

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 3

AVANCES EN MODELOS ESTADÍSTICOS NO ESTACIONARIOS

3.1 Introducción

La estadística de extremos se ha desarrollado ampliamente en los últimos años en el estudio de variables geofísicas. En el caso del oleaje, el análisis de la variabilidad de los valores extremos se ha centrado principalmente en el modelado mediante modelos estacionarios y no estacionarios de tipo paramétrico. La complejidad de estos modelos puede ser muy variada dependiendo de la escala temporal que se quiera estudiar. Ante la variedad de modelos disponibles, en este capítulo se hace un repaso y comparativa de

los modelos de extremos no estacionarios, examinando también sus condicionantes (por ejemplo tiempo computacional) y sus limitaciones.

En lo que se refiere a condicionantes en el modelado de extremos de oleaje, por un lado se ataca el problema del tiempo computacional y por otro el derivado del uso de bases de datos instrumentales con huecos en las series temporales. En lo que respecta al condicionante computacional se plantean dos métodos de selección de parámetros que se han utilizado a lo largo de esta tesis. Se explica su procedimiento y las mejoras en cuanto a tiempos de computación. En lo referido al problema de huecos en las bases de datos, hay que tener en cuenta que pueden dar lugar a resultados y conclusiones erróneas, por lo que se plantea una metodología para que la falta de datos en la serie no influya en el modelado del oleaje extremal.

A continuación se hace un estudio comparativo entre modelos de extremos. En primer lugar se comparan los modelos no estacionarios GEV y Pareto-Poisson, para modelar la misma escala temporal (en este caso variabilidad estacional e interanual). Posteriormente se estudian las diferencias en el modelado de la variabilidad de los extremos de oleaje en un punto al utilizar dos modelos GEV con distinto bloque de tiempo en la elección de los extremos.

Finalmente, se presentan, como una de las principales aportaciones de este capítulo, unas recomendaciones de uso de cada modelo de extremos en función del estudio que se quiera llevar a cabo.

3.2 Métodos automáticos de selección de parámetros

Los modelos paramétricos presentados en la revisión del estado del arte tratan de reproducir la variabilidad climática del oleaje extremo a lo largo del tiempo mediante expresiones matemáticas aditivas con funciones lineales, armónicas y covariables. Este hecho hace que la estimación de los parámetros del modelo no es una tarea fácil por varias razones. En primer lugar, la selección del número de funciones armónicas o del número de covariables a incluir en el modelo no es evidente y esto condicionará el número de posibles modelos, que puede ser realmente elevado. El hecho de que pueda haber un número tan grande de posibles modelos hace bastante inviable

(computacionalmente hablando) la comprobación de cada uno de ellos. A modo de ejemplo, un modelo con 10 parámetros permite $2^{10} = 1024$ posibles parametrizaciones. Por esta razón, una vez elegido el número de parámetros del modelo, será necesario utilizar métodos de estimación apropiados, no incluidos normalmente en los paquetes estadísticos estándar.

Uno de los objetivos en el desarrollo de esta tesis ha sido la exploración de métodos adecuados de selección del mejor modelo paramétrico de extremos. A continuación se describen los dos métodos desarrollados, los cuales difieren en el método de optimización utilizado para la maximización de la función de logaritmo de verosimilitud. Inicialmente se utilizó un procedimiento de optimización global mediante el algoritmo denominado *Shuffled Complex Evolution* (SCE-UA) (Duan et al., 1992). El algoritmo es incorporado con éxito en Menéndez et al. (2009a) y Menéndez et al. (2009b) para el caso de modelos basados en la familia de distribuciones GEV. El segundo método desarrollado presenta un gran avance en la disminución del tiempo computacional requerido para la obtención del modelo óptimo. La incorporación de nuevos parámetros al modelo está basada en el análisis de sensibilidad, y la optimización del modelo se hace mediante un algoritmo de máximos locales. Este método es presentado y contrastado en Mínguez et al. (2010) para modelos basados en la GEV. A continuación se detalla cada uno de los métodos.

3.2.1 Selección automática del modelo no estacionario

El primer requisito de este método de selección es definir los procesos que se pueden incluir en el modelo paramétrico y, por lo tanto, el número máximo de parámetros posible. Con base en la experiencia en el estudio de la variable se podrá plantear un modelo que refleje la variabilidad estacional, interanual o tendencia de largo plazo en todos los parámetros del modelo o sólo en algunos de ellos.

Una vez definidos el número máximo de parámetros que puede contener el modelo (n_p), habrá que elegir, de entre los $2^{n_p} - 1$ posibles modelos el óptimo. Para ello se impone el principio de parsimonia, por el cual se considera que el modelo óptimo será aquel que, con el menor número de parámetros, ajuste mejor los datos de la muestra.

Dada la necesidad de optimizar el proceso de búsqueda y la inviabilidad de evaluar todos los posibles modelos, este método plantea una búsqueda selectiva automatizada. Para ello se establece una codificación binaria del modelo y, partiendo del más simple, se valora la incorporación de parámetros considerando la interacción entre diferentes factores parametrizados.

Siguiendo la nomenclatura de los algoritmos genéticos se adopta una codificación binaria para definir cada modelo en función de los factores a incluir. De esta manera, los genes g_i estarán relacionados con armónicos (ciclo anual, semianual, etc), covariables y tendencias, donde $g_i = 1$ indica que el factor correspondiente es incluido en el modelo. Para comparar la calidad de los modelos se recurre a criterios estadísticos que tengan en cuenta la bondad del ajuste y el número de parámetros que se incluyen en el modelo. A continuación se definen los dos criterios utilizados en esta tesis. En primer lugar, el criterio de información de Akaike (AIC) que utiliza el valor de la función logarítmica de verosimilitud (\hat{l}) y el número de parámetros (p) mediante la expresión:

$$AIC = -2\hat{l}(p) + 2p \quad (3.1)$$

En segundo lugar se utiliza el criterio del cociente de máxima verosimilitud. Este criterio supone un modelo A obtenido de otro B mediante la incorporación a éste de $p_B - p_A$ parámetros, donde $p_A < p_B$ es el número de parámetros de cada modelo. Sea $\hat{l}(p_A) \leq \hat{l}(p_B)$ el valor máximo de las respectivas funciones logarítmicas de verosimilitud, entonces, el modelo A será rechazado si se cumple la expresión (3.2) al nivel de confianza $1 - \alpha$.

$$2(\hat{l}(p_B) - \hat{l}(p_A)) > \chi_{1-\alpha}^2(p_B - p_A) \quad (3.2)$$

La selección automática del modelo basado en estos criterios se lleva a cabo utilizando un algoritmo *stepwise* que combina otros dos procedimientos: *forward selection* y *backward elimination*. De esta manera, cada vez que se activa un gen por el procedimiento *forward*, se comprueba, mediante el procedimiento *backward*, si debe eliminarse otro activado previamente. El ajuste del modelo se lleva a cabo mediante el algoritmo SCE-UA que es

un procedimiento eficiente de optimización global pero que requiere un control de los límites de búsqueda de los parámetros a estimar y de la estabilidad de la solución final, que depende del número de evaluaciones estimadas (Menéndez et al., 2009a).

3.2.2 Selección pseudo-óptima de los parámetros del modelo no estacionario

La primera ventaja que presenta este método frente al anterior es que, una vez elegidas las variaciones temporales que se van a estudiar y definidas la parametrizaciones, no es necesario fijar el número máximo de parámetros que tendrá el modelo.

El método converge monótonamente a la solución final incorporando un único parámetro cada vez, basándose en un análisis de sensibilidad (derivadas de primer orden) a partir de la última solución obtenida. Este procedimiento reduce drásticamente el número de parametrizaciones diferentes a comprobar y, por tanto, el tiempo computacional requerido.

Los criterios de información elegidos para la comparación de la calidad de los modelos son los mismos que en el método anterior. El proceso comienza con el modelo más simple posible, que corresponde con el modelo estacionario tradicional con parámetros de la GEV constantes. El número de parámetros se incrementa en una unidad por cada iteración (excepto en el caso de armónicos que se incorporan de dos en dos), basándose en la información del análisis de sensibilidad, y el orden de incorporación de parámetros es: armónicos, covariables y tendencia de largo plazo.

La información proporcionada por el análisis de sensibilidad permite introducir en el modelo el mejor parámetro que maximiza el incremento en la función logarítmica de verosimilitud, que en este caso, es el parámetro cuya perturbación en la función logarítmica de verosimilitud es máxima. El proceso continúa incluyendo nuevos parámetros hasta que no hay mejora en el criterio de información de Akaike o en el cociente de máxima verosimilitud.

El ajuste del modelo mediante la maximización de la función logarítmica de verosimilitud se lleva a cabo mediante el algoritmo de optimización *Trust Region Reflective*

y limitando el dominio de búsqueda para hacer más robusta la estimación de los parámetros. El método completo de selección pseudo-automática está ampliamente explicado en Mínguez et al. (2010).

3.3 Influencia de la homogeneidad de las series de datos

La discontinuidad en las series de datos siempre supone una preocupación. Lógicamente, los modelos desarrollados proporcionan mejores resultados cuanto más larga y homogénea es la población de datos. Sin embargo, los datos instrumentales pocas veces cumplen estas características. Como se describe en el capítulo 2 de esta tesis, los registros de boyas son relativamente cortos (del orden de 20 años) y, pese a que miden en continuo, es inevitable encontrar huecos en las series debidos a mal funcionamiento o mantenimiento. También se encuentran problemas similares en los datos de satélite, pero en este caso la heterogeneidad en la serie de datos proviene por el aumento de misiones satelitales a lo largo del tiempo. En la Figura 3.1 se puede ver el aumento en el número de datos disponibles entre los años 1992 y 2006 debido a la incorporación de nuevos satélites, detectándose un aumento importante del número de datos en el período 2002-2005

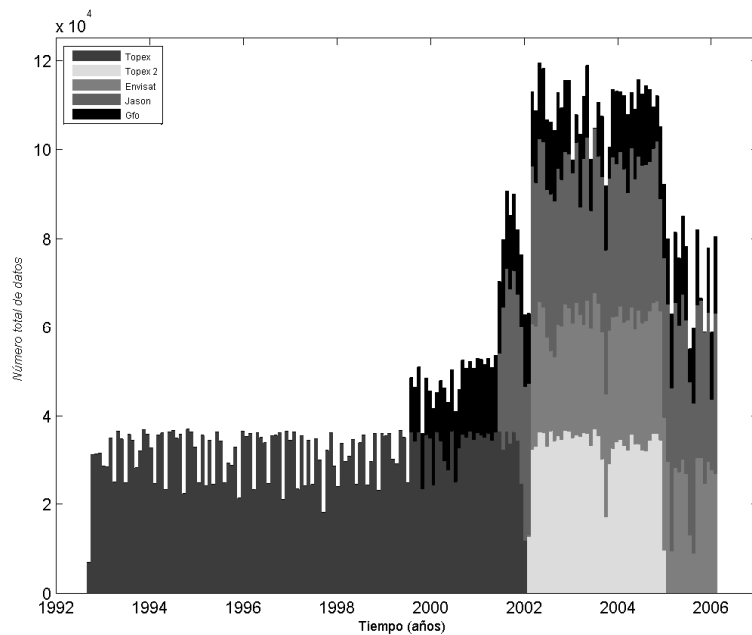


Figura 3.1 Número de medidas para cada misión satelital

La heterogeneidad descrita en las series de datos instrumentales se ha referido solamente a la escala temporal, pero las discontinuidades espaciales también son importantes. El estudio del régimen extremal en grandes áreas se lleva a cabo discretizando la zona y agregando los datos disponibles. La dispersión de las boyas a lo largo del mundo produce discontinuidades en las zonas de estudio. En concreto, este problema es bastante importante en el Hemisferio Sur, donde la red instrumental de medida es bastante escasa. Como ya se ha explicado, la solución a estos problemas viene dada por las series numéricas de oleaje pero, debido a la importancia y riqueza de las series instrumentales, su uso sigue siendo imprescindible. Por esta razón se ha introducido en los modelos de extremos una modificación que tiene en cuenta la heterogeneidad en las muestras de datos. En concreto se ha introducido un factor de escala en la función de distribución GEV.

Sea una secuencia de N observaciones aleatorias e independientes X_1, \dots, X_N y una función de distribución común asociada F , tal que $F(x) = \Pr(X_i \leq x)$. La función de distribución de la muestra de extremos $M_N = \max\{X_1, \dots, X_N\}$ podrá derivarse teóricamente considerando la N -ésima potencia de F :

$$G_N(x; \mu^*, \psi^*, \xi^*) = \{F(x)\}^N \quad (3.3)$$

Sea ahora otra secuencia de observaciones aleatorias e independientes con N_1 datos, X_1, \dots, X_{N_1} , con la misma función de distribución común asociada F . En este caso la función de distribución de la nueva muestra de extremos $M_{N_1} = \max\{X_1, \dots, X_{N_1}\}$ también puede derivarse teóricamente de la forma:

$$G_{N_1}(x; \mu, \psi, \xi) = \{F(x)\}^{N_1} \quad (3.4)$$

El hecho de que ambas secuencias de datos tengan la misma función de distribución, F , asociada lleva a la igualdad de las funciones de distribución de las muestras de extremos de la siguiente forma:

$$\{G_{N_1}(x)\}^{1/N_1} = \{G_N(x)\}^{1/N} \quad (3.5)$$

El objetivo es encontrar la expresión de la función de distribución de la familia GEV de la muestra de datos N_1 :

$$G_{N_1}(x) = \{G_N(x)\}^{N_1/N} = \exp \left\{ - \left[1 + \xi^* \left(\frac{x - \mu^*}{\psi^*} \right) \right]_+^{-1/\xi^*} \right\}^{N_1/N} \quad (3.6)$$

Utilizando las propiedades de la función exponencial, el factor N_1/N puede introducirse dentro de la expresión de la GEV a través del factor de escala $k_1 = N_1/N$:

$$G_{N_1}(x) = \exp \left\{ -k_1 \left[1 + \xi^* \left(\frac{x - \mu^*}{\psi^*} \right) \right]_+^{-1/\xi^*} \right\} \quad (3.7)$$

Este razonamiento es análogo para distintos tamaños de la secuencia de datos, obteniendo distintos factores $k_2 = N_2/N, \dots, k_n = N_n/N$. Por lo tanto se puede establecer un factor de escala $k_i = N_i/N$ para la muestra de tamaño N_i :

$$G_{N_i}(x) = \exp \left\{ -k_i \left[1 + \xi^* \left(\frac{x - \mu^*}{\psi^*} \right) \right]_+^{-1/\xi^*} \right\} \quad (3.8)$$

y asumiendo continuidad en el proceso se puede establecer un k variable a lo largo del tiempo de la forma $k(t) = n(t)/N$.

$$G_{N_i}(x) = \exp \left\{ -k(t) \left[1 + \xi^*(t) \left(\frac{x - \mu^*(t)}{\psi^*(t)} \right) \right]_+^{-1/\xi^*(t)} \right\} \quad (3.9)$$

donde los parámetros de la función de distribución GEV de referencia (para máximos de una muestra de tamaño N) están dados por las expresiones:

$$\begin{aligned}\mu^*(t) &= \mu(t) + \frac{\psi(t)}{\xi(t)} [k(t)^{\xi(t)} - 1] \\ \psi^*(t) &= \psi(t) k(t)^{\xi(t)} \\ \xi^*(t) &= \xi(t)\end{aligned}\tag{3.10}$$

Al aplicar esta corrección a la función de distribución GEV las posibles tendencias inducidas por el aumento o disminución en la densidad de datos a lo largo del tiempo quedan solventados. A modo de ejemplo, se ha aplicado el factor de escala sobre cuatro poblaciones de datos sintéticas distintas (Figura 3.2) y se ha ajustado un modelo con $\mu(t) = \mu_0 + \beta_{LT} \cdot t$, ψ_0 y ξ_0 . En la primera de ellas la población es homogénea ($N = \text{cte}$), por lo que la tendencia esperable del modelo debe ser 0 o muy próxima a 0. La segunda población (N creciente linealmente) cuenta con mayor número de datos a lo largo del tiempo, lo que estaría introduciendo una tendencia positiva, al haber más probabilidad de encontrar valores mayores en los últimos años. Sin embargo, este podría ser el caso del aumento de satélites a lo largo del tiempo, lo que no tiene por qué implicar un aumento en la altura de ola. Las dos últimas poblaciones generadas tienen distinto número de datos, N , a lo largo del tiempo pero de forma escalonada. La primera de ellas aumenta su población en un determinado momento y la segunda la disminuye. Esta variación escalonada en la población de datos de estas series también introduce una tendencia ficticia. La aplicación del factor de escala sobre los modelos ajustados a estas poblaciones evita la aparición de tendencias ligadas al aumento o disminución en la población de datos. En la figura se puede ver como la tendencia, β_{LT} , obtenida por un modelo de extremos GEV con el factor de escala presentado es prácticamente 0 en todos los casos.

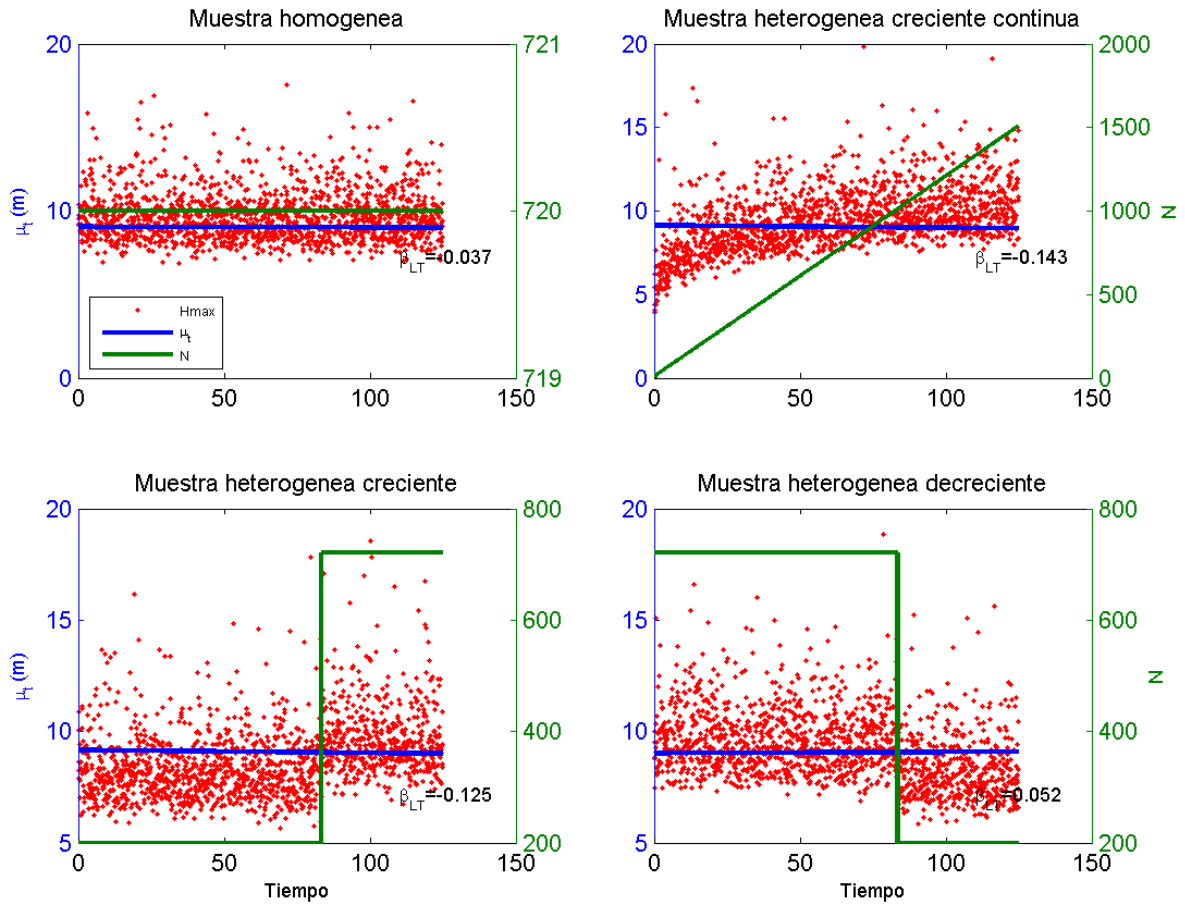


Figura 3.2 Ajuste del modelo GEV con factor de escala a distintas poblaciones de extremos y tendencias obtenidas

Como se puede ver, la aplicación del factor de escala permite utilizar poblaciones de datos con huecos o con aumento o disminución de datos a lo largo del tiempo evitando los resultados erróneos derivados de este hecho.

3.4 Comparación entre modelos no estacionarios GEV y Pareto-Poisson

El uso de las distintas variantes de los modelos no estacionarios GEV y POT es ampliamente aceptado por la comunidad científica. Así por ejemplo, Wang et al. (2004) utiliza un modelo GEV anual no estacionario para estudiar los extremos de oleaje en el Atlántico Norte, Menéndez et al. (2009a) utiliza un modelo GEV mensual no estacionario para el análisis de la estacionalidad en los extremos de oleaje en la costa española, Wang et al. (2008) utiliza un modelo GEV no estacionario con r -máximos en un bloque de tiempo para proyectar los eventos extremos de precipitación en invierno en Norte América y Caires et al. (2006) utiliza el modelo POT no estacionario para

estudiar los extremos de oleaje y además lo compara con el GEV anual utilizado por Wang et al. (2006). Con el objetivo de conocer las diferencias en los resultados de los modelos GEV y POT, en este apartado se compara el uso y resultados de dos de estos modelos. Para ello se aplican al estudio de la variabilidad climática del régimen extremal de oleaje en un punto situado en el Atlántico Norte ($15^{\circ}\text{W}, 55^{\circ}\text{N}$, ver Figura 3.3).

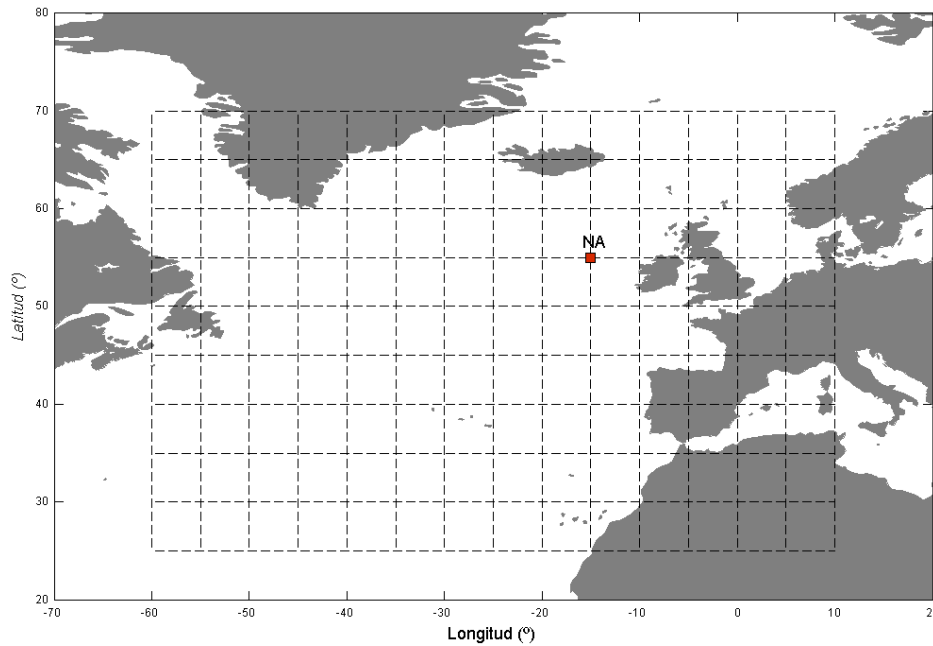


Figura 3.3 Localización del punto NA ($15^{\circ}\text{W}, 55^{\circ}\text{N}$)

El punto de estudio denominado NA se encuentra en la parte oriental del Atlántico Norte, en las coordenadas ($15^{\circ}\text{W}, 55^{\circ}\text{N}$) y proviene de la base de datos de reanálisis GOW (IH Cantabria), con un registro horario de 61 años (1948-2008). Atendiendo a su localización cabe pensar en una zona cuyo oleaje extremal presente una variabilidad climática importante. Por un lado, la proximidad a la borrasca de Islandia hace que esté sometido a la influencia directa de la NAO y otros patrones de variabilidad atmosférica del Atlántico Norte, lo que probablemente de lugar a una variabilidad interanual importante del oleaje extremo. Por otro lado, la variación invierno-verano del oleaje extremo en esta zona queda patente en Izaguirre et al. (2010). Así, un análisis descriptivo de las alturas de ola significantes máximas mensuales de la serie mediante un gráfico de cajas (Figura 3.4), demuestra claramente la variabilidad estacional que presentan los extremos de oleaje en ese punto.

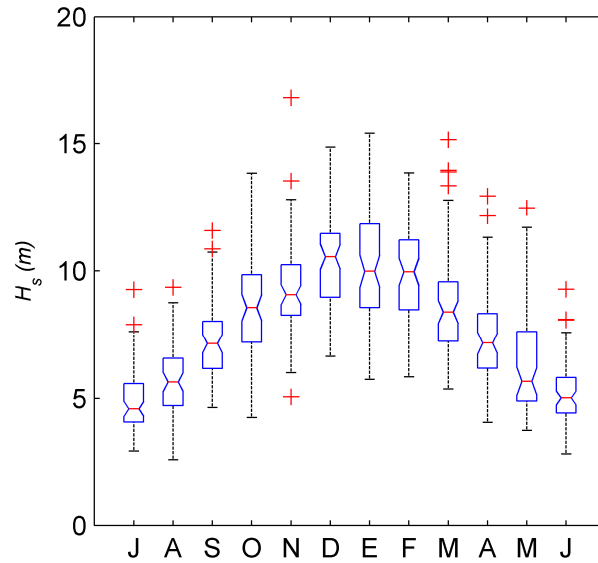


Figura 3.4 Boxplot de la H_s máxima mensual en el punto 15W55N

El gráfico de cajas (*boxplot*) representa en un gráfico los valores de los cuartiles bajo (25%), medio (50%) y alto (75%) mediante las líneas de las cajas trapezoidales. Las líneas verticales negras se extienden hasta el rango intercuartílico 1.5 o hasta el rango de los datos (el que sea menor) y por últimos, las cruces representan valores inusuales.

Se plantea por tanto, como objetivo, aplicar y comparar dos modelos no estacionarios capaces de estudiar la variabilidad inter e intra anual de los extremos de oleaje en este punto.

El primer modelo elegido proviene del método de selección de extremos de máximos en un bloque de tiempo. Dado que una de las escalas a estudiar es la intra anual, el bloque de tiempo elegido ha sido el mes (máximos mensuales), de forma que un año queda representado por doce valores. El primer modelo no estacionario es por tanto el modelo GEV mensual.

El segundo modelo está basado en el método de selección de extremos de excedencias sobre un umbral. Para poder representar adecuadamente la información del oleaje extremo contenida en un año se ha optado por elegir un umbral variable a lo largo del tiempo ($u(t)$), de forma que los datos estén repartidos homogéneamente a lo largo del año. El segundo modelo no estacionario utilizado, basado en este método, es el modelo de Pareto-Poisson.

Para comparar el comportamiento y los resultados de ambos modelos se va a utilizar para los dos la misma parametrización.

3.4.1 Parametrización de los modelos

Las ecuaciones (1.13) ligan los parámetros de las funciones de Pareto-Poisson con los de la GEV. Haciendo uso de esta relación, se puede ajustar el modelo POT a través de la estimación de los parámetros de la GEV. De esta manera, los parámetros a estimar en ambos modelos son los mismos, por lo que se propone la misma parametrización para ambos.

La parametrización de la variabilidad estacional se lleva a cabo mediante funciones armónicas que se pueden incorporar en los parámetros de localización, escala y forma. En principio, el parámetro de localización, que sería el equivalente a la moda de la distribución y el parámetro de escala (equiparable a la dispersión) son los más susceptibles de incorporar ondas para modelar la estacionalidad. Sin embargo, se deja la posibilidad de que el parámetro de forma, que define la cola de la distribución, pueda incorporar alguna onda, y por lo tanto, variar a lo largo del año (aunque este hecho es menos frecuente).

En cuanto a la variabilidad interanual, se han elegido como predictores las anomalías de los campos de presiones en superficie medios mensuales de la zona marcada con un rectángulo en la Figura 3.3. Una vez procesados, mediante un análisis por componentes principales, se introducen los 10 primeros modos temporales como posibles covariables del modelo en los parámetros de localización y escala.

Con estas consideraciones la parametrización es de la siguiente forma:

$$\begin{aligned}
 \mu(t) &= \beta_0 + \sum_{i=1}^{P_\mu} [\beta_{2i-1} \cos(2i\pi t) + \beta_{2i} \sin(2i\pi t)] + \sum_{j=1}^{C_\mu} \beta_{PC_j} PC_j(t) \\
 \psi(t) &= \alpha_0 + \sum_{i=1}^{P_\psi} [\alpha_{2i-1} \cos(2i\pi t) + \alpha_{2i} \sin(2i\pi t)] + \sum_{j=1}^{C_\psi} \alpha_{PC_j} PC_j(t) \\
 \xi(t) &= \gamma_0 + \sum_{i=1}^{P_\xi} [\gamma_{2i-1} \cos(2i\pi t) + \gamma_{2i} \sin(2i\pi t)]
 \end{aligned}
 \tag{3.11}$$

donde P_μ , P_ψ y P_ξ son el número de armónicos a considerar en cada parámetro, β_i , α_i y γ_i son las amplitudes de los armónicos, C_μ y C_ψ es el número de modos temporales considerados en cada parámetro y β_{PC_j} y α_{PC_j} representa la influencia en la altura de ola extremal por unidad de covariable.

3.4.2 Modelo GEV mensual

El método del máximo mensual utiliza los valores máximos de oleaje en un bloque de tiempo de longitud 1 mes. Aplicando este método a la serie de oleaje, la muestra de estudio cuenta con 736 valores. En la Figura 3.5 se representa la serie temporal de los máximos mensuales a modelar (unidas por una línea negra).

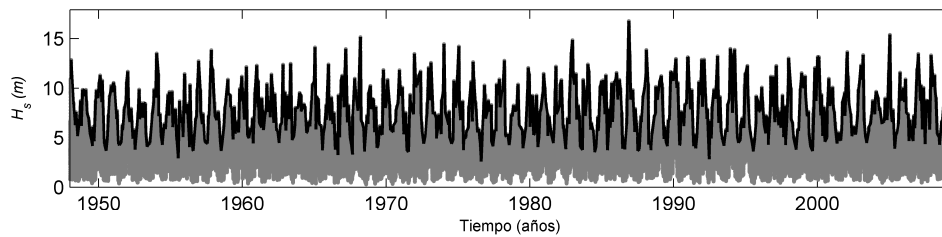


Figura 3.5 Serie temporal de oleaje en el punto 15°W, 55°N (puntos grises) y máximos mensuales (línea negra)

El mejor modelo encontrado en el ajuste introduce una onda en el parámetro de localización y una onda en el parámetro de escala para modelar la estacionalidad (la media y la dispersión varían con el ciclo anual). En cuanto al parámetro de forma, que define la cola de la distribución, permanece constante a lo largo del tiempo y toma el valor $\hat{\xi} = -0.08$, valor negativo próximo a 0, lo que indica que la cola de la distribución está ligeramente acotada (comportamiento Weibull). En la Figura 3.6 se observa la

variación a lo largo del año de los parámetros de localización y escala debido al efecto de cada onda.

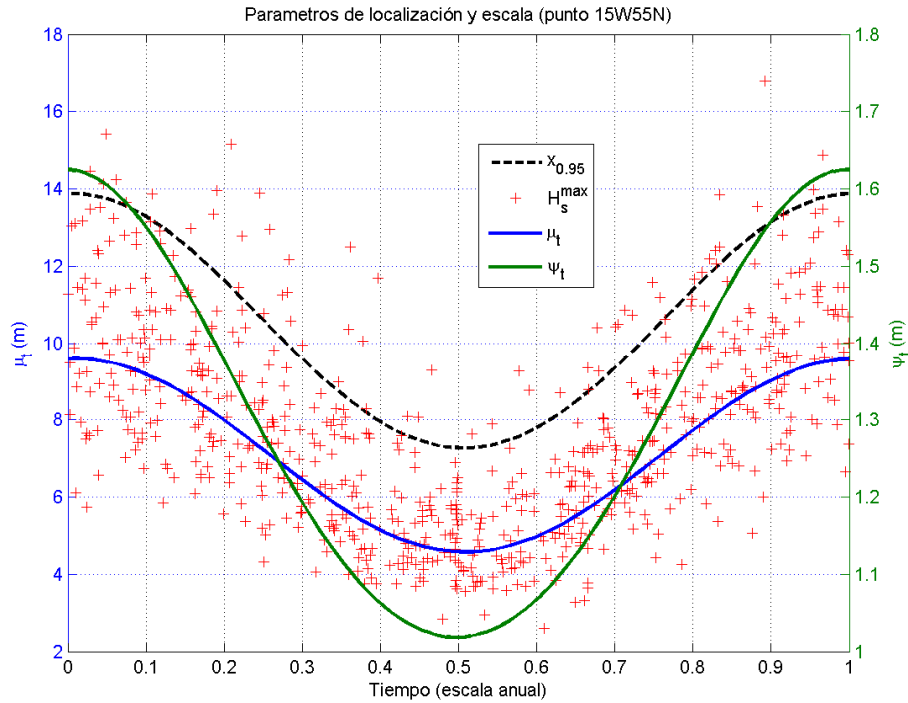


Figura 3.6 Máximos mensuales, parámetros de localización y escala y cuantil asociado a un período de retorno de 20 años

La variabilidad interanual queda definida en el modelo mediante la incorporación de cinco covariables en el parámetro de localización y dos en el parámetro de escala. Las variaciones que producen años con oleaje extremo más severo y años con oleaje extremo más suave están ligadas, principalmente, al primer modo de oscilación de las anomalías de los campos de presiones, el cual tiene una gran correlación con la NAO (0.7). Este modo es significativo en ambos parámetros, explicando 56 cm/unidad de PC del valor del parámetro de localización y 12 cm/unidad de PC del valor del parámetro de escala. En la Tabla 3.1 se pueden ver los valores que toman los estimadores de los parámetros, con sus errores estándar entre paréntesis.

β_0	β_1	β_2	β_{PC1}	β_{PC2}	β_{PC3}	β_{PC5}	β_{PC7}
7.08	2.50	0.13	0.56	0.36	-0.23	-0.065	0.15
(0.05)	(0.07)	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
α_0	α_1	α_2	α_{PC1}	α_{PC6}			
0.25	0.23	-0.003	0.12	-0.072			
(0.03)	(0.04)	(0.03)	(0.02)	(0.02)			
γ_0							
-0.08							
(0.02)							

Tabla 3.1 Valores de los estimadores (m) y sus errores estándar del modelo GEV mensual en el punto NA

Como puede verse en los dos paneles de la Figura 3.7 el ajuste del modelo es muy bueno. El panel de la izquierda muestra el diagnóstico gráfico de probabilidades y el de la derecha el de cuantiles. En ambos casos los puntos quedan muy próximos a la diagonal, lo que representa un buen ajuste del modelo a los datos de la muestra.

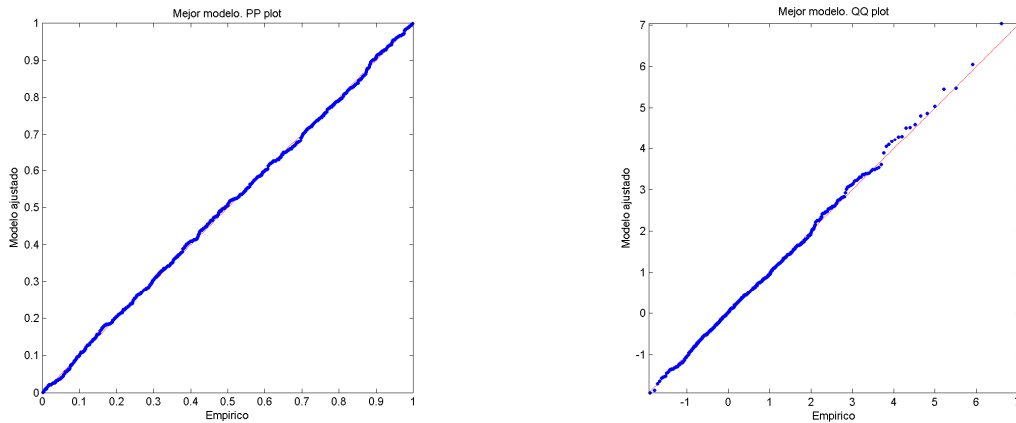


Figura 3.7 Diagnóstico gráfico PP y QQ plot

Como ya se ha expuesto, el objetivo del modelado no estacionario del oleaje extremo es poder inferir valores asociados a probabilidades de ocurrencia baja, que serán decisivos para conocer ciertos procesos de variabilidad climática natural. Con esta información se podrá gestionar y diseñar mejor actuaciones como la fase de ejecución de una obra marítima o el estudio de regeneración de una playa. En el apartado 1.2.1.3 se explicó como calcular el cuantil asociado a un período de retorno para el caso de máximos mensuales. En la Figura 3.8 se muestra la serie temporal del cuantil instantáneo de período de retorno 50 años.

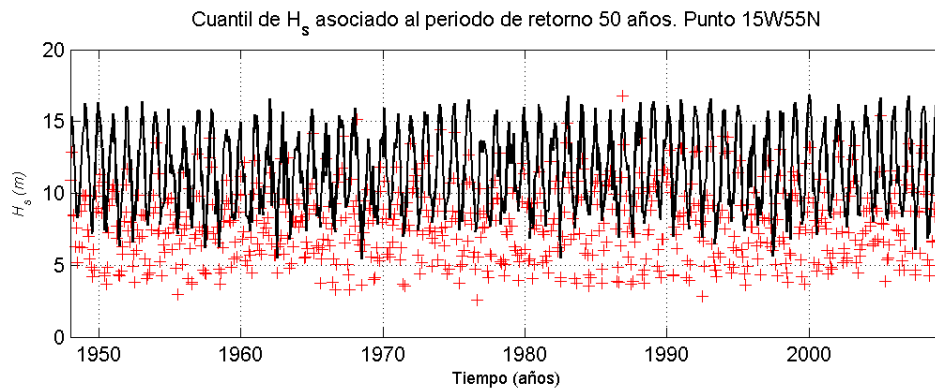


Figura 3.8 Cuantil instantáneo asociado a periodo de retorno 50 años (línea negra) y máximos mensuales (cruces rojas)

Como se puede observar hay una clara variabilidad, con años de oleaje extremo en invierno en los que se alcanzan los 17 m y otros más suaves con valores de H_{s50} en torno a 13 m. Este hecho está asociado a la variabilidad interanual que induce la NAO en esta zona.

3.4.3 Modelo POT con umbral variable

La aplicación del modelo POT para la obtención de la muestra de datos requiere fijar dos parámetros importantes. Por un lado, el valor del umbral y por otro, el tiempo mínimo de independencia entre temporales (Δt).

En numerosos estudios el valor del umbral se toma constante a lo largo del tiempo (Méndez et al., 2006, Caires et al., 2006), pero, puesto que el objetivo es estudiar la variabilidad intra e inter anual de los extremos del oleaje, en este caso, se propone un umbral variable a lo largo del año $u(t)$. Para construir este umbral variable se ha ajustado un modelo de regresión sencillo a los valores de H_s correspondientes al percentil del 97 % (Caires et al., 2006) de cada mes. El modelo de regresión ajustado tiene la siguiente expresión (en m):

$$u(t) = 7 + 2.46 \cos(2\pi t) + 0.37 \sin(2\pi t) \tag{3.12}$$

En cuanto a Δt , se ha optado por una duración de 3 días entre eventos para que éstos puedan ser considerados independientes. Tras varias pruebas con distintos valores de la independencia entre temporales basados en la bibliografía, se ha obtenido que este valor es el que mejor desagrupa las excedencias en la zona de estudio.

La muestra de datos obtenida cuenta con 691 valores que se representan en la Figura 3.9 (puntos grises) junto con los valores de $H_{97\%}$ mensuales (asteriscos) y el umbral (línea roja).

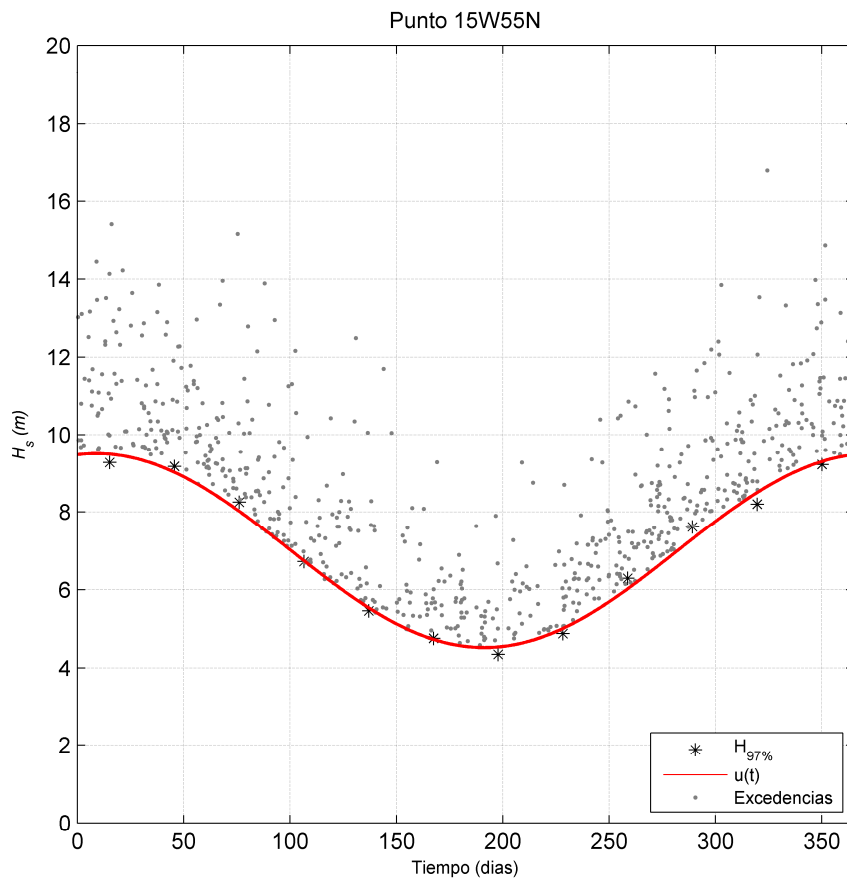


Figura 3.9 Umbral variable y excedencias de oleaje en el punto 15W55N

El modelo óptimo obtenido mediante el ajuste por máxima verosimilitud modela la estacionalidad a través del parámetro de localización, en el que introduce dos ondas (ciclo anual y semianual). Los parámetros de escala y forma permanecen constantes, tomando el parámetro de forma un valor negativo ($\hat{\xi} = -0.18$). Es decir, el modelo considera la dispersión y la cola de la distribución constantes a lo largo del tiempo. En la Figura 3.10 está representada la variación a lo largo del año de los parámetros de

localización y escala, así como las excedencias utilizadas para ajustar el modelo y la variación a lo largo del año del cuantil asociado al período de retorno 20 años.

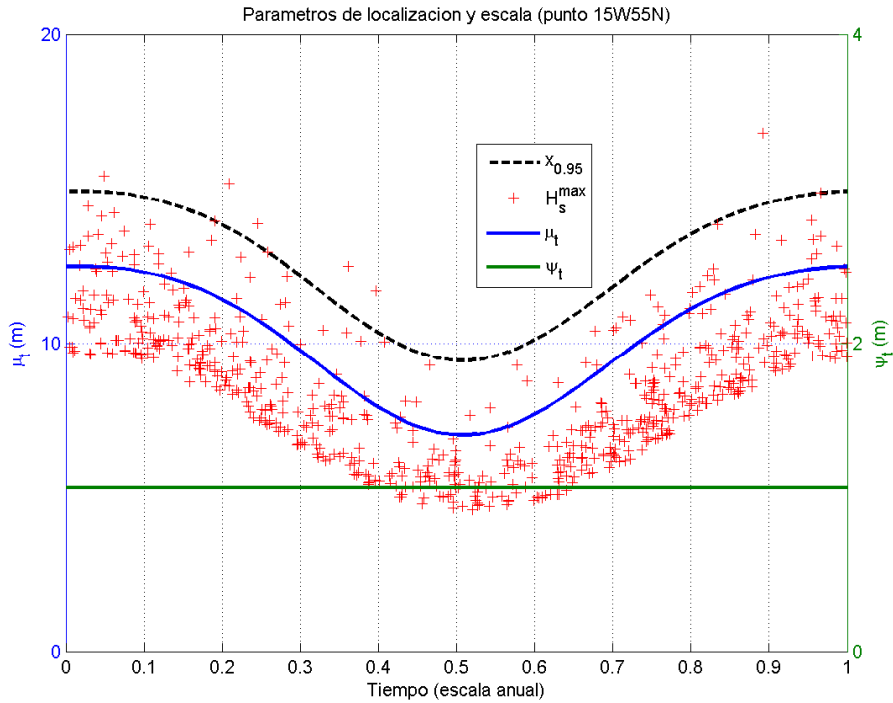


Figura 3.10 Excedencias sobre umbral variable, parámetros de localización y escala y cuantil asociado a un período de retorno de 20 años

En cuanto a la variabilidad interanual, asociada a la variación en las anomalías de presión, el modelo POT introduce en el parámetro de localización las mismas covariables que el modelo GEV mensual, mientras que en el parámetro de escala solo incluye la covariable correspondiente al modo 8 de oscilación. Esto significa que prácticamente toda la variabilidad interanual del oleaje extremo queda modelada a través del parámetro representativo de la media de la distribución, donde los dos primeros modos de oscilación tienen un gran peso, explicando hasta 1.14 y 0.84 m/unidad de PC de altura de ola (ambos modos presentan gran correlación con los conocidos patrones climáticos del Atlántico norte, NAO y EA). En la Tabla 3.2 se muestran los valores de los estimadores con sus errores estándar (en paréntesis).

β_0	β_1	β_2	β_{PC1}	β_{PC2}	β_{PC3}	β_{PC5}	β_{PC7}
10.12	2.73	0.17	1.14	0.84	-0.44	-0.18	0.26
(0.1)	(0.09)	(0.09)	(0.1)	(0.09)	(0.08)	(0.08)	(0.08)
α_0	α_1	α_2	α_{PC8}				
1.06	-0.37	-0.016	-0.06				
(0.06)	(0.09)	(0.09)	(0.02)				
γ_0							
-0.18							
(0.03)							

Tabla 3.2 Valores de los estimadores (m) y sus errores estándar del modelo Pareto-Poisson en el punto NA

En cuanto al diagnóstico del modelo, en este caso hay que comprobar la bondad de ajuste en el modelo de frecuencias y de intensidades, lo cual se lleva a cabo mediante los diagnósticos gráficos PP y QQ *plot* (Figura 3.11).

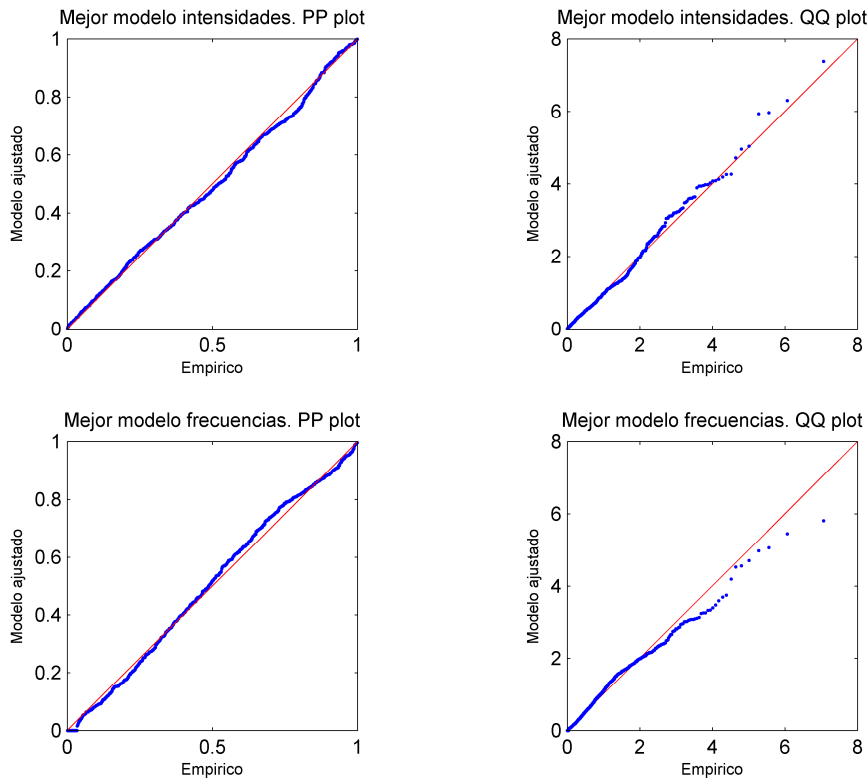


Figura 3.11 Diagnóstico gráfico PP y QQ *plot* para el modelo de intensidades y frecuencias

Como puede verse el punteado se ajusta bastante bien a la diagonal, lo que demuestra la buena pericia del modelo ajustado.

Por último, en la Figura 3.12 se muestra la serie temporal del cuantil asociado al período de retorno 50 años y las excedencias utilizadas en el ajuste del modelo. Como se puede ver la serie reproduce las variaciones estacionales e interanuales, alcanzándose altura de ola de hasta 17.5 m en los años mas severos.

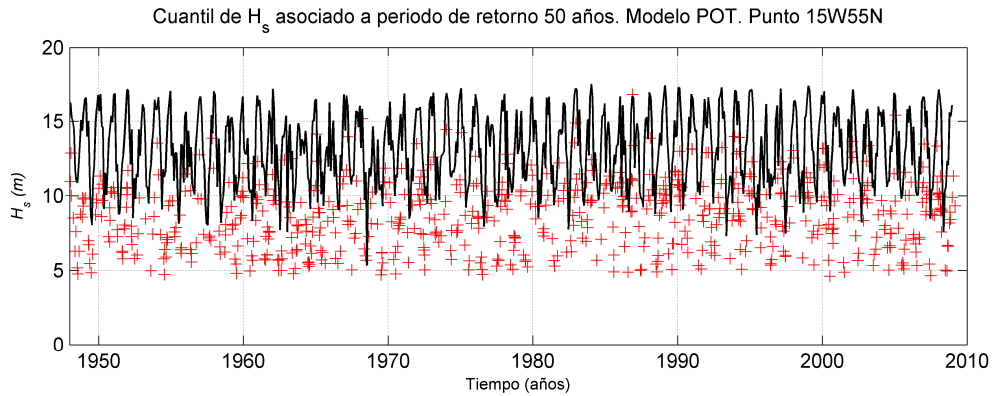


Figura 3.12 Cuantil asociado a período de retorno 50 años (línea negra) y excedencias de altura de la (cruces rojas)

3.4.4 Comparación entre los modelos GEV mensual y POT con umbral variable

La primera discordancia entre ambos modelos está en la población de datos usada para el ajuste de cada uno de ellos. El modelo GEV basado en máximos mensuales utiliza 12 valores extremos al año (736 datos). Sin embargo, el método POT con umbral variable no exige un número fijo de datos al año, sino la superación de una determinada excedencia en cada instante del año, con lo que para esta aplicación se obtienen 691 datos. Esto da lugar a poblaciones de datos distintas, como puede observarse en la Figura 3.13.

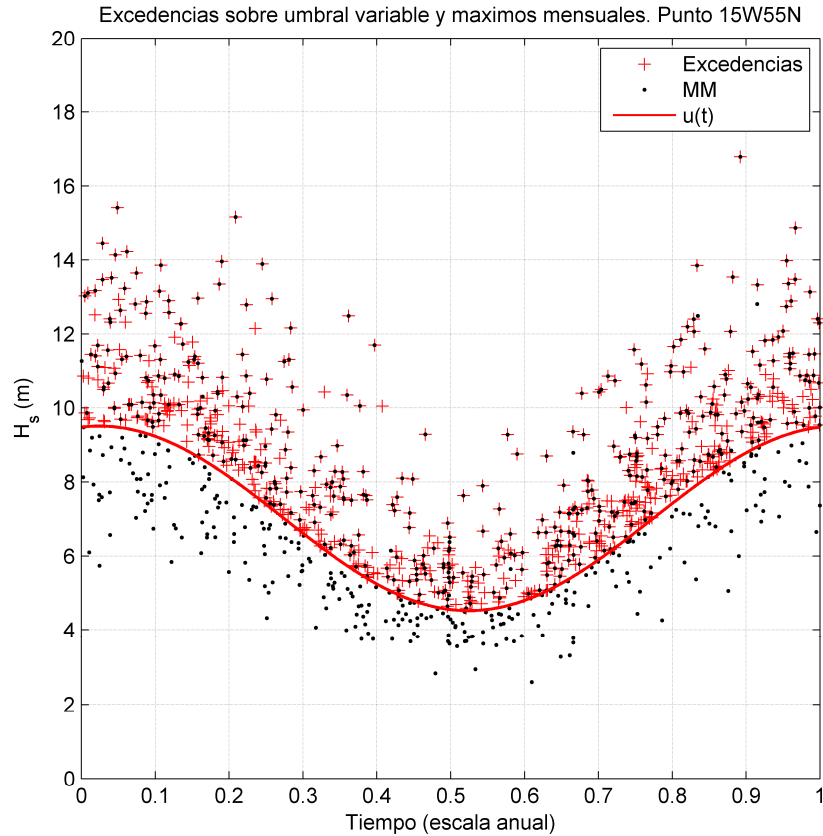


Figura 3.13 Poblaciones de datos de los modelos GEV mensual y POT con umbral variable

El método de excedencias genera una población con mayor densidad de datos en los meses de invierno que en los meses de verano, donde el número de temporales es mucho menor. Sin embargo, con el método de máximos mensuales la población es más homogénea a lo largo del año, y, por lo tanto, representa mejor la variación estacional. Según el tipo de comportamiento que se quiera modelar un método será más apropiado que otro.

La segunda diferencia a tener en cuenta radica en la unidad temporal de estudio. En el modelo GEV mensual la unidad temporal es el mes y por tanto los valores obtenidos de los parámetros y cuantiles son mensuales. Sin embargo, el modelo POT trabaja con escala anual y, por lo tanto, los resultados están referidos al año.

Para poder comparar los valores de los cuantiles inferidos por cada modelo, hay que referirlos a la misma escala de tiempo, es decir, o mensual o anual. En este caso se ha optado por transformar la escala anual del modelo POT a la escala mensual del modelo

GEV. Para ello es necesario afectar a los parámetros de un factor de escala k , mediante un razonamiento análogo al expuesto en el apartado 3.2, de tal manera que

$$\begin{aligned}\mu_m(t) &= \mu(t) + \frac{\psi(t)}{\xi(t)} \left[\left(\frac{k_y}{k_m} \right)^{\xi(t)} - 1 \right] \\ \psi_m(t) &= \psi(t) \left(\frac{k_y}{k_m} \right)^{\xi(t)} \\ \xi_m(t) &= \xi(t)\end{aligned}\tag{3.13}$$

O en el caso de $\xi = 0$:

$$\begin{aligned}\mu_m(t) &= \mu(t) + \psi(t) \log \left(\frac{k_y}{k_m} \right) \\ \psi_m(t) &= \psi(t)\end{aligned}\tag{3.14}$$

donde $1/k_y = 1$ año y $1/k_m$ es la longitud del bloque de tiempo, en este caso $1/12$. $\mu(t)$, $\psi(t)$ y $\xi(t)$ son los parámetros de la distribución a escala anual y $\mu_m(t)$, $\psi_m(t)$ y $\xi_m(t)$ son los parámetros de la distribución a escala mensual.

Con los parámetros del modelo POT expresados en escala mensual se podrán obtener los cuantiles asociados a períodos de retorno directamente comparables con los obtenidos mediante el modelo GEV mensual. En la Figura 3.14 se muestra la serie temporal del cuantil asociado a 50 años de período de retorno obtenido mediante el modelo GEV mensual (línea roja) y mediante el modelo POT (línea azul). Mediante puntos y cruces se representan los valores de las muestras de datos utilizadas en el ajuste de los modelos.

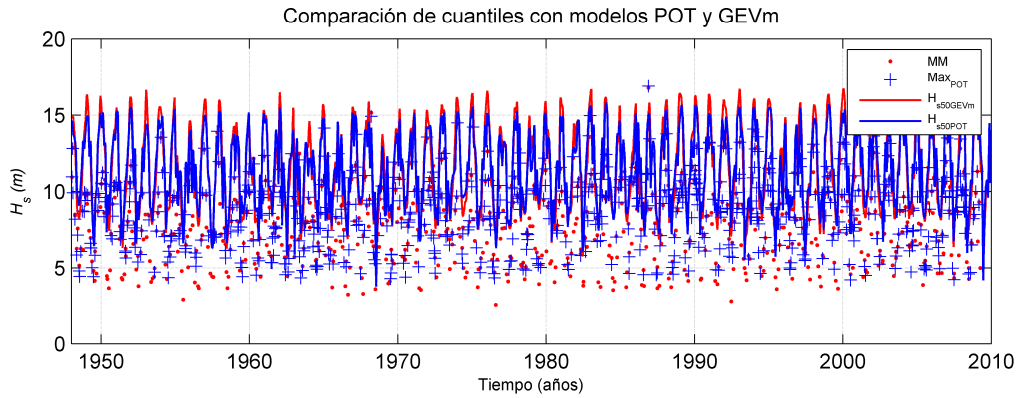


Figura 3.14 Comparación de cuantiles de período de retorno 50 años obtenidos con el modelo GEV mensual y POT

Como se puede observar, ambos modelos proporcionan resultados similares, pese a proceder de poblaciones distintas. En este punto se puede ver que el modelo GEV mensual proporciona valores un poco mayores de $H_{s,50}$, lo cual es debido a la forma de la cola de la distribución. El modelo GEV mensual tiene un valor del parámetro de forma de $\hat{\xi} = -0.08$, comportamiento prácticamente Gumbel, sin embargo, el modelo POT encuentra que la distribución de su población de datos tiene una cola acotada tipo Weibull ($\hat{\xi} = -0.18$) y, por lo tanto, hay una limitación en las alturas de ola más grandes. Para ver más clara la diferencia entre ambas series se muestra, en la Figura 3.15, un zoom del gráfico anterior para el período 1960-1990.

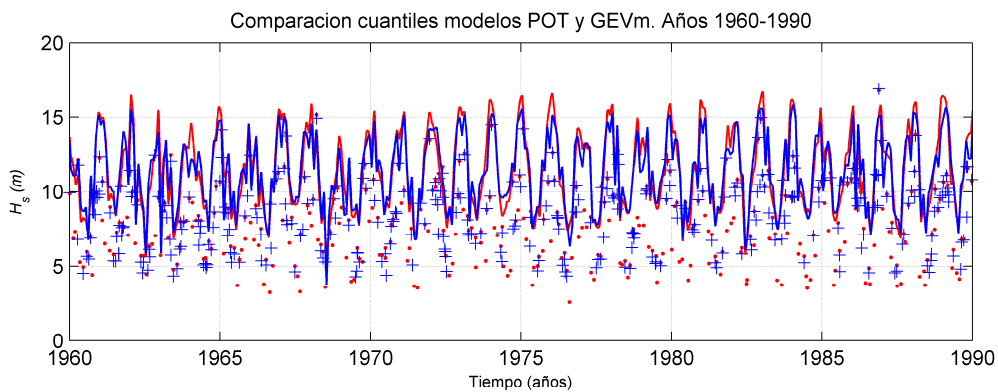


Figura 3.15 Comparación de cuantiles de período de retorno 50 años obtenidos con el modelo GEV mensual y POT. Años 1960-1990

3.5 Comparación entre modelos no estacionarios GEV mensual y GEV semanal

El modelo de extremos elegido para los estudios de esta tesis es el derivado de la familia GEV con máximos en un bloque de tiempo. Dependiendo de la variabilidad temporal que se quiera modelar será más apropiado elegir un bloque de tiempo u otro. En el apartado anterior se aplicó un modelo GEV mensual para caracterizar el régimen extremal de oleaje en un punto del Atlántico norte obteniendo buenos resultados. Sin embargo, en este apartado se plantea la posibilidad de utilizar un modelo GEV con bloque de tiempo semanal. El objetivo es comparar el modelo GEV mensual y GEV semanal aplicando ambos modelos en un punto de Cádiz (Figura 3.16).

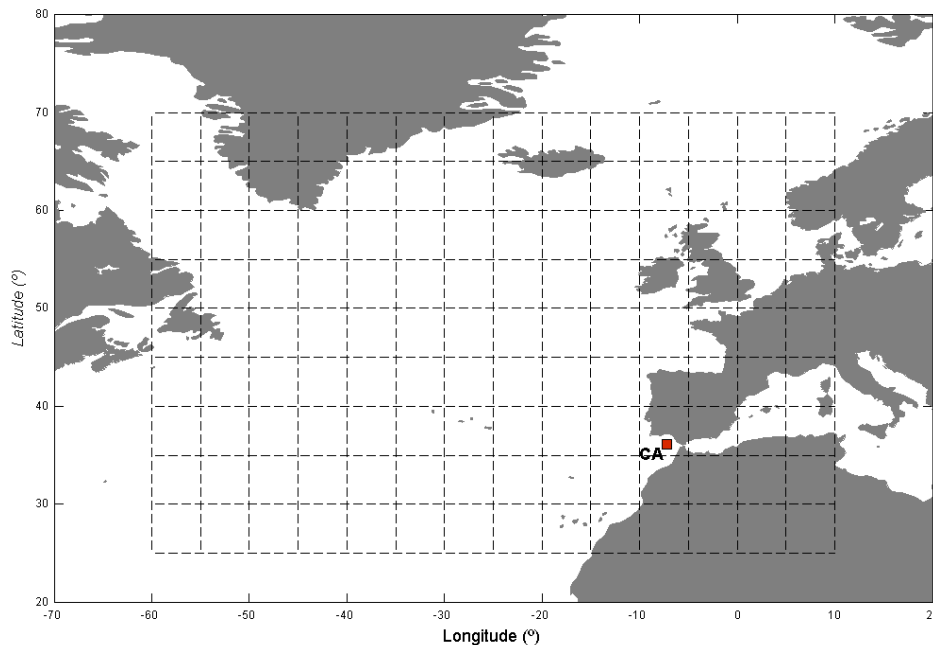


Figura 3.16 Localización del punto CA en el Golfo de Cádiz

El punto de estudio se encuentra situado en el Golfo de Cádiz, una zona con un clima marítimo complejo debido a su proximidad al estrecho de Gibraltar y al Cabo de San Vicente, en Portugal. Por un lado recibe mares energéticos generados en el océano Atlántico, pero por otro, también es zona de generación de oleaje por viento. El régimen de vientos de la zona muestra dos familias claras: ponientes, provenientes del oeste y levantes, provenientes del este, que son los que dan lugar a oleajes de corto período. La serie de oleaje a estudiar proviene de la base de datos de reanálisis GOW con registro horario de 60 años (1948-2008). Es importante señalar que este reanálisis está cerrado

en el contorno, en la zona del Estrecho de Gibraltar, no modelando de forma adecuada los oleajes de levante, por lo que los resultados obtenidos se deben tomar en cuenta tan sólo desde el punto de vista matemático estadístico.

La aplicación de herramientas de estadística descriptiva a los datos permite tener una idea inicial del comportamiento de los oleajes extremos de la zona. En la Figura 3.17 se muestra la serie temporal de altura de ola significativa en el punto de estudio (puntos grises) y los máximos anuales (círculos negros). Se puede ver una variación invierno-verano dentro del año, y también se aprecia muy bien una importante variabilidad interanual, registrándose alturas de ola de más de 7 m en los años 1966, 1973 y 1979, mientras que en los años 1949 y 1984 no se superaron los 3 m de altura.

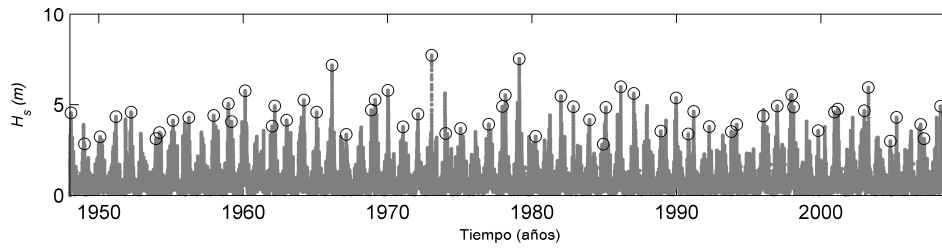


Figura 3.17 Serie temporal de H_s (puntos grises) y máximos anuales (círculos negros)

Entrando en la escala del año y aplicando el gráfico *boxplot* a los máximos mensuales (Figura 3.18) se puede ver como la evolución de los valores máximos no es homogénea a lo largo del año, mostrando una onda asimétrica. Los valores más suaves se registran en los meses de verano, comenzando éstos en Mayo-Junio y prolongándose hasta Septiembre-Octubre. El paso de verano a invierno es brusco, alcanzándose las mayores alturas de ola en Diciembre-Enero-Febrero, mientras que la primavera es más suave y larga.

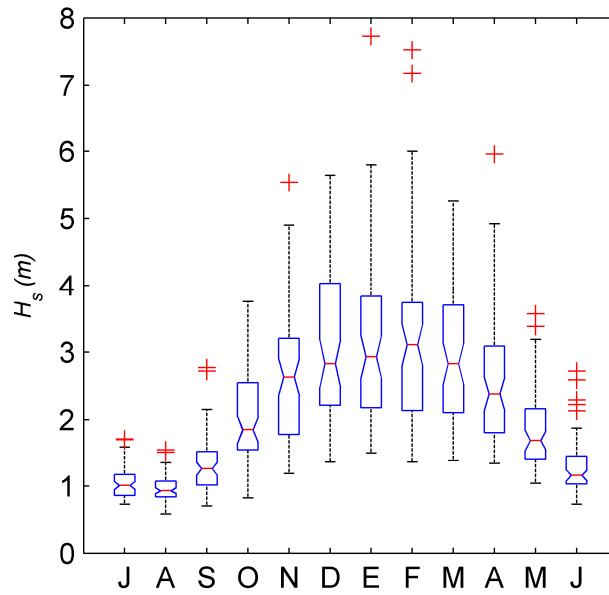


Figura 3.18 *Boxplot* de la H_s máxima mensual en el punto 15W55N

Por último, el oleaje en generación de la zona tiene una escala de tiempo más pequeña que el oleaje formado en zonas alejadas del océano. Por esta razón, se propone utilizar el modelo de máximos semanales para intentar modelar mejor esta variabilidad temporal asociada al oleaje.

3.5.1 Parametrización de los modelos

En ambos modelos se va a intentar modelar las variaciones temporales de la misma forma. La variabilidad estacional se va a modelar mediante funciones armónicas en cada uno de los parámetros de la GEV y la variabilidad a lo largo de los años se modelará a través de las anomalías de los campos de presiones a nivel del mar. Es en este punto donde difieren ambos modelos.

Se asume que la variabilidad interanual está vinculada a la fluctuación de los campos de presiones y esta información se incorpora en los parámetros de localización y escala de la GEV mediante covariables. El modelo GEV mensual trabaja con la escala de tiempo mensual y, por tanto, la información de las covariables se introduce a través de los 10 primeros modos principales de las anomalías de los campos de presión medios mensuales. Sin embargo, el modelo GEV semanal utiliza una unidad temporal menor, por lo que introduce más variabilidad a la muestra de datos, siendo necesario incluir

esta información en las covariables mediante los modos principales de los campos de anomalías de presión medias semanales.

La parametrización que se propone para ambos modelos es la siguiente:

$$\begin{aligned}
 \mu(t) &= \beta_0 + \sum_{i=1}^{P_\mu} [\beta_{2i-1} \cos(2i\pi t) + \beta_{2i} \sin(2i\pi t)] + \sum_{j=1}^{C_\mu} \beta_{PC_j} PC_j(t) \\
 \psi(t) &= \alpha_0 + \sum_{i=1}^{P_\psi} [\alpha_{2i-1} \cos(2i\pi t) + \alpha_{2i} \sin(2i\pi t)] + \sum_{j=1}^{C_\psi} \alpha_{PC_j} PC_j(t) \\
 \xi(t) &= \gamma_0 + \sum_{i=1}^{P_\xi} [\gamma_{2i-1} \cos(2i\pi t) + \gamma_{2i} \sin(2i\pi t)]
 \end{aligned} \tag{3.15}$$

siendo las PCs del modelo GEV mensual las correspondientes a las anomalías de presión medias mensuales y las del modelo GEV semanal las correspondientes a las anomalías de presión medias semanales.

3.5.2 Modelo GEV mensual

Mediante la selección de máximos mensuales la muestra de datos cuenta con 732 valores para ajustar el mejor modelo GEV mensual. El gráfico *boxplot* de la Figura 3.17 ya proporciona una información de la compleja distribución de los datos a lo largo del año.

Como se ha descrito en el apartado anterior, el Golfo de Cádiz se caracteriza por un complejo clima marítimo con dos familias principales de oleaje provenientes del oeste (ponientes) y del este (levantes). Sin embargo, la representación de la población de datos utilizada en el modelo GEV mensual en una rosa de oleaje (Figura 3.19), hace ver que los eventos escogidos son principalmente los provenientes del tercer y cuarto cuadrante (ponientes), 674 eventos frente a 58 eventos de levante (debido a la limitación que supone el contorno en el Estrecho de Gibraltar). Atendiendo al criterio de máximos mensuales, la selección de estos eventos, que corresponden a los más energéticos, puede ser adecuada, pero, sin embargo, se está obviando una variabilidad importante generada por los oleajes de levante.

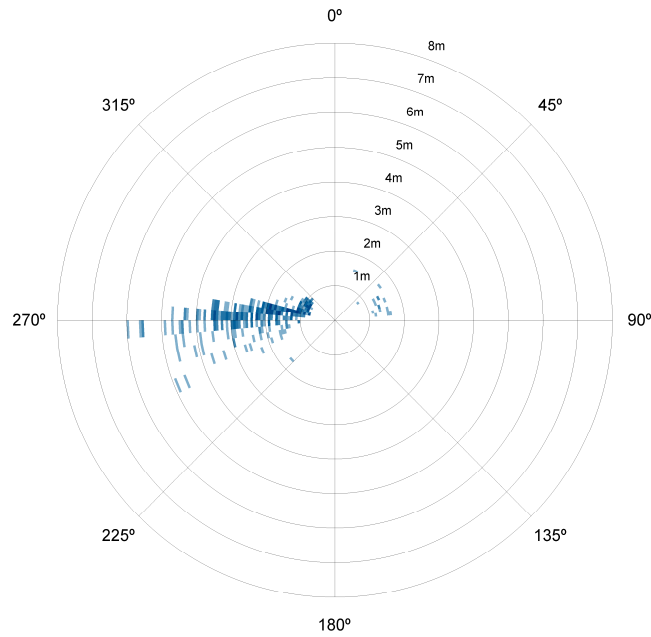


Figura 3.19 Rosa de oleaje de los eventos extremos de máximos mensuales. Cádiz. Este reanálisis de oleaje está cerrado en el contorno Este (Estrecho de Gibraltar)

Con esta información el modelo representa la variación invierno-verano incorporando una onda en el parámetro de localización y 4 ondas en el parámetro de escala. El parámetro de forma permanece constante a lo largo del tiempo y toma el valor $\hat{\xi} = 0.07$ que muestra un comportamiento prácticamente Gumbel de máximos. Este comportamiento de la cola se puede explicar a través de la variedad de eventos extremos, con oleajes de más de 7 m y otros que no llegan al metro de altura. La combinación de ondas encontrada para los parámetros de localización y escala intenta reproducir la variabilidad intra anual explicada en el apartado anterior. En la Figura 3.20 se muestra la evolución a lo largo del año de los parámetros de localización y escala y del cuantil asociado a 20 años de período de retorno.

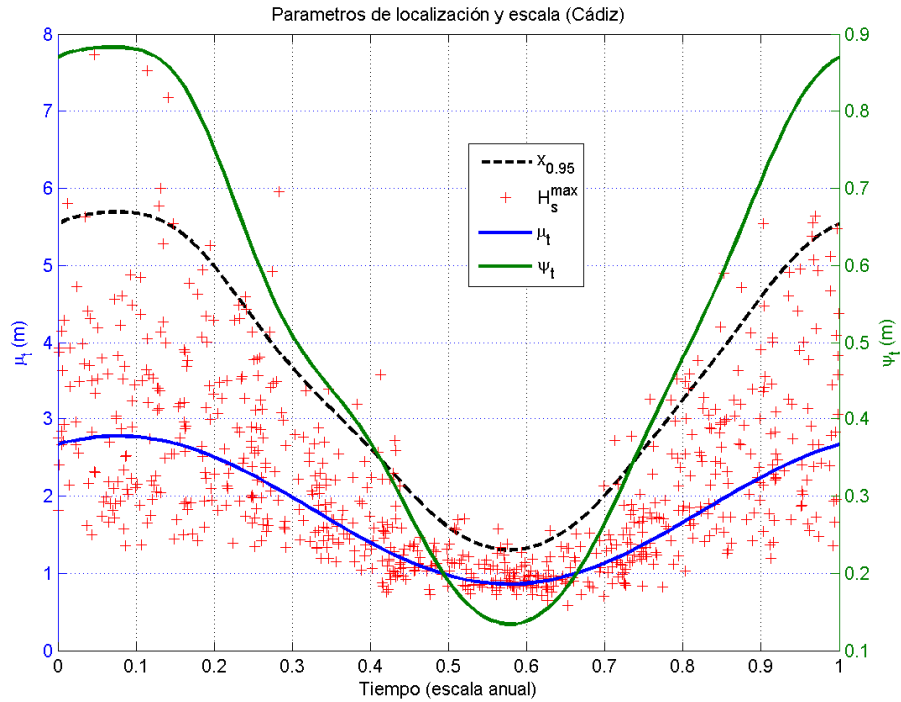


Figura 3.20 Máximos mensuales, parámetros de localización y escala y cuantil asociado a un período de retorno de 20 años

La variación de las alturas de ola a lo largo de los años queda caracterizada a través del parámetro de localización. El modelo encuentra que los modos principales de oscilación de los campos de las anomalías medias de presión más influyentes en el oleaje extremo son los 3 primeros y el séptimo. Sin embargo, ninguno de ellos aporta una gran contribución a la altura de ola extrema. La mayor aportación proviene del primer modo que explica -0.14 m/unidad de PC en el parámetro de localización, lo que en septiembre de 1998 (mes en que el valor de la PCI tomó el valor -3.37) fue una contribución de 47 cm. El modelo no encuentra significativo el aporte de las PCs en el parámetro de escala, lo que quiere decir que la muestra de datos no presenta una dispersión inter anual muy grande. En la Tabla 3.3 se muestran los valores que toman los parámetros y sus errores estándar.

β_0	β_1	β_2	β_{PC1}	β_{PC2}	β_{PC3}	β_{PC7}		
1.81	0.84	0.44	-0.14	0.04	0.03	-0.05		
(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)		
α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
-0.82	0.77	0.035	-0.08	-0.17	-0.008	0.09	0.01	-0.05
(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
γ_0								
0.07								
(0.03)								

Tabla 3.3 Valores de los estimadores (m) y sus errores estándar del modelo GEV mensual en Cádiz

El diagnóstico gráfico del modelo (Figura 3.21) muestra un buen ajuste de los datos, salvo para los cuantiles elevados en los que el modelo ajustado los infravalora.

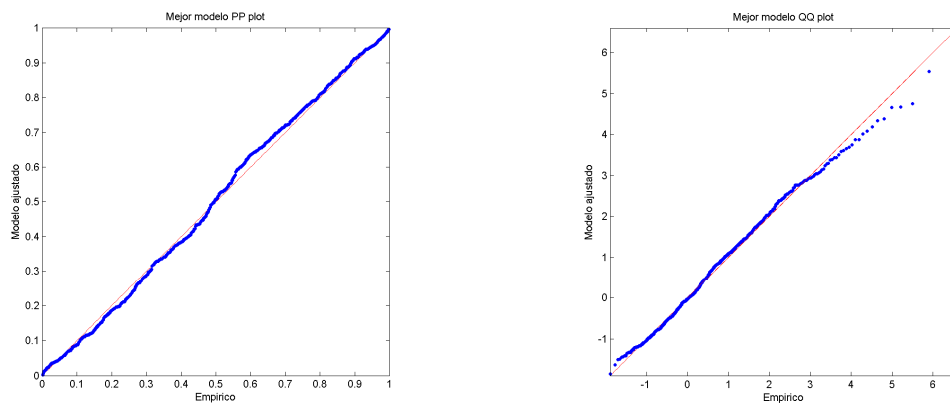


Figura 3.21 Diagnóstico gráfico PP y QQ plot

Por último, la Figura 3.22 muestra la evolución temporal del cuantil asociado al período de retorno 50 años junto a la población de datos. En el gráfico se puede ver como la población de datos se encuentra mayoritariamente alrededor de lo 3 m de altura; hay datos repartidos a lo largo de los años entre los 3 y 6 m y solo 3 eventos de toda la serie superan los 7 m. Ante esta población de datos, con gran variabilidad dentro del año, el modelo intenta adaptarse reproduciendo una gran variabilidad estacional. Sin embargo, la población no muestra una variabilidad importante a lo largo de los años, y pese a que el modelo reproduce años con oleajes más severos y más suaves, la diferencia entre éstos no es muy acusada.

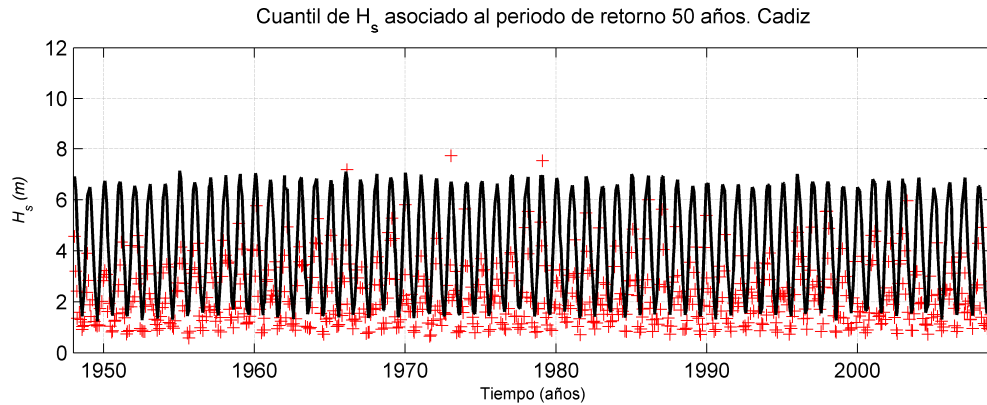


Figura 3.22 Cuantil instantáneo asociado a período de retorno 50 años (línea negra) y máximos mensuales (cruces rojas)

3.5.3 Modelo GEV semanal

La población de datos obtenida a partir del máximo mensual cuenta con 3183 valores. Con la disminución de escala temporal se intenta describir mejor la variabilidad del clima marítimo extremo de la zona, pero no hay que perder de vista que al tomar un número elevado de eventos, puede que algunos no cumplan las hipótesis de la teoría de extremos y no pertenezcan a la cola de la distribución. Además, analizando la muestra de datos, se puede ver, que aunque ha aumentado el número de eventos de levante, ha aumentado en igual medida el número de oleajes de poniente.

La escala semanal permite visualizar mejor el campo de presiones que genera el evento extremo. En la Figura 3.23 se muestran los campos de presiones semanales que generaron un temporal de poniente (panel de la izquierda) y uno de levante (panel de la derecha). Como se puede ver en el panel de la izquierda, el campo de presiones se encuentra en situación NAO positiva con el centro de bajas presiones desplazado hacia el oeste y el centro de altas un poco más al sur de las Azores. Esta situación provocó la generación de un oleaje al sur de Groenlandia que llegó al Golfo de Cádiz con una altura de 7.73 m. El panel de la derecha el campo de presiones muestra un anticiclón sobre la Península Ibérica que, en el golfo de Cádiz, genera vientos procedentes del este. Estos vientos proceden de tierra, dando lugar a un oleaje en generación que alcanza los 2.14 m de altura.

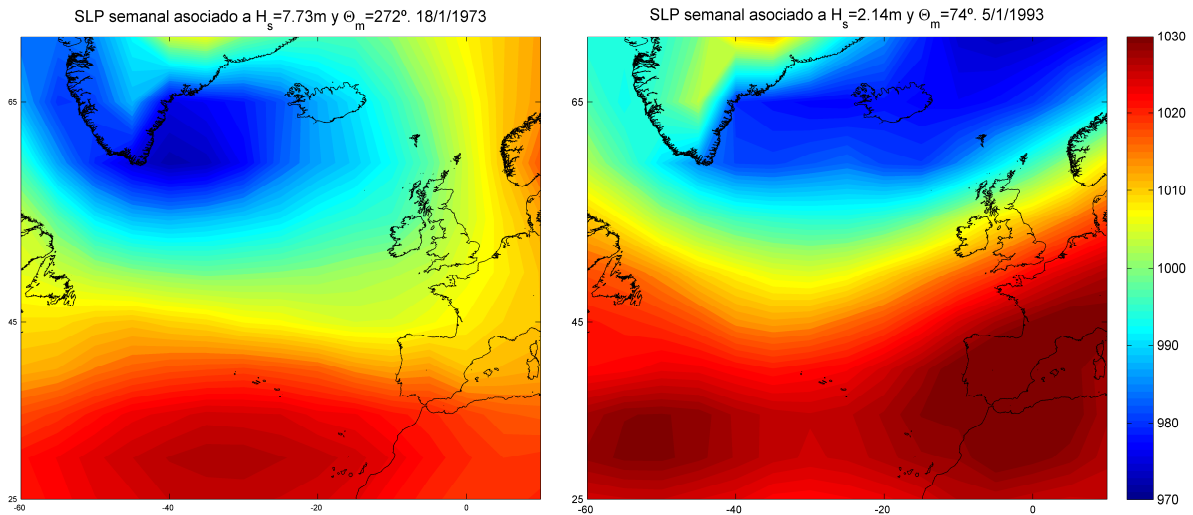


Figura 3.23 Ejemplos de campos de presiones medias semanales

Al tener más número de valores en la muestra y, por lo tanto, mayor variabilidad, el modelo obtiene una combinación de parámetros más compleja que en el modelo anterior. En este caso el mejor modelo incorpora 3 ondas en el parámetro de localización, 3 ondas en el parámetro de escala y 2 en el parámetro de forma. Mediante esta combinación de funciones armónicas trata de reproducir la variabilidad de invierno a verano (Figura 3.24).

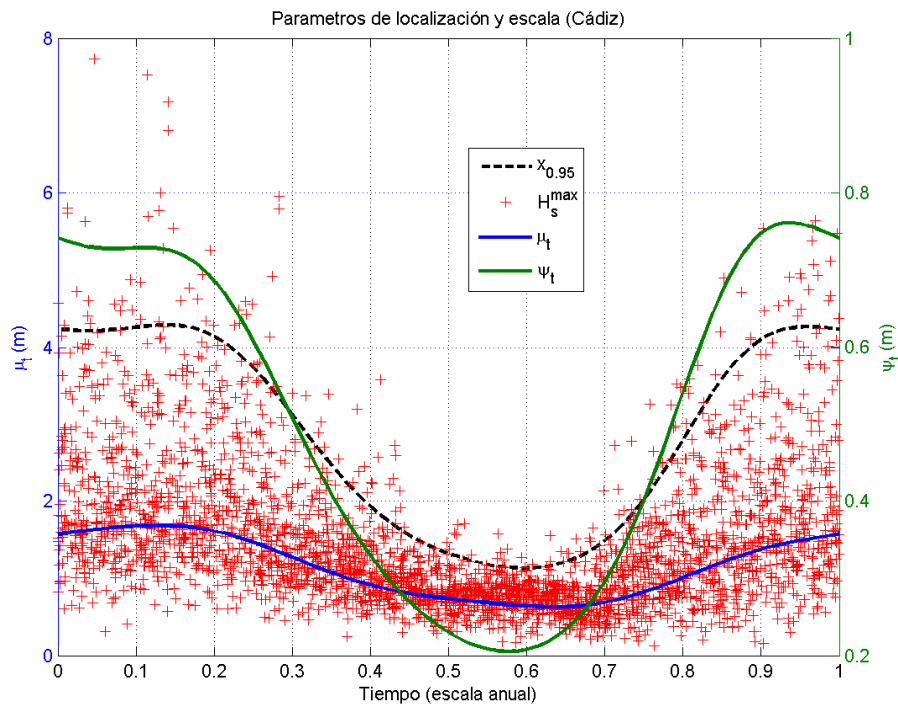


Figura 3.24 Máximos semanales, parámetros de localización y escala y cuantil asociado a un período de retorno de 20 años

En este caso el parámetro de forma es variable a lo largo del año (Figura 3.25), tomando valores positivos en los meses de invierno y negativos (Weibull) en los meses de verano. En los meses de invierno la muestra cuenta con valores de altura de ola elevados, debido a los oleajes del noroeste, pero también cuenta con valores más pequeños debidos a ciertos oleajes del este. Esta variabilidad en los valores queda representada por una cola tendida de tipo Fréchet. Sin embargo, en los meses de verano, el oleaje proviene fundamentalmente del Este, generado por vientos de levante. Como es un oleaje en generación con una limitación por *fetch* las alturas de ola no superan cierto umbral, por lo que el comportamiento de la cola de la distribución es Weibull, es decir, cola acotada.

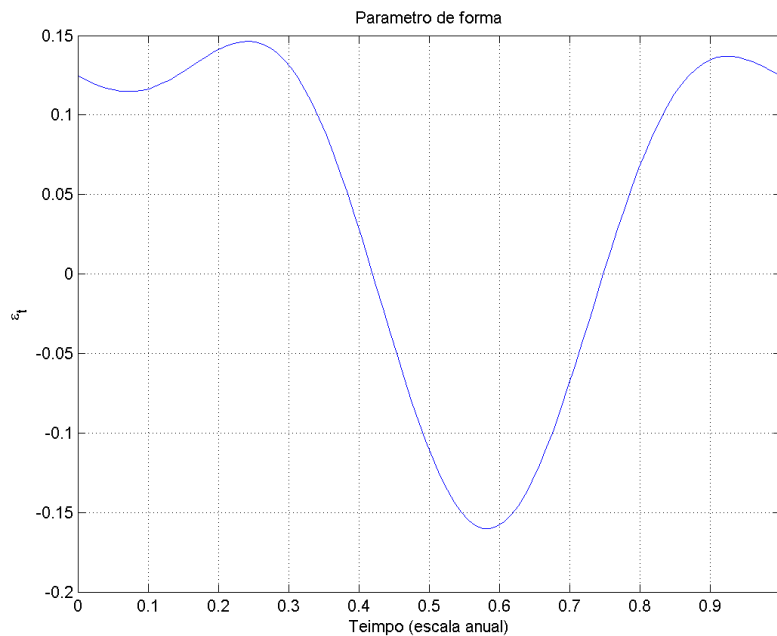


Figura 3.25 Variación anual del parámetro de forma

La variabilidad interanual que presenta la muestra es modelada a través de los parámetros de localización y escala. El modelo encuentra 9 componentes principales significativas en ambos parámetros, en el parámetro de localización los 9 primeros y en el parámetro de escala todos salvo el séptimo. Al igual que en el modelo GEV mensual ninguna de las covariables produce un aporte muy importante en los parámetros. Sin embargo, la cantidad de datos de la muestra y la variabilidad que éstos aportan hace que el modelo encuentre una combinación de parámetros más compleja para reproducir esta información. En la Tabla 3.4 se muestran los valores que toman los parámetros y sus errores estándar.

β_0	β_1	β_2	β_3	β_4	β_5	β_6		
1.14	0.44	0.29	0.005	0.008	-0.03	-0.03		
(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)		
β_{PC1}	β_{PC2}	β_{PC3}	β_{PC4}	β_{PC5}	β_{PC6}	β_{PC7}	β_{PC8}	β_{PC9}
-0.12	0.1	-0.12	0.09	0.06	-0.06	-0.02	-0.07	0.02
(0.008)	(0.01)	(0.009)	(0.008)	(0.008)	(0.009)	(0.007)	(0.007)	(0.007)
α_0	α_1	α_2	α_3	α_4	α_5	α_6		
-0.79	0.62	0.19	-0.08	-0.12	-0.04	-0.009		
(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)		
α_{PC1}	α_{PC2}	α_{PC3}	α_{PC4}	α_{PC5}	α_{PC6}	α_{PC8}	α_{PC9}	α_{PC10}
-0.15	0.14	-0.03	0.03	0.1	-0.1	-0.08	0.02	0.05
(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
γ_0	γ_1	γ_2	γ_3	γ_4				
0.04	0.11	0.07	-0.03	-0.05				
(0.01)	(0.01)	(0.01)	(0.01)	(0.01)				

Tabla 3.4 Valores de los estimadores (m) y sus errores estándar del modelo GEV semanal en Cádiz

En la Figura 3.26 el diagnóstico gráfico muestra un buen ajuste del modelo.

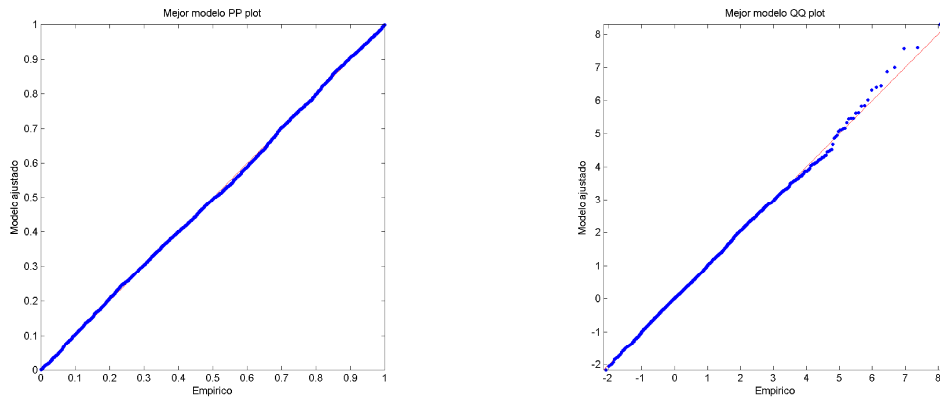


Figura 3.26 Diagnóstico gráfico PP y QQ plot

A partir de los valores de los parámetros obtenidos por el modelo se puede inferir en el cuantil asociado a 50 años de período de retorno a escala semanal (Figura 3.27).

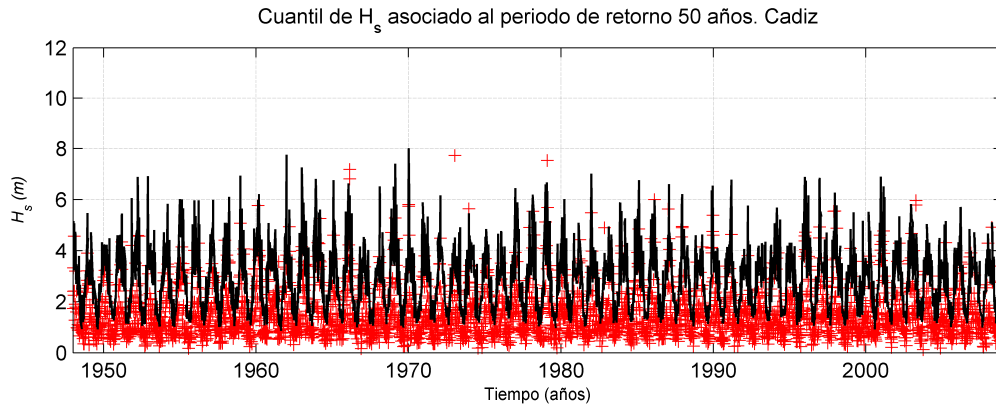


Figura 3.27 Cuantil instantáneo asociado a período de retorno 50 años (línea negra) y máximos semanales (cruces rojas)

En este caso se puede ver como la variabilidad interanual del cuantil es bastante acusada, con años de valor del cuantil de hasta 7 m y otros que se quedan en torno a los 4 m. Esta variación es debida a la combinación de covariables en los parámetros de localización y escala.

3.5.4 Comparación entre los modelos GEV mensual y GEV semanal

La diferencia entre ambos modelos es la escala de tiempo escogida para la selección de máximos. El hecho de que una escala sea mensual y otra semanal da lugar a poblaciones de datos distintas. La población de datos del modelo GEV mensual cuenta con 732 valores y la del modelo GEV semanal tiene un total de 3183. La población del modelo GEV semanal tiene los mismos valores que la del GEV mensual y además una serie de extremos más que incorporan más información acerca del clima marítimo extremo de la zona pero también más dispersión a la muestra. Además hay que tener en cuenta que los eventos seleccionados cumplan las hipótesis de la teoría de extremos.

La optimización de ambos modelos se ha llevado a cabo mediante el mismo procedimiento y con el criterio de información de Akaike. El modelo GEV mensual modela la estacionalidad mediante 5 ondas (una en localización y 4 en escala), mientras que el GEV semanal necesita 8 ondas para reproducir la variabilidad invierno-verano que presenta la muestra de datos, lo que representa un número elevado de parámetros. En cuanto a las variaciones a mayor escala, la muestra de máximos semanales parece que proporciona una información más completa que permite al modelo reproducir años

con mayores alturas de ola frente a otros mucho más suaves. Sin embargo el modelo GEV mensual no tiene suficiente información para reproducir este comportamiento.

Las series temporales de los cuantiles proporcionados por ambos modelos se pueden comparar siempre que se haga a la misma escala. En este caso se va a reescalar el cuantil semanal a la escala mensual haciendo uso del razonamiento presentado en el apartado 1.2.1.4. La comparación se muestra en la Figura 3.28.

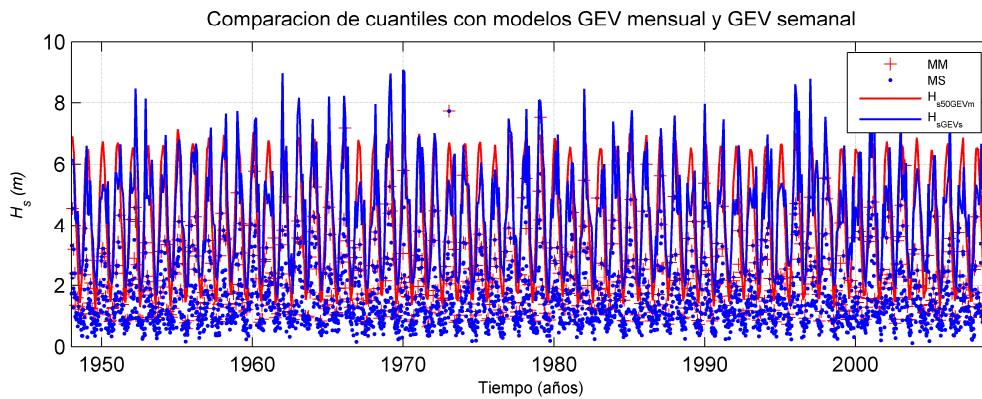


Figura 3.28 Comparación de cuantiles de periodo de retorno 50 años obtenidos con el modelo GEV mensual y GEV semanal

Como puede verse, al transformar el cuantil del modelo GEV semanal a escala mensual, la tendencia es similar, pero la variabilidad interanual es mucho más acentuada en el modelo semanal. La Figura 3.29 muestra un zoom para los años 1960-1990 donde se ve como el modelo GEV semanal se adapta mejor a los valores más extremos de altura de ola.

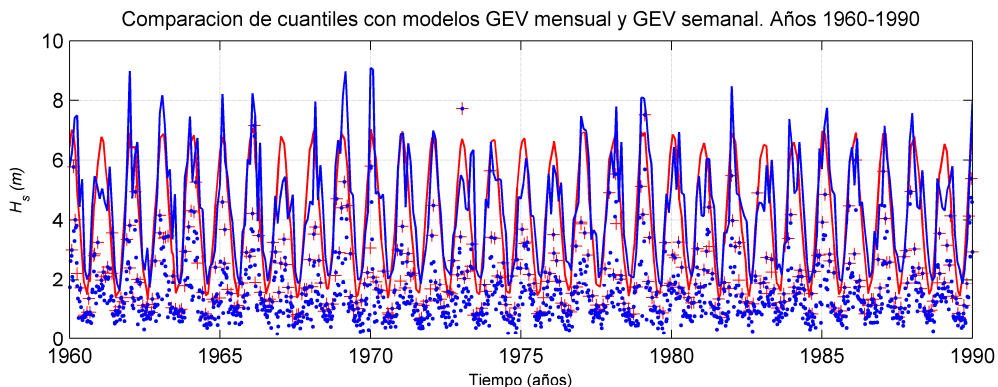


Figura 3.29 Comparación de cuantiles de periodo de retorno 50 años obtenidos con el modelo GEV mensual y GEV semanal. Años 1960-1990

Esta comparativa evidencia pros y contras de ambos modelos. Por un lado, el modelo GEV semanal trabaja con una muestra de datos más amplia y que representa mucho mejor la variabilidad del oleaje de la zona. Sin embargo, hay que ser cauto en su utilización puesto que algunos de los datos pueden no cumplir las hipótesis de la teoría de extremos.

Se puede concluir que es adecuado utilizar el modelo GEV semanal en zonas con clima marítimo complejo donde los estados bimodales del oleaje son importantes, como por ejemplo el Golfo de Cádiz o el Mar de Alborán.

Por otro lado, se ha podido apreciar como el modelo basado en datos semanales es capaz de encontrar mayor variabilidad en los extremos de oleaje. Este hecho se debe sin duda a que la escala semanal se aproxima más a la escala temporal de la generación y propagación de los eventos extremos de oleaje a lo largo del océano Atlántico (entre 1 y 3 días). Estas conclusiones han conducido al desarrollo del Capítulo 6, en el cual se va a definir el predictor de eventos extremos a una escala temporal de 3 días (campo de presiones medio de 3 días en la zona de generación y propagación del temporal de oleaje).

3.6 Recomendaciones de uso de los distintos modelos de extremos

Como se ha visto hasta ahora en el estudio de los valores extremos de oleaje no hay un único modelo de extremos a aplicar. Desde los modelos estacionarios de extremos, que son los utilizados tradicionalmente en los estudios de ingeniería, hasta los más sofisticados que se presentan en este capítulo hay una gama de posibilidades que se puede utilizar según el objetivo del estudio que se quiera llevar a cabo.

A continuación se muestra un cuadro con una clasificación de los modelos de extremos presentados en esta tesis y unas recomendaciones de uso en función del objetivo del análisis de extremos (Tabla 3.5).

La clasificación está hecha con base en si los modelos son estacionarios (constantes a lo largo del tiempo) o no estacionarios (variables a lo largo del tiempo). Dentro de cada una de estas posibilidades se presentan los modelos derivados de la familia de distribuciones GEV, con las distintas posibilidades de selección de extremos: r -máximos en un bloque de tiempo (r -max), máximos anuales (MA), máximos mensuales (MM) y máximos semanales (MS) y los derivados de Pareto-Poisson (P-P) con umbral constante ($u=cte$) y variable a lo largo del tiempo ($u=u(t)$) para la selección de máximos con el método POT.

Los objetivos de los estudios que se plantean son los derivados de algunos de los problemas a los que se enfrentan los ingenieros de costas, oceanógrafos, biólogos o gestores.

Objetivos del estudio	Modelos estacionarios			Modelos no estacionarios				
	GEV		GPD-P	GEV			GPD-P	
	r -max	MMA	$u=cte$	r -max	MMM	MMS	$u=cte$	$u=u(t)$
Diseño de obra marítima (Nivel 1)	✓	✓	✓					
Diseño de obra marítima (Nivel 3)					✓		✓	✓
Diseño del plan de obra anual					✓			✓
Diseño de obras de internada				✓	✓	✓		✓
Diseño de rutas de barcos					✓			✓
Diseño de barcos	✓	✓	✓					
Diseño de período de regeneración de playas				✓	✓	✓		✓
Gestión de la costa, limitación de deslindes...		✓				✓	✓	
Diseño de mecanismos para extracción de energía	✓	✓	✓					
Estudio de comunidades de organismos					✓	✓		

Tabla 3.5 Clasificación de los modelos de extremos y recomendaciones de uso

La comparativa de modelos hecha en este capítulo muestra la validez de uso de distintos modelos, según el estudio de variabilidad temporal que se quiera hacer. Por esta razón, en el cuadro 3.5 no se propone un único modelo para el cálculo del régimen extremal para una determinada problemática, sino que se proponen varias opciones para cada caso.

En el ámbito de las obras marítimas, el diseño se ha realizado tradicionalmente mediante modelos estadísticos estacionarios, sin embargo, la ROM 0.0 establece tres niveles de diseño. El más básico (tradicional) es el Nivel 1, en el que la altura de ola de cálculo se estima mediante régimen extremal estacionario (parámetros constantes) calculado a partir de los tres modelos estacionarios propuestos. El nivel más sofisticado es el Nivel 3, que contempla un cálculo de un régimen mucho más completo y que podrá realizarse a partir de los modelos no estacionarios marcados en el cuadro 3.5. Una vez comenzados los trabajos de construcción, los diseños del plan de obra y obras de invernada dependerán fundamentalmente de las condiciones extremas de oleaje a lo largo del año, por lo que los modelos a utilizar son los no estacionarios.

En cuanto al diseño de barcos, se puede optar por el régimen extremal a partir de modelos estacionarios (ya que interesan las condiciones más desfavorables para el cálculo, al igual que para los dispositivos de extracción de energía) y en lo que respecta a rutas marítimas habrá que tener en cuenta las variaciones del régimen extremal dentro del año y entre años, por lo que se proponen los modelos no estacionarios señalados en el cuadro.

En lo que respecta a la gestión de costas el interés puede radicar en las variaciones interanuales o en el largo plazo, por lo que el cálculo del régimen extremal puede hacerse mediante un modelo estacionario (largo plazo) y modelos no estacionarios (variaciones interanuales y largo plazo).

Por último, en el estudio de organismos marinos (comunidades bentónicas, peces, mamíferos...) suelen interesar las variaciones estacionales e interanuales, por lo que se propone el uso de los modelos no estacionarios GEV.