson, hay que tener en cuenta que es necesario escalar el período de retorno a la unidad anual. Se considera el factor de escala como el número medio de excedencias sobre un umbral por año:

$$\lambda = \frac{n_u}{N} \quad (1.25)$$

donde  $n_u$  es el número total de valores extremos obtenidos a partir del método POT y  $N$  es el número de años de la muestra observada. La obtención del cuantil asociado a un período de retorno  $R$  viene dada por la expresión:

$$X_R = \begin{cases} u - \frac{\sigma}{\xi} \left[ 1 - (\lambda R)^\xi \right] & \text{para } \xi \neq 0 \\ u + \sigma \log(\lambda R) & \text{para } \xi = 0 \end{cases} \quad (1.26)$$

Los valores obtenidos de los cuantiles estarán sujetos a la precisión del modelo, ya que no hay que perder de vista, que estos resultados, tienden a un límite asintótico cuando la muestra de datos tiende a infinito. Por esta razón, junto con el valor del cuantil, es necesario informar de su probabilidad de ocurrencia, lo que viene representado por las bandas de confianza.

Para calcular las bandas de confianza es necesario conocer los errores estándar de los estimadores. El cálculo de los errores estándar de los estimadores se llevará a cabo mediante el método de Fisher y el de las bandas de confianza mediante el método Delta (Rice, 1994). Ambos métodos están explicados en Coles (2001) y Menéndez (2008).

## Modelos no estacionarios de extremos

Como se ha descrito anteriormente, el clima marítimo, desde el punto de vista estadístico, se puede entender en términos de media y su variabilidad a lo largo de un período de tiempo. Esta variabilidad tiene diversas escalas, que van desde los cambios estacionales de verano-invierno a ciclos más amplios de tipo interanual o decadal. Esta variabilidad en el clima marítimo viene propiciada por la variabilidad del clima atmosférico y los efectos planetarios que en él ocurren, considerándose los 30 años el período en el que se pueden establecer unas características estadísticas del clima.

La variabilidad del clima marítimo, tanto de régimen medio como de régimen extremal, es determinante en los procesos de sedimentación y erosión en playas, comunidades biológicas, pesquerías... Así, las playas tienen una variación de su forma en planta y perfil a lo largo del año, produciéndose también, pérdida o depósitos de arena en ciclos más elevados de tiempo (del orden de años). En cuanto a las actuaciones humanas en la costa, el conocimiento de la variabilidad del clima marítimo permitirá gestionar mejor las fases de construcción de obras marítimas, la operatividad en puertos, la explotación y obtención de energía del oleaje, la operatividad en estructuras offshore, el diseño de competiciones deportivas de tipo náutico, el diseño de rutas marítimas...

Es por estas razones por lo que se presenta como algo fundamental el estudio de la variabilidad del clima marítimo y más en concreto de los eventos extremos. Sin embargo, los modelos estacionarios presentados en el apartado 1.2.2 no parecen apropiados para este análisis, puesto que están basados en las hipótesis de partida 2 y 3 de la TVE (las observaciones son *iid* y la variable de estudio es estacionaria). Para solventar el problema de la no estacionariedad en extremos se propone utilizar una extensión de la TVE que se basa en considerar la función de distribución de extremos variable a lo largo del tiempo y sus estadísticos dependientes del tiempo. Hay diversas aplicaciones de este método, como por ejemplo la propuesta por Smith (1986), quien estudió el régimen extremal del nivel del mar en Venecia introduciendo una tendencia de largo plazo y la posible influencia del ciclo nodal. Casson y Coles (2000) estudian los huracanes mediante un modelo de intensidad de extremos sobre centros de presión, introduciendo en los parámetros de la GEV covariables que proporcionan información sobre la vida del huracán. Katz et al. (2002) estudia variaciones estacionales de

precipitación y caudales introduciendo funciones sinusoidales en los parámetros de la GEV. Cairns et al. (2006) modela las variaciones de oleaje mediante anomalías de presión y viento, introduciendo su influencia mediante covariables en un modelo de Pareto-Poisson. Méndez et al. (2006) estudia la tendencia de largo plazo en los eventos extremos de oleaje de una serie de boyas del Pacífico mediante un modelo POT dependiente del tiempo y por último, Menéndez et al. (2009a) introduce funciones armónicas para estudiar la estacionalidad de los eventos extremos de oleaje a lo largo de la costa española.

- *Modelos paramétricos*

La forma de estudiar la variabilidad del clima marítimo extremal se va a llevar a cabo mediante el modelado estadístico que incluya la variación temporal, continua y suave de los parámetros. De esta forma, la función de distribución de extremos varía a lo largo del tiempo.

El estudio de las diferentes variaciones temporales se llevará a cabo mediante una aproximación paramétrica. Este método consiste en proponer a priori una familia de modelos en los cuales se encuentre la distribución real de los datos. En ellos se simulan las variaciones temporales mediante parámetros en expresiones matemáticas que se resuelven por regresión. La estimación de los parámetros permite cuantificar las variaciones y su afcción a los estadísticos de la distribución. Las variaciones temporales consideradas podrán afectar al oleaje de forma independiente o como combinación de ellas. Por lo tanto, queda proponer un modo de agregación de las diferentes variaciones temporales, que será aditivo o multiplicativo.

A continuación, se muestran las expresiones paramétricas más utilizadas en el estudio de las variaciones estacionales, interanuales y la tendencia de largo plazo.

- Variación estacional

La variación estacional del oleaje es la variación producida a lo largo del año, caracterizada principalmente por el ciclo anual, definido por la variación invierno-verano (estación seca-humedad, según la latitud). Según el lugar de estudio puede

presentarse también un ciclo semianual caracterizado por un patrón climático sistemático a largo del año, como ocurre por ejemplo en el Mediterráneo con la llamada “gota fría”.

Menéndez et al. (2009) utiliza funciones armónicas para la parametrización de la variación estacional del oleaje. De esta forma, los parámetros en los que se vaya a incluir esta variación tendrán la expresión:

$$\theta_i(t) = \beta_0 + \sum_{i=1}^{P_\theta} [\beta_{2i-1} \cos(2i\pi t) + \beta_{2i} \sin(2i\pi t)] \quad (1.27)$$

Para  $i=1$  se representa el ciclo anual mediante los parámetros  $\beta_1$  y  $\beta_2$ , para  $i=2$  el ciclo semianual con los parámetros de regresión  $\beta_3$  y  $\beta_4$  y así sucesivamente.  $P_\theta$  representa el número de armónicos a considerar.

- Variación interanual

La observación de una serie temporal de altura de ola permite ver, aparte de la variación estacional descrita en el apartado anterior, una variación interanual, con períodos de tiempo (años) que registran mayor altura de ola, frente a otros en los que esta altura de ola es menor. La Figura 1.5 muestra una serie de altura de ola significativa representativa de la costa de La Coruña, donde se han representado con círculos los máximos mensuales. Aquí se puede ver como hay años con oleaje más severo y otros más calmados.

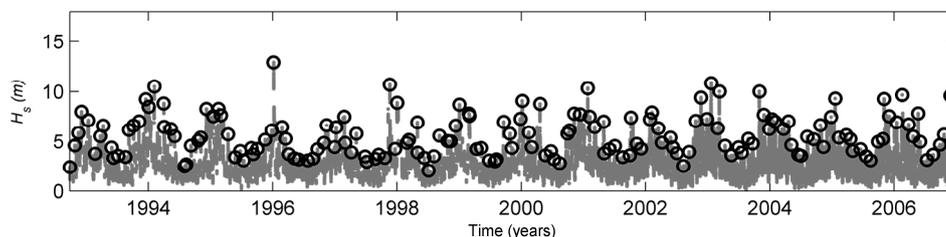


Figura 1.5 Serie temporal de oleaje procedente de satélite

Numerosos autores han explicado este tipo de fluctuaciones mediante índices climáticos. Los índices climáticos tratan de sintetizar un estado de la atmósfera frente a

su estado medio (es decir, muestran una anomalía). En concreto aquellos que pueden afectar al estado del océano normalmente están relacionados con la presión a nivel del mar, como la Oscilación del Atlántico Norte (NAO), o con la temperatura superficial del agua, como es el caso del índice del NIÑO 3. Por lo tanto, aparte de los índices convencionales, la variación interanual también se puede intentar explicar mediante un análisis detallado de la situación atmosférica.

Así, Woolf y Challenor (2002) explican la variabilidad del clima marítimo del Atlántico Norte mediante la NAO, Lionello y Sanna (2005) relacionan la variabilidad del clima marítimo en el Mediterráneo con la NAO y el monzón y Hemer et al. (2010) correlaciona la variabilidad interanual del régimen medio de altura de ola en Australia con diferentes índices influyentes en esa región, como por ejemplo el SOI (*Southern Oscillation Index*). Por otro lado, Sasaki y Hibiya (2007) explican la variabilidad interanual de la altura de ola del 90% en el noroeste del Pacífico mediante los campos de vientos a 10 m de altura, los campos de presiones a nivel del mar y la temperatura superficial del agua.

Por lo tanto, la variabilidad interanual de los valores extremos se considera que está relacionada con la variabilidad de otra variable a la que se denomina covariable. La forma de incorporar esta información en los parámetros del modelo estadístico será mediante un término lineal de la forma:

$$\theta_i = \beta_0 \cdot \text{covariable}(t) \tag{1.28}$$

- Tendencias de largo plazo

En lo que se refiere a la escala de tiempo de mayor período, se plantea estudiar la posible tendencia de largo plazo que muestre la población de datos. Las conclusiones obtenidas para esta escala temporal deberán ser utilizadas con precaución, puesto que si la muestra de datos a estudiar no es suficientemente grande, se pueden encontrar tendencias que en realidad pertenecen a ciclos de tiempo más cortos, por ejemplo decadales. No obstante, el estudio de la tendencia de largo plazo en series suficientemente largas permite conocer la evolución de los extremos del oleaje y

estudiar si su aumento o disminución (en intensidad y frecuencia) es debido a causas naturales o al efecto antropogénico del hombre y su consecuencia en el cambio climático.

Gulev y Grigorieva (2006) estudian la tendencia de largo plazo de la altura de ola significativa de viento (*sea*) y de fondo (*swell*) en invierno, a partir de datos visuales de barcos en ruta (el estudio se centra en el Pacífico Norte y Atlántico Norte). Menéndez et al. (2008) estudian la tendencia de largo plazo de los extremos de oleaje en varias boyas de la costa Pacífica estadounidense y el estudio de Hemer et al. (2010) se centra en el análisis de la tendencia de largo plazo mensual y estacional de la altura de ola significativa en el Hemisferio Sur.

La expresión más utilizada para su parametrización es una función exponencial:

$$\theta_i(t) = \beta_0 e^{\beta_T t} \tag{1.29}$$

Para variaciones pequeñas de  $\beta_T$  ( $\beta_T \rightarrow 0$ ) la expresión exponencial se puede asemejar a una relación lineal de la forma:

$$\beta_0 e^{\beta_T t} \approx \beta_0 (1 + \beta_T t) \tag{1.30}$$

De esta forma el producto  $100 \times \beta_T$  puede aproximarse al porcentaje de variación. Dado que cabe esperar variaciones pequeñas en el estudio de los extremos de oleaje, puesto que los registros no son muy largos, se puede utilizar esta aproximación.

- *Modelo GEV no estacionario*

En este apartado se van a redefinir la función de distribución y densidad para el modelo GEV, así como las expresiones que se vean afectadas para el ajuste, diagnóstico y cálculo de cuantiles.

La función de distribución del modelo GEV no estacionario es equivalente a la expresión (1.4) introduciendo la dependencia temporal de los parámetros, es decir,  $\mu(t), \psi(t), \xi(t)$ . Lo mismo ocurre con la función de densidad y la expresión (1.5).

Sean  $\{x_1, \dots, x_n\}$  las  $N$  observaciones de máximos en un bloque de tiempo ocurridas en los instantes  $\{t_1, \dots, t_n\}$ , la función logarítmica de verosimilitud se expresará como:

$$l(x; \theta|t_i) = \sum_{i=1}^N \log g(x_i; \mu(t_i), \psi(t_i), \xi(t_i)) \quad (1.31)$$

siendo  $g(x; \mu, \psi, \xi)$  la función de densidad de la GEV, ecuación (3.5). De esta forma la función logarítmica de verosimilitud será:

$$l(x; \theta|t_i) = - \sum_{i=1}^N \left\{ \log \psi(t_i) + \left(1 + \frac{1}{\xi(t_i)}\right) \log \left[ 1 + \xi(t_i) \left( \frac{x_i - \mu(t_i)}{\psi(t_i)} \right) \right] + \left[ 1 + \xi(t_i) \left( \frac{x_i - \mu(t_i)}{\psi(t_i)} \right) \right]_+^{-1/\xi(t_i)} \right\} \quad (1.32)$$

siendo  $[a]_+$  valores reales positivos. Para el caso de Gumbel ( $\xi=0$ ) la función logarítmica de verosimilitud se simplifica a:

$$l(x; \theta|t_i) = \sum_{i=1}^N \left\{ -\log \psi(t_i) - \frac{x_i - \mu(t_i)}{\psi(t_i)} - \exp \left[ -\frac{x_i - \mu(t_i)}{\psi(t_i)} \right] \right\} \quad (1.33)$$

siendo en ambos casos para todo  $t_i$ ,  $\psi(t_i) > 0$ .

El diagnóstico gráfico presentado en el apartado 1.2.3 es aplicable cuando las variables son *iid*, sin embargo, para el caso de no estacionariedad este tipo de diagnóstico también es aplicable realizando ciertas modificaciones que se basan en la estandarización de los datos.

Una estandarización apropiada para los modelos GEV no estacionarios es transformar la distribución del modelo estadístico en la distribución Gumbel. Esto permite representar el proceso linealizado, de forma que resulte más cómodo describir

cualitativamente la bondad del ajuste. Al estadístico resultante del cambio de variable se le denota como  $W$  y tiene la expresión:

$$W(t) = \frac{1}{\hat{\xi}(t)} \log \left[ 1 + \hat{\xi}(t) \left( \frac{x - \hat{\mu}(t)}{\hat{\psi}(t)} \right) \right] \quad (1.34)$$

donde se representa cada una de las  $N$  observaciones de máximos  $\{x_1, \dots, x_n\}$  en un bloque de tiempo ocurridos en los instantes  $\{t_1, \dots, t_n\}$ .

La probabilidad asociada a la distribución de Gumbel a partir del estadístico  $W(t)$  se obtiene mediante la relación:

$$\Pr(W \leq x) = \exp(-\exp^{-x}) \quad (1.35)$$

El diagnóstico mediante papeles probabilísticos consistirá en representar mediante puntos la probabilidad asociada a los valores de la muestra frente a la estimada mediante el estadístico  $W$ :

$$\left\{ \tilde{F}(x_{(i)}), \exp[-\exp(-W_{(i)})] \right\} \quad (1.36)$$

El diagnóstico mediante el método de cuantiles se obtendrá representando los valores ordenados de la muestra (con su transformación logarítmica) frente a los cuantiles obtenidos con la distribución del modelo de extremos y escalados a la distribución Gumbel:

$$\left\{ -\log(-\log(x_{(i)})), W_{(i)} \right\} \quad (1.37)$$

Por último, la no estacionariedad asumida en el modelo de extremos influye también en el cálculo de cuantiles asociados a un período de retorno, pudiéndose obtener valores instantáneos del cuantil. Además, esta metodología permite calcular cuantiles

agregados a períodos de tiempo que cubren un rango superior a la variabilidad producida.

El cálculo del cuantil asociado a un período de retorno  $R$  se obtendrá a partir de la ecuación explícita de la variable  $X$  en la función de distribución no estacionaria de la GEV:

$$X_R(\mu(t), \psi(t), \xi(t)) = \begin{cases} \mu(t) - \frac{\psi(t)}{\xi(t)} \left[ 1 - \left\{ -\log \left( 1 - \frac{1}{R} \right) \right\}^{-\xi} \right] & \text{para } \xi(t) \neq 0 \\ \mu(t) - \psi(t) \log \left\{ -\log \left( 1 - \frac{1}{R} \right) \right\} & \text{para } \xi(t) = 0 \end{cases} \quad (1.38)$$

La obtención del cuantil agregado a un período de tiempo elevado conlleva mayor complejidad. En particular, el cuantil anual  $\bar{X}_R[t_1, t_2]$ , asociado a un período de retorno  $R$  y para un intervalo  $[t_1, t_2]$  se puede obtener a partir de la ecuación implícita:

$$1 - \frac{1}{R} = \exp \left\{ -k_m \int_{t_1}^{t_2} \left[ 1 + \xi(t) \left( \frac{\bar{X}_R[t_1, t_2] - \mu(t)}{\psi(t)} \right) \right]^{-1/\xi(t)} dt \right\} \quad (1.39)$$

donde  $1/k_m$  es la longitud del bloque de tiempo seleccionado para definir los máximos (por ejemplo, si se utilizan máximos mensuales será  $1/k_m = 1/12$  años). Además, si  $t_1 = 0$  y  $t_2 = 1$  se obtendrá el cuantil del máximo anual cuando la escala temporal de trabajo está en años. Si por ejemplo,  $t_1 = \frac{i-1}{12}$  y  $t_2 = \frac{i}{12}$ , se obtendrá el cuantil del máximo anual correspondiente al mes  $i$ -ésimo.

Para el método de máximos anuales y escala temporal anual, la ecuación (3.39) puede simplificarse de la forma:

$$1 - \frac{1}{R} = \exp \left\{ - \int_{t_1}^{t_2} \left[ 1 + \xi^*(t) \left( \frac{\bar{X}_R[t_1, t_2] - \mu^*(t)}{\psi^*(t)} \right) \right]^{-1/\xi^*(t)} dt \right\} \quad (1.40)$$

donde  $\mu^*$ ,  $\psi^*$  y  $\xi^*$  son los parámetros específicos para máximos anuales. Cuando se trabaja con bloques temporales de una escala menor al año los parámetros de la GEV pueden reescalarsse a los parámetros de máximos anuales:

$$\begin{aligned} \mu^*(t) &= \mu(t) + \frac{\psi(t)}{\xi(t)} (k_m^{\xi(t)} - 1) \\ \psi^*(t) &= \psi(t) k_m^{\xi(t)} \\ \xi^*(t) &= \xi(t) \end{aligned} \quad (1.41)$$

facilitando así la comparación entre modelos que analizan máximos de bloques temporales diferentes.

- *Modelo de Pareto no estacionario*

La muestra de máximos que resulta de aplicar el método POT a la población de datos, tiende a la función no estacionaria de Pareto, con función de distribución  $H(y; \sigma(t), \xi(t))$ , resultado de introducir en la función (1.8) la dependencia temporal de los parámetros.

Dados los  $N$  valores extremos  $\{y_1, \dots, y_n\}$  observados en los instantes  $\{t_1, \dots, t_n\}$  durante el intervalo de tiempo  $T$ , que exceden un determinado umbral y caracterizan cada pico de la muestra observada, la función logarítmica de verosimilitud se obtiene a partir de:

$$l(y; \theta | t_i) = \sum_{i=1}^N \log(h(y_i; \sigma(t_i), \xi(t_i))) \quad (1.42)$$

donde  $h(y_i; \sigma(t_i), \xi(t_i))$  es la función de densidad de Pareto. Desarrollando la ecuación (1.42) se obtiene:

$$l(y; \theta | t_i) = \begin{cases} \sum_{i=1}^N \left\{ - \left( 1 + \frac{1}{\xi(t_i)} \right) \log \left( 1 + \xi(t_i) \frac{y_i}{\sigma(t_i)} \right) \right\} & \text{para } \xi(t) \neq 0 \\ \sum_{i=1}^N \left\{ - \log \sigma(t_i) - \left( \frac{y_i}{\sigma(t_i)} \right) \right\} & \text{para } \xi(t) = 0 \end{cases} \quad (1.43)$$

Una vez ajustado el modelo de Pareto no estacionario, el diagnóstico gráfico del modelo mediante *PP plot* y *QQ plot* requiere la transformación de la distribución de Pareto en una distribución estándar exponencial.

El estadístico  $W$  resulta:

$$W(t) = \frac{1}{\hat{\xi}(t)} \log \left[ 1 + \hat{\xi}(t) \frac{y_u - u}{\hat{\sigma}(t)} \right] \quad (1.44)$$

donde  $y_u$  representa cada uno de los  $N$  valores extremos que exceden un umbral  $u$ , observados en los instantes  $t_i$ .

Los pares de valores de probabilidad a representar en el papel probabilístico serán:

$$\left\{ \frac{i}{n+1}, 1 - \exp(-W_{(i)}) \right\} \quad (1.45)$$

Para obtener los gráficos de cuantiles se representarán los valores ordenados de la muestra frente a los cuantiles obtenidos con la distribución teórica. Los valores de la muestra tienen que ser afectados por un escalado:

$$\left\{ - \log \left( 1 - \frac{i}{N+1} \right), W_{(i)} \right\} \quad (1.46)$$

- *Modelo Pareto-Poisson no estacionario*

En el modelado de la frecuencia se asume un proceso de Poisson no homogéneo. Para un instante  $t_i$ , la probabilidad de que ocurra un evento extremo condicionado al instante anterior  $t_{i-1}$ , vendrá dada por la función de distribución exponencial que define un proceso de Poisson:

$$t_i \sim F(x; \lambda(t)) \tag{1.47}$$

donde  $\lambda(t)$  es el parámetro de Poisson definido por:

$$\lambda(t_i) = \int_{t_{i-1}}^{t_i} v(t) dt \tag{1.48}$$

donde  $v(t)$  es la tasa de ocurrencia de un proceso de Poisson no homogéneo, variable en el tiempo. El parámetro de Poisson es el número de sucesos extremos que ocurren en un determinado intervalo de tiempo, y será variable en función de la posición del intervalo de tiempo en la secuencia temporal.

Sean  $\{y_1, \dots, y_n\}$  los  $N$  valores extremos observados en los instantes  $\{t_1, \dots, t_n\}$  durante el intervalo de tiempo  $T$ , que exceden un determinado umbral y caracterizan cada pico de la muestra observada. La función logarítmica de verosimilitud se obtiene a partir de:

$$l(y; \theta | t_i) = \sum_{i=1}^N \log(p(N; \lambda(t_i)) h(y_i; \sigma(t_i), \xi(t_i))) \tag{1.49}$$

donde  $p(N; \lambda(t_i))$  define un proceso de Poisson no homogéneo y  $h(y_i; \sigma(t_i), \xi(t_i))$  es la función de densidad de Pareto. Desarrollando los términos de la ecuación se obtiene:

$$l(y; \theta | t_i) = \begin{cases} \sum_{i=1}^N \left\{ \eta_{t_i} \log \lambda(t_i) - \frac{\lambda(t_i)}{T_s} - \eta_{t_i} \log \sigma(t_i) - \eta_{t_i} \left( 1 + \frac{1}{\xi(t_i)} \right) \log \left( 1 + \xi(t_i) \frac{y_i}{\sigma(t_i)} \right) \right\} & \text{para } \xi(t) \neq 0 \\ \sum_{i=1}^N \left\{ \eta_{t_i} \log \lambda(t_i) - \frac{\lambda(t_i)}{T_s} - \eta_{t_i} \log \sigma(t_i) - \eta_{t_i} \left( \frac{y_i}{\sigma(t_i)} \right) \right\} & \text{para } \xi(t) = 0 \end{cases} \quad (1.50)$$

$\eta_{t_i}$  toma valor 1 en los instantes  $t_i$  donde haya ocurrido un evento extremal y 0 en el resto de situaciones.  $T_s$  es una constante para escalar la unidad temporal de trabajo (generalmente anual) y la resolución con que se definen los instantes en los que ocurren excedencias (en el caso de que los  $t_i$  definan los días en los que ocurrió evento,  $T_s = 365.25$ ).

En las condiciones de no estacionariedad se sigue manteniendo la relación entre los parámetros de la familia de distribuciones GEV, Poisson y Pareto sustituyendo la dependencia temporal en las ecuaciones (1.13).

Una vez ajustado el modelo de Pareto-Poisson no estacionario, el diagnóstico gráfico del modelo mediante *PP plot* y *QQ plot* se llevará a cabo analizando por un lado las intensidades y por otro las frecuencias.

Para el caso de intensidades se utilizará el estadístico  $W$  definido en la ecuación (1.44). Los pares de datos del gráfico de probabilidades vendrán dado por la expresión (1.45) y los del gráfico de cuantiles por la (1.46).

Para diagnosticar el modelado de las frecuencias mediante un proceso de Poisson se define un estadístico  $Z$ :

$$Z_k = \int_{t_{k-1}}^{t_k} \hat{v}(t) dt \quad (1.51)$$

que es la tasa acumulada entre eventos extremales consecutivos y  $\hat{v}(t)$  es la tasa de ocurrencia estimada mediante el modelo. Si el modelo fuera adecuado, los valores  $Z_1, \dots, Z_N$  deben ser una variable aleatoria distribuida uniformemente.

En el método de papeles probabilísticos se representan las probabilidades asociadas a los valores de la muestra frente a la probabilidad de ocurrencia del estadístico  $Z$ :

$$\left\{ \frac{i}{N+1}, 1 - e^{-Z_{(i)}} \right\} \quad (1.52)$$

La representación del gráfico de cuantiles se realiza mediante el punteado:

$$\left\{ -\log\left(1 - \frac{i}{N+1}\right), Z_{(i)} \right\} \quad (1.53)$$

La obtención de cuantiles asociados a períodos de retorno en modelos Pareto-Poisson no estacionarios pasa por resolver la ecuación explícita (1.26) introduciendo la variación temporal de los parámetros,  $\lambda(t)$ ,  $\sigma(t)$  y  $\xi(t)$ .

Al igual que el modelo GEV no estacionario, el modelo Pareto-Poisson no estacionario también permite estimar el cuantil ( $\bar{X}_R[t_1, t_2]$ ) asociado a un período de retorno  $R$  agregado a un período de tiempo  $[t_1, t_2]$ . Para ello hay que resolver la ecuación implícita:

$$1 - \frac{1}{R} = \exp \left\{ - \int_{t_1}^{t_2} \left[ \lambda(t) \left( 1 + \xi(t) \frac{\lambda(t)^{-\xi(t)} (\bar{X}_R[t_1, t_2] - u)}{\sigma(t)} \right) \right]_{-}^{-1/\xi(t)} dt \right\} \quad (1.54)$$

### 1.2.1.2 Técnicas de minería de datos y de clasificación

El oleaje depende directamente de las variaciones en la atmósfera. Por esta razón, en esta tesis, el estudio estadístico del oleaje se ha llevado a cabo a partir del análisis propio del oleaje y de la situación atmosférica a través de los campos de presión en superficie, que, de manera indirecta, indica la intensidad y la dirección del campo de vientos. Para poder extraer la información atmosférica adecuada, eliminando las dimensiones redundantes, se han utilizado dos técnicas de minería de datos y clasificación utilizadas frecuentemente en meteorología.

## Componentes Principales

El análisis por componentes principales (*Principal Components*, PCs), o también llamado en la literatura sobre meteorología funciones empíricas ortogonales (*Empirical orthogonal Functions*, EOFs) es una técnica utilizada para reducir la dimensionalidad de una muestra de datos que presenta redundancia de información, reteniendo la máxima variabilidad existente.

La compresión de la información se consigue mediante la proyección de los datos en un espacio de menor dimensión que el original conservando el máximo de la varianza de la muestra. La base del nuevo espacio se forma con aquellos vectores donde la muestra proyectada presenta mayor varianza.

Sea una muestra de  $N$  observaciones de un vector  $x_k$ :

$$x_k = (x_{k1}, \dots, x_{kn})^T \quad \text{para } k = 1, \dots, N \quad (1.55)$$

o bien expresado el conjunto total de observaciones en forma matricial:

$$\mathcal{X} = \begin{pmatrix} x_{11} & \cdots & x_{N1} \\ \vdots & & \vdots \\ x_{1n} & \cdots & x_{Nn} \end{pmatrix} \quad (1.56)$$

definido en un espacio  $n$ -dimensional con base canónica  $\{e_1, \dots, e_n\}$ . Se desea obtener un subespacio de dimensión  $d < n$ , definido por una nueva base  $\{f_1, \dots, f_d\}$ , donde los vectores  $f_j$  de esta nueva base son una combinación lineal de los vectores  $e_i$  de la base canónica. El criterio para obtener este subespacio es que la muestra proyectada tenga una varianza máxima. Esto es equivalente a buscar un conjunto de vectores  $f_i$  con una orientación tal que la suma de los cuadrados de las proyecciones de todas las observaciones sobre estos  $f_i$  sea máximo. De esta manera, los datos iniciales se pueden expresar de forma aproximada mediante la proyección en una nueva base:

$$\mathcal{X} = \sum_{i=1}^n e_i x_i \approx \bar{\mathcal{X}} = \sum_{i=1}^d f_i \bar{x}_i \quad (1.57)$$

Matemáticamente se trata de optimizar la matriz de varianzas-covarianzas de la muestra de datos. Los autovectores (EOFs) de esta matriz son los nuevos  $f_i$  y los correspondientes autovalores ( $\lambda_i$ ) indican la varianza explicada. Los coeficientes de cada elemento de la muestra en la nueva base se denominan componentes principales (PCs). Los patrones espaciales de los EOFs se pueden interpretar como los modos de variabilidad del campo de los datos, y como la nueva base es ortogonal, están intercorrelacionados.

Mediante este análisis se obtienen  $n$  vectores, cada uno de los cuales representa una varianza de los datos. Normalmente los primeros vectores ( $d$ ) explican una varianza suficiente como para poder prescindir del resto de vectores. Por tanto, el campo original se puede expresar como:

$$\mathcal{X} \approx \sum_{i=1}^d EOF_i(s) \cdot PC_i(t) \quad (1.58)$$

donde  $s$  es la dimensión de los vectores de entrada ( $s=1, \dots, n$ ) y  $t$  es la dimensión temporal ( $t=1, \dots, N$ , siendo  $N$  el número total de observaciones).

Este método se suele aplicar a los campos de presiones para obtener los fenómenos sinópticos más relevantes que explican la variabilidad climática. También se usa en la definición de índices climáticos, como por ejemplo la NAO o el SAM (*Southern Annular Mode*).

### **Redes neuronales auto-organizativas (SOM)**

Las redes auto-organizativas son una técnica de clasificación que detecta automáticamente patrones o clases en el conjunto de datos y los proyecta en un espacio bidimensional con organización espacial que permite una visualización muy intuitiva de los resultados. Por esta razón, esta técnica está especialmente indicada para trabajar

en espacios de alta dimensionalidad. Además tiene la propiedad de vecindad que mantiene unidos en el espacio de los datos aquellos centroides vecinos en la red bidimensional de proyección.

La técnica de las SOM es frecuentemente utilizada en meteorología, siendo una de las aplicaciones más relevantes la realizada por Hewitson y Crane (2002). En este estudio utilizaron las SOM para detectar patrones sinópticos de circulación atmosférica y relacionarlos con distintas variables características del clima local. También se ha utilizado esta técnica para detectar patrones de la temperatura superficial del agua y variabilidad de las corrientes y Camus (2009) lo utilizó para caracterizar el clima marítimo multidimensional.

En esta técnica se parte de una muestra de datos  $\mathcal{X} = \{x_1, \dots, x_N\}$  compuesta por  $N$  vectores  $n$ -dimensionales, donde  $N$  es el número total de datos a clasificar y  $n$  es la dimensión de los vectores de la base de entrada  $x_k = \{x_{1k}, \dots, x_{nk}\}$ . Se aplica la SOM para obtener un número de prototipos o centroides igual a  $M$ .

El algoritmo de aprendizaje tiene una inicialización aleatoria de los prototipos  $\{v_1^0, \dots, v_M^0\}$ . Los prototipos se van adaptando de forma iterativa mediante la minimización de una distancia global entre grupos  $d(j)$ , definida como la distancia entre los vectores de entrada  $x_i$  y el vector prototipo correspondiente  $v_j$  para cada grupo:

$$\sum_{j=1}^M d(j) = \sum_{j=1}^M \sum_{x_i \in j} \|x_i - v_j\|^2 \quad (1.59)$$

donde  $M$  es el número total de grupos y  $\| \cdot \|$  es la norma euclidiana.

El algoritmo de entrenamiento procede iterativamente en ciclos, en los cuales se recorren todos los datos de entrada. En cada ciclo de entrenamiento, se considera cada vector de entrada  $x_i$  y se calcula el prototipo “ganador”  $v_{w(i)}$  como aquel más próximo al vector de entrada:

$$\|v_{w(i)} - x_i\| = \min\{\|v_j - x_i\|, j = 1, \dots, M\} \quad (1.60)$$

donde  $1 \leq w(i) \leq M$  indica el índice del prototipo ganador.

El proceso de entrenamiento de este algoritmo de clasificación incluye un núcleo espacial de vecindad en la rejilla de proyección que produce, por un lado, que el centroide ganador se desplace hacia el vector de entrada, y por otro, que los centroides vecinos en la rejilla 2D sean modificados de la siguiente forma:

$$v_j = v_j + \alpha h[w(i), j] \cdot (x_i - v_j) \quad \text{para } j = 1, \dots, M \quad (1.61)$$

donde  $0 \leq \alpha \leq 1$  es la tasa de aprendizaje que controla la velocidad del proceso de aprendizaje. Valores pequeños de  $\alpha$  dan lugar a aprendizajes lentos y suaves, mientras que valores elevados suponen un aprendizaje rápido e inestable.

Por tanto, el núcleo de vecindad permite la proyección de la relación topológica en el espacio multidimensional en la rejilla 2D. Esto significa que patrones similares en el espacio multidimensional de los datos de entrada estarán próximos en la rejilla, es decir hay organización espacial. Los mapas auto-organizativos pueden tener una estructura rectangular o hexagonal, siendo 4 ó 6 el número de vecinos respectivamente. En cuanto a la forma de mapa, éste puede ser una rejilla, un cilindro o un toroide. Cada grupo  $j$  obtenido con esta clasificación está definido por dos vectores, uno en el espacio de los datos, denominado centroide  $v_j$  (de la misma dimensión que los datos de entrada) y otro en el espacio bidimensional de proyección que define su correspondiente posición en la rejilla  $(m_j, n_j)$ .

Otra cualidad de esta técnica es que la proyección en el espacio 2D permite transformar las distribuciones en el espacio real de los datos en funciones de densidad de probabilidad sobre la rejilla. La frecuencia de presentación o probabilidad de cada centroide se define como la proporción de datos de partida correspondientes a cada centroide respecto a la cantidad total de datos de partida.

Una descripción más amplia y detallada de esta técnica se puede encontrar en Kohonen (2000).

## 1.2.2 Variabilidad espacial y temporal del clima marítimo

### 1.2.2.1 Escala global y regional

El estudio de la variabilidad espacial del clima marítimo a escala global y regional ha estado condicionado a la disponibilidad de bases de datos con suficiente cobertura. La creciente importancia de bases de datos con cobertura global y fiable como son los datos de satélite (Panchang et al., 1999; Queffeulou, 2005) o los modelos numéricos (ERA-40, WASA) hace posible el estudio del clima marítimo a grandes escalas espaciales. Las bases de datos visuales, procedentes de medidas de barcos en ruta, aunque no tienen una cobertura homogénea a lo largo de todo el mundo, han permitido también estudiar la variabilidad espacial del oleaje, principalmente en el Pacífico y Atlántico Norte (Gulev y Grigorieva, 2006). Por último, el estudio del clima marítimo en determinadas regiones también ha sido posible gracias a la disponibilidad de boyas, pero estas zonas están normalmente próximas a la costa y se limitan, principalmente, al Hemisferio Norte.

Dependiendo de la longitud del registro de la base de datos utilizada, los estudios se han centrado en el análisis de distintas escalas temporales. Así por ejemplo, Chen et al. (2004) estudian la estructura global de los extremos de oleaje y viento a partir de datos del satélite TOPEX. Alves y Young (2004) estiman la altura de ola de período de retorno 100 años ( $H_{s_{100}}$ ) a partir de un modelo POT de extremos y datos procedentes de distintas misiones satelitales, mientras que Caires y Sterl (2005) estiman el cuantil de altura de ola y viento de 100 años de período de retorno también con un modelo POT pero a partir de los datos ERA-40. A escala más regional Ruggiero et al. (1997) estudia las olas extremas de la costa oeste americana a partir del registro de cuatro boyas.

En cuanto a la escala estacional, el estudio a nivel global es más común en el régimen medio del oleaje. Young (1999) estudia, a partir de datos de satélite y datos numéricos la variabilidad estacional de las condiciones de oleaje y viento en el globo. A nivel regional,

Woolf y Challenor (2002) estudian las variaciones mes a mes del clima marítimo en el Atlántico Norte (en términos de altura de ola significativa) a partir de medidas de varios satélites y Hemer et al. (2007) analiza las variaciones mensuales del clima marítimo en Australia con la base de datos C-ERA-40. Lionello y Sanna (2005) estudian la variabilidad estacional del clima marítimo del Mar Mediterráneo a partir de datos de boyas, satélite y modelo numérico. En lo que se refiere a extremos de oleaje, Menéndez et al. (2009a) estudia la variabilidad estacional de los extremos de oleaje en cinco boyas del litoral español mediante un modelo de extremos no estacionario basado en máximos mensuales.

La variabilidad interanual del clima marítimo también ha sido estudiada por numerosos autores. Woolf y Challenor (2002) analizan la relación entre la variabilidad del clima marítimo y los patrones de presión a gran escala en el Atlántico Norte durante el invierno. Los resultados de este estudio muestran una fuerte correlación entre las anomalías de altura de ola y la Oscilación del Atlántico Norte (NAO). Estos análisis demuestran que gran parte de la variabilidad interanual del clima marítimo en el Atlántico Norte puede ser explicada a partir de la variabilidad de los campos de presiones a gran escala. Por otro lado, Gulev y Grigorieva (2004) estudian la variabilidad interanual del clima marítimo del Atlántico y Pacífico Norte (mediante observaciones de barcos en ruta) correlacionando la altura de ola de invierno con índices climáticos característicos de ambas zonas, como la NAO, el *Southern Oscillation Index* (SOI) o el *North Pacific Index* (NPI). En su estudio sobre la variabilidad del clima marítimo del Mar Mediterráneo, Lionello y Sanna (2005) encuentran correlaciones entre la variabilidad interanual del oleaje, la NAO y el Monzón Indio. Kako y Kubota (2006) examinan la relación entre la variabilidad interanual de la altura de ola significativa de invierno en el Pacífico Norte y los eventos de Niño. Sasaki y Hibiya (2007) estudian la variabilidad del clima marítimo durante los meses de verano (Junio-Agosto) y su relación con la circulación ciclónica en el oeste del Pacífico Norte a partir de datos de reanálisis (ERA-40) y datos de satélite (TOPEX/Poseidon). Menéndez et al. (2008) estudia la variabilidad interanual de los extremos de oleaje en la costa pacífica de Estados Unidos encontrando una fuerte relación con el índice Niño 3.4, representativo del fenómeno El Niño Oscilación del Sur (ENSO) y con el índice *Pacific North America* (PNA), relacionado con la intensidad de los vientos del oeste. En el Hemisferio Sur, Harley et al. (2009) estudian la variabilidad interanual del clima

marítimo del área de Sydney (Australia), encontrando correlaciones con el fenómeno del ENSO y con el *Southern Annular Mode* (SAM), el cual es un descriptor de las diferencias de presión zonales del Hemisferio Sur entre las latitudes medias y sub-antárticas. Hemer et al. (2010) estudian la variabilidad del clima marítimo en el Hemisferio Sur mediante datos de satélite, boyas y modelo numérico, encontrando correlaciones entre la media mensual de altura de ola significativa y los índices climáticos SAM y SOI.

Por último, la variabilidad a largo plazo del clima marítimo también ha sido estudiada. Como ejemplos cabe destacar el estudio de Allan y Komar (2000) en el que detectan un incremento de la altura de ola de las boyas analizadas en el Pacífico Norte oriental, así como el estudio de Graham y Diaz (2001) donde encuentran tendencias positivas de largo plazo en la actividad tormentosa del invierno del Pacífico Norte. Wang y Swail (2001) analizan los cambios en los extremos de oleaje en el Hemisferio Norte, a partir de datos numéricos, los cuales relacionan con los cambios en la NAO a partir de datos de reanálisis y las tendencias de largo plazo del oleaje de viento (Gulev y Grigorieva 2004) a partir de los datos de barcos en ruta, encontrando tendencias positivas en el Pacífico y Atlántico Norte y negativas en el Pacífico Central occidental y la zona de Australia. Estos autores también estudiaron las tendencias de largo plazo de la altura de ola de invierno de oleaje de viento y de fondo mediante la misma base de datos (Gulev y Grigorieva 2006) encontrando tendencias positivas en el oleaje de viento y de fondo en el Pacífico Norte y tendencias positivas en el Atlántico Norte oriental por encima de la latitud 45° en el oleaje de fondo y tendencias negativas en el oleaje de viento en la zona comprendida entre Islandia y las Islas Británicas. Méndez et al. (2006) propone un modelo de extremos no estacionario basado en el método POT con el que estudia la tendencia de largo plazo de los extremos de oleaje de una boya frente a las costas del estado de Washington en Estados Unidos. Menéndez et al. (2008) estudia la tendencia de largo plazo de los extremos de oleaje a partir de los datos de una serie de boyas del Pacífico Norte y mediante un modelo no estacionario de extremos basado en el método POT. En sus resultados encuentran tendencias de largo plazo positivas en una serie de boyas frente a las costas de California y tendencias negativas en boyas de la costa Canadiense. Por último, Hemer et al. (2010) estudia las tendencias de largo plazo de la altura de ola media mensual de datos de satélite en el Hemisferio Sur.

### 1.2.2.2 Escala local

El estudio de la variabilidad climática de los extremos de oleaje a escala local requiere, en primer lugar, de la transferencia de las dinámicas del oleaje de aguas profundas a aguas someras. Esta transferencia se puede realizar mediante un *downscaling* estadístico (estableciendo relaciones estadísticas entre las observaciones de oleaje en aguas profundas y someras, por ejemplo Kalra et al. (2005) o Browne et al. (2007)), *downscaling* dinámico (aumentando la resolución mediante el anidamiento de modelos de generación y propagación del oleaje) o mediante una combinación de los dos.

En la literatura se pueden encontrar varias metodologías de la combinación de *downscaling* estadístico con *downscaling* dinámico para la transferencia de las dinámicas. Groeneweg (2006) propone una metodología en la que las condiciones de oleaje en aguas profundas son propagadas a la costa mediante una función de transferencia. La definición de las funciones de transferencia se realiza mediante un barrido de velocidades y direcciones de viento y nivel del mar en varias estaciones de la zona de estudio. A cada una de las condiciones de velocidad y dirección del viento se le asignan unas condiciones de oleaje en profundidades indefinidas, que constituyen las condiciones de contorno del modelo de propagación. Stansby et al. (2006) define otra función de transferencia basándose en el “Shore Protection Manual” para calcular las condiciones de un estado de mar a partir de datos de viento. Estas condiciones de oleaje calculadas establecen el número de casos a propagar, dividiendo la altura de ola y el período en una serie de intervalos para distintos sectores de dirección y varios niveles de mar. Una vez definidos los casos, se propagan mediante el modelo de propagación TOMAWAC (basado en la ecuación de balance de energía) y por último, se puede obtener cualquier situación de oleaje en profundidades indefinidas trasladándolo a la malla del modelo numérico.

Herman et al. (2009) plantea una metodología para la transferencia de las dinámicas basada en el *downscaling* dinámico de las condiciones de oleaje en indefinidas con la obtención de los principales modos de variación de los campos de oleaje (mediante la técnica de componentes principales) y el cálculo de un modelo estadístico para la obtención de esos modos a partir de los datos de clima marítimo en indefinidas.

Camus (2009) desarrolla una metodología de propagación del clima marítimo que tiene en cuenta la variabilidad espacial desde aguas profundas a aguas someras basada en el método de selección MaxDiss y el método de interpolación de funciones de base radial (RBF). El clima marítimo con variabilidad espacial se define mediante las condiciones de oleaje en el contorno de propagación y las condiciones de viento sobre toda el área. La aplicación del algoritmo de selección MaxDiss y de la técnica RBF se simplifica con una reducción de la dimensionalidad de los datos mediante la técnica de componentes principales.

### 1.2.3 Proyecciones de clima marítimo extremal

El estudio de la variabilidad a largo plazo del clima marítimo extremal mediante sus proyecciones a distintos escenarios de cambio climático puede realizarse mediante *downscaling* estadístico o mediante *downscaling* dinámico (combinado con un posterior análisis estadístico).

Wang et al. (2004) se basa en las relaciones entre la presión a nivel del mar y la altura de ola significativa (representadas mediante modelos de regresión) para obtener los valores de altura de ola en el Atlántico Norte para tres escenarios distintos de cambio climático mediante un análisis estadístico de redundancia. Los cambios proyectados en la altura de ola se analizan mediante un análisis de tendencias. En sus resultados encuentran que, en general, el calentamiento global está asociado con una mayor frecuencia de la fase positiva de la NAO, lo que genera mayor número de tormentas y un incremento en la altura de ola del noreste atlántico. Caires et al. (2006) utiliza un modelo no homogéneo de Poisson para modelar los valores extremos de oleaje a partir de los datos numéricos ERA-40. Los parámetros de la distribución están expresados en función de las anomalías medias estacionales de presión a nivel del mar en cada punto de análisis. Utilizando las proyecciones de presión a nivel del mar para tres escenarios de cambio climático se obtienen las proyecciones de los parámetros de Poisson. Mediante el análisis de estas proyecciones se determinan las tendencias de largo plazo, los cuantiles asociados a períodos de retorno y sus correspondientes incertidumbres. Por último Wang et al. (2006) obtiene las proyecciones de extremos y valores medios estacionales de oleaje a partir de las proyecciones de campos de presiones a nivel del mar para tres escenarios de cambio climático. En base a los datos de oleaje en el período

1960-1990 de ERA-40 establece un modelo GEV con sus parámetros expresados en términos de anomalías y gradientes de presión estacional, modelo que utilizará para proyectar el oleaje. También caracteriza las incertidumbres asociadas a los distintos modelos climáticos y escenarios y establece tendencias en el oleaje mediante modelos de regresión.

El downscaling estadístico se basa en la ejecución de modelos numéricos de generación de oleaje alimentados con las condiciones atmosféricas de los escenarios y haciendo el posterior análisis estadístico. Grabemann et al. (2008) analizan los cambios en los extremos de oleaje en el Mar de Norte simulando las condiciones de oleaje (mediante el modelo WAM) en el período 2071-2100 para los escenarios A2 y B2 y haciendo un posterior análisis estadístico a las diferencias de oleaje entre ese período y el período 1960-1990.

### 1.3 Oportunidad de la tesis

Tras la revisión del estado del arte se ha concluido que existen varias carencias metodológicas que condicionan el análisis de los procesos asociados a la variabilidad del oleaje a las diferentes escalas temporales y espaciales de interés. Entre ellos se consideran los siguientes:

#### *En cuanto a los métodos estadísticos:*

Los modelos utilizados usualmente en la caracterización del régimen extremal de oleaje son los modelos, estacionarios o no estacionarios, basados en la distribución GEV y en el método de selección de extremos POT. En lo que se refiere a los modelos GEV el bloque de tiempo escogido para la selección de extremos suele ser de un año o un mes, pero no se han encontrado otros intervalos de tiempo que puedan estar ligadas a procesos físicos de la escala regional. En cuanto al modelo POT, generalmente se utiliza un valor constante del umbral, pero no se han encontrado estudios con umbral variable para caracterizar una determinada escala temporal. Tampoco se ha encontrado ninguna comparativa entre modelos que determine cual es el más apropiado según la escala de variabilidad de interés.

Por otro lado, el uso de bases de datos instrumentales lleva asociado un porcentaje de falta de datos. En los estudios revisados no se ha encontrado ninguna metodología que tenga en cuenta esta carencia de datos.

### *En cuanto a la variabilidad espacial y temporal:*

En la literatura se han encontrado estudios del oleaje extremal a nivel global mediante datos de satélite y reanálisis. Sin embargo, se aprecia una falta de análisis estacional a nivel global, mediante datos instrumentales. En lo que se refiere a la variabilidad interanual, se ha estudiado la influencia de distintos índices climáticos en los extremos de oleaje de distintas partes del mundo, pero siempre de forma regional y no de forma global.

En la escala regional el estudio de la variabilidad interanual de los extremos de oleaje está enfocado a la relación con índices climáticos conocidos, mientras que no se explora la relación con procesos atmosféricos de escala sinóptica. Por otro lado, no se ha encontrado un análisis de la variabilidad espacial de los extremos de oleaje en el continente sudamericano, incluyendo la variabilidad estacional y el largo plazo.

Respecto a la escala local, se han encontrado aplicaciones de las distintas metodologías de transferencia de las dinámicas de aguas profundas a someras. Sin embargo, no se ha encontrado un análisis de la variación de los valores extremos de oleaje de aguas profundas a aguas reducidas.

### *En cuanto a las proyecciones de clima marítimo extremal*

Los estudios revisados, en lo que se refiere a las técnicas de *downscaling* estadístico, establecen relaciones entre los extremos de oleaje en un punto y las anomalías y gradientes de presión en ese mismo punto. Además el estudio estacional se hace de manera independiente, obteniendo un modelo de extremos distinto para cada estación. En vista de esto, y dado que el oleaje en un punto responde a las variaciones en las condiciones atmosféricas de un área de influencia, se considera necesario realizar el estudio de la relación entre los extremos de oleaje y las situaciones sinópticas de la zona de influencia para poder proyectar a distintos escenarios de cambio climático. Por otro

lado, tampoco se encuentra un análisis estacional utilizando toda la población de datos de oleaje.

## 1.4 Objetivos y metodología

El objetivo general de esta tesis es profundizar en el conocimiento de la variabilidad climática de los valores extremos del oleaje. La finalidad es conocer la variabilidad espacial y temporal del clima marítimo extremal en aguas profundas y someras y poder proyectar el régimen extremal de oleaje a distintos escenarios de cambio climático.

Una vez revisado el estado del conocimiento de la teoría estadística de valores extremos y su aplicación al estudio del oleaje y su variabilidad climática se establecen como objetivos específicos de esta tesis:

- Ahondar en el estudio de los modelos estadísticos no estacionarios comparando distintos modelos; aportando avances que permitan modelar mejor la variabilidad climática y estableciendo una serie de recomendaciones en el uso de los modelos disponibles. También se pretende solventar la existencia de “huecos” en las series temporales de oleaje para evitar su influencia en el modelado de los extremos y explorar distintos métodos que permitan disminuir el tiempo computacional del ajuste de los modelos no estacionarios.
- Estudiar la variabilidad climática del clima marítimo extremal en aguas profundas. Para ello se plantea un estudio a nivel global mediante datos instrumentales y un modelo estadístico capaz de modelar la estacionalidad y la variabilidad interanual de los extremos de oleaje de todo el globo. Por otro lado, se aborda el estudio a nivel regional para poder incluir procesos físicos de mesoescala que afectan al oleaje. El estudio a nivel regional se lleva a cabo en el sur de Europa y el sur de América, estudiando distintas escalas de variabilidad temporal en cada zona de estudio.
- Estudiar la variabilidad del clima marítimo extremal en aguas someras. Para ello se hará uso de la metodología de transferencia de las dinámicas de aguas profundas a aguas someras desarrollada por IH Cantabria y se estudiará la variabilidad espacial y temporal

del clima marítimo extremal en aguas profundas y someras del Golfo de Cádiz y el litoral Mediterráneo español.

- Establecer una metodología para obtener el régimen extremal de oleaje, a corto, medio y largo plazo, en función de patrones sinópticos y que permita proyectarlo a distintos escenarios de cambio climático. Una vez validada la metodología se proyectará el clima marítimo extremal a diversas zonas del Atlántico Norte y a distintos escenarios de cambio climático.

En la Figura 1.6 se muestra un esquema conceptual de la tesis donde quedan plasmados los motivos que llevan al estudio de la variabilidad climática de extremos de oleaje, qué se estudia, por qué y cómo.

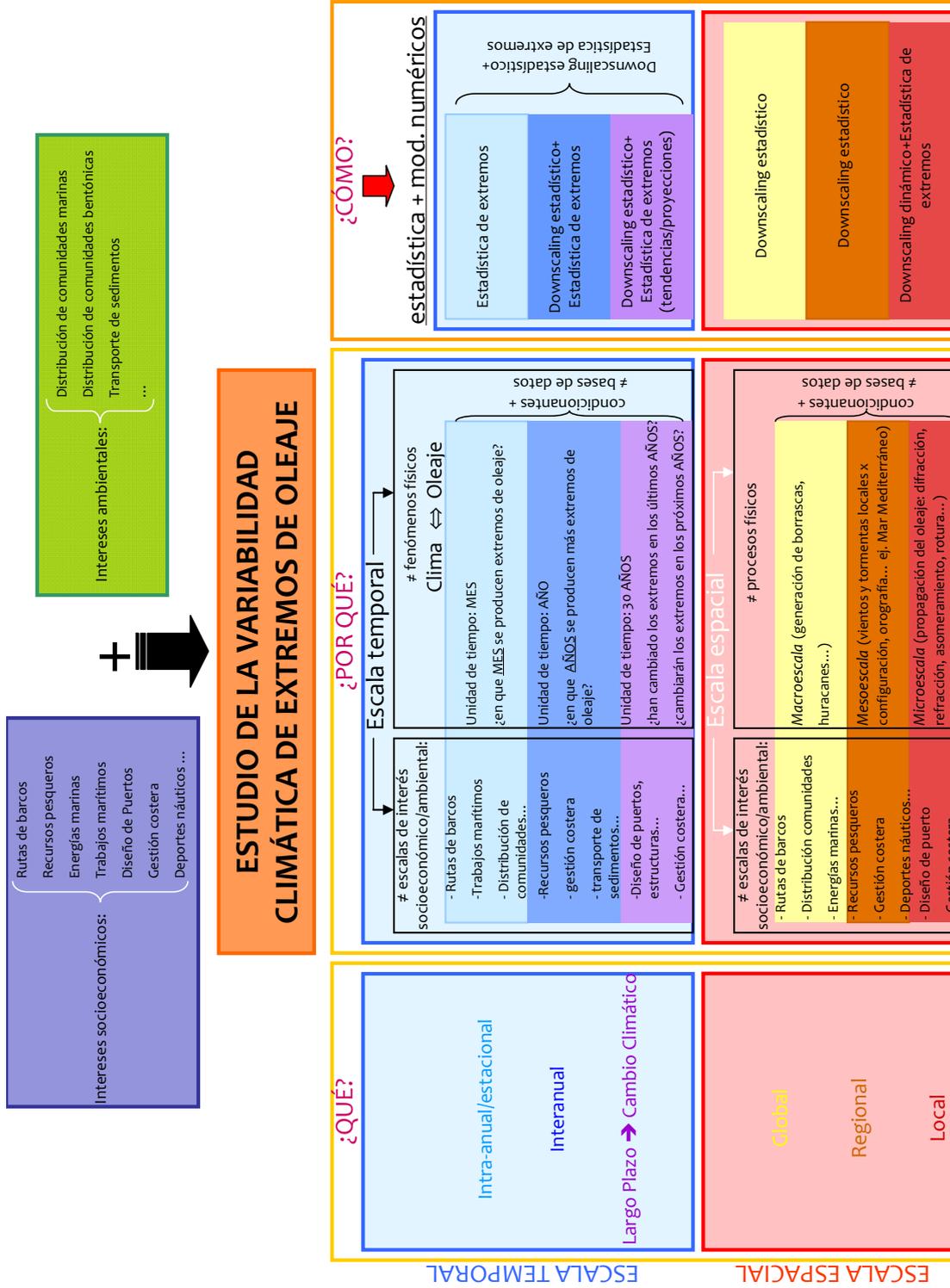


Figura 1.6 Esquema conceptual de la tesis

## 1.5 Organización de la tesis

Esta tesis se ha estructurado en 7 capítulos. En los capítulos 3, 4, 5 y 6 se presentan las aportaciones de la tesis que dan respuesta a los objetivos específicos planteados.

En el Capítulo 1 se establece un marco de la tesis en el que se pone de manifiesto la importancia y necesidad del estudio de los eventos extremos del oleaje y en especial de su variabilidad a distintas escalas. A continuación, se revisa el estado del conocimiento en cuanto a la estadística extremos y las técnicas de clasificación y su uso para caracterizar la variabilidad del clima marítimo. Finalmente, se analizan las carencias del estado del conocimiento y se plantea el objetivo general de la tesis y unos objetivos específicos.

En el Capítulo 2 se hace un esquema general de las bases de datos utilizadas y se detallan algunos aspectos sobre cada una de ellas.

El Capítulo 3 recoge los avances en los modelos estadísticos no estacionarios aportados en esta tesis. En primer lugar, se describen los dos métodos automáticos de selección de parámetros utilizados en la tesis. A continuación, se estudia la influencia de la homogeneidad en las series de datos y su tratamiento en los modelos estadísticos y seguidamente, se hace una comparación entre distintos modelos estadísticos no estacionarios para el estudio del clima marítimo extremal. Por último, se establecen unas recomendaciones de uso de los distintos modelos estadísticos de extremos para el estudio del clima marítimo.

En el Capítulo 4 se estudia la variabilidad temporal y espacial del clima marítimo extremal a distintas escalas. Por un lado, se estudia la variabilidad estacional e interanual del clima marítimo extremal global y del sur de Europa y, por otro, se estudia la variabilidad estacional y de largo plazo del clima marítimo extremal de América del sur.

En el Capítulo 5 se hace un análisis del clima marítimo extremal en aguas someras. En primer lugar, se plantea la metodología seguida en la transferencia de las dinámicas de aguas profundas a aguas someras. A continuación, se presenta el modelo estadístico

utilizado y, por último, los resultados de variabilidad estacional y de largo plazo obtenidos en el Golfo de Cádiz y el litoral Mediterráneo español.

En el Capítulo 6 se plantea una metodología para determinar el régimen extremal de oleaje en función de patrones de tiempo y a distintas escalas temporales (corto, largo y muy largo plazo). Utilizando la técnica de redes auto-organizativas se hace un análisis de las situaciones atmosféricas de la zona de influencia de un punto de estudio estableciendo una serie de tipos de tiempo de mesoescala. A continuación, se vinculan los oleajes extremos con los tipos de tiempo, estableciendo el régimen extremal de oleaje para cada situación sinóptica mediante un modelo GEV (corto plazo). En el largo plazo (escala estacional) se plantea una metodología en la que se vinculan los tipos de tiempo con los eventos extremos en un punto para generar el régimen extremal de ese lugar concreto. La metodología permite obtener el régimen extremal total y por estaciones del punto de estudio. Por último, se presenta una metodología para obtener el régimen extremal en un punto y poder proyectarlo a distintos escenarios de cambio climático. Esta metodología utiliza información de las situaciones sinópticas de la zona de estudio y de aquellas que generan extremos de oleaje. Una vez validada la metodología se aplica en tres puntos del Atlántico Norte y se obtienen las proyecciones de clima marítimo extremal para tres escenarios de cambio climático proporcionados por el Panel Intergubernamental de Cambio Climático (IPCC).

El Capítulo 7 incluye una serie de conclusiones y en él se indican unas futuras líneas de investigación.

Por último, se recoge toda la bibliografía utilizada en esta tesis.