

UNIVERSIDAD DE CANTABRIA

**DEPARTAMENTO DE CIENCIAS Y TÉCNICAS
DEL AGUA Y DEL MEDIO AMBIENTE**

TESIS DOCTORAL

**METODOLOGÍAS DE CALIBRACIÓN DE BASES DE
DATOS DE REANÁLISIS DE CLIMA MARÍTIMO**

Presentada por: ANTONIO TOMÁS SAMPEDRO

**Dirigida por: FERNANDO J. MÉNDEZ INCERA
IÑIGO J. LOSADA RODRÍGUEZ**

Mayo, 2009

CAPÍTULO 5

CALIBRACIÓN PUNTUAL PARAMÉTRICA

5.1. Introducción.

Como ya se ha definido en el capítulo 4, de entre los tipos de calibración definidos en profundidades indefinidas, la calibración puntual corrige los oleajes del modelado numérico (SIMAR-44) con datos instrumentales (boyas y/o satélites) cuando ambos climas marítimos son similares. En la figura 5.1 se representa el esquema de la calibración puntual para el supuesto 2 (de los 12 definidos en el capítulo 4), que coincide con la definición de calibración puntual realizada, aunque también puede utilizarse esta metodología de calibración puntual conjuntamente con la propagación o la retropropagación (por ejemplo en los supuestos 5, 8 y 11).



Figura 5.1. Representación de la ubicación de la información disponible y necesaria para la calibración puntual con datos en profundidades indefinidas (supuesto N° 2 del capítulo 4).

Existen dos metodologías de calibración puntual (ver figura 4.2 del capítulo 4): la paramétrica y la no paramétrica. En este capítulo se describen los dos métodos de calibración puntual paramétrica, definidos en profundidades indefinidas: la calibración puntual paramétrica por datos coincidentes o *scatter* y la calibración puntual paramétrica por cuantiles o *QQ*. A su vez, de cada una de ellas se define la metodología escalar y direccional.

5.2. Calibración puntual paramétrica por datos coincidentes o *scatter*.

Este tipo de calibración puntual paramétrica, definido a partir de los diagramas de dispersión o *scatter plot*, es el que más comúnmente se utiliza. Existen calibraciones de diferentes

variables oceanográficas, basadas en distintas técnicas de regresión, que tienen gran implantación y desarrollo histórico (como se ha puesto de manifiesto en el capítulo 3). Aquí no se va a volver a explicar o aplicar todos los diferentes tipos o técnicas de regresión, sino que se van a utilizar casos típicos de aplicación para poder comparar los resultados obtenidos entre las distintas metodologías de calibración puntual. Aunque esta metodología es aplicable a cualquier variable geofísica y en cualquier ubicación.

Por ello, únicamente se va a calibrar H_s de SIMAR-44 (ver apartado 2.4.5) con los datos de la boya de Mahón (ver apartado 2.4.3), validando los resultados con datos de satélites (ver apartado 2.4.4, en concreto GEOSAT, TOPEX/POSEIDON, GFO, JASON-1 y ENVISAT), ya que este caso se va a resolver mediante la aplicación de todas las distintas metodologías de calibración puntual.

En lo que sí se ha innovado (en lo que a calibración *scatter* se refiere) es en implementar la metodología direccional a las regresiones. Seguidamente se va a explicar dicha metodología direccional así como la metodología escalar tradicional.

5.2.1. Metodología escalar.

La metodología escalar de calibración *scatter* busca las relaciones de calibración a través de la regresión clásica, en esta aplicación se han definido dos modelos de regresión entre los datos coincidentes de la boya de Mahón (X , datos instrumentales) y los de SIMAR-44 (Y , datos de modelado numérico). El primero es un modelo no lineal de tipo potencial ($Y = \beta X^\gamma$, ver apartado 3.2.3.1.3 del capítulo 3) y el segundo es un modelo lineal de la recta que pasa por el origen ($Y = \beta X$, ver apartado 3.2.3.1.2 del capítulo 3).

En la figura 5.2 se presenta un croquis para explicar cómo a partir del ajuste del modelo de regresión clásico (que asume todo el error en la variable Y) se determina la relación de calibración. El modelo de regresión no lineal (con vector de parámetros $\Theta = \{\beta, \gamma\}$) y el lineal (con vector de parámetros $\Theta = \{\beta\}$) no incluyen el término independiente, para evitar inhomogeneidades en la calibración de valores de Y ; por lo que se obtienen unas relaciones de calibración de SIMAR-44 (Y_{CAL}) a partir de sus datos originales (Y) del tipo $Y_{CAL} = bY^c$ y $Y_{CAL} = bY$, respectivamente para el modelo no lineal y lineal.

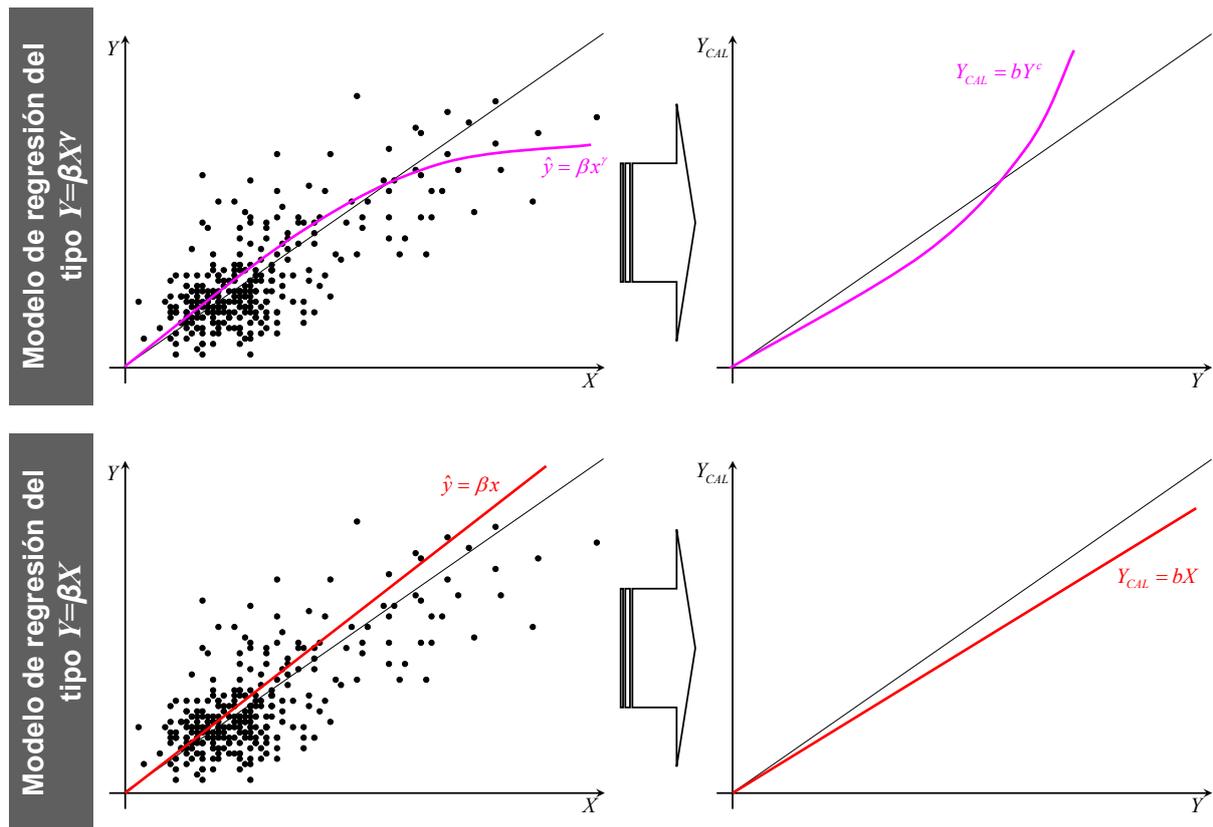


Figura 5.2. Croquis de la determinación de las relaciones de calibración a partir de las regresiones no lineal ($Y = \beta X^\gamma$, arriba) y lineal ($Y = \beta X$, abajo).

Para el modelo de regresión no lineal, a partir de β y γ es sencillo despejar la relación de calibración de Y (Y_{CAL}), calculando los valores de los parámetros b y c , asumiendo que la población de datos calibrados debe coincidir con la población de datos instrumentales. De manera análoga se produce para el modelo de regresión lineal. En la ecuación 5.1 se muestran ambas relaciones, no lineal a la izquierda y lineal a la derecha:

$$\left. \begin{array}{l} Y = \beta X^\gamma \\ Y_{CAL} = b Y^c \end{array} \right\} \rightarrow Y_{CAL} \equiv X \rightarrow \left\{ \begin{array}{l} b = \left(\frac{1}{\beta} \right)^{\frac{1}{\gamma}} \\ c = \frac{1}{\gamma} \end{array} \right. \quad \parallel \parallel \quad \left. \begin{array}{l} Y = \beta X \\ Y_{CAL} = b Y \end{array} \right\} \rightarrow Y_{CAL} \equiv X \rightarrow \left\{ b = \frac{1}{\beta} \right.$$

(5.1)

En la figura 5.3 se muestran los ajustes de ambos modelos de regresión, así como las relaciones de calibración despejadas $Y_{CAL} = b Y^c$ y $Y_{CAL} = b Y$, que permiten calibrar H_s de SIMAR-44 a partir de la variable original sin calibrar, para la posición de la boya de Mahón. Así mismo se muestra el intervalo de confianza del 95% de los parámetros estimados. Se puede observar que el cambio en el modelo de regresión utilizado implica grandes variaciones en la calibración de los oleajes más energéticos.

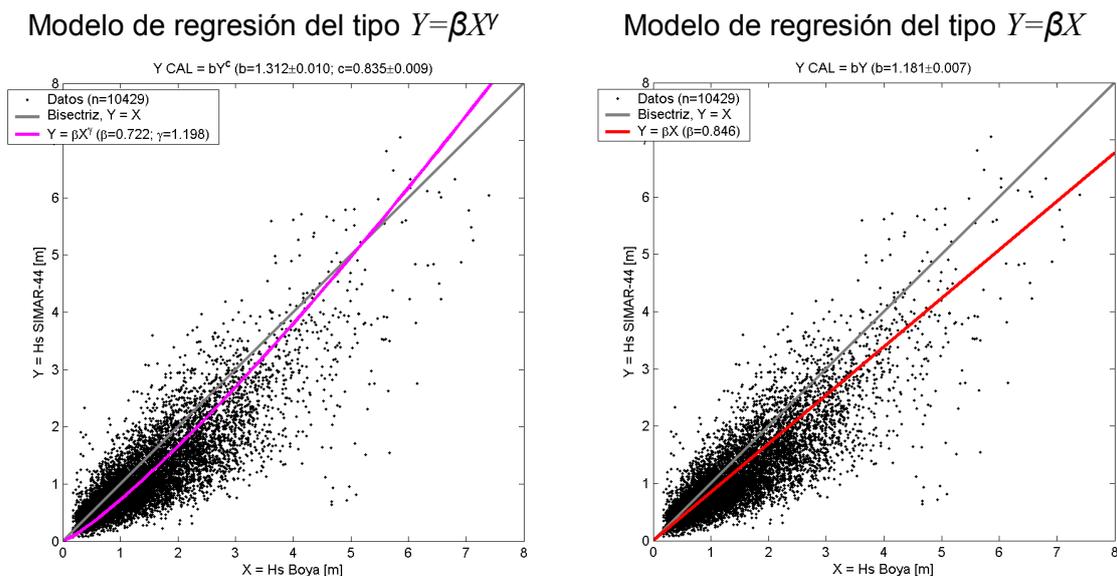


Figura 5.3. Calibración escalar mediante regresiones clásicas de H_s de SIMAR-44 con datos coincidentes de la boya de Mahón, utilizando un modelo de regresión no lineal de tipo potencial (izquierda) y un modelo de regresión lineal, recta que pasa por el origen (derecha).

A partir de las relaciones de calibración de la figura 5.3 se han calibrado los datos de SIMAR-44 de la posición de la boya de Mahón. En la figura 5.4 se presenta la verificación y validación de las calibraciones realizadas con los dos modelos de regresión estudiados. Se comparan los resultados mediante diagramas de dispersión (*scatter plot*) y los parámetros de diagnóstico *BIAS*, *RMS*, ρ y *SI* de la población de datos de SIMAR-44 antes y después de calibrar con los datos instrumentales; también se comparan los regímenes medios mediante la comparación de los cuantiles (antes y después de calibrar de SIMAR-44 con los instrumentales) equiespaciados en $-\log[-\log(\text{Pr})]$ desde el de $\text{Pr}=15\%$ hasta el mayor que verifica $\text{Pr} < 1 - 5/n$. En la verificación de las calibraciones se comparan los datos calibrados (SIMAR-44) con los datos utilizados para calibrar (boya de Mahón), en cambio para la validación se utilizan los datos de satélites que distan menos de 0.5° de la posición de la boya de Mahón. En la verificación se mejoran los parámetros cuantitativos medios (ver parámetros de ajuste del *scatter plot*), pero la parte medio-alta de los regímenes medios (ver *QQ plot*) se calibra de forma incorrecta, fundamentalmente para el modelo de regresión lineal. Esto es debido a que las regresiones se ajustan fundamentalmente donde se concentra la mayor cantidad de información, los valores medios de oleaje. Parece no tenerse control sobre la calibración del régimen medio-alto con las regresiones, ya sean lineales o no lineales. Es por ello que en la validación con datos de satélites los regímenes medios mejoran, pues no se tienen valores de satélites muy energéticos, aunque algunos de los parámetros medios empeoran.

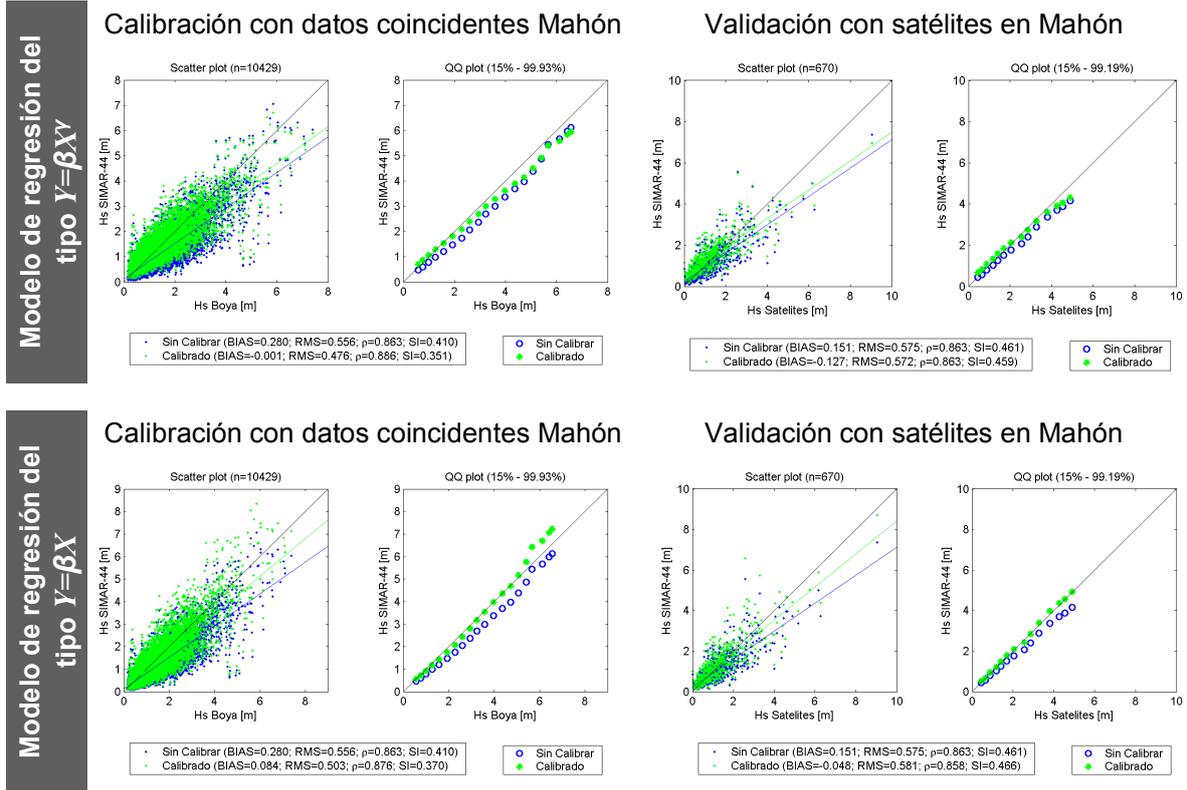


Figura 5.4. Calibración escalar mediante regresiones clásicas de H_s de SIMAR-44 con datos coincidentes de la boya de Mahón (izquierda), validando con datos de satélites (derecha). Utilizando un modelo de regresión no lineal de tipo potencial (arriba) y un modelo de regresión lineal, recta que pasa por el origen (abajo).

5.2.2. Metodología direccional.

En la aplicación de la metodología direccional de calibración *scatter* se utilizan los mismos tipos de relaciones de calibración que los de la metodología escalar, un modelo no lineal de tipo potencial ($Y_{CAL}(\theta) = b(\theta)Y(\theta)^{c(\theta)}$) y un modelo lineal de la recta que pasa por el origen ($Y_{CAL}(\theta) = b(\theta)Y(\theta)$). Para la calibración direccional se permite determinar relaciones de calibración distintas en función de la dirección de procedencia del oleaje, por eso los parámetros que definen la relación de calibración dependen de la dirección θ , definiéndose en la ecuación 5.2 $b(\theta)$ y $c(\theta)$ como una superposición de ondas sinusoidales (permitiendo la variación de b y c en función de θ).

$$\begin{aligned}
 b(\theta) &= b_0 + \sum_{i=1}^{n_\theta} \left[b_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + b_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right] \\
 c(\theta) &= c_0 + \sum_{i=1}^{n_\theta} \left[c_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + c_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right]
 \end{aligned}
 \tag{5.2}$$

donde b_0 y c_0 son parámetros que representan el valor medio sobre el que oscilan las ondas sinusoidales y se aproximan a los parámetros de las calibraciones escalares. El resto de

parámetros $(b_{2i-1}, b_{2i}, c_{2i-1}$ y $c_{2i})$ modifican las ondas sinusoidales, si no existiesen variaciones direccionales, el valor de todos estos parámetros sería 0 y b_0 y c_0 coincidirían con los valores de la metodología escalar. n_θ es el número de ondas a considerar, que ha sido $n_\theta = 6$, con el que se aseguran periodos mínimos de oscilación de las ondas de 60° .

Para el modelo no lineal de tipo potencial $(Y_{CAL}(\theta) = b(\theta)Y(\theta)^{c(\theta)})$ el vector de parámetros a estimar es por lo tanto $\Theta = \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}, c_0, c_1, \dots, c_{2n_\theta-1}, c_{2n_\theta}\}$, en cambio para el modelo lineal de la recta que pasa por el origen $(Y_{CAL}(\theta) = b(\theta)Y(\theta))$ el vector de parámetros a estimar se reduce a $\Theta = \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}\}$.

En la metodología de calibración direccional no se utilizará la regresión clásica para determinar los parámetros de los modelos de calibración (como se hizo con la metodología escalar *scatter*), pues resultaría muy complicado, sino que se minimizará por OLS la función objetivo $J(\Theta)$ dada en la ecuación 5.3, siendo Θ el vector de parámetros a estimar.

$$J(\Theta) = \sum_{i=1}^n [H_{s\ BOYA}(t_i) - H_{s\ CAL}(t_i; \Theta)]^2 \quad (5.3)$$

En esta expresión se calcula la diferencia entre los n pares de datos de H_s coincidentes de la boya de Mahón ($H_{s\ BOYA}$) y la expresión paramétrica de SIMAR-44 calibrada ($H_{s\ CAL}$). Donde $H_{s\ CAL}$ está particularizada para el caso no lineal en la ecuación 5.4 y para el caso lineal en la ecuación 5.5.

$$\begin{aligned} H_{s\ CAL}(\theta) &= b(\theta)H_s(\theta)^{c(\theta)} \\ \begin{cases} b(\theta) = b_0 + \sum_{i=1}^{n_\theta} \left[b_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + b_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right] \\ c(\theta) = c_0 + \sum_{i=1}^{n_\theta} \left[c_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + c_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right] \end{cases} \\ \hat{\Theta} &= \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}, c_0, c_1, \dots, c_{2n_\theta-1}, c_{2n_\theta}\} \end{aligned} \quad (5.4)$$

$$\begin{aligned} H_{s\ CAL}(\theta) &= b(\theta)H_s(\theta) \\ b(\theta) &= b_0 + \sum_{i=1}^{n_\theta} \left[b_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + b_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right] \\ \hat{\Theta} &= \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}\} \end{aligned} \quad (5.5)$$

En total el vector de parámetros objetivo a estimar ($\hat{\Theta}$) tiene 26 parámetros para el caso no lineal ($2(1+2n_\theta)$) y 13 parámetros para el caso lineal ($1+2n_\theta$), pues el número de ondas utilizadas es $n_\theta = 6$, lo que implica un complejo y costoso proceso de estimación con las técnicas de regresión clásicas.

Para resolver el problema de minimización de la ecuación 5.3, en cada una de las dos expresiones paramétricas presentadas en las ecuaciones 5.4 y 5.5, se utiliza el algoritmo de optimización global denominado SCEM-UA (*Shuffled Complex Evolution Metropolis*) desarrollado por la Universidad de Arizona y la Universidad de Ámsterdam presentado en Vrugt *et al.* (2003). Este algoritmo es una adaptación del método de optimización global SCE-UA (*Shuffled Complex Evolution*) desarrollado por Duan *et al.* (1992) y muy utilizado en esta tesis para la minimización de diversas funciones objetivo. Ambos métodos permiten la resolución de problemas no lineales y han sido utilizados con éxito en numerosas aplicaciones en el ámbito de la calibración de modelos hidrológicos de cuenca. La mayor diferencia radica en que el método SCE-UA calcula un único conjunto de parámetros solución del problema, mientras que el algoritmo SCEM-UA proporciona la función de densidad de cada uno de ellos, lo que permite calcular el valor más probable y el intervalo de confianza asociado a cada parámetro.

La metodología SCEM-UA combina el algoritmo *Metropolis Hastings* (que es una cadena de Markov con Monte Carlo), la búsqueda aleatoria controlada, la evolución competitiva y el mezclado de equipos. De tal forma que el espacio de búsqueda de los parámetros va evolucionando hacia los valores óptimos, generándose al final una muestra aleatoria de dichos valores con los que se puede determinar su función de densidad.

Con la implementación del algoritmo SCEM-UA a la resolución del problema de minimización de la calibración direccional *scatter* (ecuación 5.3) para las dos parametrizaciones propuestas, no lineal y lineal (ecuación 5.4 y 5.5), se resuelve de forma eficiente el complicado problema de calibración puntual paramétrica direccional por datos coincidentes.

Como ejemplo, se muestra los resultados de la calibración direccional de H_s del punto SIMAR-44 en la posición de la boya de Mahón, con los datos de dicha boya, para el caso no lineal y lineal presentados. En la parte superior de la figura 5.5 se muestra la representación gráfica de $b(\theta)$ y $c(\theta)$, con sus intervalos de confianza del 95%, y en su parte inferior se representa la fracción de corrección que impone cada calibración hasta el mayor cuantil de SIMAR-44 en cada sector direccional. Se puede verificar que los parámetros medios de calibración sobre los que oscilan las diferentes ondas; b_0 y c_0 , tienen valores muy similares a

los parámetros de la calibración escalar (ver b y c de la figura 5.3); siendo respectivamente dichos valores para la regresión no lineal $b_0 = 1.30$ y $c_0 = 0.85$ y $b = 1.31$ y $c = 0.84$ y para la lineal $b_0 = 1.18$ y $b = 1.18$. A pesar de ello, se puede observar que el cambio en el modelo de regresión utilizado implica enormes variaciones en la calibración resultante.

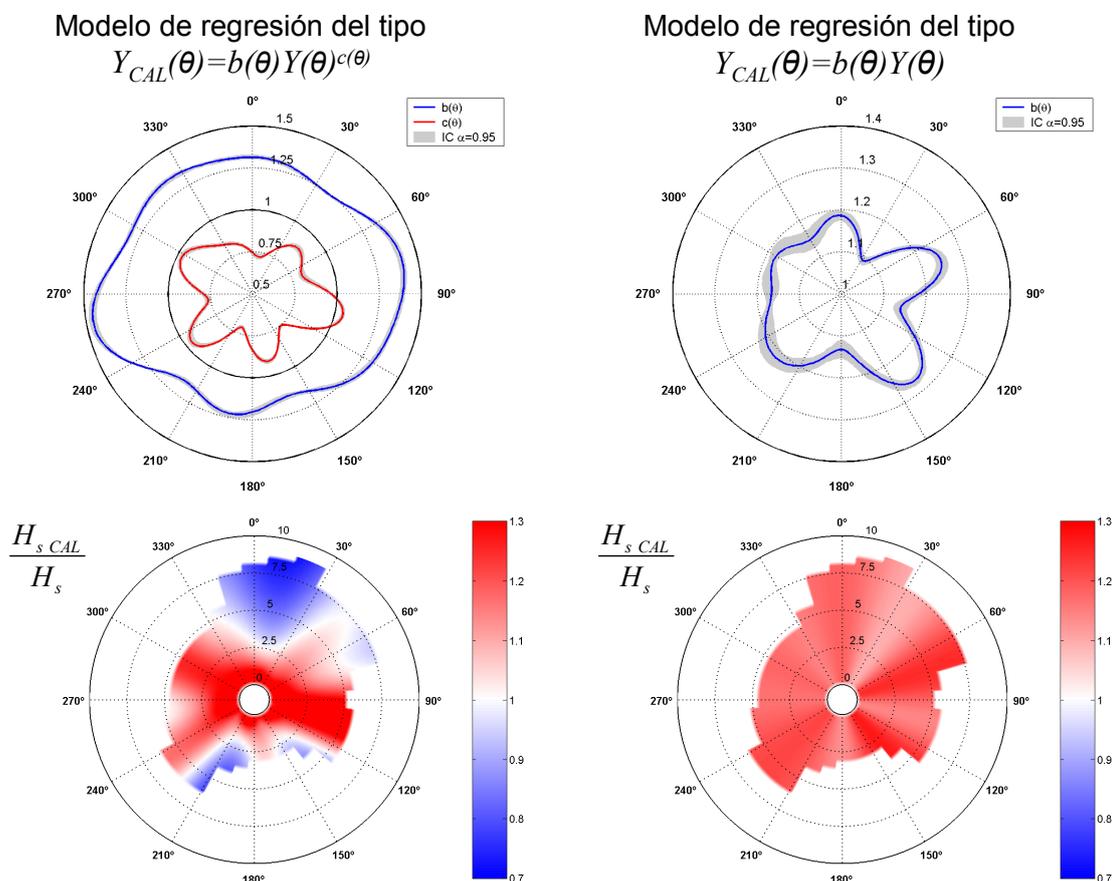


Figura 5.5. Calibración direccional mediante regresiones clásicas de H_s de SIMAR-44 con datos coincidentes de la boya de Mahón, utilizando un modelo de regresión no lineal de tipo potencial (izquierda) y un modelo de regresión lineal, recta que pasa por el origen (derecha). Se representan los parámetros del modelo de calibración (arriba) y la fracción de corrección que impone dicha calibración sobre H_s (abajo).

En la figura 5.6 se presenta la verificación y validación de las calibraciones realizadas con los dos modelos de regresión estudiados. Para la validación se utilizan los datos de satélites que distan menos de 0.5° de la posición de la boya de Mahón. En la verificación se mejoran los parámetros cuantitativos medios (aunque no tanto como con la calibración escalar, ver figura 5.4), pero la parte medio-alta de los regímenes medios se calibra de forma incorrecta (sobreestimación con la regresión lineal y subestimación con la regresión no lineal). Esto es debido a que las regresiones se ajustan fundamentalmente donde se concentra la mayor cantidad de información, los valores medios de oleaje. Es por ello que en la validación con datos de satélites los regímenes medios mejoran, pues no se tienen valores de satélites muy energéticos, aunque muchos de los parámetros medios empeoran.

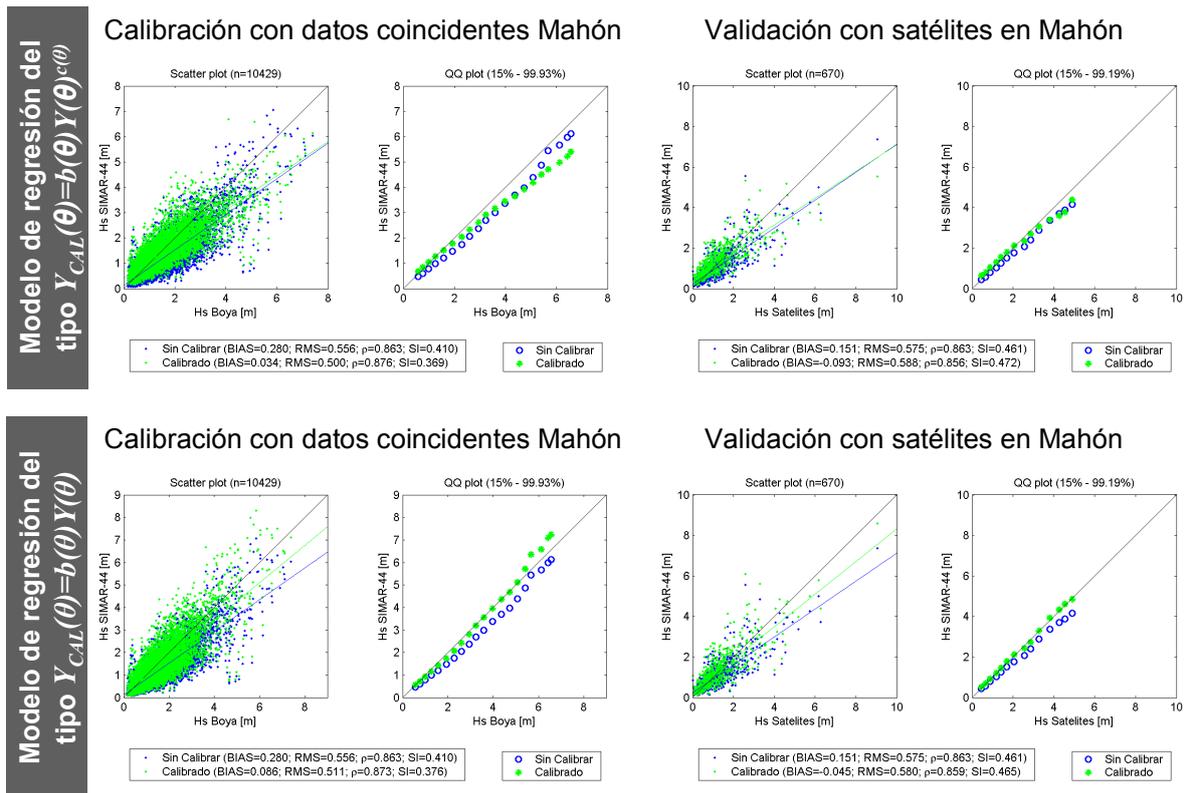


Figura 5.6. Calibración direccional mediante regresiones clásicas de H_s de SIMAR-44 con datos coincidentes de la boya de Mahón (izquierda), validando con datos de satélites (derecha). Utilizando un modelo de regresión no lineal de tipo potencial (arriba) y un modelo de regresión lineal, recta que pasa por el origen (abajo).

5.3. Calibración puntual paramétrica por cuantiles o QQ.

Una vez explicadas las calibraciones puntuales paramétricas por datos coincidentes, como ya se ha descrito, la calibración por cuantiles es la otra metodología de calibración puntual paramétrica (ver figura 4.2). Las relaciones paramétricas que se buscan con la calibración *QQ* están basadas en la comparación de regímenes medios de oleaje o funciones de distribución, pero a la hora de materializar dichas comparaciones, la elección de los cuantiles de referencia no es arbitraria y de ella depende la correcta calibración de, sobre todo, la rama medio-alta del régimen medio. Tras la revisión de distintas formas de comparar los regímenes medios entre dos variables, probabilidad-probabilidad, cuantil-cuantil,... del apartado 3.2.4.2 del capítulo 3, se ha considerado que la manera más eficiente de realizar dichas comparaciones es seleccionando cuantiles equiespaciados en $-\log[-\log(\text{Pr})]$ hasta la mayor probabilidad que verifica $\text{Pr} < 1 - 5/n$.

Por la propia definición de los cuantiles, que agregan los datos de una variable para determinar el valor de dicha variable que tiene una probabilidad de no excedencia dada, se pierde la información temporal de los datos; esto puede aprovecharse para calibrar una

variable con datos no coincidentes en el tiempo, pero debe asegurarse que éstos recojan información de las mismas escalas hiperanuales de la variable a calibrar. Es por ello que se pueden definir los cuantiles a partir de pares de datos coincidentes o no.

En general, para la definición de la curva que relaciona los regímenes medios de oleaje de dos variables, la que se va a calibrar, Y (por ejemplo de modelado numérico) y la de referencia, X (por ejemplo instrumental), se suelen utilizar relaciones potenciales del tipo $Y = \beta X^\gamma$, la cual no incluye el término independiente para evitar inhomogeneidades en la calibración de valores de Y próximos a cero. A lo largo de esta tesis se va a utilizar exhaustivamente este tipo de relación paramétrica, estimándose los valores de β y γ por el método de regresión clásica a partir de los cuantiles de X e Y , en lugar de sus poblaciones totales de datos.

A continuación se va a aplicar la metodología de calibración QQ para corregir los datos de H_s de SIMAR-44 en la posición de la boya de Mahón con información de dicha boya y validando dichas calibraciones con información de satélites (al igual que se realizó con la calibración *scatter*). Se ha elegido esta aplicación para poder comparar con los resultados del resto de metodologías de calibración, pero la metodología es aplicable a cualquier variable geofísica y en cualquier ubicación.

De hecho, el método de calibración QQ es el que tiene mayor potencial de aplicación, ya que es el único método de calibración que puede aplicarse cuando no existen datos instrumentales coincidentes a la base de datos a calibrar, pues puede implementarse para datos coincidentes o no, seguidamente se mostrarán ejemplos de ambos. Esta metodología puede utilizarse como tradicionalmente, es decir agregando todos los datos para calibrar escalarmente una variable. Pero también es posible realizar calibraciones direccionales, como se muestra a continuación, definiendo los regímenes medios direccionales agregando los datos por sectores direccionales. Aunque de manera análoga se pueden definir regímenes medios atendiendo a otros criterios, como por ejemplo el tiempo, con lo que se agregarían los datos por meses, trimestres,...

5.3.1. Metodología escalar.

En este apartado se va a aplicar la metodología escalar de calibración QQ mediante la definición de 30 cuantiles que están equiespaciados en $-\log[-\log(\text{Pr})]$, desde una probabilidad del 15% hasta el 99.999%, utilizando únicamente los cuantiles que verifican $\text{Pr} < 1 - 5/n$. Se va a utilizar una relación paramétrica potencial de calibración como la especificada en la calibración escalar *scatter*, pero ajustándola a los cuantiles de los datos en lugar de a los datos coincidentes.

En la figura 5.7 se presenta un croquis para explicar cómo a partir del ajuste del modelo de regresión clásico entre los cuantiles de los datos instrumentales (X) y los de modelado numérico (Y), $Y = \beta X^\gamma$ (con vector de parámetros $\Theta = \{\beta, \gamma\}$), se determina la relación con la que se calibran los datos del modelo numérico (Y_{CAL}) a partir de sus datos originales (Y), $Y_{CAL} = bY^c$.

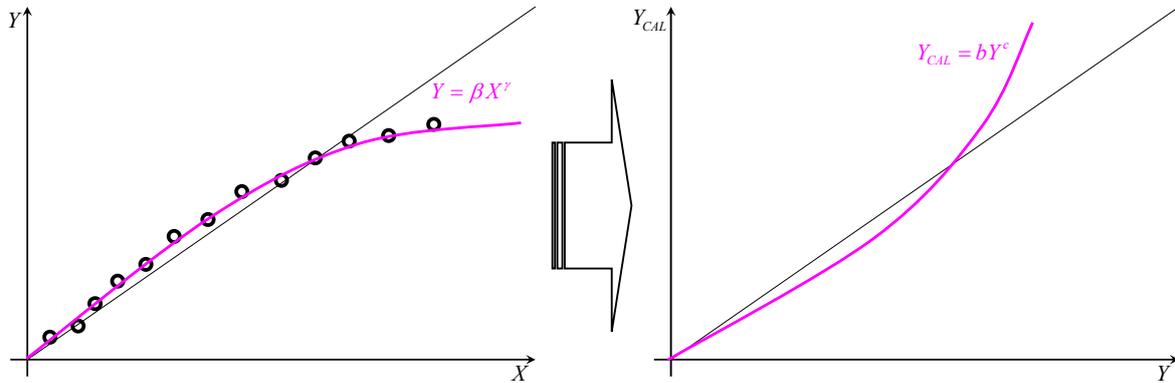


Figura 5.7. Croquis de la determinación de la relación de calibración a partir del ajuste de la expresión potencial $Y = \beta X^\gamma$ a los cuantiles de los datos.

Para el modelo de regresión potencial, a partir de β y γ es sencillo despejar la relación de calibración de Y (Y_{CAL}), calculando los valores de los parámetros b y c , asumiendo que su régimen medio debe coincidir con el de los datos instrumentales. De manera análoga a como se presentó en la ecuación 5.1 de la calibración escalar *scatter*, en la ecuación 5.6 se muestran de nuevo dichas relaciones:

$$\left. \begin{array}{l} Y = \beta X^\gamma \\ Y_{CAL} = bY^c \end{array} \right\} \rightarrow Y_{CAL} \equiv X \rightarrow \left\{ \begin{array}{l} b = \left(\frac{1}{\beta} \right)^{\frac{1}{\gamma}} \\ c = \frac{1}{\gamma} \end{array} \right. \quad (5.6)$$

Los datos sobre los que se calculan los cuantiles para definir las curvas de calibración pueden ser o bien la población completa de datos de cada variable, o bien los pares de datos coincidentes en el tiempo de ambas variables. En la figura 5.8 se presentan ambas alternativas para calibrar los datos H_s de SIMAR-44 con los datos de la boya de Mahón.

Calibración con datos coincidentes Mahón Calibración con todos los datos de Mahón

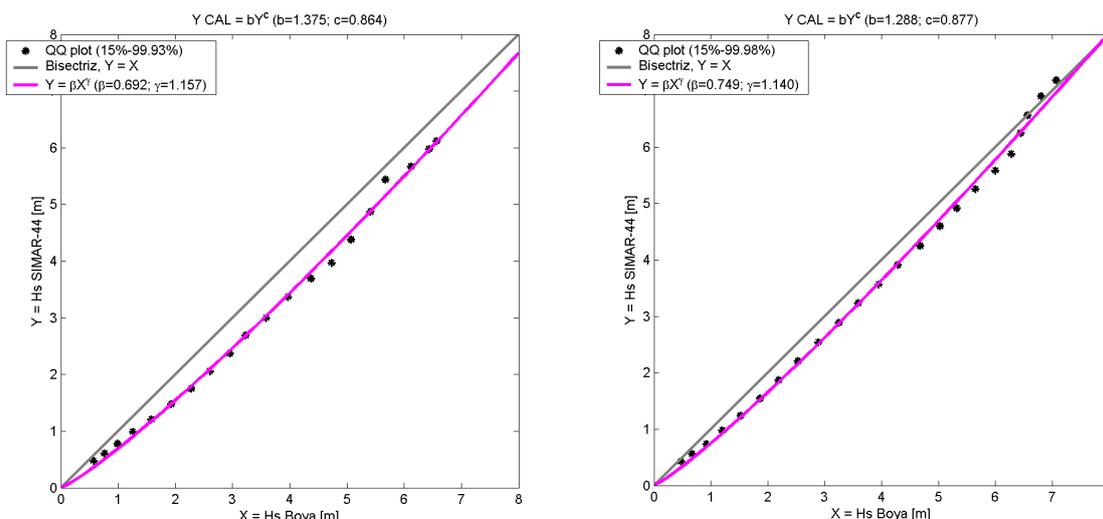


Figura 5.8. Calibración *QQ* escalar de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha).

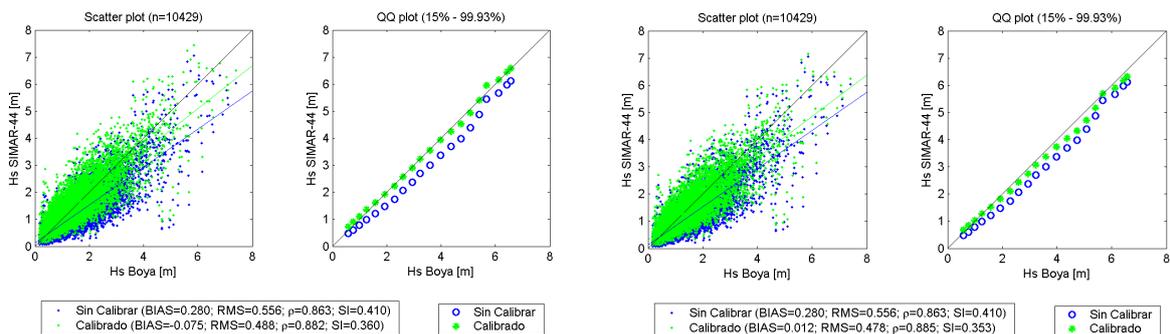
Puede observarse en la figura 5.8 que se llega a definir el cuantil del 99.98% al utilizar todos los datos, en cambio cuando se usan los datos coincidentes sólo se llega hasta el cuantil del 99.93%, pues la población de datos es menor. Se comprueba que se obtienen curvas de calibración distintas, aunque presentan aproximadamente las mismas tendencias (con la calibración se aumentan los valores, fundamentalmente los medios, siendo acrecentados en menor proporción los valores medio-altos). También se puede comprobar que dichas curvas de calibración son similares a las de la calibración no lineal escalar *scatter* ($b = 1.31$ y $c = 0.84$).

De las dos curvas ajustadas, a priori, la que produce resultados más fiables es la que utiliza sólo los datos coincidentes. Aunque cuando hay suficientes datos, como en este caso, la calibración definiendo los cuantiles con todos los datos también proporciona resultados satisfactorios; dicha alternativa se podrá utilizar en lugar de la que usa tan sólo datos coincidentes cuando éstos sean insuficientes como para definir correctamente el régimen medio (como mínimo hasta probabilidades del orden del 97-99%, lo que implica tener al menos entre 200 y 500 datos para definir dichos cuantiles, aunque este criterio puede ser modificado en función de las necesidades de calibración).

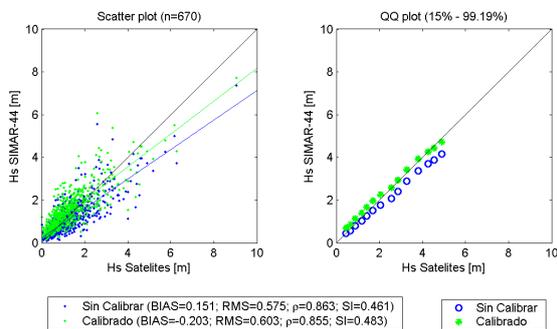
A partir de las relaciones de calibración de la figura 5.8 ($Y_{CAL} = bY^c$) se han calibrado los datos de SIMAR-44; en la figura 5.9 se presenta la verificación y validación de dichas calibraciones. Se comprueba que ambas calibraciones son correctas, mejorando la calidad de los datos. En la verificación con datos coincidentes se mejoran los parámetros cuantitativos medios y el comportamiento de los regímenes de las dos calibraciones, siendo la comparación de cuantiles de la calibración con datos coincidentes prácticamente coincidente con la

bisectriz. Por otro lado y comparando con los resultados de la calibración escalar *scatter* (ver figura 5.4) se pueden considerar estos resultados mejores, aunque alguno de los parámetros medios de diagnóstico empeora. A pesar de ello, en la validación con datos de satélites, aunque los regímenes medios mejoran, algunos de los parámetros de la calibración con datos coincidentes empeoran, esto es debido a que con la calibración se aumentan excesivamente los valores más pequeños de H_s . De hecho dichos valores de SIMAR-44 sin calibrar se aproximan bastante a los de la boya de Mahón.

Calibración con datos coincidentes Mahón Calibración con todos los datos de Mahón



Validación con satélites en Mahón



Validación con satélites en Mahón

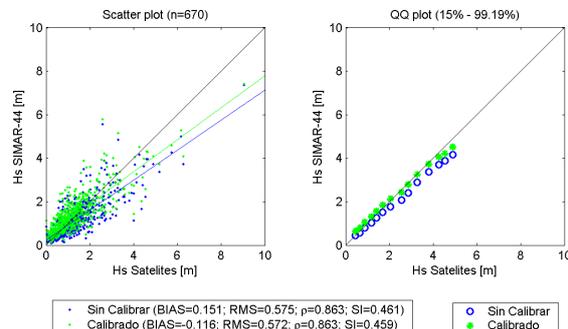


Figura 5.9. Calibración *QQ* escalar de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Verificando con los datos coincidentes de la boya (arriba) y validando con datos de satélites (abajo).

5.3.2. Metodología direccional.

Además de la tradicional metodología escalar de calibración *QQ* mediante relaciones paramétricas potenciales de calibración, que calcula una relación de calibración común para todas las direcciones de procedencia de oleaje; también se puede agregar los datos en distintos sectores direccionales, definiendo regímenes medios direccionales que permiten determinar relaciones de calibración diferentes en función de la dirección del oleaje.

En la figura 5.10 se representa un croquis de la calibración direccional con *QQ-plot*, de un cuadrante de oleaje de 90°. En la parte izquierda se puede apreciar la definición de los cuantiles para el sector direccional centrado en la dirección media de procedencia del oleaje $\theta = 0^\circ$, que agrega oleajes de un determinado sector direccional ($\pm\Delta\theta$), denominando $\varphi = 0^\circ$ a dicho sector direccional ($\varphi \equiv \theta \pm \Delta\theta$ ¹); siendo Y la variable a calibrar, X la instrumental de referencia e Y_{CAL} la variable Y calibrada. Se puede definir una relación de calibración potencial del tipo $Y_{CAL} = bY^c$ para cada sector direccional; por ejemplo, para $\varphi = 0^\circ$ sería $Y_{CAL}(\varphi = 0) = b_{\varphi 0}Y(\varphi = 0)^{c_{\varphi 0}}$, con $b_{\varphi 0}$ y $c_{\varphi 0}$ los parámetros que definen la relación de calibración para todos los datos que estén en el sector direccional $\varphi = 0^\circ$ (ver figura 5.10, izquierda). Pero al pasar al siguiente sector, la relación de calibración puede ser distinta, por lo que puede haber inhomogeneidades en los resultados de la calibración de los datos con direcciones justo en el límite entre un sector y otro, por ello deben hacerse muchos sectores para que los cambios entre sectores consecutivos sean pequeños o definir una relación de calibración direccional continua (como se hizo en la calibración direccional *scatter*), de forma que $b(\theta)$ y $c(\theta)$ sean funciones paramétricas circulares y continuas en θ (ver figura 5.10, derecha). Esta última opción ha sido la utilizada en esta tesis.

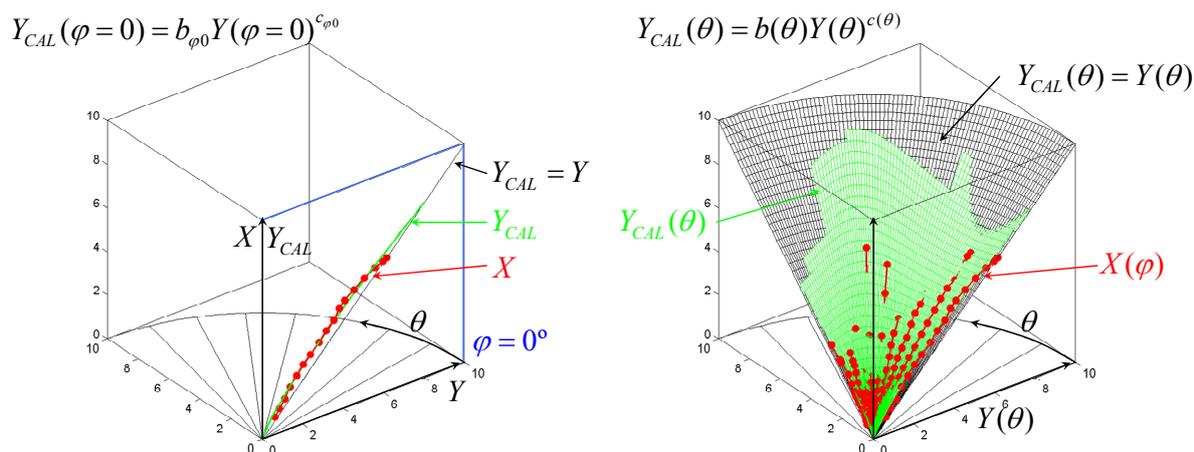


Figura 5.10. Esquema de calibración *QQ* direccional, para un sólo sector direccional, $\varphi = 0^\circ$ (izquierda) y para los sectores de un cuadrante de 90° (derecha), señalando la recta o el cono bisectriz respectivamente.

Pero previamente a la definición de relaciones de calibración continuas con θ hay que definir los sectores direccionales de comparación. Para establecer la cantidad de sectores óptima con los que se debe representar todas las direcciones, así como la amplitud o intervalo de direcciones que se agregan en cada uno, se han realizado multitud de posibles combinaciones. De las distintas opciones posibles se busca la que defina suficientemente la variación direccional de oleaje, lo que implica que se definan gran cantidad de sectores; que cada sector

¹ Los sectores direccionales φ es un variable discreta cuyos valores son los centros de los sectores direccionales que se definen a partir de la variable continua θ (dirección media de procedencia del oleaje), $\theta \pm \Delta\theta$.

sea representativo de su dirección, lo que supone que la amplitud de cada sector sea suficientemente pequeña; pero también que en cada uno se agregue gran cantidad de datos para definir con la mayor precisión posible los cuantiles más extremos, lo que implica que la amplitud de cada sector debe ser bastante grande. A pesar de que los condicionantes son contrapuestos, se ha llegado a una solución de compromiso en la que, con los datos de oleaje utilizados en esta tesis, se satisfacen estas distintas premisas.

Así, finalmente se ha decidido utilizar sectores direccionales que recogen direcciones de una amplitud de 22.5° ($\pm\Delta\theta = 22.5^\circ/2 = 11.25^\circ$) y que se definen cada 11.25° , siendo en total 32 sectores ($360^\circ/11.25^\circ=32$); lo que supone que cada dato se utiliza dos veces, una vez en cada uno de los dos sectores más próximos. En la figura 5.11 se muestra esta sectorización, marcando el centro de cada uno de los 32 sectores ($n_\phi = 32$), así como que las direcciones comprendidas entre dos centros son utilizadas por los dos sectores adyacentes, pues distan menos de 11.25° del centro del sector.

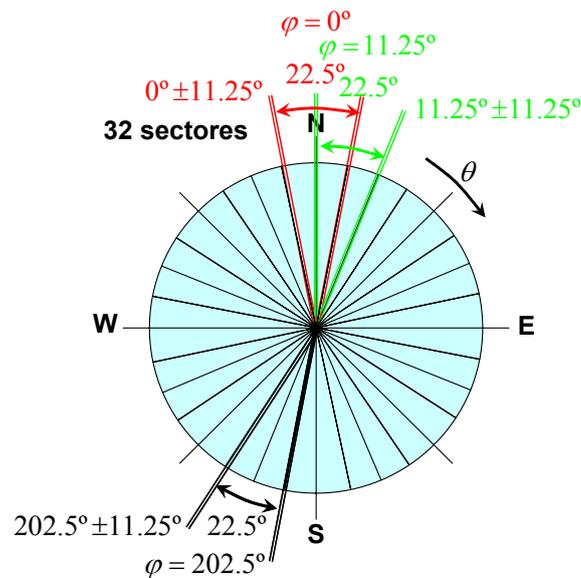


Figura 5.11. Esquema de la sectorización direccional utilizada, con 32 sectores de 22.5° de amplitud.

Otra decisión que ha motivado la realización de numerosas pruebas ha sido la definición de las funciones paramétricas circulares y continuas de los parámetros que caracterizan las calibraciones potenciales direccionales, $b(\theta)$ y $c(\theta)$. Dichas funciones se construyen como una superposición de ondas sinusoidales (al igual que en la calibración direccional *scatter*), que se describen en la ecuación 5.7:

$$\begin{aligned}
 b(\theta) &= b_0 + \sum_{i=1}^{n_\theta} \left[b_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + b_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right] \\
 c(\theta) &= c_0 + \sum_{i=1}^{n_\theta} \left[c_{2i-1} \cos\left(\frac{2\pi i}{360} \theta\right) + c_{2i} \sin\left(\frac{2\pi i}{360} \theta\right) \right]
 \end{aligned}
 \tag{5.7}$$

donde b_0 y c_0 son parámetros que representan el valor medio sobre el que oscilan las ondas sinusoidales y se aproximan a los parámetros de las calibraciones potenciales escalares. El resto de parámetros (b_{2i-1} , b_{2i} , c_{2i-1} y c_{2i}) modifican las ondas sinusoidales, si no existiesen variaciones direccionales, el valor de todos estos parámetros sería 0 y b_0 y c_0 coincidirían con los valores de la metodología escalar. n_θ es el número de ondas a considerar, que en general ha sido $n_\theta = 6$, con el que se aseguran periodos mínimos de oscilación de las ondas de 60° .

En total las funciones paramétricas de la ecuación 5.7 ($b(\theta)$ y $c(\theta)$) tienen $2(1+2n_\theta)$ parámetros a determinar (con $n_\theta = 6$ se tienen en total 26 parámetros), lo que implica un complejo y costoso proceso de estimación con las técnicas de regresión clásicas. Por ello ha sido necesario implementar el algoritmo de optimización denominado *shuffled complex evolution* (SCE-UA) para estimar por el método OLS los numerosos parámetros del modelo de calibración ($\hat{\Theta} = \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}, c_0, c_1, \dots, c_{2n_\theta-1}, c_{2n_\theta}\}$).

El mecanismo de optimización SCE-UA ha sido desarrollado por Duan *et al.* (1992) en la Universidad de Arizona y su eficiencia ha sido ampliamente reconocida ante problemas de calibración de modelos hidrológicos con un elevado número de parámetros y una alta no linealidad. El funcionamiento básico del algoritmo SCE está inspirado en los principios de selección natural y la genética y es una combinación de procedimientos deterministas y aleatorios. Se parte de diferentes puntos de búsqueda (individuos) que se organizan por equipos (*complex*). De esta manera, la búsqueda de la solución global óptima se plantea como un proceso evolutivo (*evolution*) basado en la reproducción (cruce, mutación, recombinación) existiendo, además, mezcla de equipos (*shuffled*).

La necesidad de la utilización de un método de optimización se fundamenta en dos requerimientos. Por un lado, ante la complejidad de las diferentes funciones objetivo a minimizar (en los siguientes apartados se describirán con más detalle) pues es muy frecuente la aparición de multitud de soluciones locales. Es más, el número de soluciones locales aumenta considerablemente a medida que se incrementan las dimensiones de búsqueda ($2(1+2n_\theta)$ es el número de parámetros a estimar, que en general serán 26), de forma que los algoritmos de optimización más sencillos presentan dificultades para encontrar la solución global óptima. Por otro lado, la metodología planteada requiere no solo de la búsqueda de la solución global óptima para un modelo de calibración determinado, sino del estudio y

comparación de los resultados estimados entre diferentes modelos de calibración. Este hecho, hace necesaria la utilización de un algoritmo eficiente que reduzca las operaciones de búsqueda y con ello el tiempo computacional requerido.

Así, la metodología de calibración direccional QQ va a aplicarse a la calibración de H_s de SIMAR-44 con la boya de Mahón, buscando siempre un modelo o relación de calibración como el de la ecuación 5.8:

$$H_{s\text{ CAL}}(\theta) = b(\theta)H_s(\theta)^{c(\theta)} \quad (5.8)$$

siendo $b(\theta)$ y $c(\theta)$ las definidas en la ecuación 5.7. Pero para definir la función objetivo a minimizar por OLS con el algoritmo SCE-UA se pueden estudiar varias alternativas, como por ejemplo, comparar los cuantiles del oleaje de la boya y SIMAR-44 directamente (sin ponderar), ponderándolos y ponderándolos con la condición de que los cuantiles de los oleajes más extremos varíen lo menos posible, si no se dispone de información instrumental para ello. Estas tres alternativas se van a describir a continuación.

5.3.2.1. Calibración sin ponderación de cuantiles.

Dentro de la metodología de calibración direccional QQ inicialmente se va a describir la alternativa de comparar los cuantiles de H_s de la boya de Mahón con los de SIMAR-44 para esa posición, sin ningún tipo de ponderación, siendo $J(\Theta)$ la función objetivo a minimizar por OLS, definida en la ecuación 5.9, en la que Θ es el vector de los 26 parámetros a estimar:

$$J(\Theta) = \sum_{i=1}^{n_\varphi} \sum_{j=1}^{n_{pr}} \left[H_{s\text{ BOYA}}(\varphi_i, Pr_j) - H_{s\text{ CAL}}(\varphi_i, Pr_j; \Theta) \right]^2 \quad (5.9)$$

En esta ecuación $H_{s\text{ CAL}}$ es la definida en las ecuaciones 5.8 y 5.7, que depende del vector de parámetros de búsqueda ($\hat{\Theta} = \{b_0, b_1, \dots, b_{2n_\varphi-1}, b_{2n_\varphi}, c_0, c_1, \dots, c_{2n_\varphi-1}, c_{2n_\varphi}\}$), n_φ es el número de sectores direccionales, que son 32, como se definió en la figura 5.11 y n_{pr} es el número de cuantiles de comparación válidos para cada sector direccional.

La minimización de la ecuación 5.9 se resuelve con el algoritmo SCE-UA, calculando los parámetros que caracterizan el modelo de calibración. Pero se obtienen distintos resultados en función de si se introducen los cuantiles de todos los datos o si sólo se utilizan los datos coincidentes para determinar dichos cuantiles. En la parte superior de la figura 5.12 se muestra la representación gráfica de $b(\theta)$ y $c(\theta)$ para ambas posibilidades; en su parte

inferior se representa la fracción de corrección que impone cada calibración hasta el mayor cuantil de SIMAR-44 en cada sector direccional.

Con los gráficos de la figura 5.12 se puede comprobar como varían las correcciones con la dirección, siendo bastante distintas la calibración realizada con cuantiles de todos los datos y de sólo los datos comunes, teniendo esta última oscilaciones más suaves, aunque para las direcciones más energéticas (el cuadrante Norte) se obtiene el mismo patrón de calibración para ambas opciones. También se puede verificar que los parámetros medios de calibración sobre los que oscilan las diferentes ondas, b_0 y c_0 , tienen valores similares a los parámetros de la calibración escalar (ver b y c de la figura 5.8); siendo respectivamente dichos valores para los cuantiles de datos coincidentes $b_0 = 1.31$ y $c_0 = 0.93$ y $b = 1.38$ y $c = 0.86$ y para cuantiles determinados a partir de todos los datos $b_0 = 1.22$ y $c_0 = 0.93$ y $b = 1.29$ y $c_0 = 0.88$.

Calibración con datos coincidentes Mahón Calibración con todos los datos de Mahón

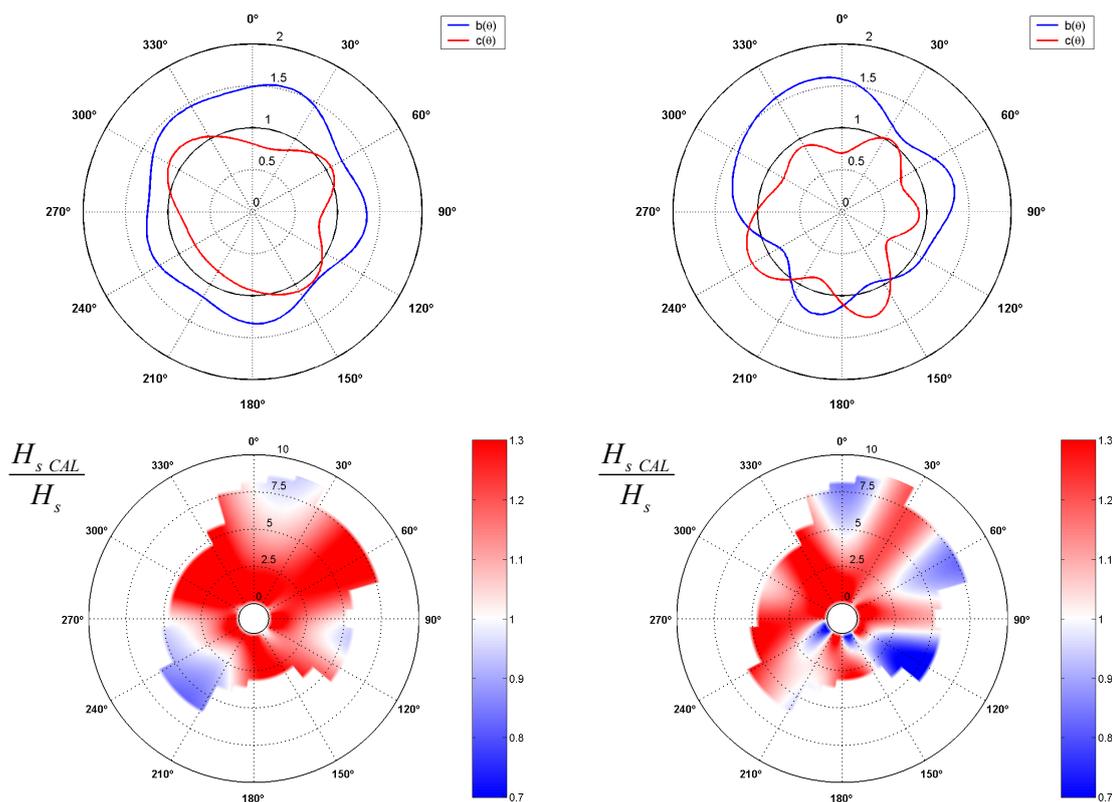
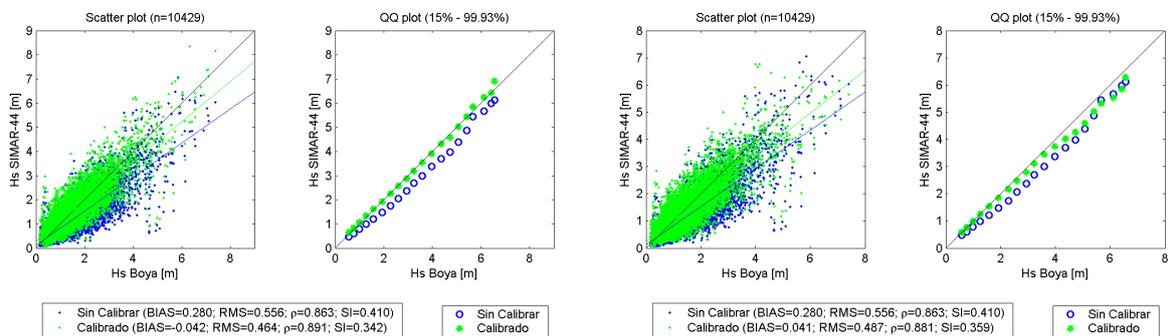


Figura 5.12. Calibración QQ direccional de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Se representan los parámetros del modelo de calibración (arriba) y la fracción de corrección que impone dicha calibración sobre H_s (abajo).

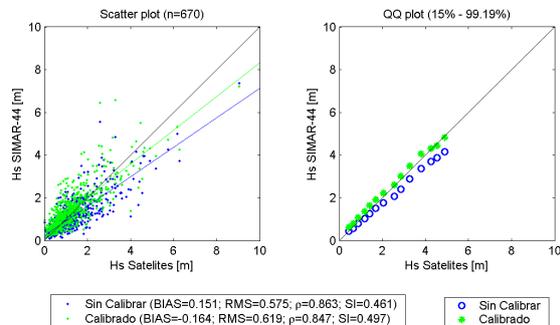
Finalmente, y como siempre, se validan las calibraciones realizadas; en la figura 5.13 se presenta la verificación con los datos coincidentes de la boya de Mahón y la validación con

datos coincidentes de satélites. Se comprueba que ambas calibraciones son correctas, mejorando la calidad de los datos. En la verificación de la calibración con cuantiles de datos comunes se mejora respecto a la calibración escalar (ver figura 5.9) aunque los cuantiles más extremos se sobreestiman; por el contrario, con la calibración de todos los datos no se llega a los resultados logrados con la calibración escalar. En la validación con datos de satélites, aunque los regímenes medios mejoran, los parámetros de la calibración con datos coincidentes empeoran. Esto es debido a que, al igual que ocurría con la calibración escalar, con la calibración se aumentan excesivamente los valores más pequeños de H_s , de hecho dichos valores sin calibrar son prácticamente coincidentes con los de las boyas.

Calibración con datos coincidentes Mahón Calibración con todos los datos de Mahón



Validación con satélites en Mahón



Validación con satélites en Mahón

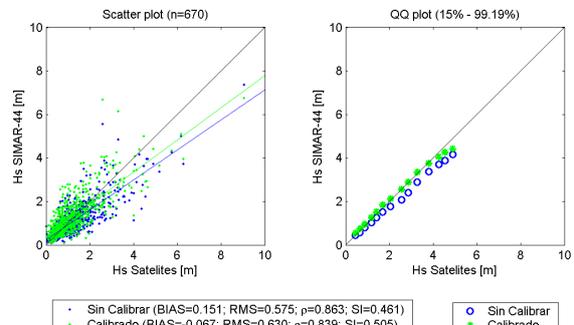


Figura 5.13. Calibración QQ direccional de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Verificando con los datos coincidentes de la boya (arriba) y validando con datos de satélites (abajo).

5.3.2.2. Calibración con ponderación de cuantiles.

Una vez desarrollada la metodología de calibración direccional QQ sin ponderación de cuantiles, se va crear una alternativa que pondera los cuantiles para dar más peso a los que se definen a partir de una mayor cantidad de información. Así una primera forma intuitiva de ponderación sería multiplicar a los cuantiles de cada sector direccional por el número de datos que contiene ese sector (n). Pero al igual que se puede multiplicar por n , se pueden

multiplicar por distintas potencias de n , así de manera ilustrativa en la figura 5.14 se representa el peso, W , en función del número n de elementos para $W = n^{\frac{1}{2}}$, $W = n$ y $W = n^2$.

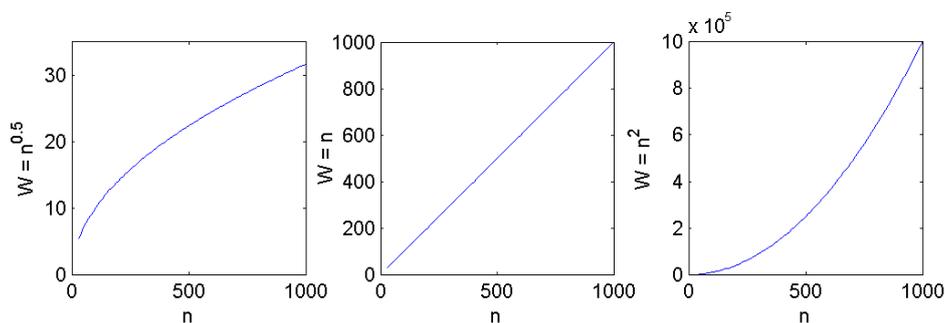


Figura 5.14. Esquema de la repercusión de aplicar ponderaciones proporcionales al número de elementos n elevado a distintas potencias.

En la figura 5.14 se puede observar grandes incrementos del peso con el aumento de n cuando la potencia utilizada es mayor que uno, por lo que parece razonable utilizar pesos del orden de $W = n^{\frac{1}{2}}$ para no penalizar demasiado la falta de datos y que con la ponderación se calibren bien solamente los sectores direccionales con más datos. Pero no sólo se debe tener en cuenta la cantidad de datos sino también el cuantil, pues cuantiles cercanos al 50% deben tener más peso que otros más extremos, con probabilidades de ocurrencia mucho menores y cuya estimación tiene muchas más incertidumbres. De hecho con los gráficos de Cloper y Pearson (tomadas de Castillo, 1993) se puede mostrar el intervalo de confianza en la estimación de un determinado cuantil. Así en la figura 5.15 se representan los intervalos de confianza para el 95% de los cuantiles para tres muestras de datos (entre $Pr > 5/n$ y $Pr < 1 - 5/n$). Comprobándose que a medida que se aumenta el número de datos y que el cuantil está más próximo al del 50%, menor es el intervalo de confianza.

$$Pr \pm \Delta Pr = Pr \pm z_{\alpha} \sqrt{\frac{Pr(1-Pr)}{n}}$$

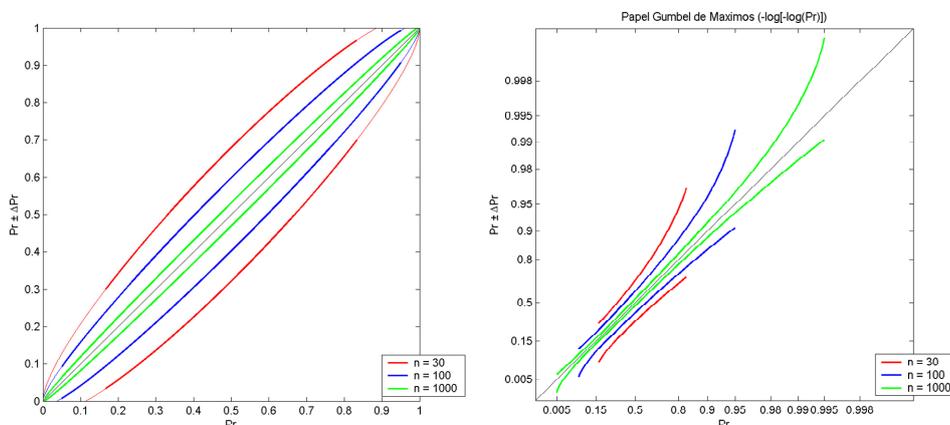


Figura 5.15. Representación de los intervalos de confianza para los cuantiles cuando se tienen muestras de datos con 30, 100 y 1000 elementos. A escala natural (izquierda) y a escala del papel probabilístico Gumbel de máximos ($-\log[-\log(Pr)]$).

A partir de los comentarios realizados de las figuras 5.14 y 5.15 se han planteado muchas funciones peso, pero la que se ha considerado más coherente y efectiva para los resultados que se buscan en esta tesis es la presentada en la ecuación 5.10:

$$W(\varphi, \text{Pr}) = \sqrt{n(\varphi) \text{Pr}(1 - \text{Pr})} \quad (5.10)$$

Con ella se disminuye el peso de manera moderada de los cuantiles extremos y los definidos a partir de pocos datos. En la figura 5.16 se puede ver una representación de la función peso de la ecuación 5.10 en la que se puede constatar estos efectos. De hecho, como se presentó en el apartado 3.2.2.2, el método de los momentos ponderados estudia el ajuste de las colas de funciones de distribución con expresiones similares; coincidiendo el momento $M_{0.5,0.5,0.5}$ con la esperanza del peso propuesto.

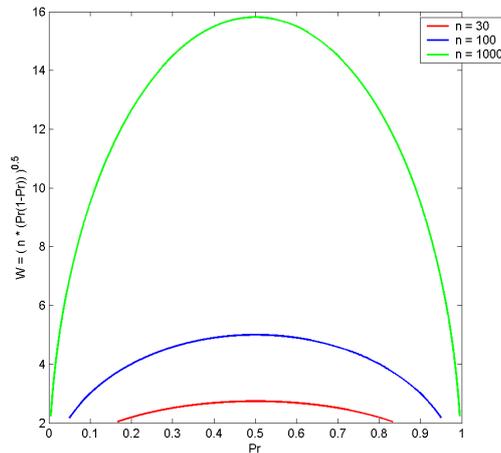


Figura 5.16. Representación de la función peso $W = \sqrt{n \text{Pr}(1 - \text{Pr})}$ en función de los cuantiles y de tres muestras de 30, 100 y 1000 datos cada una.

Con la introducción de la función peso de la ecuación 5.10, se pondera la función objetivo definida en la ecuación 5.11, $J(\Theta)$, en la que Θ es el vector de los 26 parámetros a estimar:

$$J(\Theta) = \frac{\sum_{i=1}^{n_\varphi} \sum_{j=1}^{n_{\text{Pr}}} [H_{s \text{ BOYA}}(\varphi_i, \text{Pr}_j) - H_{s \text{ CAL}}(\varphi_i, \text{Pr}_j; \Theta)]^2 W(\varphi_i, \text{Pr}_j)}{\sum_{i=1}^{n_\varphi} \sum_{j=1}^{n_{\text{Pr}}} W(\varphi_i, \text{Pr}_j)} \quad (5.11)$$

En esta ecuación $H_{s \text{ CAL}}$ es la definida en las ecuaciones 5.8 y 5.7, que depende del vector de parámetros de búsqueda ($\hat{\Theta} = \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}, c_0, c_1, \dots, c_{2n_\theta-1}, c_{2n_\theta}\}$), n_φ es el número de sectores direccionales (32) y n_{Pr} es el número de cuantiles de comparación válidos para cada sector direccional.

La minimización por OLS de la ecuación 5.11 se resuelve con el algoritmo SCE-UA, calculando los parámetros que caracterizan el modelo de calibración. Pero se obtienen distintos resultados en función de si se introducen los cuantiles de todos los datos o si se utilizan los cuantiles de solo los datos coincidentes. En la figura 5.17 se muestra la representación gráfica de $b(\theta)$ y $c(\theta)$ para ambas posibilidades y la fracción de corrección que impone cada calibración, dichos gráficos son bastante similares a los obtenidos para la calibración sin ponderación (ver figura 5.12), lo que muestra la robustez del método de calibración y las soluciones encontradas por el algoritmo de optimización SCE-UA.

Calibración con datos coincidentes Mahón Calibración con todos los datos de Mahón

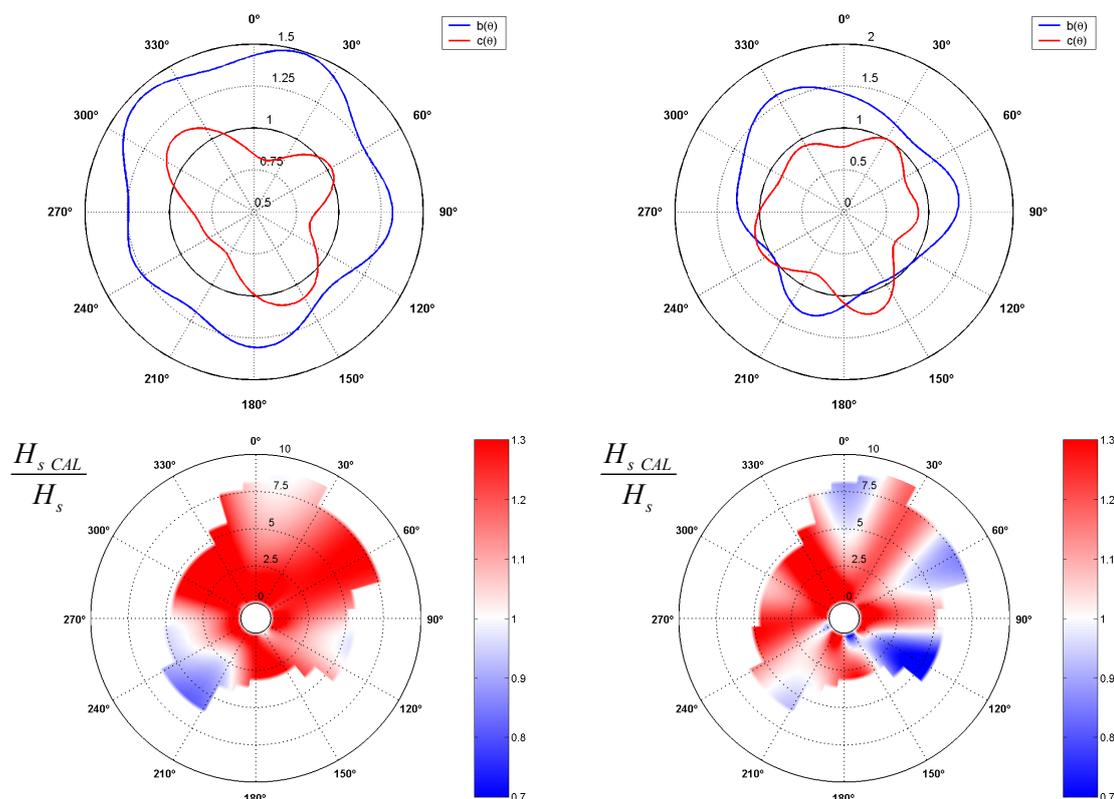


Figura 5.17. Calibración QQ direccional ponderada de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Se representan los parámetros del modelo de calibración (arriba) y la fracción de corrección que impone dicha calibración sobre H_s (abajo).

Una vez estimados los parámetros de la calibración con la ponderación propuesta y para poder contrastar y comparar los resultados obtenidos con los de la calibración sin ponderación, se seleccionan dos sectores direccionales para verificar que las calibraciones obtenidas son correctas. Para las dos direcciones elegidas ($\theta = 22.5^\circ$ y $\theta = 135^\circ$) en la figura 5.18 se muestran los cuantiles de los sectores direccionales de la boya de Mahón y la relación de calibración de SIMAR-44 particularizadas para ambas direcciones y evaluadas hasta el mayor cuantil de SIMAR-44 para cada sector direccional.

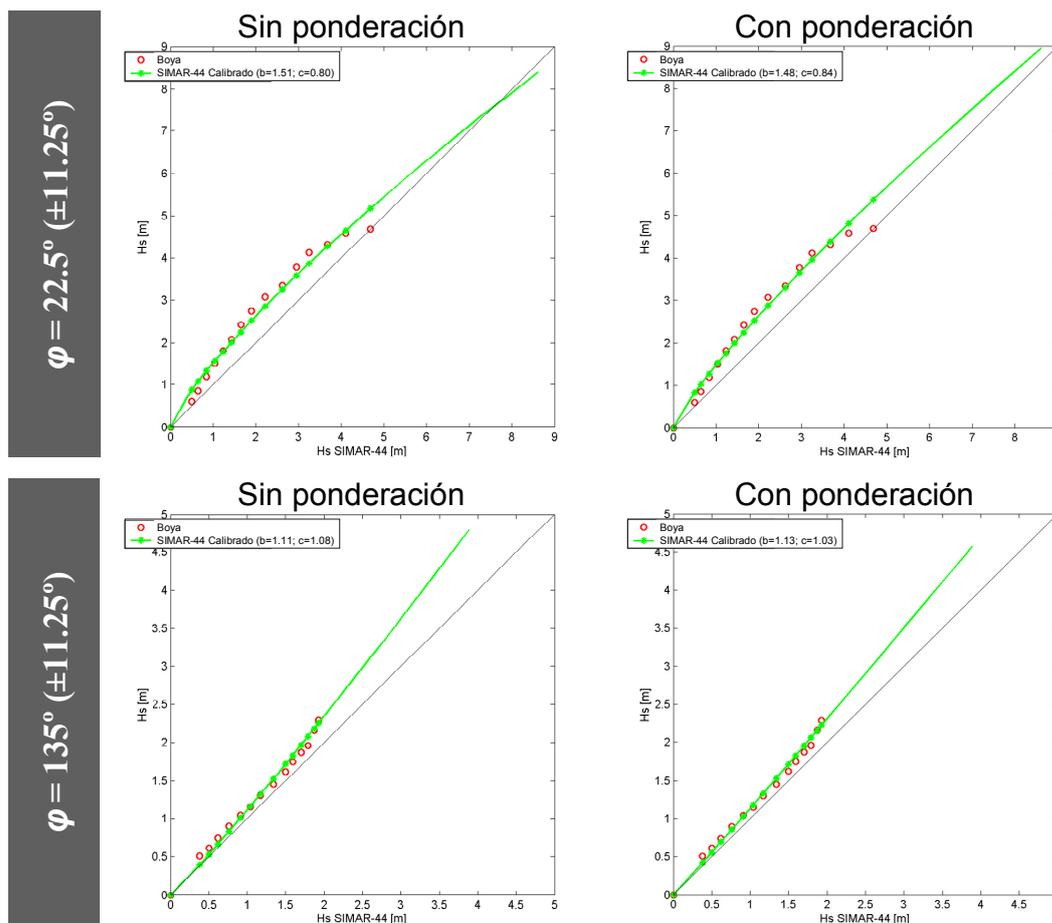


Figura 5.18. Comparación de la calibración QQ direccional sin ponderar (izquierda) y ponderada (derecha) de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes, para $\theta = 22.5^\circ$ (arriba) y para $\theta = 135^\circ$ (abajo).

Puede comprobarse en la figura 5.18 que las relaciones de calibración se ajustan muy bien a los cuantiles registrados por la boya de Mahón y que las expresiones sin ponderar y ponderadas son muy similares. Existen pequeñas diferencias debidas a la ponderación introducida, que hace que en este caso las relaciones de calibración se ajusten más a los cuantiles medios de la boya; estas pequeñas variaciones se hacen notorias en la calibración de los oleajes más energéticos.

Finalmente se validan las calibraciones realizadas; en la figura 5.19 se presenta la verificación con los datos coincidentes de la boya de Mahón y la validación con datos coincidentes de satélites. Se comprueba que ambas calibraciones son correctas, mejorando todos los resultados respecto de la calibración sin ponderar (ver figura 5.13). En la verificación de la calibración con cuantiles de todos los datos se consigue mejorar respecto a la calibración escalar (ver figura 5.9) aunque los cuantiles más extremos de la calibración con datos comunes se siguen sobreestimando. En la validación con datos de satélites, aunque se mejora

respecto a la calibración sin ponderar, se siguen observando las tendencias explicadas anteriormente de los valores más pequeños de H_s .

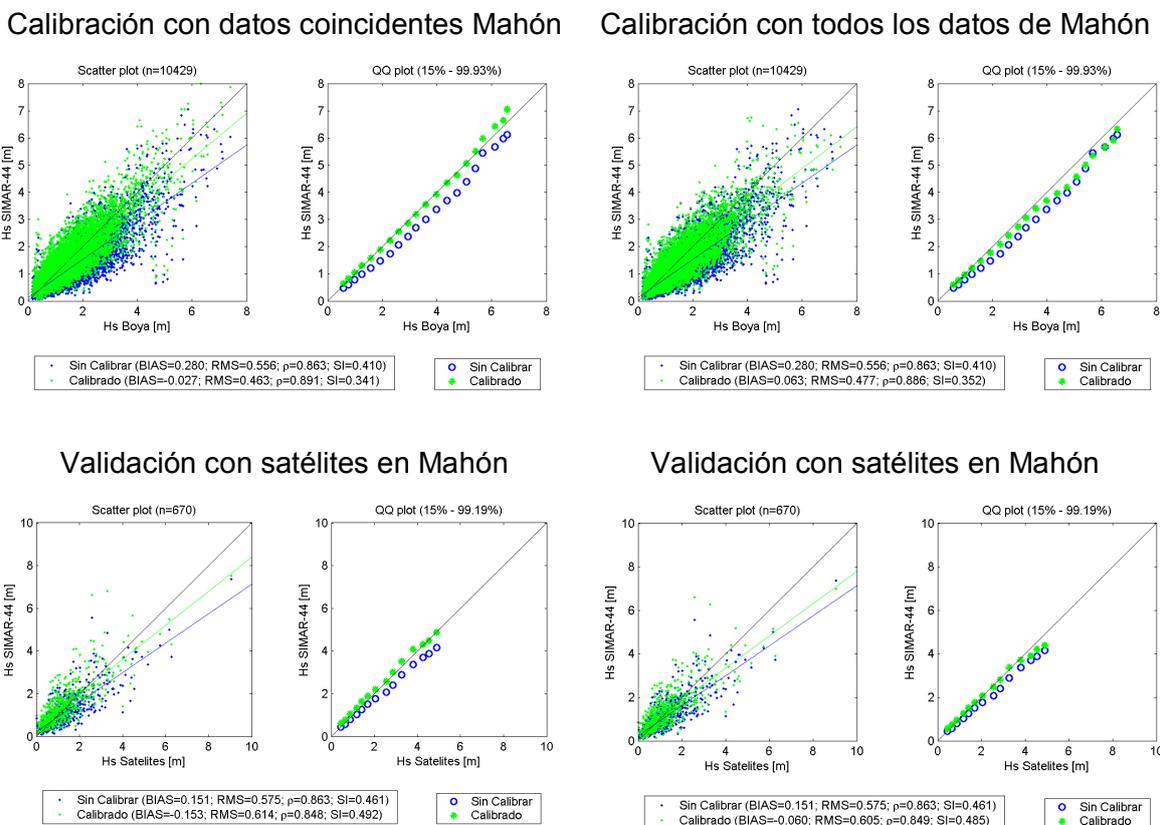


Figura 5.19. Calibración QQ direccional ponderada de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Verificando con los datos coincidentes de la boya (arriba) y validando con datos de satélites (abajo).

5.3.2.3. Calibración con ponderación de cuantiles y mínima variación del régimen extremal.

Como se ha comprobado en los ejemplos anteriores de calibración direccional QQ , los oleajes más energéticos sufren correcciones a veces indebidas provocadas por el método de calibración, pues siguen la tendencia de corrección de oleajes instrumentales menos energéticos. Con estos métodos no se impone ninguna restricción para evitar la modificación de los oleajes cuando no existe información fiable para ello.

En este apartado se propone una metodología de calibración direccional QQ con la ponderación presentada en la ecuación 5.10, pero además se introduce un factor que hace que la variación del régimen extremal del oleaje sea mínima (denotándose en adelante MVRExt). De manera que si existe información instrumental suficiente, los valores más energéticos del oleaje se calibrarán con base en ello, pero si no permanecerán similares a su estado original.

Para lograr que los oleajes más energéticos varíen poco respecto de su estado original sin calibrar se introduce una nueva condición en cada sector direccional, que se añade al resto de las condiciones de variación mínima de los cuantiles de SIMAR-44 calibrado con respecto a los cuantiles de la boya ($H_{s\ BOYA}(\varphi, Pr) - H_{s\ CAL}(\varphi, Pr; \Theta)$). Esta nueva condición impone que, en cada sector direccional, la variación del mayor cuantil de SIMAR-44 calibrado con respecto a antes de calibrar sea mínima ($H_s(\varphi, Pr_{\max}) - H_{s\ CAL}(\varphi, Pr_{\max}; \Theta)$). Con lo que la nueva función objetivo a minimizar, $J(\Theta)$, se define en la ecuación 5.12:

$$J(\Theta) = \frac{\sum_{i=1}^{n_\varphi} \sum_{j=1}^{n_{Pr}} [H_{s\ BOYA}(\varphi_i, Pr_j) - H_{s\ CAL}(\varphi_i, Pr_j; \Theta)]^2 W(\varphi_i, Pr_j)}{\sum_{i=1}^{n_\varphi} \sum_{j=1}^{n_{Pr}} W(\varphi_i, Pr_j) + \sum_{j=1}^{n_\varphi} W(\varphi_i, Pr_{\max})} + \dots$$

$$\dots + \frac{\sum_{i=1}^{n_\varphi} [H_s(\varphi_i, Pr_{\max}) - H_{s\ CAL}(\varphi_i, Pr_{\max}; \Theta)]^2 W(\varphi_i, Pr_{\max})}{\sum_{i=1}^{n_\varphi} \sum_{j=1}^{n_{Pr}} W(\varphi_i, Pr_j) + \sum_{i=1}^{n_\varphi} W(\varphi_i, Pr_{\max})}$$

(5.12)

En esta ecuación $H_{s\ CAL}$ es la definida en las ecuaciones 5.8 y 5.7, que depende del vector de parámetros de búsqueda ($\hat{\Theta} = \{b_0, b_1, \dots, b_{2n_\theta-1}, b_{2n_\theta}, c_0, c_1, \dots, c_{2n_\theta-1}, c_{2n_\theta}\}$), n_φ es el número de sectores direccionales (32) y n_{Pr} es el número de cuantiles de comparación válidos para cada sector direccional. El primer sumando es análogo al de la ecuación 5.11 (con ponderación) y el segundo sumando introduce la mínima variación del régimen extremal (MVRExt), que penaliza la separación de los datos calibrados más energéticos respecto de su posición original, antes de la calibración.

La minimización por OLS de la ecuación 5.12 también se resuelve con el algoritmo SCE-UA, calculando los parámetros que caracterizan este modelo de calibración; planteándose el problema de las mismas dos formas que anteriormente, introduciendo los cuantiles de todos los datos y utilizando los cuantiles de solo los datos coincidentes. En la figura 5.20 se muestra la representación gráfica de $b(\theta)$ y $c(\theta)$ para ambas posibilidades y la fracción de corrección que impone cada calibración, dichos gráficos son bastante similares a los obtenidos para la calibración sin ponderación y con ponderación (ver figura 5.12 y 5.17 respectivamente).

Calibración con datos coincidentes Mahón Calibración con todos los datos de Mahón

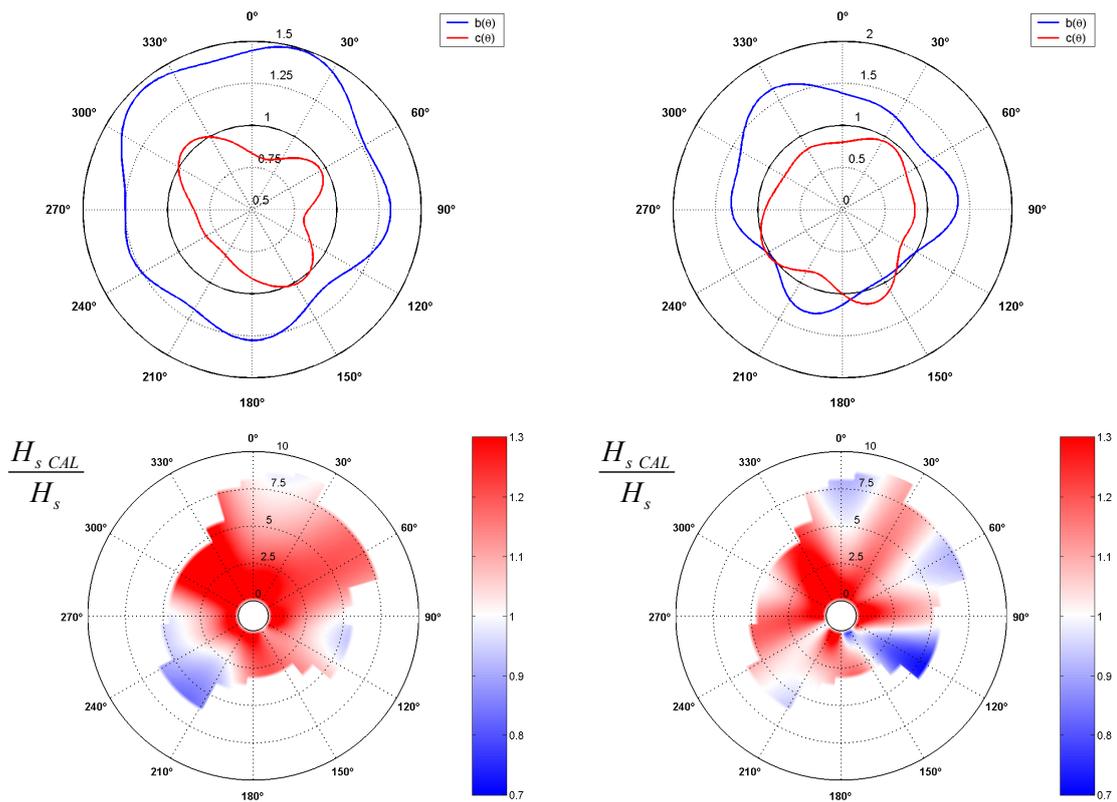


Figura 5.20. Calibración QQ direccional ponderada y con MVRExt de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Se representan los parámetros del modelo de calibración (arriba) y la fracción de corrección que impone dicha calibración sobre H_s (abajo).

Una vez estimados los parámetros de la calibración con la ponderación y con MVRExt propuesta y para poder contrastar con los resultados obtenidos con la calibración sin ponderación y con ponderación de la figura 5.18, se representan en la figura 5.21 las calibraciones sin ponderación y con ponderación y con MVRExt para $\theta = 22.5^\circ$ y $\theta = 135^\circ$.

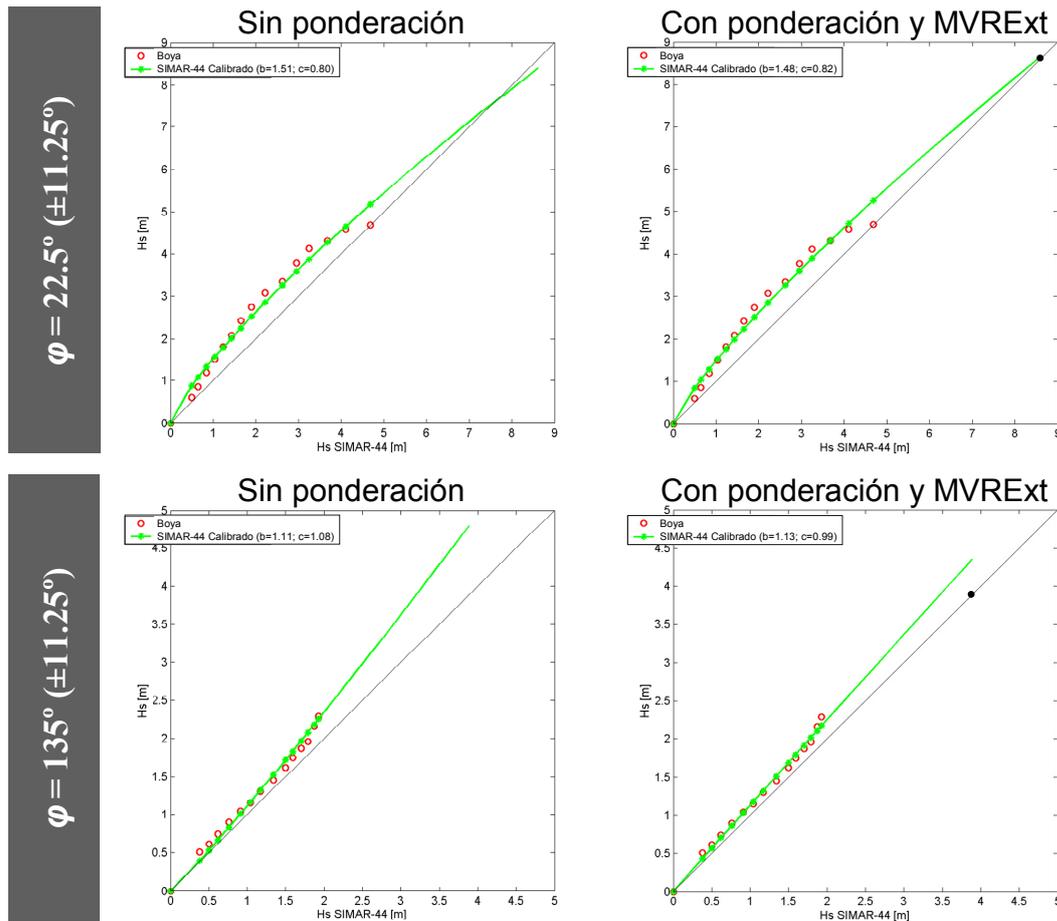


Figura 5.21. Comparación de la calibración QQ direccional sin ponderar (izquierda) y ponderada y con MVRExt (derecha) de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes, para $\theta = 22.5^\circ$ (arriba) y para $\theta = 135^\circ$ (abajo).

Puede comprobarse en la figura 5.21 que las relaciones de calibración se ajustan muy bien a los cuantiles registrados por la boya de Mahón y que las expresiones sin ponderar y ponderadas y con MVRExt son similares, pero la calibración de los oleajes más energéticos es menos arriesgada, pues se aproxima más al mayor cuantil de SIMAR-44 sin calibrar de cada sector direccional (marcados con un punto negro).

Para concluir, y poder comparar con los resultados anteriores, en la figura 5.22 se presenta la verificación con los datos coincidentes de la boya de Mahón y la validación con datos coincidentes de satélites. Se comprueba que ambas calibraciones son correctas y mejoran todos los resultados respecto del resto de calibraciones direccionales QQ (sin ponderar, ver figura 5.22 y ponderado, ver figura 5.19). Se puede observar en estos resultados que los cuantiles más extremos de la calibración se sitúan siempre entre la bisectriz y los cuantiles sin calibrar, corrigiéndose los problemas de sobreestimación de la calibración con cuantiles de datos comunes. En la validación con datos de satélites se mejora con respecto a los dos casos anteriores (sin y con ponderación) y aunque los regímenes medios mejoran, los parámetros de

la calibración con datos coincidentes empeoran, sin llegar a los resultados de la calibración escalar QQ ; esto es debido a que con la calibración se aumentan excesivamente los valores más pequeños de H_s , de hecho dichos valores sin calibrar son prácticamente coincidentes con los de la boya.

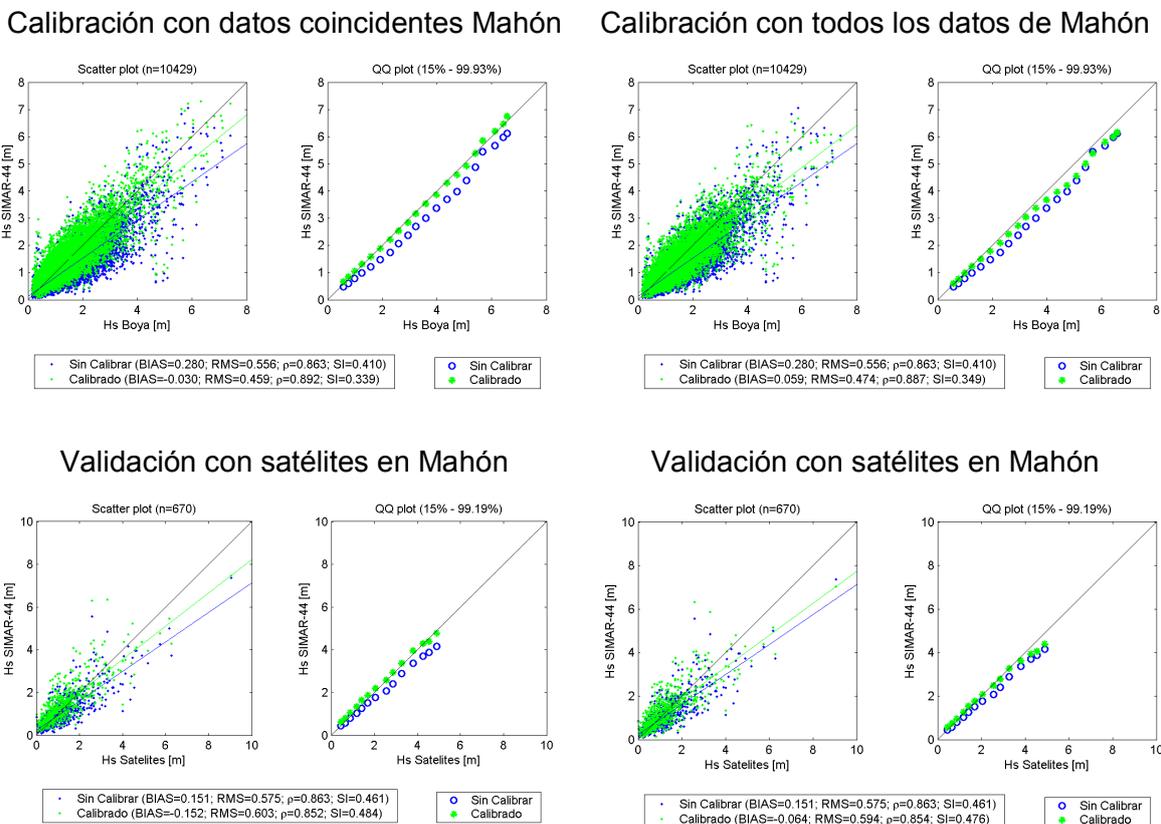


Figura 5.22. Calibración QQ direccional ponderada con mínima variación del régimen extremal de H_s de SIMAR-44 con datos de la boya de Mahón, utilizando cuantiles de los pares de datos coincidentes (izquierda) y con todos los datos disponibles (derecha). Verificando con los datos coincidentes de la boya (arriba) y validando con datos de satélites (abajo).

Finalmente, en la figura 5.23 se muestran dos series temporales de los datos de H_s de SIMAR-44 sin calibrar y calibrados y la serie de datos de la boya de Mahón; para la calibración utilizando cuantiles de los pares de datos coincidentes y para la calibración utilizando todos los datos disponibles. Se puede comprobar el buen ajuste logrado tras las calibraciones, siendo los resultados de la calibración con datos coincidentes más próximos a los registros de la boya, fundamentalmente para los oleajes más energéticos.

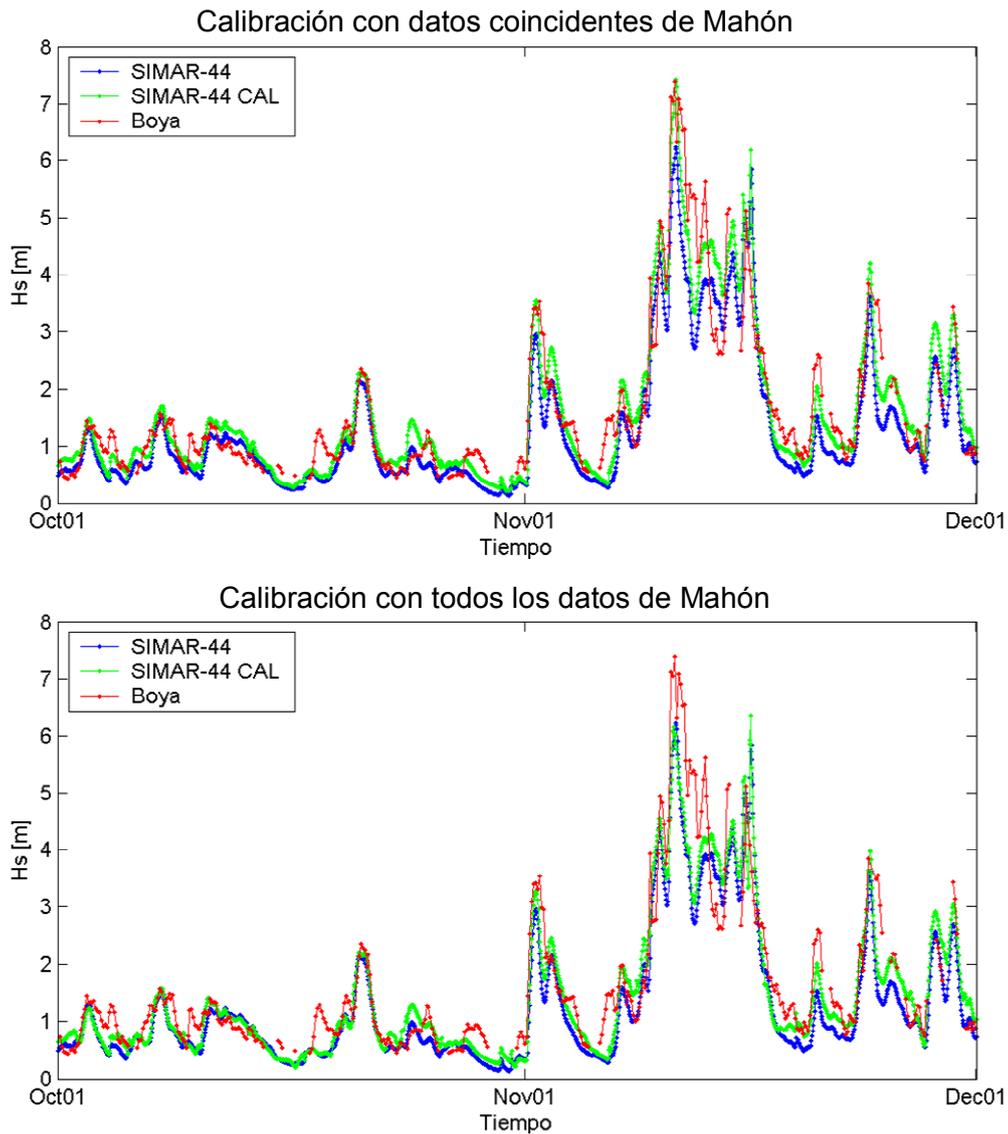


Figura 5.23. Serie temporal de dos meses de datos de H_s de la boya de Mahón y de SIMAR-44 sin calibrar y calibrados con QQ direccional ponderado y MVRExt, utilizando cuantiles de los pares de datos coincidentes (arriba) y con todos los datos disponibles (abajo).

5.4. Conclusiones.

A continuación se resume una serie de conclusiones a las que se ha llegado en el presente capítulo 5:

- La metodología de calibración puntual paramétrica de diagramas de dispersión, o *scatter plot* se define a partir de pares de datos coincidentes, pudiendo estimar el intervalo de confianza de los parámetros de la calibración. Se ha implementado relaciones paramétricas lineales y no lineales, de forma escalar para todos los datos o introduciendo la variación de θ para definir la calibración direccional.

- El método de calibración puntual paramétrico *scatter* no es capaz de corregir las fases temporales de las series de oleaje. Además, como se compara dato a dato, si las series no son coincidentes y sincronizadas, se pueden obtener relaciones de calibración incorrectas.
- El principal problema que afecta a la calibración puntual *scatter*, es el hecho de que no permite corregir la rama medio-alta del régimen del oleaje si tiene una tendencia distinta a la de los valores medios, pues las regresiones se ajustan mejor dónde se acumula la mayor cantidad de información.
- La calibración puntual paramétrica *QQ* define la relación potencial de calibración a partir de comparaciones entre los cuantiles de datos instrumentales y de datos de reanálisis. Debido a que esta metodología de calibración agrega la información para definir los cuantiles, no es necesario que los datos sean coincidentes, aunque sí muy conveniente, sobre todo si se tienen pocos datos. Con esta metodología de calibración se consigue corregir los regímenes medios (intensidad del oleaje), pero si las series temporales de oleaje no están sincronizadas o en fase, no es posible corregir el desfase entre las series. Con esta metodología de calibración se puede calibrar los regímenes extremos, pues introduce la tendencia de corrección del régimen medio-alto para calibrar los datos más extremos. Para esta metodología no se ha definido el intervalo de confianza de los datos calibrados, por lo que no se evalúa la incertidumbre que se comete en la calibración de cada dato.
- La metodología de calibración puntual paramétrica *QQ* se ha aplicado en su versión tradicional, definiendo los cuantiles con todos los datos o sólo con los coincidentes, de forma que se calibra escalarmente la base de datos. También se ha desarrollado una metodología direccional que agrega los datos por direcciones para definir unas relaciones paramétricas direccionales y continuas en θ . Se han aplicado distintos métodos de calibración direccional con distintas ponderaciones de los cuantiles, para corregir lo mejor posible el régimen medio-alto del oleaje.
- Para definir la metodología de calibración puntual *QQ* se han definido una serie de estándares de aplicación, como por ejemplo, la elección de 30 cuantiles equiespaciados en $-\log[-\log(\text{Pr})]$ desde la probabilidad del 15% a la del 99.999%, utilizando únicamente los que verifican $\text{Pr} < 1 - 5/n$. Se ha ajustado siempre relaciones paramétricas potenciales del tipo $Y = bX^c$. Se ha utilizado distintas ponderaciones, pero con la que mejores resultados se han obtenido es utilizando pesos del tipo $W = \sqrt{n \text{Pr}(1 - \text{Pr})}$ e introduciendo un término de MVRExt que minimiza la

variación del mayor cuantil (régimen extremal) de la base de datos sin calibrar. También se ha estandarizado la definición de los sectores direccionales de agregación, tomando sectores de amplitudes de 22.5° centrados cada 11.25° .

- Para las distintas metodologías de calibración puntual direccional, se ha utilizado una parametrización continua en θ de los parámetros de la relación potencial, $b(\theta)$ y $c(\theta)$, que superpone 6 ondas sinusoidales sobre sus valores medios.
- Ambas metodologías de calibración puntual paramétricas, *scatter* y *QQ*, se han aplicado únicamente con los datos de H_s de la boya de Mahón, validando con datos de satélites. Aunque las metodologías presentadas son bastante generales como para poder ser fácilmente aplicables a cualquier variable geofísica.
- Para resolver los complicados problemas de minimización de las funciones objetivo con las que se estiman los parámetros de las relaciones de calibración, ha sido de gran ayuda la implementación de algoritmos de optimización de problemas de alta dimensionalidad como el SCE-UA o el SCEM-UA.

