UNIVERSIDAD DE CANTABRIA

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS

CLIMATE-BASED MONTE CARLO SIMULATION TECHNIQUES FOR COASTAL AND OCEAN APPLICATION

A thesis by:

Yanira Guanche García

Directed by:

Roberto Mínguez Solana

Fernando J. Méndez Incera

SANTANDER, 2013

CLIMATE-BASED MONTE CARLO SIMULATION TECHNIQUES FOR COASTAL AND OCEAN APPLICATION

ISBN: unknown

Imprime: Universidad de Cantabria
Avda. Castros, s/n. 39005 Santander

Impreso en España *Printed in Spain*

A mi familia.

Agradecimientos

Al acabar la tesis son muchas las personas y muchos los momentos de los que te acuerdas. Empezando desde el nivel personal quiero agradecer a mi familia, que son los principales responsables de que a día de hoy esté donde estoy. A mis padres Julián y Nieves, por su esfuerzo y dedicación que nos ha servido a los tres como modelo a seguir. A mis hermanos, Raquel y Raúl, por guiarme e ir abriéndome el camino.

Quiero extender mi agradecimiento a mis directores de tesis, Roberto Mínguez y Fernando Méndez: gracias por haberme dado la oportunidad de trabajar con ustedes, y de haberle dado sentido a esta tesis. Espero haber aprendido un poco de ambos, de Fernando a no perder el sentido físico de lo que se está haciendo y de Roberto a saber sustentarlo matemática y estadísticamente. Ambos forman un equipo con el que es un placer trabajar. Pero no sería justo pensar que esto es trabajo de una, ni siquiera de tres personas. Este trabajo ha sido posible gracias a mucha gente y a sus aportaciones; empezando por todo el grupo de clima: Paula, Melisa, Borja, Antonio, Cristina, Jorge, Ana, Omar, Gabo,... Por todas esas bases de datos que sirven de punto de inicio a tanta gente como yo, por todas las herramientas que han ido desarrollando con el tiempo y por no haber dudado en ayudarme.

En el instituto hay mucha más gente a quien querría transmitir mi agradecimiento, entre todos hacen del IH un buen lugar donde trabajar. A Gabi por su predisposición para siempre echarme una mano y especialmente a Arantza, porque durante todo este tiempo hemos compartido muchas risas y alguna que otra lágrima.

Diese Zeilen sind für Dich, weil Du mir während der Dissertation stets zur Seite standst. Du warst derjenige, der mir selbst in schwierigen Phasen ein Lächeln ins Gesicht zauberte und bist derjenige, mit dem ich das Erreichte teilen möchte. Danke Chris!

Finalmente quiero agradecer a Raúl Medina por hacer que todo esto funcione y a Iñigo Losada por haberme prestado apoyo cuando lo he necesitado.

Yanira Guanche García

Junio 2013

Contents

List of Figures

List of Tables

CHAPTER

0

Resumen

Con el objetivo de lograr una mayor difusión dentro de la comunidad científica, esta tesis doctoral ha sido redactada en inglés. Sin embargo, para cumplir los requisitos impuestos por la Universidad de Cantabria, previo al cuerpo principal de la tesis se adjunta un resumen de la misma en español. En este resumen se intenta sintetizar el contenido de la tesis, pero dada la naturaleza propia de un resumen, este apartado no supone una traducción literal.

I Introducción

El diseño de infraestructuras marítimas, costeras o en aguas profundas; así como el análisis de cualquier proceso costero gobernado por el oleaje requiere del conocimiento del clima marítimo en la zona de estudio. Dependiendo del análisis o el tipo de estructura a diseñar es necesaria la disposición del régimen extremal, el régimen medio o ambos. Por lo tanto la existencia de datos que sirvan al ingeniero como base con la que iniciar sus diseños se torna imprescindible.

Las bases de datos de clima marítimo existentes pueden diferenciarse claramente en 4 tipos:

1. *Registros instrumentales (boyas)*

2. *Datos provenientes de satélite*

3. *Registros de observaciones visuales*

4. *Series generadas mediante modelado numérico (reanálisis)*

Dependiendo del estudio que se esté realizando, o la estructura que se esté diseñando, se pueden utilizar unas u otras bases de datos. Asimismo, hay que tener en cuenta las diferentes escalas temporales en la caracterización del clima marítimo. Los estudios se pueden clasificar en 4 niveles en función de su alcance temporal:

1. *Corto plazo*: estudios que requieren de predicciones del oleaje del orden de 3 días.

2. *Medio plazo*: diseños en los que es necesario un análisis del oleaje para periodos de hasta 6 meses.

3. *Largo plazo*: infraestructuras para las que se precisa de la definición del clima marítimo con registros históricos de larga longitud ($>$20 años).

4. *Muy largo plazo*: estudios en los que en necesario tener en cuenta las proyecciones del clima marítimo bajo los diferentes escenarios de cambio climático.

Esta tesis se centra en estudios de largo o muy largo plazo con la utilización de datos de reanálisis.

I.A **Medidas de fiabilidad estructural**

El diseño de cualquier estructura debe realizarse de tal forma que durante la vida útil de la misma no se superen unos estados límites predeterminados. Los estados límites pueden clasificarse según su grado de relevancia en:

1. *Parada operativa.* Interrupción del uso normal de la estructura producida generalmente por agentes atmosféricos.

2. *Servicio.* Interrupción del uso normal de la estructura por deformaciones, vibraciones o daños superficiales.

3. *Daño.* Paradas que requieran de reparaciones de importancia para evitar el colapso de la estructura.

4. *Último.* Colapso de toda o parte de la estructura.

En función de cómo se se determine la fiabilidad de la estructura, se pueden diferenciar tres tipos de diseño:

1. *Diseño determinista.*

   El diseño determinista forma parte de los métodos clásicos, basados en coeficientes de seguridad, que tienen en cuenta la aleatoriedad de las variables de forma implícita.

   - **Coeficiente de seguridad global**

     Se basa en dividir el espacio n-dimensional de las variables implicadas en dos zonas. Dicha división se hace respecto a un determinado estado límite. En la región segura se cumplen los condicionantes de proyecto mientras que en la región de fallo la estructura deja de cumplir las funciones para las que se diseñó.

     La inclusión de un coeficiente de seguridad F se traduce en un desplazamiento de la curva que separa ambas regiones, de forma que se gana una zona extra de seguridad (ver Figura 1).

   - **Coeficiente de seguridad parcial**

     Los coeficientes de seguridad parciales surgen como evolución del coeficiente de seguridad global. Se basan en la determinación de diferentes coeficientes de seguridad aplicables a las cargas y/o elementos de la estructura. Estos coeficientes se dividen en coeficientes de minoración para las resistencias de la estructura y coeficientes de mayoración para las cargas actuantes.

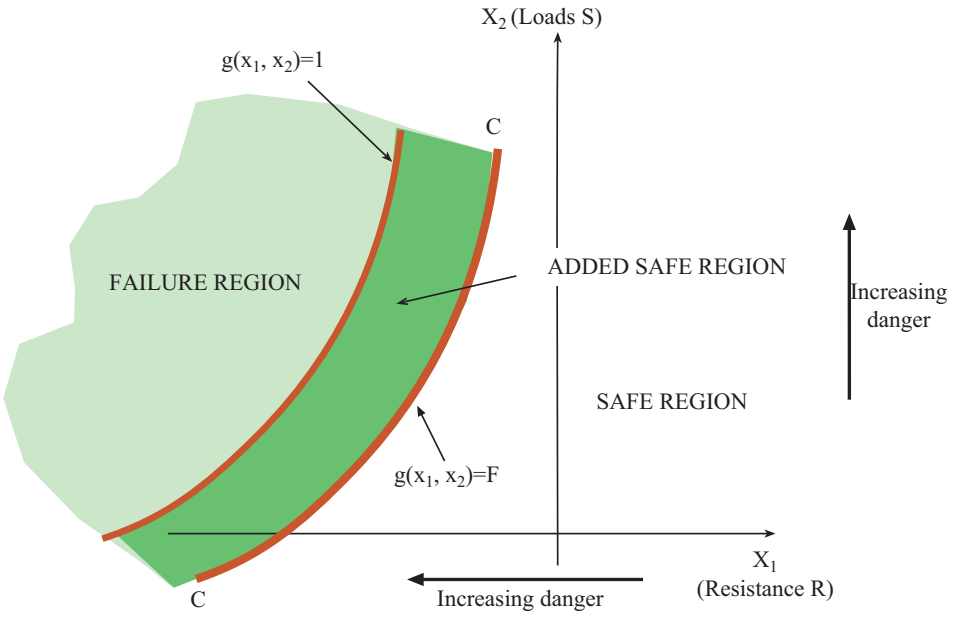

Figure 1: Aumento de la región segura mediante la inclusión de un coeficiente de seguridad

La aplicación de coeficientes de seguridad (totales o parciales) en el diseño de estructuras supone una primera aproximación en la estimación de la fiabilidad estructural. Sin embargo, su aplicación conlleva incertidumbres ya que la determinación de los coeficientes puede no ser única y los valores representativos de las variables aleatorias implicadas pueden variar.

2. *Diseño parcialmente probabilista.*

   Los métodos de diseño parcialmente probabilista son aquellos en los que se recurre al concepto de periodo de retorno. La definición del periodo de retorno depende en gran medida de las hipótesis consideradas sobre las distribuciones de los sucesos. La utilización del periodo de retorno en el diseño de estructuras es válida siempre y cuando se tenga en cuenta que:

   - La definición del periodo de retorno depende de la escala de tiempo utilizada.
   - No se considera la posibilidad de ocurrencia de más de un evento dentro de un mismo intervalo de tiempo. Por ello esta aproximación sólo es

válida cuando los sucesos tengan una probabilidad de ocurrencia muy baja dentro del periodo unidad.

3. *Diseño probabilista.*

   Los métodos explicados anteriormente son herramientas de gran utilidad en el diseño de estructuras, ahora bien, no son capaces de tener en cuenta que incluso para un tiempo definido las variables tienen un nivel de incertidumbre asociado. La incertidumbre asociada a la resistencia R y a cada esfuerzo S se puede representar mediante sus funciones de densidad f_R y f_S. Tanto las cargas actuantes sobre la estructura como la resistencia de la misma pueden ser variables con el tiempo debido al deterioro de los materiales, esto implica que las funciones de densidad se ensanchen con el tiempo. Aún así, en la mayoría de casos se suponen constantes tanto las cargas como las resistencias.

   Al suponer constantes en el tiempo las funciones de densidad de las incertidumbres asociadas a las resistencias $f_R(r)$ y a los esfuerzos $f_S(s)$ es posible determinar la función de densidad conjunta $f_{RS}(r, s)$. Si ambas variables pueden considerarse independientes, la función de densidad conjunta puede estimarse como el producto de ambas marginales: $f_{RS}(r, s) = f_R(r)f_s(s)$. De esta forma, la probabilidad de fallo se puede estimar mediante la ecuación 1

$$p_f = \int_D \int f_{RS}(r, s)drds = \int_{-\infty}^{\infty} \int_{-\infty}^{s} f_{RS}(r, s)drsr \tag{1}$$

   Sin embargo, en la mayoría de diseños de estructuras no es posible simplificar el problema a dos variables R y S, por lo que R y S han de ser sustituidas por las funciones $h_R(x_1, x_2, ..., x_n)$ y $h_S(x_1, x_2, ..., x_n)$, de manera que la función de densidad conjunta resulta:

$$f(\mathrm{x}) = f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n; \Theta), \tag{2}$$

   donde Θ es un vector paramétrico que define las distribuciones de las variables. Con ello, la probabilidad de fallo se obtiene mediante:

$$fp_f(\Theta) = \int_{g^*(x_1,x_2,...,x_n)\leq 0} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n; \Theta)dx_1dx_2...dx_n \quad (3)$$

siendo $g^*(x_1, x_2, ..., x_n)$ la ecuación de estado límite, y que no es más que la integral de la función de densidad conjunta de las variables que intervienen en la región de fallo.

I.B **Métodos con los que determinar la fiabilidad**

En el ámbito del diseño de infraestructuras marítimas, las Recomendaciones de Obras Martimas (R.O.M.) de Puertos del Estado proporcionan un conjunto de normativas y criterios técnicos de aplicación en las cuatro fases en las que se puede dividir la vida útil de una infraestructura marítima:

1. Fase de planificación y diseño.
2. Fase de construcción.
3. Fase de explotación y mantenimiento.
4. Fase de reutilización y desmantelamiento.

La R.O.M. se divide en diferentes capítulos, siendo la *ROM 0.0, Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias*, en el que especificamente se definen los diferentes métodos de verificación y diseño aplicables en el ámbito de las obras marítimas. En este documento se dividen los métodos de fiabilidad en tres niveles, Nivel I, Nivel II y Nivel III. Mediante su aplicación se puede abordar un diseño capaz de garantizar la seguridad, el servicio y la explotación de la infraestructura a lo largo de su vida útil.

- **Nivel I**

  Los Métodos de Nivel I son aquellos en los que la fiabilidad se estima basándose en la determinación de coeficientes de seguridad. Existen dos ramas principales dentro de los métodos de Nivel I, los métodos basados en coeficientes de seguridad globales y los métodos basados en la estimación de coeficientes parciales para cada una de las variables involucradas.

- **Nivel II**

  Los Métodos de Nivel II y III se basan en la estimación de la probabilidad de fallo, p_f. Normalmente, la obtención de una solución analítica de la integral que define p_f no es posible. Mediante los Métodos de Nivel II se trata de obtener una aproximación de la misma transformando el integrando para trabajar con variables gaussianas independientes. De esta forma se pueden estimar los dos primeros momentos de la función de distribución conjunta y con ellos ya se procede al diseño. Esto es posible siempre y cuando las variabes no estén correladas, cuando esto no ocurra previamente habrá que transformarlas en variables independientes ([74]).

  Los Métodos de Nivel I también son conocidos como FOSM (First Order Second Moment). Los métodos FOSM tratan de aproximar la ecuación del estado límite mediante una recta y determinar el punto de diseño como el punto de dicha recta para el cual la distancia sea mínima. La distancia β en la Figura 2 representa el coeficiente de fiabilidad propuesto por Hasofer y Lind ([74]) para determinar si el sistema es seguro o no.

- **Nivel III**

  Finalmente, los Métodos de Nivel III son aquellos que tratan de resolver la ecuación de la p_f mediante técnicas de integración (analítica o numéricamente) o mediante técnicas de simulación. Sólo existe solución analítica de la ecuación anterior para casos concretos, por lo que, normalmente se recurre a técnicas de simulación. Los Métodos de Nivel III basados en estimar la región de fallo pueden dividirse entre métodos FORM (First Order Reliability Methods) y SORM (Second Order Reliability Methods). Por otro lado, los Métodos de Nivel III que consisten en la generación sintética de las variables implicadas están generalmente basados en técnicas de como la de Monte Carlo ([112]).

  Los métodos FORM son una aproximación lineal teniendo en cuenta la distribución real de las variables. El proceso se lleva a cabo de forma similar que en los métodos FOSM pero con la principal diferencia en la forma de transformar las variables al espacio normal multidimensional. En el caso en que

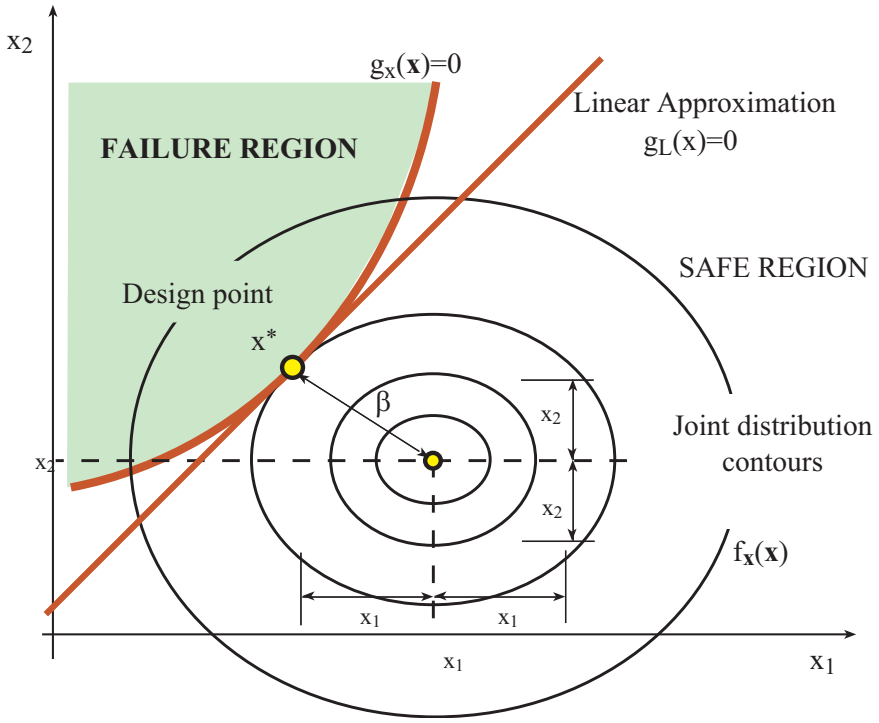

Figure 2: Representación gráfica de superficie de estado límite $G_{\mathbf{X}}(\mathbf{X}) = 0$ y su aproximación lineal para un caso 2-D.

las variables no sean independientes, éstas pueden transformarse mediante la transformación de Rosenblatt ([140]).

Existen casos en los que no es posible aproximar la ecuación de estado límite a una recta, ya sea porque dicha ecuaci ón define una curva muy pronunciada o bien porque al pasar las variables al espacio normal la curvatura de dicha curva se incrementa. En estos casos el problema puede solucionarse mediante la aplicación de mtodos SORM. Estos métodos de segundo orden se basan en la aproximación de las ecuación de estado límite mediante funciones parabólicas o esféricas en el entorno del punto de diseño.

I.C **Objetivos**

El objetivo general de esta tesis es la generación de series sintéticas que sirvan como base para la aplicación al diseño de estructuras marítimas dentro de los

métodos de diseño de Nivel III. Para una correcta simulación de las variables meteo-oceánicas involucradas en el diseño de obras marítimas hay que profundizar en su comportamiento y variabilidad. Asímismo, la gestión de series temporales multivariadas de largo periodo requiere de la utilización de diferentes técnicas estadísticas.

El objetivo general propuesto puede disgregarse en dos objetivos parciales en los que se puede resumir el contenido de esta tesis.

1. Ser capaces de generar series sintéticas multivariadas de las principales variables que definen el oleaje y las condiciones atmosféricas. Para ello se abordan distintas técnicas que hacen posible tener en cuenta las diferentes escalas y los distintos procesos involucrados.

2. Transformar las series temporales de las variables meteo-oceánicas simuladas en series temporales de parámetros de los que dependan los principales modos de fallo de las estructuras.

La consecución de los objetivos propuestos requiere la exploración y el desarrollo de diferentes técnicas y algoritmos y su consecuente validación.

I.D **Metodología**

La metodología general para la generación de series sintéticas aplicables al diseño de estructuras marítimas que se plantea está esquematizada en la Figura 3. Con los datos históricos de oleaje y presiones atmosféricas se simulan un número elevado de vidas útiles de las variables de las cuales depende la estructura. Ésta simulación conlleva varios pasos de manera que se tengan en cuenta las diferentes escalas temporales y espaciales involucradas. Una vez se disponga de series sintéticas de las variables se transfieren éstas a la estructura en forma de parámetros relativos a los principales modos de fallo de cada estructura (estimación de carga actuantes sobre la estructura, erosión,...). Esta transferencia se hace mediante una clusterización de los datos simulados en grupos con características semejantes, una estimación de los parámetros relacionados con los modos de fallos de los datos clusterizados y una posterior interpolación para reconstruir series temporales de los mismos. Con estas

series temporales es posible analizar la integridad estructural de los elementos que componen la infraestructura así como los periodos de retorno de las cargas. En la Figura 3 se señala el alcance de los objetivos parciales planteados previamente para la consecución de la metodología general.

Simulación

Para llevar a cabo la metodología propuesta se precisa de la utilización de diferentes técnicas y algoritmos. Para simular correctamente estados de mar multivariados es necesario conocer las condiciones atmosféricas asociadas, ya que éstas son el princpipal forzador del oleaje. Asímismo, es necesaria la inclusión de la estacionalidad y el carácter autorregresivo tanto del oleaje como de las presiones atmosféricas. Es por ello que la simulación de estados de mar se hace en tres pasos, los cuales están esquematizados en la Figura 4:

1. Se simulan las presiones atmosféricas en la zona de estudio. De esta forma se tiene en cuenta la escala espacial de influencia en el oleaje al cual está sometida la estructura a diseñar.

2. Se simulan las condiciones medias de oleaje diario frente a la estructura. En este paso la escala espacial se reduce a la localización exacta de la obra.

3. Se simulan los estados de mar horarios en la infraestructura. Con este tercer paso se aumenta la resolución temporal de la simulación.

Transferencia

Por otro lado, una vez se dispone de series sintéticas de estados de mar en la zona de localización de la obra se estiman las cargas y/o parámetros de los que van a depender los modos de fallo de la obra. Este proceso se sintetiza en tres pasos:

1. Clasificación de los estados de mar horarios en un número determinado de patrones representativos (clusterización).

2. Estimación de las cargas producidas por los patrones representativos. Esta estimación puede realizarse mediante el uso de formulaciones semiempíricas, modelado numérico o modelado físico.

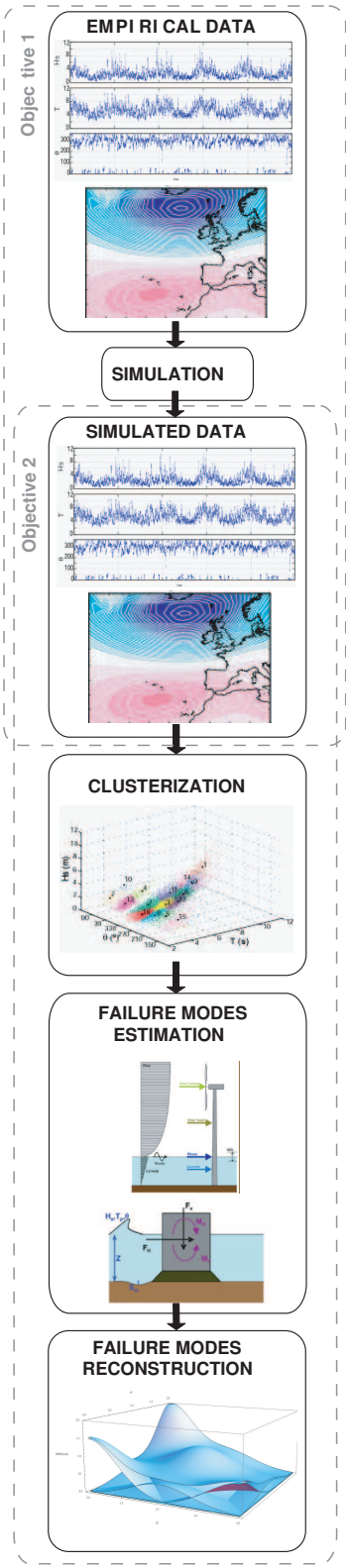

Figure 3: Esquema de la metodología.

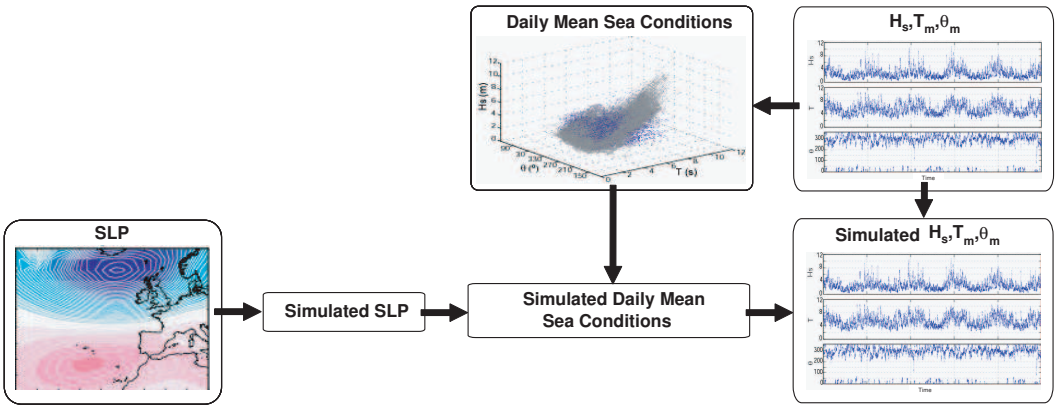

Figure 4: Esquema del proceso de simulación.

3. Reconstrucción de las series temporales de cargas o parámetros que hayan sido estimados previamente mediante una técnica de interpolación.

La metodología expuesta se desarrolla tomando como base de las simulaciones las distribuciones empíricas de las variables. Esto supone una limitación al no tener en cuenta el diferente comportamiento para los regímenes medio y extremal y al no dejar libertad de simulación en los extremos. Este aspecto se intenta solventar con el desarrollo de una técnica de simulación univariada que tiene en cuenta dicha diferencia de comportamiento y es capaz de reproducir bien tanto los eventos extremos con el régimen medio. Asímismo, la cola superior se ajusta a un modelo de extremos de forma que en el proceso de simulación los eventos simulados no estén costreñidos a los máximos registrados históricamente.

En este trabajo de tesis se explica en detalle cada uno de los pasos expuestos en esta metodología y que son necesarios para la consecución de los objetivos planteados en el apartado anterior.

La tesis está compuesta por 7 capítulos con los que se profundiza en los objetivos anteriormente especificados. El cuerpo central de la tesis está compuesto por los Capítulos (2-6) que se corresponden con una serie de artículos ya publicados o en proceso de revisión. Cada artículo ha sido modificado sólamente en cuestión de formato para poder adaptarlo al presente documento.

La secuencia de los mismos traza el camino a seguir planteado en el apartado anterior. Los Capítulos 2 y 3, solventan el problema de la simulación multiva-

riada de variables meteo-oceánicas. Para ello inicialmente se desarrolla un modelo logístico capaz de simular sistemas multivariados agregados (Capítulo 2). A continuación, con base en el modelo desarrollado previamente y una técnica existente en la literatura ([120]) se desarrolla un método de simulación de Monte Carlo de estados de mar trivariados condicionado a las condiciones atmosféricas (Capítulo 3). Por otro lado, los Capítulos 4 y 5 presentan un modelo híbrido con el que estimar de una forma eficaz los parámetros de diseño de la estructura marítima a partir de las series sintéticas de oleaje creadas previamente, ahorrando carga computacional. Este modelo se aplica a dos tipos de estructuras diferentes: un dique vertical (Capítulo 4) y una turbina eólica offshore (Capítulo 5). Finalmente, el Capítulo 6 explora la posibilidad de tratar conjuntamente los regímenes medio y extremal en las variables de diseño. Por último, en el Capítulo 7 se recogen las conclusiones obtenidas al trabajo realizado y se plantean futuras líneas de investigación.

II Modelo logístico autorregresivo aplicado a patrones de circulación atmosférica

La utilización de modelos logísticos autorregresivos no es un tema nuevo. Estos modelos son conocidos dentro del ámbito de la investigación farmacológica y médica. En el capítulo 2 se profundiza en la aplicación de este tipo de modelos al campo de la modelización de patrones de circulación atmosférica o también llamados tipos de tiempo. El uso de estos modelos permite la consideración simultánea de procesos de diferente naturaleza, como pueden ser la estacionalidad, la variabilidad interanual, las tendencias de largo plazo y procesos autorregresivos.

II.A Modelo logístico autorregresivo

Los modelos de regresión tracidionales además de asumir respuestas regidas por distribuciones normales, no son capaces de tratar con estados categóricos, es por eso que se recurre al uso de modelos logísticos. Asímismo, mediante el correcto tratamiento de las covariables (variables predictoras) en dichos modelos, es posible la inclusión de estados previos del sistema, lo cual es de gran utilidad en procesos claramente autorregresivos como la circulación atmosférica. Para la aplicación del modelo a patrones de circulación atmosférica, los campos de presiones son promediados diariamente y clasificados según el algoritmo de clasificación K-Medias en n_{wt} grupos. Asignando arbitrariamente un valor a cada tipo de tiempo se obtiene la secuencia temporal de los patrones de circulación diarios.

Sea $Y_t;\ t = 1, \ldots, n$ el patrón de circulación para cada instante t, de tal forma que los posibles valores de Y_t estén entre 1 y n_{wt}; y considerando $\boldsymbol{X}_t;\ t = 1, \ldots, n$ como un vector de covariables para cada instante t, el modelo toma la forma:

$$\ln\left(\frac{\text{Prob}(Y_t = i|Y_{t-1}, \ldots, Y_{t-d}, \boldsymbol{X}_t)}{\text{Prob}(Y_t = i^*|Y_{t-1}, \ldots, Y_{t-d}, \boldsymbol{X}_t)}\right) = \alpha_i + \boldsymbol{X}_t\boldsymbol{\beta}_i + \textstyle\sum_{j=1}^{d} Y_{t-j}\gamma_{ij}; \quad \forall i = 1, \ldots, n_{wt}|i \neq i*, \tag{4}$$

donde $\boldsymbol{\alpha}_i$ es un término constante, $\boldsymbol{\beta}_i$ $(n_c \times 1)$ y γ_{ij} $(j = 1, \ldots, d)$ se corresponden, para cada patrón de circulación i, con el vector asociado a las covariables y a

los d estados previos considerados, respectivamente. De acuerdo con la expresión anterior, la probabilidad condicionada de cualquier tipo de tiempo se define por:

$$\text{Prob}(Y_t = i|Y_{t-1},\ldots,Y_{t-d},\boldsymbol{X}_t) = \frac{\exp\left(\alpha_i + \boldsymbol{X}_t\boldsymbol{\beta}_i + \sum_{j=1}^{d} Y_{t-j}\gamma_{ij}\right)}{\sum_{k=1}^{n_{wt}} \exp\left(\alpha_k + \boldsymbol{X}_t\boldsymbol{\beta}_k + \sum_{j=1}^{d} Y_{t-j}\gamma_{kj}\right)};$$

$$\forall i = 1,\ldots,n_{wt}. \tag{5}$$

Como covariables de influencia en el modelo se puede considerar:

- Estacionalidad en términos harmónicos

$$\pi^S = \beta_0^S + \beta_1^S \cos(wt) + \beta_2^S \sin(wt), \tag{6}$$

- Diferentes covariables como la variabilidad interanual

$$\pi^C = \boldsymbol{X}\boldsymbol{\beta}^C = (X_1,\ldots,X_{n_c}) \begin{pmatrix} \beta_1^C \\ \vdots \\ \beta_{n_c}^C \end{pmatrix} = \sum_{i=1}^{n_c} X_i\beta_i^C, \tag{7}$$

- Tendencias de largo plazo

$$\pi^{LT} = \beta^{LT}t, \tag{8}$$

- Términos autorregresivos

$$\pi^{AR_d} = \sum_{j=1}^{d} Y_{t-j}\gamma_j, \tag{9}$$

Introduciendo estas covariables en el modelo, se obtiene:

$$\text{Prob}(Y_t = i|Y_{t-1},\ldots,Y_{t-d},\boldsymbol{X}_t) = \frac{\exp\left(\pi_i^S + \pi_i^C + \pi_i^{LT} + \pi_i^{AR}\right)}{\sum_{k=1}^{n_{wt}} \exp\left(\pi_k^S + \pi_k^C + \pi_k^{LT} + \pi_k^{AR}\right)};$$

$$\forall i = 1,\ldots,n_{wt}. \tag{10}$$

Para poder lidiar con las diferentes escalas temporales en el modelo: anual, mensual y diaria; se transforman todos los parámetros a la escala de menor duración temporal, en nuestro caso se reduce todo a la escala diaria. Las covariables con escalas mayores se repiten diariamente, por ejemplo una covariable con escala mensual se considera como 30 días con el mismo valor de dicha covariable.

II.B **Caso práctico**

Para mostrar la aplicabilidad del modelo propuesto, una vez se explica el tratamiento requerido para cada tipo de covariable, en el capítulo 2 se plantea y resuelve detalladamente un ejemplo práctico. Dicho ejemplo se trata de la modelización y posterior simulación de patrones de circulación atmosférica en el Atlántico Norte. Los pasos seguidos pueden sintetizarte en:

1. Clasificación de las presiones atmosféricas en patrones sinópticos de circulación.

   Esta clasificación se realiza mediante la técnica de K-medias. En este caso, para facilitar la representación gráfica y su consecuente comprensión, se clasificaron las presiones atmosféricas diarias en 9 tipos de tiempo, $n_{wt} = 9$. En la Figura 5 se muestran los 9 tipos de tiempo representativos de los campos de presiones en el Atlántico Norte.

2. Definición de las variables de influencia en el modelo.

   En el caso propuesto se analizó la influencia de: i) estacionalidad, definida tanto por harmónicos como por un proceso autorregresivo de periodo anual, ii)variabilidad interanual, incluyendo las anomalías de presiones mensuales calculadas respecto a su valor medio, iii) tendencias de largo plazo y iv) procesos autorregresivos.

   En la Figura 6, se muestra el ajuste del modelo teniendo en cuenta como variables de influencia la estacionalidad, la variabilidad interanual, la tendencia de largo plazo y el primer orden del proceso autorregresivo (estado anterior al estado actual). Mediante las barras de colores se muestra la probabilidad de ocurrencia observada en el registro histórico de cada uno de los 9 patrones de

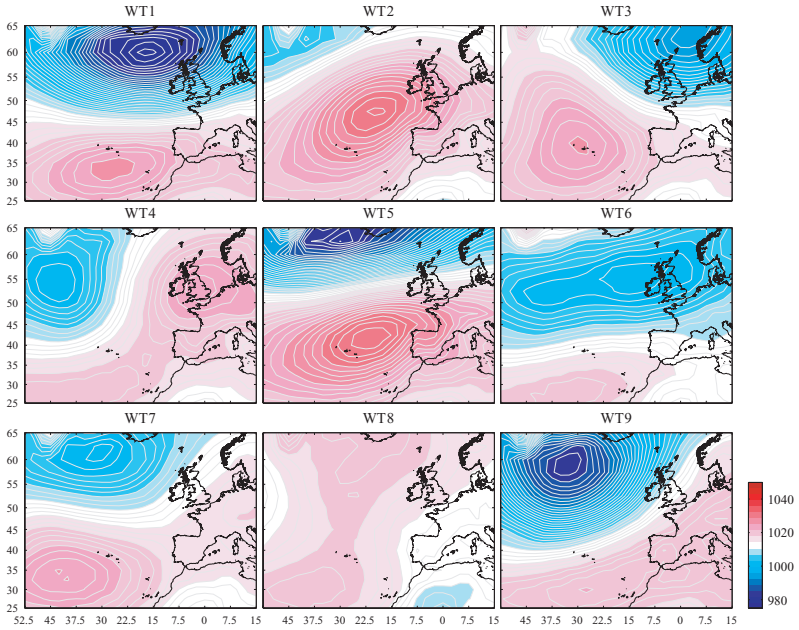

Figure 5: Patrones de circulación sinópticos obtenidos con la clasificación K-Medias.

circulación en los que se han clasificado las presiones atmosféricas, mientras que las líneas negras representan las probabilidades estimadas por el modelo. Se representan las probabilidades agregadas mensualmente a lo largo de un periodo de 20 años.

3. Una vez definidas las variables de influencia significativa en el modelo, es posible la realización de simulaciones de Monte Carlo de diferentes escenarios.

4. La validación de las simulaciones, y por tanto de la habilidad del modelo para reproducir patrones de circulación atmosférica, se centra en tres aspectos: i) correcta reproducción de la distribución de ocurrencias de cada uno de los patrones (Figura 7 (a)), ii) capacidad de simulación de las transiciones entre los diferentes grupos (Figura 7 (b)) y iii) análisis de las persistencias de cada grupo (Figura 8).

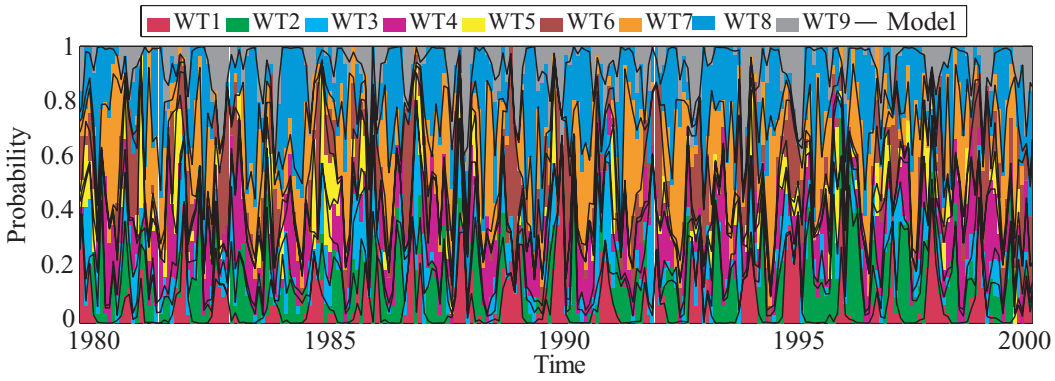

Figure 6: Ajuste del modelo logístico autorregresivo teniendo en cuenta la estacionalidad, la variabilidad interanual, la tendencia de largo plazo y un proceso autoregresivo de orden 1.

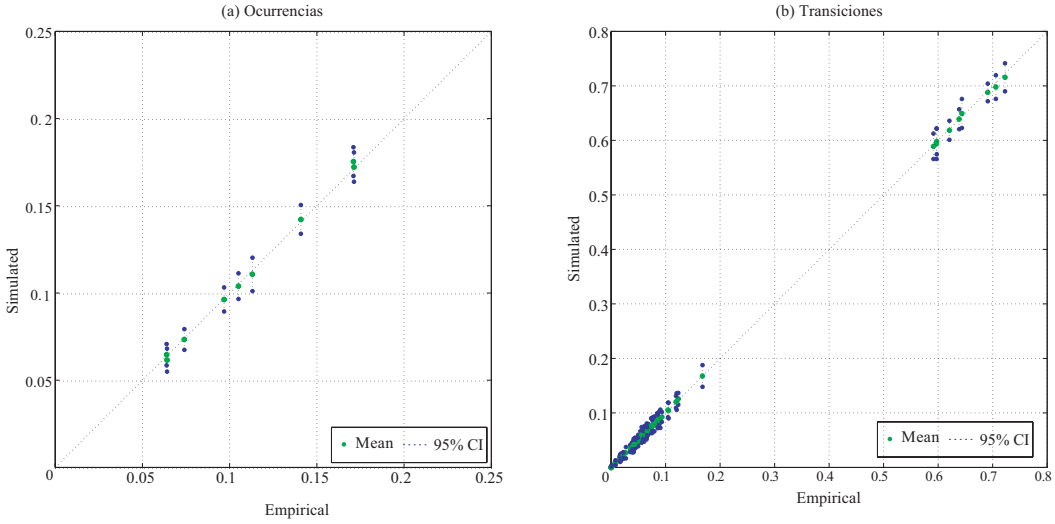

Figure 7: (a) Scatter plot de las probabilidades de ocurrencia de cada tipo de tiempo empíricas vs. simuladas. (b) Scatter plot de las probabilidades de transición entre tipos de tiempo empíricas vs. simuladas.

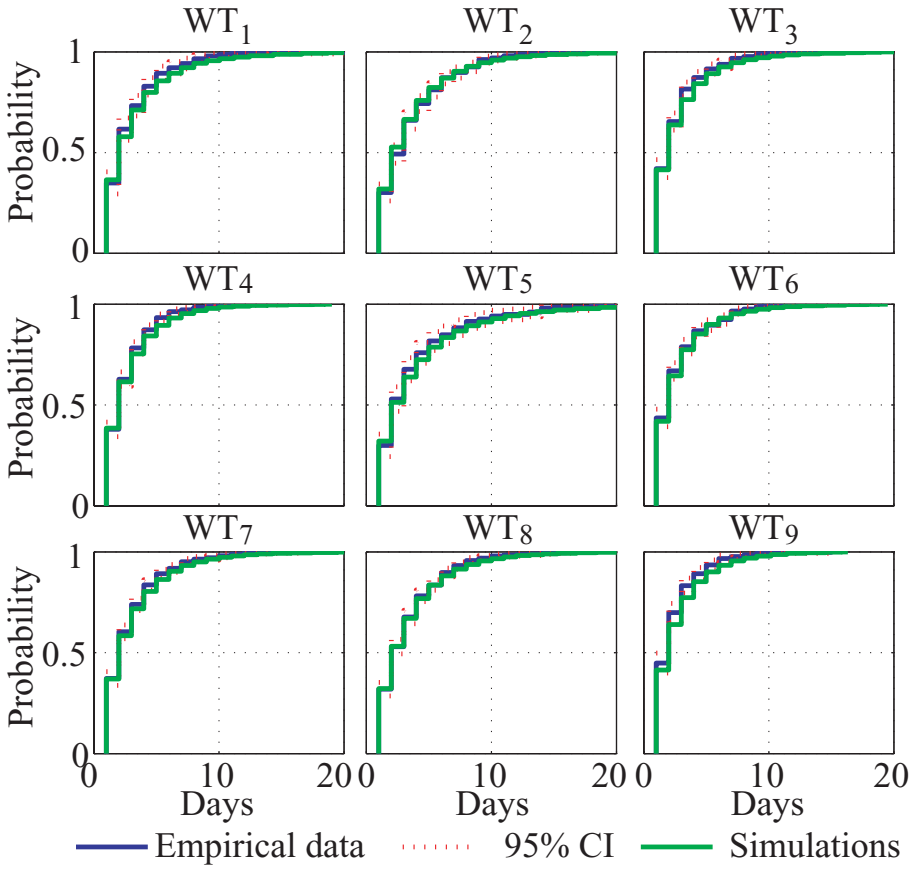

Figure 8: Distribución acumulada empírica de las persistencias de cada uno de los 9 tipos de tiempo: datos históricos vs. simulados.

III Simulación de Monte Carlo de estados de mar trivariados condicionada a la circulación atmosférica

Una correcta caracterización del clima marítimo es vital para entender todos aquellos procesos gobernados por el oleaje, así como para el diseño de cualquier estructura costera u offshore. Allí donde la disposición de bases de datos de largo plazo no sea posible, la generación de series de datos simulados se presenta como una buena alternativa. En el Capítulo 3 se presenta una metodología para simular series de estados de mar trivariados a escala horaria que sean capaces de mantener las características estadísticas de los datos históricos.

III.A Metodología

En la Figura 9 se esquematiza la metodología llevada a cabo. El proceso completo puede dividirse en tres pasos interconectados: i) simulación de campos de presiones diarias, ii) simulación de condiciones medias diarias de oleaje y iii) simulación de estados de mar horarios condicionados a las condiciones medias diarias previamente simuladas.

Para el desarrollo de la metodología propuesta es necesario disponer de datos de campos de presiones (SLP) y de oleaje. Los campos de presiones son el predictor del oleaje en el punto objetivo, siendo este último el predictando. Asímismo se necesitan series temporales de las variables que van a ser simuladas en la misma localización.

El primer paso del proceso es la simulación de los campos de presiones agregados diariamente. Los campos de presiones medias diarias, descompuestos en componentes principales se simulan mediante una técnica de simulación multivariada ([120]). Esta técnica permite tener en cuenta tanto la autocorrelación como las correlaciones cruzadas entre variables, por eso es necesaria la descomposición en componentes principales.

Los datos de oleaje se analizan con dos escalas temporales distintas. Por un lado se agregan a una escala diaria para poderlos relacionar con los campos de presiones simulados previamente, y por otro lado se simulan a escala horaria de modo que no se pierde información.

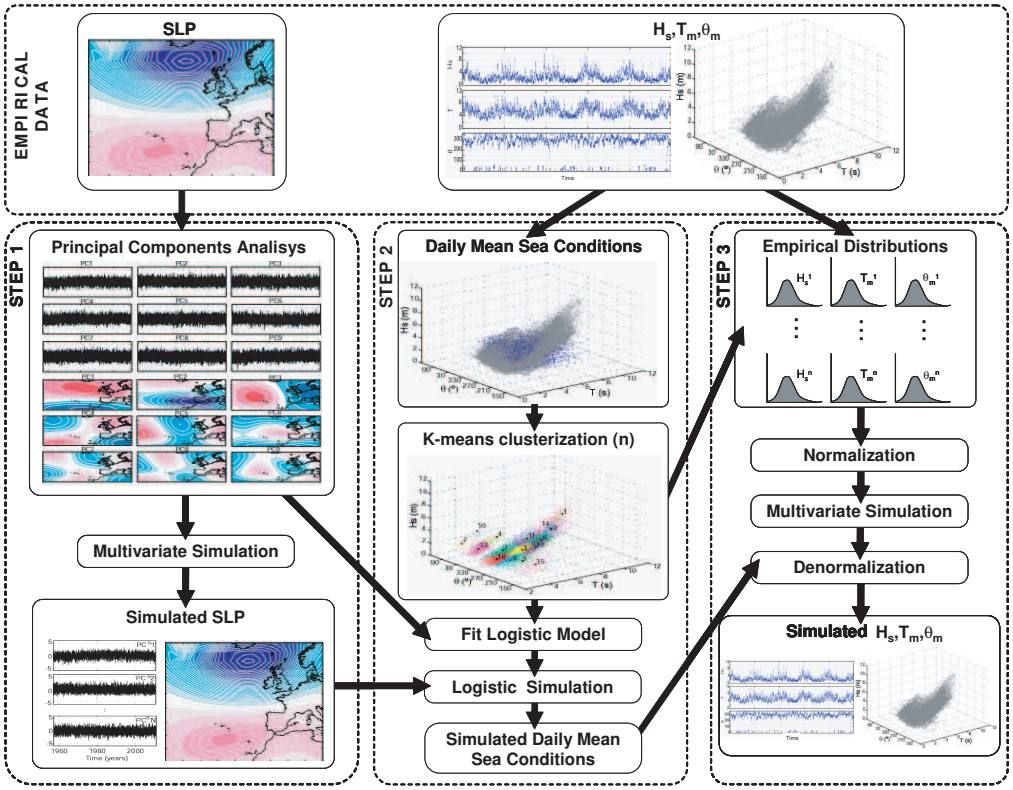

Figure 9: Esquema de la metodología.

Por lo tanto, como segundo paso se simulan los datos agregados diariamente. Esta simulación se lleva a cabo con el modelo logístico presentado en el Capítulo 2. Para ello, previamente el oleaje se clasifica según estados de mar representativos mediante un algoritmo de clasificación como es K-Medias. Una vez el modelo se ajusta, tomando como covariables explicativas las componentes principales de los campos de presiones simulados en el paso anterior, es posible realizar simulaciones de las condiciones medias diarias del oleaje.

Finalmente, como tercer paso del proceso se simulan las variables que definen el oleaje con una resolución temporal horaria. Previo a la simulación se relacionan las series horarias con los estados medios diarios históricos. Con esto se obtienen distribuciones empíricas de H_s, T y θ para cada tipo de condición media diaria. Usando dichas distribuciones empíricas se normalizan las variables. Con ello se

posibilita la utilización de la misma técnica de simulación multivariada usada en el primer paso ([120]). Esta técnica tiene en cuenta las autocorrelación existente en H_s, T y θ y las correlaciones cruzadas entre ellas. Una vez simuladas las tres variables, se desnormalizan teniendo en cuenta las series simuladas de condiciones medias diarias generadas en el paso 2.

III.B **Caso práctico**

En el capítulo 3 se explica detalladamente la metodología propuesta anteriormente mediante su aplicación a un caso práctico. La localización del caso práctico se sitúa en un punto frente al dique de Langosteira, en el puerto de La Coruña. Para ello se utilizan datos prodecentes de reanálisis tanto para las series de oleaje, [19] y [136], como para los campos de presiones, [90].

Simulación de campos de presiones diarias

Los campos de presiones medios diarios, descompuestos en componentes principales (14 componentes principales que representan más del 92% de la varianza), son simulados mediante la técnica de simulación multivariada propuesta por [120]. Para ello, se normalizan las componentes principales y para cada una de ellas se ajusta un modelo de autoregresión y media móvil, ARMA, univariado. Se estudia la posible correlación cruzada existente entre los residuos derivados de estos modelos, tanto simultáneamente como con distintos desfases entre las series. Con las correlaciones existentes entre residuos se determina la matriz de varianza-covarianza, G. A continuación se generan residuos aleatorios, los cuales tras correlacionarlos mediante la matriz G, son los que se introducen en la simulación de cada una de las componentes principales.

En la Figura 10 se muestra la comparación entre las distribuciones de probabilidad empíricas y simuladas para cada una de las componentes principales

Simulación de condiciones medias diarias de oleaje

Para la simulación de condiciones medias diarias de oleaje, las series históricas de H_s, T y θ son promediadas a escala diaria y agrupadas en 16 grupos mediante K-medias (Figura 11); con esto se obtiene la secuencia temporal de patrones de oleaje medio diario. Esa secuencia temporal se modeliza con el modelo logístico

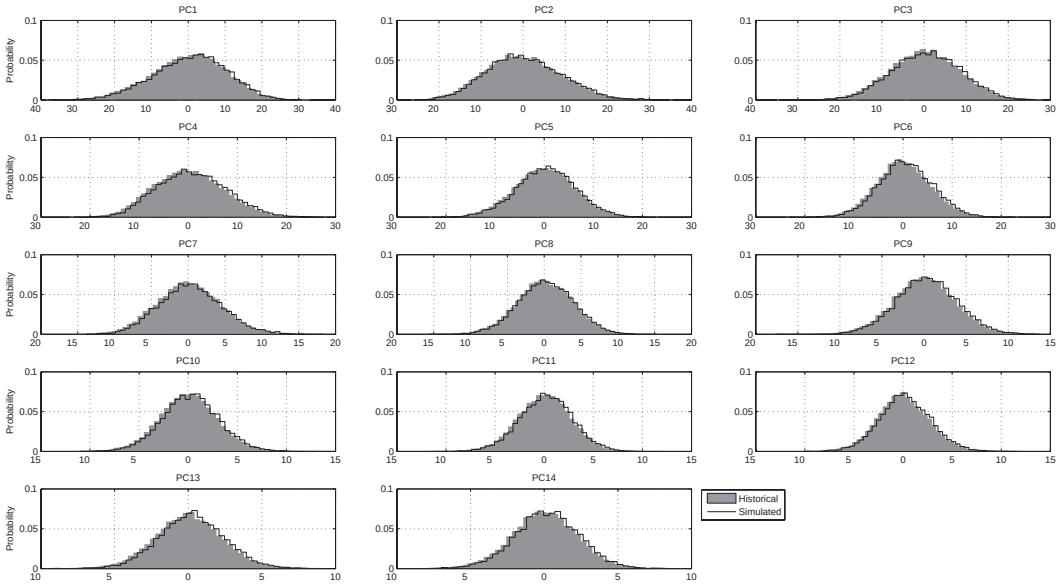

Figure 10: Distribuciones de probabilidad de cada PC: i) datos históricos (barras grises) y ii) datos simulados (líneas negras).

explicado en el capítulo anterior y teniendo en cuenta como covariables las componentes principales históricas de los campos de presiones. En la Figura 12 se muestran las probabilidades de ocurrencia de los 16 grupos de oleaje agregados a un año medio, en la parte superior se presentan los datos historicos mientras que en la parte inferior los ajustados con el modelo.

Una vez el modelo está ajustado se pueden realizar simulaciones de diferentes secuencias de condiciones medias diarias de oleaje. Para estas simulaciones se

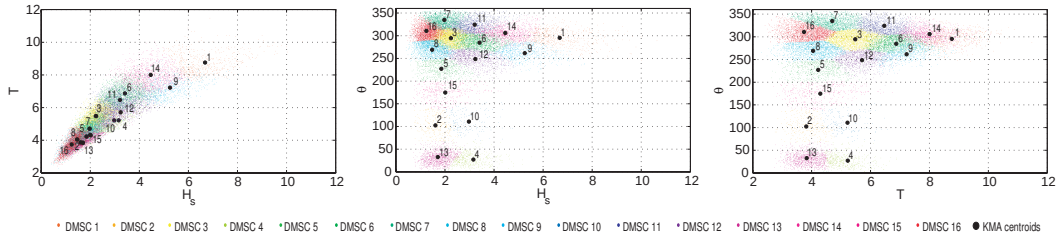

Figure 11: Clasificación de las condiciones medias diarias de oleaje mediante K-medias.

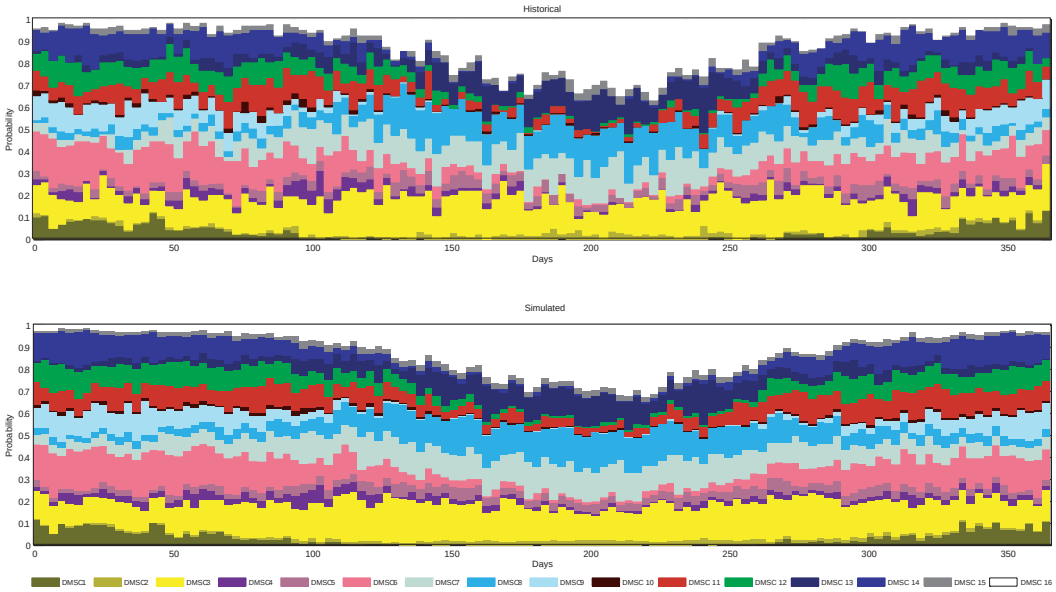

Figure 12: Gráfico de diagnóstico del ajuste del modelo.

toman como covariables las componentes principales simuladas en el paso anterior.

Simulación de estados de mar horarios condicionados a las condiciones medias diarias

Finalmente, se simulan los estados de mar horarios condicionados al oleaje medio diario. Se relacionan la serie temporal de condiciones de oleaje diario con las series horarias de las 3 variables. Con esto se obtienen las distribuciones empíricas de las tres variables para cada uno de los 16 grupos de condiciones medias diarias. Según estas distribuciones empíricas se normalizan las tres variables y una vez normalizadas se ajusta un ARMA para cada una de ellas. Se estiman los residuos, se correlacionan y se determina la matriz de varianzas-covarianzas, G. A continuación, con residuos generados aleatoriamente y posteriormente correlados con G, se simulan series de las tres variables a través de los ARMAs ajustados. Finalmente las variables simuladas se desnormalizan teniendo en cuenta la secuencia simulada en el paso 2 de condiciones medias diarias de oleaje y las distribuciones empíricas.

Las Figuras 13-15 representan la comparación entre los resultados simulados y las series históricas. En la Figura 13 se muestran las comparaciones entre las

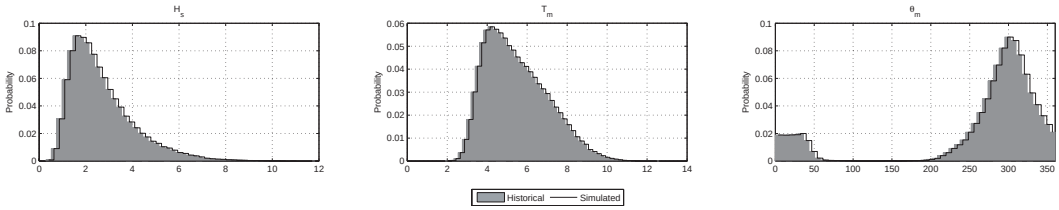

Figure 13: Distribuciones de probabilidad de H_s, T_m y θ_m : i) datos históricos (barras grises) y ii) datos simulados (líneas negras).

distribuciones de probabilidad, mientras que en la Figura 14 se comparan las distribuciones conjuntas. Finalmente, para comprobar si las series simuladas cumplen la estacionalidad inherente en el oleaje, la Figura 15 muestra cada una de las tres variables agregada en un año. Como se puede ver en las tres figuras, las series simuladas se corresponden con las series originales, lo cual valida la metodología propuesta.

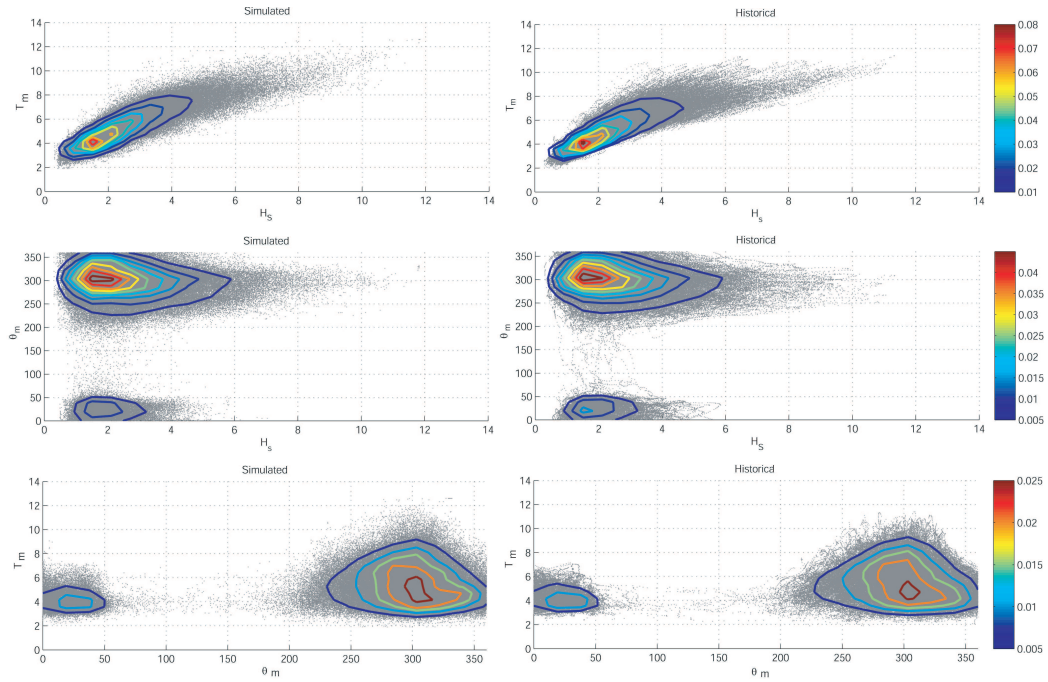

Figure 14: Comparación entre las distribuciones conjuntas históricas y simuladas.

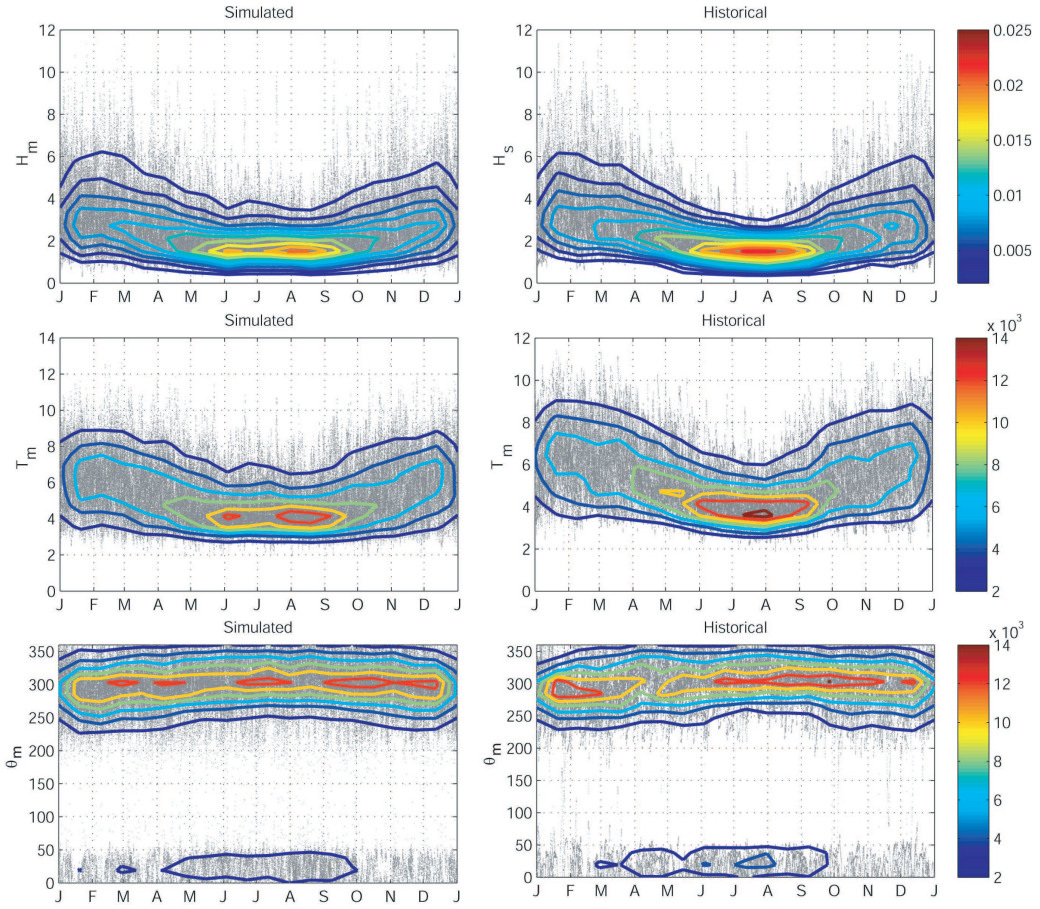

Figure 15: Comparación entre la estacionalidad de las series históricas y simuladas.

IV Metodología para transferir variables meteo-oceánicas a parámetros de diseño de estructuras

El diseño de cualquier obra marítima, costera u offshore requiere del cálculo de las cargas a las que se supone va a estar sometida la obra durante su vida útil. El diseñõ clásico se hace basado en un estado de mar de diseño, pero sin embargo la combinación simultánea de diferentes parámetros puede producir otras situaciones críticas que pueden poner en riesgo la estabilidad de la obra. Las técnicas de mayor resolución, basadas en simulación de escenarios mediante modelos numéricos o modelado físico son muy costosas y requieren mucho tiempo.

En los Capítulos 4 y 5 se presenta una metodología híbrida con la que transformar las series temporales de las variables meteo-oceánicas en series temporales de los parámetros de los que va a depender el diseño. De esta forma se pueden obtener series temporales de los cargas y/o parámetros relacionados con los modos de fallo de la estructura. Se presentan dos casos de aplicación, un dique vertical y una turbina eólica.

IV.A Metodología

La metodología propuesta combina dos técnicas estadísticas conocidas. Primero con ayuda de una algoritmo de clasificación de máxima dissimilitud (MDA), se determinan los casos representativos de las variables. Para ese conjunto de casos respresentativos se estiman las cargas, ya sea por medio de formulaciones semiempíricas, modelado numérico o modelado físico. Y por último, mediante una técnica de interpolación basada en funciones radiales (RBF), se obtienen las series temporales de las cargas o parámetros que hayan sido estimados.

El algoritmo de clasificación MDA, comparado con otros algoritmos existentes como pueden ser K-Medias o mapas autoorganizativos (SOM), es el único capaz de seleccionar casos que se encuentran en los contornos del espacio multidimensional de los datos. En la Figura 16 se muestra una comparación entre las tres técnicas mencionadas.

Por su parte, la interpolación que se realiza tras la estimación de los parámetros se lleva a cabo mediante RBF. Esta técnica consiste en aproximar el valor de los

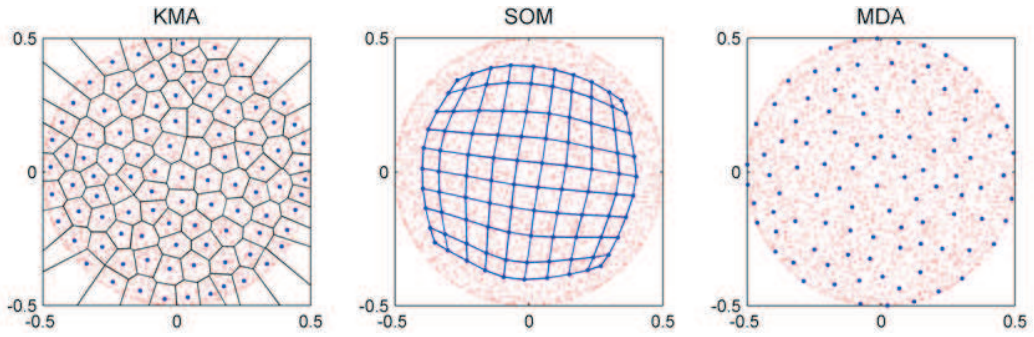

Figure 16: Comparación de los algoritmos K-medias (KMA), mapas auto–organizativos (SOM) y máxima disimilitud (MDA).

parámetros mediante una suma ponderada de funciones radialmente simétricas localizadas en los puntos de los cuales se dispone de dato empírico. Los puntos en los que se encuentran las RBF son los puntos que se han seleccionado previamente mediante el MDA.

IV.B **Caso práctico 1. Dique vertical.**

En el capítulo 4 se presenta la metodología explicada anteriormente aplicada al caso de un dique vertical. En el mismo se consideran estados de mar definidos por su altura de ola, H_s, periodo de pico T_p, dirección media del oleaje, θ_m y nivel medio del mar, Z_m. Se utilizaron datos provenientes de las bases de reanálisis ([136]) con una cobertura temporal de 60 años horarios.

En función de las variables mencionadas se estiman las cargas a las que está sometido el dique según el método de Goda-Takahashi ([63], [157]) y la erosión frente al mismo según [81].

Para poder validar los resultados obtenidos con la metodología propuesta por un lado se estimaron las cargas y la erosión sólo de los estados de mar seleccionados mediante MDA, y por otro se estimaron también para la serie completa. De esta forma se pudo hacer una comparación directa entre los resultados históricos y los obtenidos tras la interpolación con RBF.

En la Figura 17 se puede ver la evolución del error cometido en la comparación directa entre el percentil del 99% de F_h y F_v en función del número de estados

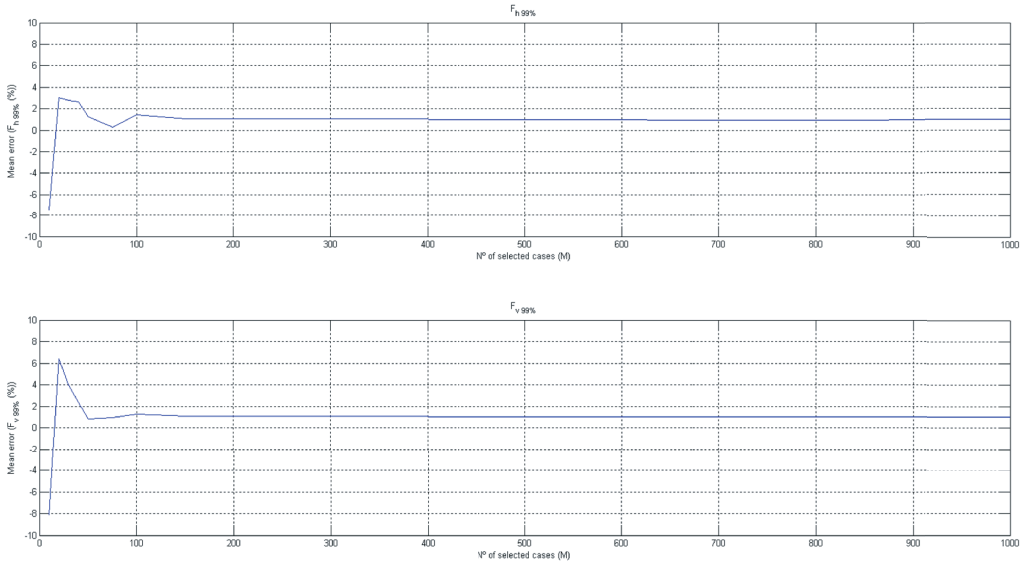

Figure 17: Evolución del error del percentil del 99% de F_h y F_v variando el número de casos seleccionados.

seleccionados con MDA y con los que a posteriori se reconstruyó la serie con RBF.

Finalmente, en la Figura 18 se muestra los resultados obtenidos de la comparación de erosión calculada empíricamente para toda la serie temporal y la erosión estimada aplicando la metodología propuesta, para 50, 100, 200 y 500 casos seleccionados con MDA.

IV.C **Caso práctico 2. Turbina eólica.**

En el capítulo 5 se presenta una extensión del caso práctico anterior aplicando la metodología propuesta al caso de una turbina eólica. Se calculan las cargas en el mástil de la turbina relacionadas con agentes ambientales. Para ello se consideran las siguientes variables meteo-oceánicas: altura de ola significante, H_s, periodo medio, T_m, dirección del oleaje, θ_{Waves} magnitud y dirección del viento V_{1-hour} y θ_{Wind}, magnitud y dirección de las corrientes de marea ,U_{Tidal} y θ_{Tidal}, y nivel medio del mar, SWL. Los datos han sido extraídos de bases de datos de reanálisis ([19],[111],[6]) con una cobertura temporal de 20 años y resolución horaria.

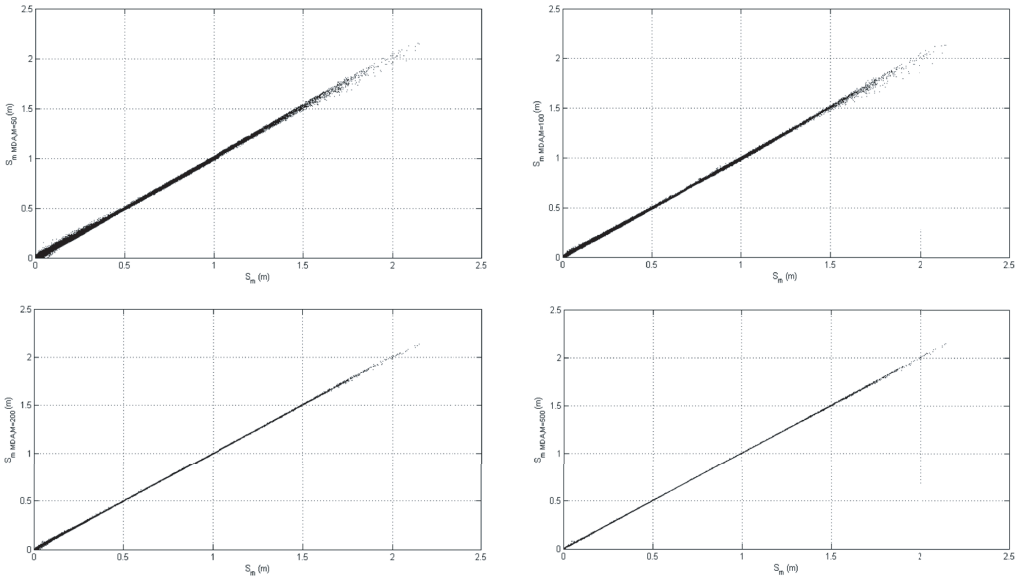

Figure 18: Erosión calculada empíricamente vs estimada según la metodología propuesta con 50, 100, 200 y 500 casos seleccionados.

Para estimar las cargas sobre el mástil de la turbina se aplica la normativa IEC 61400–3 [4].

Al igual que en el caso práctico anterior, se han estimado las cargas paralelamente para toda la serie y sólo a los casos seleccionados con MDA y posteriormente se interpola con RBF. Con esto es posible realizar una comparación directa de los resultados obtenidos con la metodología.

En la Figura 19 se muestra la comparación de los resultados obtenidos para el cálculo de F_u (componente X) y F_v (componente Y) por los dos métodos, empíricamente y mediante el uso de la metodología propuesta.

Con la disposición de series temporales de las cargas sobre el mástil de la turbina eólica es posible analizar los esfuerzos a los que se ve sometida la misma desde diferentes puntos de vista. En la Figura 20 se representan las rosas probabilísticas de los esfuerzos producidos por los diferentes agentes (olas, viento y corriente) en comparación con el esfuerzo total en todo el periodo de 20 años. Mientras que en la Figura 21 se muestran los esfuerzos totales en tres años consecutivos.

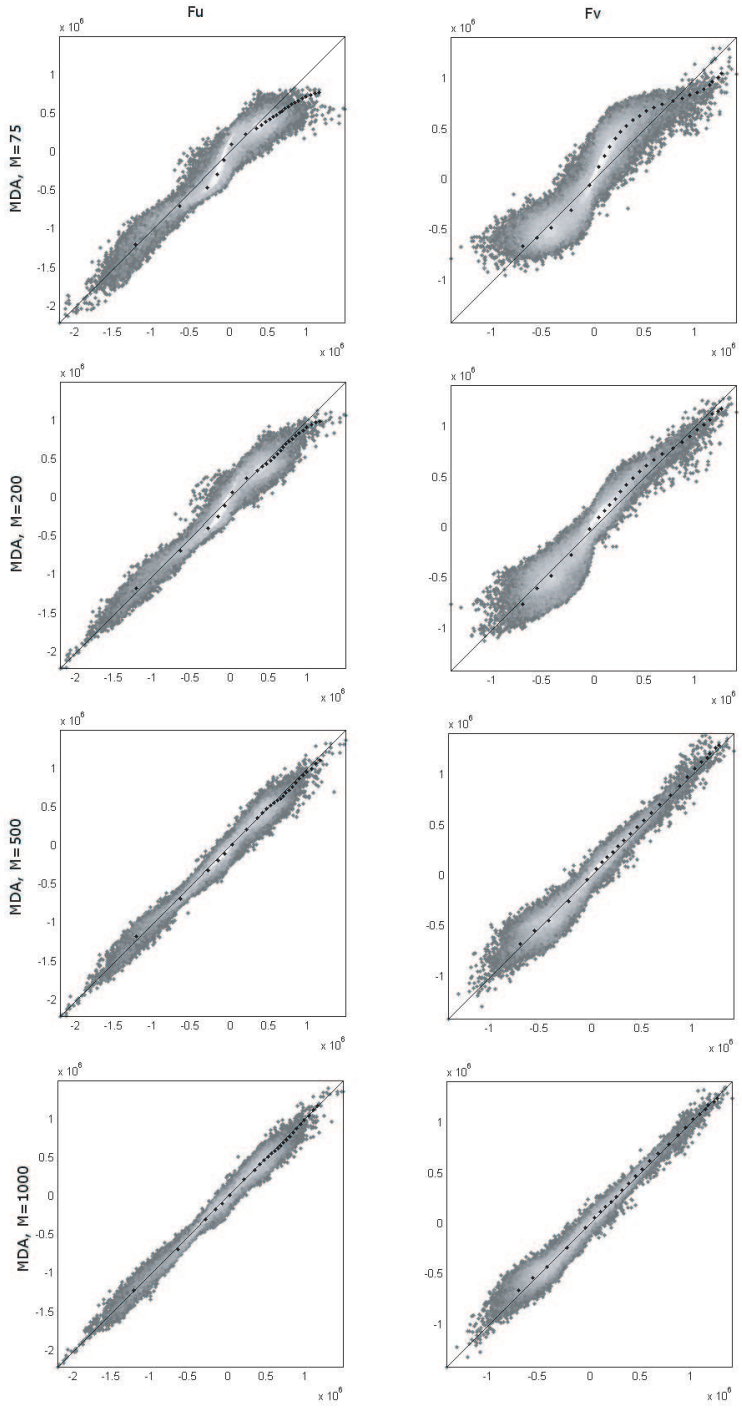

Figure 19: Scatter plot de las series temporales (calculada empíricamente vs. reconstruida con RBF) de F_u (componente X) y F_v (componente Y) considerando 75, 200, 500 y 1000 casos.

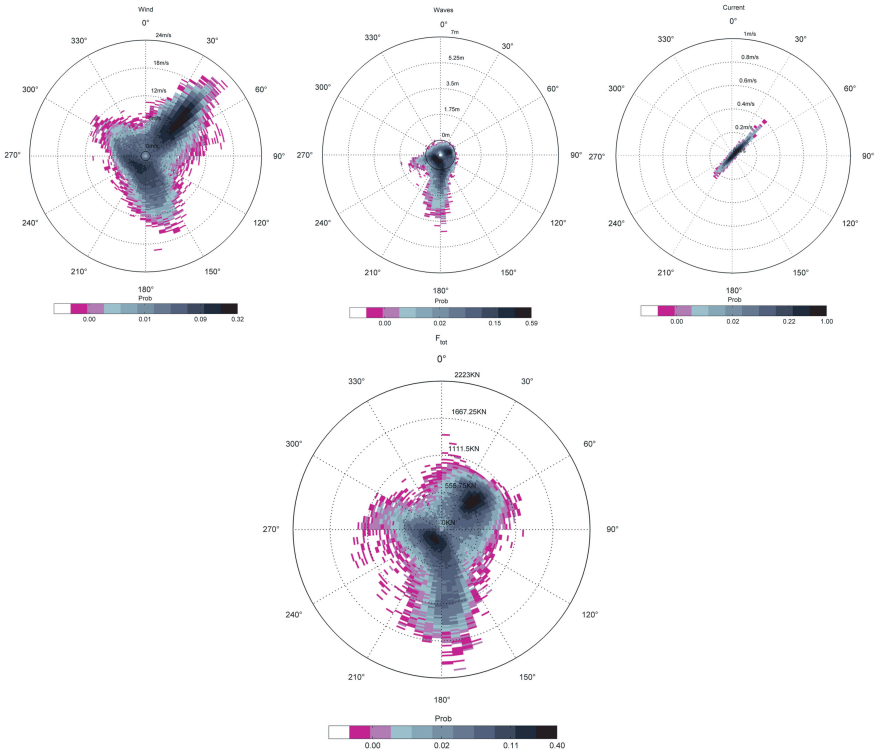

Figure 20: Rosas probabilísticas de los diferentes esfuerzos (viento– arriba izquierda, olas– centro arriba y corrientes–arriba derecha) y el total (abajo).

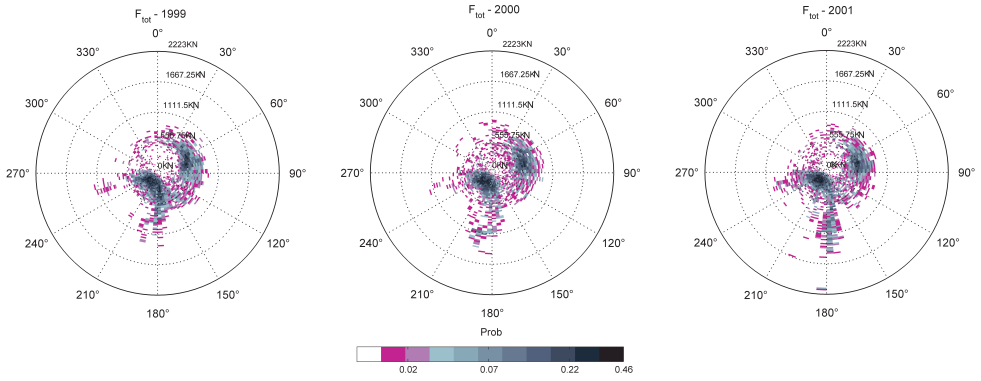

Figure 21: Rosas probabilísticas del esfuerzo total, 1999 (izquierda), 2000 (centro) and 2001 (derecha).

V Técnica de simulación conjunta del régimen medio y extremal

En los capítulos anteriores, se han tomado las distribuciones empíricas como base tanto para las simulaciones como para las interpolaciones. Esto supone una limitación en los valores máximos que se pueden reproducir, ya que están costreñidos a los máximos registrados históricamente. Además, el comportamiento de los valores extremos no es el mismo que el del régimen medio, lo cual imposibilita el ajustar las distribuciones empíricas a distribuciones conocidas. Para solventar esta limitación, el capítulo 6 propone una técnica de simulación con la que es posible tener en cuenta simultáneamente tanto el régimen medio como el extremal, manteniendo la dependencia temporal del proceso estocástico.

V.A Relación entre el régimen medio y el régimen extremal. Representación gráfica

En esta sección se presenta una forma útil de representación gráfica en la que se muestran ambos regímenes simultáneamente.

Supongamos un proceso estocástico X_t, cuya frecuencia de ocurrencia sea $f = 1/T_x$ (donde T_x es su frecuencia de muestreo: horario, 3 horario...) y su distribución de régimen medio es $F^{PT}(x)$. Si simulamos muestras de n valores del proceso estaocástico X_t y calculamos sus máximos, X_M se rige según una distribución $F^{EV}(x)$. El método para representar gráficamente ambos regímenes simultáneamente se basa en re–escalar el régimen extremal. Para ello inicialmente se calcula el periodo de retorno equivalente para el régimen medio: $T^{PT} = \frac{1}{1-F^{PT}(x)}$. Con esto ya es posible representar gráficamente el régimen medio. Para representar en la misma escala el régimen extremal se calcula su periodo de retorno, $T^{EV} = \frac{1}{1-F^{EV}(x)}$, y se re–escala teniendo en cuenta el tamaño de la muestra, n: $T_r^{EV} = n \cdot T^{EV}$. De esta forma ya se puede representar T_r^{EV} frente a x.

La Figura 22 muestra dos ejemplos de la representación gráfica propuesta. En la parte superior se muestra un ejemplo en el que la variable X_t se corresponde con una distribucion Normal(0,1), mientras que en la parte inferior se corresponde con una distribución Gamma($\theta = 5, \kappa = 10$). En ambos casos se simulan 1000

años de datos horarios y se considera que el régimen extremal esté asociado a los máximos anuales, por lo que $n = 8766$. La línea gris oscura representa el régimen medio (T^{PT}, x) mientras que los puntos negros representan los máximos anuales. Finalmente la línea gris clara representa (T_r^{EV}, x^{MAX}).

V.B Simulación simultánea del régimen medio y el régimen extremal

Para poder realizar simulaciones que simultáneamente tengan en cuenta ambos regímenes y mantengan la autocorrelación del proceso estocástico hay que realizar diferentes pasos. Primero hay que determinar hasta qué punto se supone el régimen medio y a partir de ahí se considera el régimen extremal. Para ello se determina el valor de x para el cual la distancia entre T^{PT} y T_r^{EV} es mínima. En la gráfica 23 se representa las PDFs y CDFs de ambos regímenes y cómo se determina el umbral entre ambos.

$$\underset{x}{\text{Minimize}} \quad \left(\frac{1}{1 - F^{\mathrm{PT}}(x)} - \frac{n}{1 - F^{\mathrm{EV}}(x)} \right)^2 . \tag{11}$$

Una vez se ha determinado el punto a partir del cual se usa uno u otro régimen, hay que re–escalar los valores que están por encima de ese umbral. Ese re–escalado se realiza considerando que la ditribución extremal está relacionada con el máximo de n elementos del régimen medio; y que hay una zona de solape en ambas distribuciones (Fig 23): los valores que aún pertenciendo al régimen extremal, están por debajo del umbral considerado se suponen pertenencientes al régimen medio. Por tanto, la forma de re–escalar los valores pertenecientes al régimen extremal es la siguiente:

$$x = \begin{cases} F^{\mathrm{PT}^{-1}}\left(u^{\mathrm{PT}}\right) & \text{if} \quad u^{\mathrm{PT}} \leq p_{\mathrm{lim}}^{\mathrm{PT}} \; (x \leq x_{\mathrm{lim}}) \\ F^{\mathrm{EV}^{-1}}\left(u^{\mathrm{EV}}\right) & \text{if} \quad u^{\mathrm{PT}} > p_{\mathrm{lim}}^{\mathrm{PT}} \; (x > x_{\mathrm{lim}}), \end{cases} \tag{12}$$

donde u^{EV}, es la probabilidad re–escalada y es igual a:

$$u^{\mathrm{EV}} = p_{\mathrm{lim}}^{\mathrm{EV}} + \frac{u^{\mathrm{PT}} - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}). \tag{13}$$

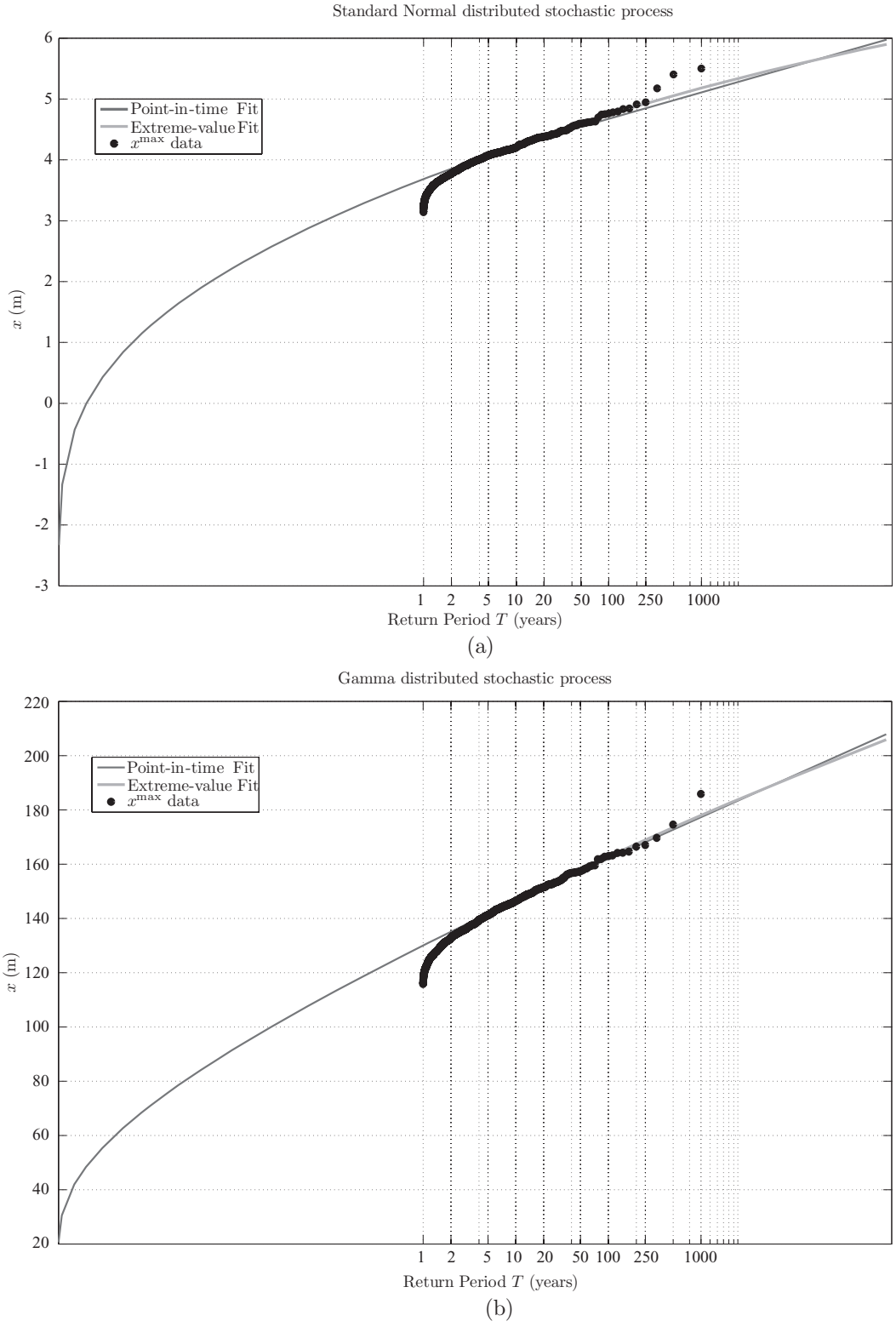

Figure 22: Representación gráfica del régimen medio y extremal: a) distribución normal y b) distribución gamma.

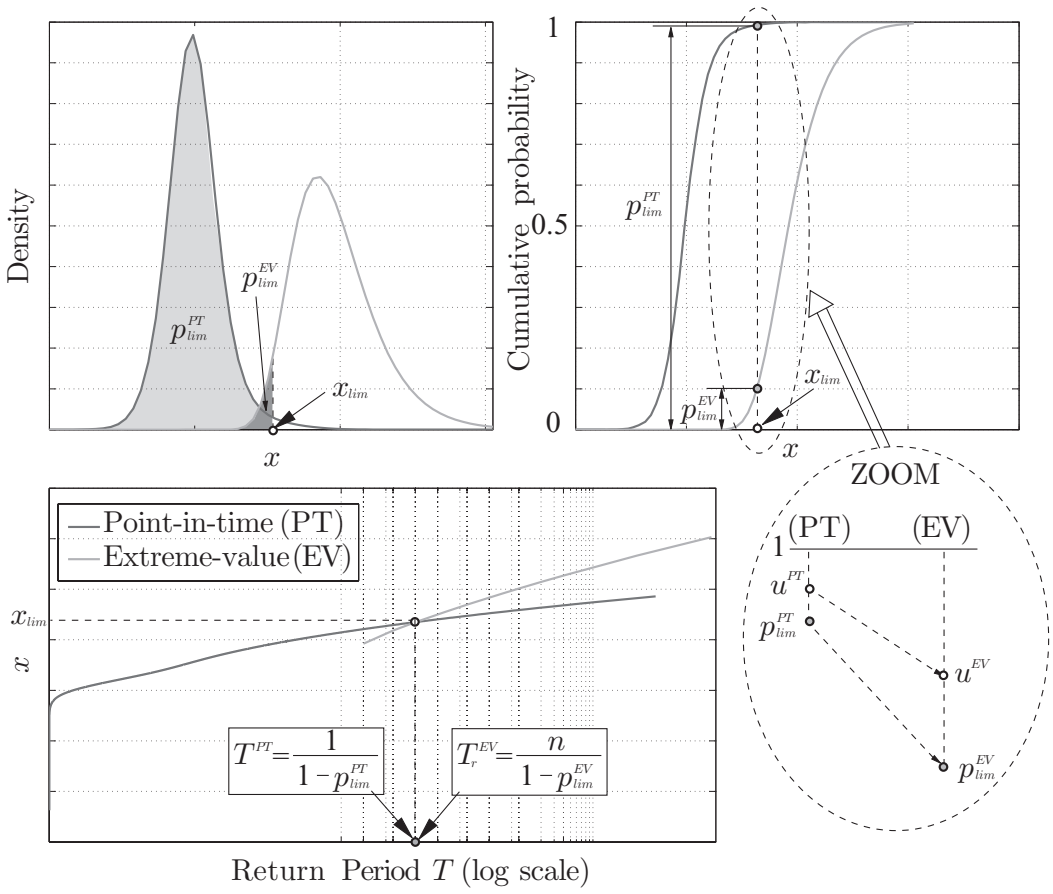

Figure 23: Representación gráfica del proceso de simulación.

Determinadas las transformaciones que hay que realizar para simular simultáneamente ambos regímenes, falta tener en cuenta la autocorrelación del proceso. Para ello se utilizará un modelo ARMA. Este modelo ARMA se ajusta a la serie temporal del proceso estocástico, de esta forma se tendrá en cuenta su autocorrelación.

Por lo tanto, el proceso completo de simulación puede sintesizarse en los siguientes puntos:

1. Utilizando el registro histórico se ajustan los regímenes medio y extremal.

2. Se normaliza el registro histórico.

3. Con el registro histórico normalizado, se ajusta un modelo ARMA y se obtienen los residuos. Estos residuos no están correlados y siguen una distribución

normal.

4. Se simulan errores independientes según una normal de media 0 y varianza la misma que se obtenga con los residuos históricos.

5. Con los residuos simulados y el ARMA ajustado, se hace la simulación.

6. Los valores simulados hasta entonces están normalizados, al desnormalizarlos se utiliza la transformación explicada en 12

V.C **Caso Práctico. Análisis de las condiciones ambientales para el diseño de turbinas eólicas offshore.**

En el capítulo 6 se, una vez se ha presentado la técnica de simulación conjunta del régimen medio y extremal, se añade un ejemplo de aplicación práctica. Dicho ejemplo es el análisis de las condiciones ambientales en el diseño de turbinas eólicas offshore, en concreto la determinación de los contornos $H_s - V$ para un periodo de retorno de 50 años (H_s-Altura de ola significante y V-Velocidad del viento a 10 m sobre el nivel del mar). Los datos han sido extraídos de bases de datos de reaálisis con una cobertura temporal de 60 años horarios.

Una vez clasificadas las alturas de ola según el valor de la velocidad del viento asociada, se ajustan las distribuciones resultantes. En la Figura 24 se muestran los histogramas y las funciones de densidad ajustadas para cada uno de los 27 grupos en los que se han dividido ambas variables. En la misma figura se muestran también los ajustes en la cola de cada una de las distribuciones en las que se han dividido los datos originales.

Una vez se ha aplicado la técnica de simulación conjunta para ambos regímenes se pueden obtener las envolventes de $H_s - V$ para un periodo de retorno de 50 años. Para poder aplicar dicha técnica es preciso transformar las variables al espacio normal mediante la transformación de Rosenblatt, de manera que se conviertan en variables independientes. El resultado de la aplicación de dicha técnica asícomo las envolventes resultantes se muestran en la Figura 25

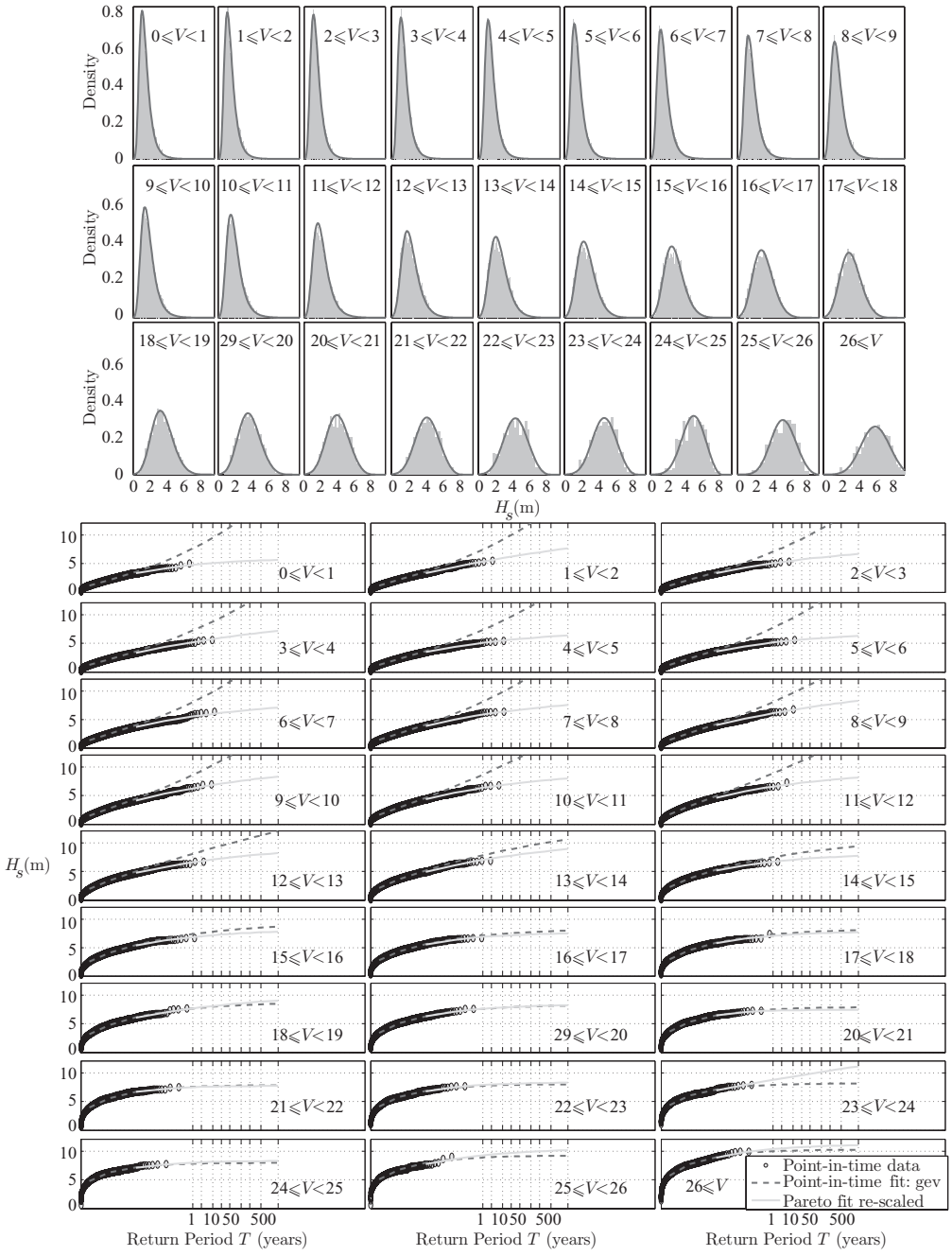

Figure 24: Histogramas y funciones de densidad ajustadas de H_s condicionada a V.

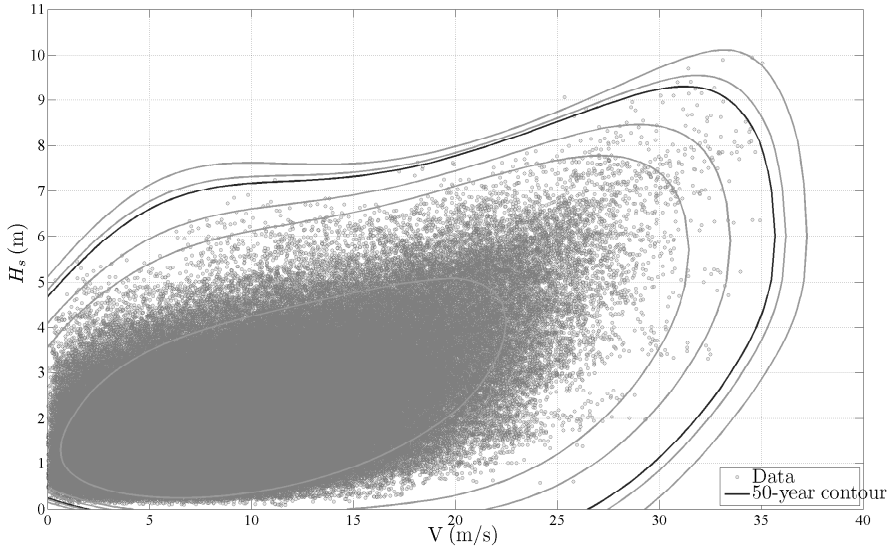

Figure 25: Envolventes de periodo de retorno 50 años.

VI Conclusiones

VI.A Resumen de Aportaciones

En este apartado se hace un resumen de las contribuciones de mayor relevancia presentadas en esta tesis:

1. Se ha presentado un modelo logístico autorregresivo con el cual es posible modelizar las condiciones atmosféricas en términos de patrones sinópticos de circulación. El carácter nominal del modelo permite tener en cuenta procesos autoregresivos así como otras covariables para tener en cuenta procesos de distinta naturaleza, como son la estacionalidad, la variabilidad interanual y las tendencias de largo plazo.

2. Se ha desarrollado una metodología para la simulación de estados de mar horarios trivariados. Para ello se ha hecho uso del modelo logístico presentado previamente y de una técnica de simulación multivariada existente en la literatura. Esta metodología tiene en cuenta procesos de diferente naturaleza con distintas escalas temporales y espaciales y permite generar series temporales de largo periodo de estados de mar horarios.

3. Se ha mostrado la posibilidad de utilización de un método híbrido que combina dos algoritmos estadísticos para estimar los parámetros de cálculo en el diseño de estructuras costeras reduciendo el tiempo computacional. Este método se ha aplicado a dos tipologías distintas de estructuras.

4. Por último se ha desarrollado una metodología de simulación que permite utilizar los regímenes medio y extremal de una variable. De esta forma se facilitan los diseños de aquellas estructuras en las que ambos regímenes son de interés.

VI.B Conclusiones

Tras revisar las aportaciones hechas dentro del trabajo expuesto se llega a las siguientes conclusiones:

1. Una correcta simulación de las variables implicadas en el diseño de estructuras marítimas requiere del conocimiento de los procesos y los forzamientos del oleaje al que va a estar sometida la misma. Es por eso que un estudio del clima marítimo conlleva el estudio de las condiciones atmosféricas asociadas al mismo, es decir, conocer el comportamiento de los vientos y las presiones en la zona de influencia del punto de estudio.

2. Además de tener en cuenta la escala espacial del proceso, es necesario el conocimiento y la inclusión de las diferentes escalas temporales de las cuales depende el oleaje. Es obvio que el oleaje es un proceso en el cual el estado actual depende de estados previos (proceso autoregresivo), sin embargo hay otras variaciones de mayor periodo que también han de ser tenidas en cuenta como son la estacionalidad, la variabilidad interanual o las tendencias de largo plazo.

3. El uso de técnicas estadísiticas permite una gestión más eficiente de bases de datos de largo periodo, de forma que su tratamiento y la estimación de parámetros a partir de ellas sea más manejable y eficaz. Esto permitiría una selección objetiva en los casos a modelar numéricamente o en laboratorio de forma que luego los resultados obtenidos se puedan extrapolar para el resto de estados de mar de la base de datos.

4. La disposición de bases de datos de largo periodo hacen posible la simulación tomando como referencia las distribuciones empíricas de las variables. Sin embargo para una correcta caracterización del régimen extremal es necesaria una correcta modelización de los eventos extremos de manera que en la simulación los eventos extremos no estén limitados al registro histórico de la base de datos.

Sobre el modelo logístico autorregresivo

- El modelo permite tener en cuenta estados previos del sistema.

- La naturaleza nominal del modelo realza el sentido físico de las clasificaciones de tipos de tiempo.

- Es posible considerar simultáneamente la influencia de covariables de distinta naturaleza.

- Los patrones atmósfericos sinópticos están definidos mediante una clasificación realizada con K-medias, una técnica apropiada para la definición de tipos de tiempo.

Sobre la metodología para generar estados de mar trivariados

- La combinación de diferentes técnicas estadísticas hace posible la consideración de diferentes escalas de tiempo y espacio.

- El procedimiento desarrollado permite reproducir estados de mar trivariados, considerando la correlación entre las variables.

- Las series temporales horarias simuladas están condicionadas a las condiciones medias dominantes, facilitando la relación existente entre el oleaje en un punto y las condiciones atmosféricas reinantes.

- Las condiciones medias se clasifican mediante K-medias. Esta técnica de clasificación genera grupos con similares características entre ellos, lo cual resulta conveniente a la hora de definir patrones sinópticos.

Sobre el procedimiento para transferir estados de mar en parámetros de diseño

- La gestión de series temporales multivariadas de largo periodo se puede realizar de una forma eficiente mediante el uso combinado de técnicas estadísticas.

- Con el algoritmo MDA es posible seleccionar estados de mar representativos. A diferencia de otras técnicas de clusterización este método incluye en los estados seleccionados estados extremos.

- La posterior interpolación mediante RBF posibilita la generación de series de largo periodo de parámetros de diseño de estructuras marítimas.

- Para la estimación de los parámetros de diseño se han empleado formulaciones semiempíricas, pero la metodología es de similar aplicación en el caso de utilizar modelos numéricos o modelado físico.

Sobre la técnica de simulación simultánea para regímenes medio y extremal

- Se ha presentado una nueva forma de representación gráfica que facilita el entendimiento de la relación entre ambas distribuciones.
- El método de simulación propuesto tiene en cuenta simultáneamente ambos regímenes y mantiene la dependencia temporal de los mismos.
- Al simular ambos regímenes a la vez se evita la decisión sobre qué régimen es necesario para cada aplicación.

VI.C **Futuras Líneas de Investigación**

Una vez se ha planteado el trabajo y las aportaciones del mismo, y se han expuesto las conlusiones a las que se ha llegado tras su realización, se plantean las siguientes líneas para futuras investigaciones.

Respecto al desarrollo de modelos logísticos con los que analizar las condiciones atmosféricas u oceánicas:

- Determinación del número óptimo de patrones sinópticos con los que describir los procesos atmosféricos y/o las dinámicas marinas. En la literatura existen discrepancias entre autores en relación con el número óptimo de tipos de tiempo a tener en cuenta. Hay que profundizar sobre este tema para establecer un criterio objetivo con el que determinar el número de patrones a considerar.
- Comparación entre los diferentes escenarios de cambio climático. Mediante el uso del modelo logístico se pueden simular los diferentes escenarios de cambio climático propuestos y así comparar las tendencias que se obtengan.

En relación con la simulación de estados de mar basada en tipos de tiempo:

- La inclusión de regímenes extremales en la técnica de simulación propuesta. Todo el trabajo mostrado en relación a la simulación multivariada se ha hecho con base en las distribuciones empíricas obtenidas de los registros históricos. Es por ello que el siguiente paso es la inclusión en la metodología la posibilidad de simular eventos extremos. Esta extensión a la metodología actual no es trivial, ya que requiere un estudio detallado de los límites entre ambos regímenes y un análisis de las correlaciones existentes entre las diferentes variables.

- Habilitar la metodología para añadir más variables. Dependiendo de la estructura a diseñar o el proceso costero a aestudiar, puede ser necesario el disponer de unas u otras variables. Para conseguir adaptar el método actual es preciso analizar las correlaciones entre variables y la propia autocorrelación que presenten las variables. Además, la adición de otras covariables que expliquen los procesos a parte de las tendencias, la variabilidad interanual o la estacionalidad pueden ser necesarias.

- Generalización del método propuesto. Previo a las modificaciones mencionadas sería necesario una generalización de la técnica de simulación propuesta. Su aplicabilidad en diferentes localizaciones donde los procesos involucrados sean diferentes tiene que ser probada.

- Simulación de eventos extremos basada en patrones climáticos. En la literatura existen métodos con los que generar series sintéticas de eventos extremos ([75]) que también distinguen entre ambos regímenes. Sería interesante en el ámbito del diseño marítimo el ser capaces de combinar esas técnicas de simulación con los métodos basados en patrones de circulación propuestos en esta tesis. De esta forma, la influencia de diferentes variables con distintas escalas temporales y espaciales podrían ser tenidas en cuenta.

En relación a la transferencia de estados de mar a parámetros de diseño:

- Aplicación de la combinación MDA-RBF al diseño de ensayos de laboratorio. La metodología propuesta ha sido aplicada a formulaciones semi

empíricas, pero su aplicación a la hora de crear planes de ensayo en laboratorio sería de gran utilidad. Para realizarlo, habría que realizar un análisis con el que se pueda determinar el número mínimo de ensayos a realizar en el laboratorio y la repetitibilidad alcanzable en el mismo.

- Determinación de las debilidades del proceso. Es necesario una mayor profundización para determinar qué parte de la metodología conlleva una mayor incertidumbre asociada: el número de casos extrídos por el algoritmo MDA o la técnica de interpolación. Una vez se haya determinado qué parte es la más débil del proceso se podría estimar la incertidumbre asociada a un cierto número de casos.

En el área de la simulación simultánea de los regímenes medio y extremal:

- Inclusión del extremal index. La inclusión de un índice de transición entre ambos regímenes debe ser estudiada con mayor detalle. En la metodología propuesta aquí no se ha considerado este hecho, de forma que la transición entre ambos regímenes puede producir resultados no precisos para periodos de retorno bajos.

CHAPTER

1

Introduction

1 Introduction

Wave climate characterization is vital to understand wave-driven coastal processes and to design coastal and offshore structures. In each case, mean or extreme conditions, or even both may be needed. The availability of data from which to start the design process is therefore essential.

Existent wave climate databases can be divided into four categories: instrumental records (buoys), satellite data, visual data and time series generated by numerical models (reanalysis). Each presents advantages and disadvantages. Buoy records, despite being high quality data and being able to record different variables, are local measurements, and their density is limited and may prove insufficient. Visual observations are inevitably conditioned by the subjectivity of the observer, and the zones where data of this kind are available are restricted to the main maritime waterways. On the other hand, in recent decades many satellites have been developed which are able to provide wave data with a high spatial density, but have a temporal density which is restricted by the frequency of the satellite overpass. Finally, thanks to advances in computational fields and improvements in calculation times, numerical models have been developed which are able to generate historical met-ocean time series. These are known as reanalysis databases or hindcast. These

data are very useful because of their high spatial density and their continuity over time. However, before they can be used they need to be corrected. To correct them, the other kinds of data (buoys, satellite) become crucial.

Depending on the study or design to be developed, one type of data or another, or combinations, may be used. Moreover, the different temporal scales must be taken into account to properly characterize the wave climate. With reference to these temporal scales, studies may be classified into four types:

1. *Short term*: Studies requiring forecasts of 3 days or less.

2. *Middle term*: Designs for which an analysis of the wave climate for periods of up to 6 months is needed.

3. *Long term*: Infrastructures for which the design requires the use of long term historic data ($>$20 years)

4. *Very long term*: Studies that require considering wave climate projections under different climate change scenarios.

Even when wave data in the study area is available, often its time duration is inadequate. In these cases, simulation techniques may be used to allow for a probabilistic–design. For applications using these design methods, wave climates must be accurately generated using synthetic methods. To achieve this, the different temporal and spatial scales involved must be taken into account. For the temporal scales, the autoregressive nature of the wave climate, such as its seasonality must be considered. As regards the spatial scales, the designer has to be aware of the structure's location as well as the atmospheric and oceanic processes which are characteristic of the area. Because of this, in wave climate studies special attention is usually paid to atmospheric conditions, since they are the principal determinant of the wave generating force.

There are different reliability measurements and methods used to estimate these atmospheric forces. The next section examines the different reliability measurements existent, and the basic concepts associated with each. Following the methods, the various levels of design applied to maritime structures are explained.

2 Structural reliability measurements

The design of any structure must be done so as to ensure that during its *service life*, predetermined *limit states* are not exceeded. By service life is meant the period of time during which the structure must maintain acceptable safety, functionality and aspect conditions without requiring any rehabilitation operation. The limit state refers to the threshold above which the structure is not able to accomplish any of the functions for which it was designed ([106]). Limit states can be classified into:

1. *Operational stop.* Any interruption of normal use resulting from atmospheric agents.
2. *Service.* Interruption of normal use due to deformations, vibrations or superficial damages.
3. *Damage.* Shutdown which requires important repairs to prevent the collapse of the structure.
4. *Ultimate.* Partial or total collapse of the structure

Regarding the reliability measurement, the design can be classified into three different types:

1. *Deterministic design.*
2. *Partially probabilistic design.*
3. *Probabilistic design.*

2.1 Deterministic design

Deterministic design belongs to classical methods, based on safety factors that implicitly consider variable randomness.

- **Global safety factor**

  This consists of dividing the n-dimensional space of the involved variables into two areas. This division corresponds to a specified limit state. In the

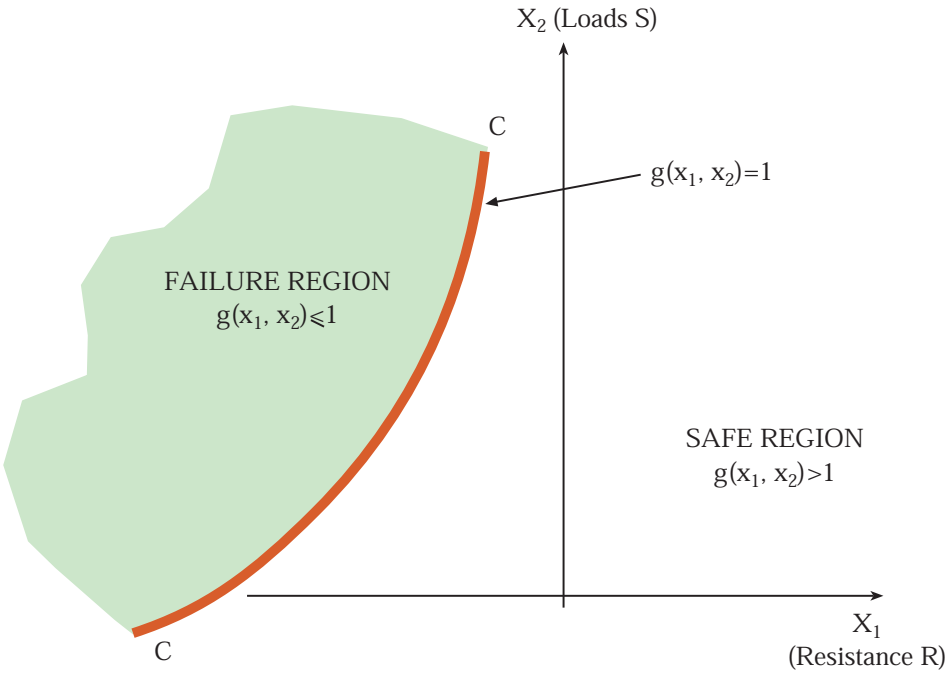

Figure 1.1: Scheme showing the safe and failure regions and the limit state equation for a bidimensional example.

safe region, project conditions are achieved, while in the failure region the structure no longer ensures the functions for which it was designed.

In Figure 1.1 the red line shows the locus where the safe region equals the failure region. In order to increase the safe region, a safety factor is applied, as in Figure 1.2. Let us consider the safe region as $S \equiv \{g(x_1, x_2) > 1\}$ and the failure region as $F \equiv \{g(x_1, x_2) \leq 1\}$, and $h_S(x_1, x_2)$ and $h_F(x_1, x_2)$ as the magnitudes that correspond to safety and failure respectively. Thus, the red line would be the one that ensures $g(x_1, x_2) = 1$. To guarantee that the limit state is not surpassed a safety factor is added:

$$g^*(x_1, x_2) = \frac{h_R(x_1, x_2)}{h_S(x_1, x_2)} - F > 0 \tag{1.1}$$

The addition of a safety factor F, results in a displacement of the curve that divides both regions, thereby gaining an extra region of safety.

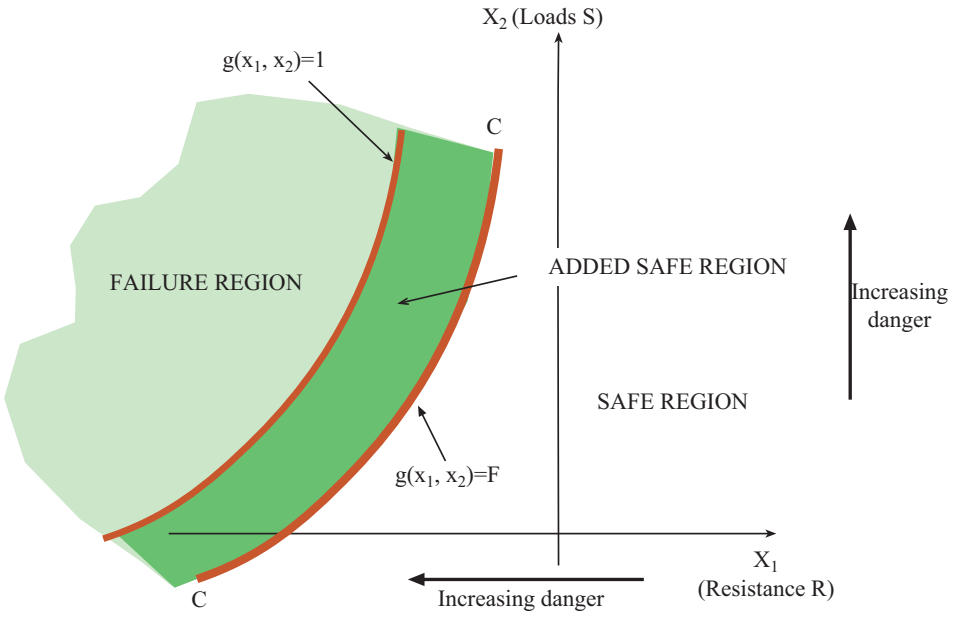

Figure 1.2: Increase of the safety region by adding a safety factor.

- **Partial safety factor**

  Partial safety factors arise from the evolution of the global safety factor. Based on the estimation of different safety factors applicable to the load and/or elements that conform the structure, these coefficients are divided into reduction factors for the resistances, and magnification factors for the loads.

The application of safety factors (globally or partially defined) in structure design can be seen as a first approach in structural reliability estimation. Nevertheless, its application involves uncertainties because the factors determination may not be unique, and the representative values of the random variables may vary.

2.2 Partially probabilistic design

In order to reduce the temporal uncertainties related to the occurrence of natural phenomena, the concept of a *return period* is utilized. The return period of a process is the mean time between two statistically independent sequential events.

$$T = E(X) \tag{1.2}$$

where $E(X)$ stands for the expected value of X. The definition of the return period depends on the events distributions. If it is assumed that the time between events, X, is a random variable with a geometric distribution, the probability of the time between two events will be:

$$P[X = x] = p(1-p)^{(x-1)} x = 1, 2, ... \tag{1.3}$$

So, the expected value of X, and consequently the return period, is:

$$E[X] = \sum_{x=1}^{\infty} xp(1-p)^{(x-1)} = \frac{p}{(1-(1-p))^2} = \frac{1}{p} \tag{1.4}$$

From equation 1.4 it follows that the return period is the inverse of the event occurrence probability in the period of time defined. Using the return period when designing structures is valid as long as it is considered that:

- The return period definition depends on the time scale used.
- The possibility of more than one event within the same period of time is not taken into account. Thus, this approach is only viable when the occurrence probability of the events is very low in the period of time defined.

2.3 **Probabilistic design**

The methods explained previously are very useful tools for structural design, but they are not able to take into account that even for a defined period of time, the variables present uncertainties. The uncertainty related to the resistance R and to every load S, may be represented by their density functions f_R and f_S. Both loads and resistances can vary over time due to for example, material spoiling. This implies an increase in density function widths (Figure 1.3). Even so, in most of the cases, loads and resistances are considered to remain constant over time.

Considering the density functions of the uncertainties related to resistances $f_R(r)$, and loads $f_S(s)$, constant in time, allows the determination of the joint density function $f_{RS}(r, s)$. If both variables are independent, the joint distribution is

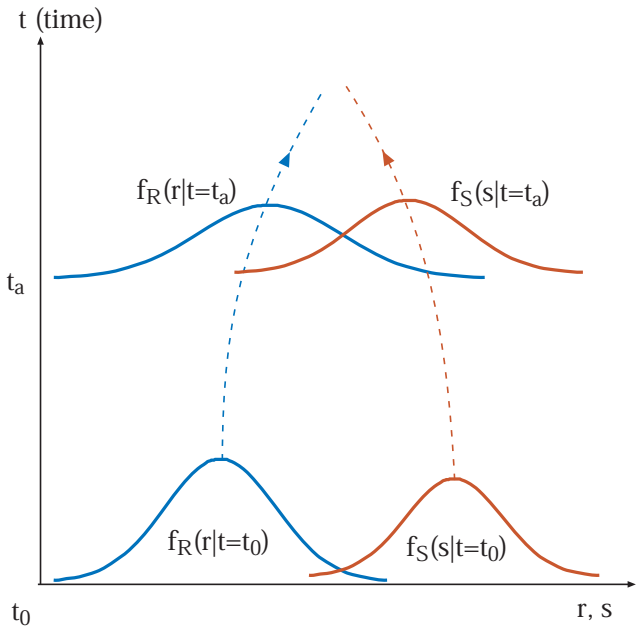

Figure 1.3: Reliability varying with time.

the product of the marginals, i.e. $f_{RS}(r,s) = f_R(r)f_s(s)$. In Figure 1.4 all these concepts are shown. Failure probability can be then defined as:

$$p_f = \int_D \int f_{RS}(r,s)drds = \int_{-\infty}^{\infty} \int_{-\infty}^{s} f_{RS}(r,s)drsr. \tag{1.5}$$

Taking into account that the joint distribution is the product of the marginals, equation 1.5 can be reformulated as:

$$p_f = \int_D \int f_{RS}(r,s)drds = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{s} f_R(r)dr \right) f_S(s)ds. \tag{1.6}$$

In equation 1.6, the integration of $f_R(r)$ between $-\infty$ and s, is the density function particularized for $r = s$, thus:

$$p_f = \int_{-\infty}^{\infty} F_R(s)f_S(s)ds. \tag{1.7}$$

Equation 1.7 is known as the convolution integration. F_R represents the probability of $R \leq s$, while $f_S(s)$ represents the probability of $S = s$. If it is desired, p_f can be expressed in terms of the resistance:

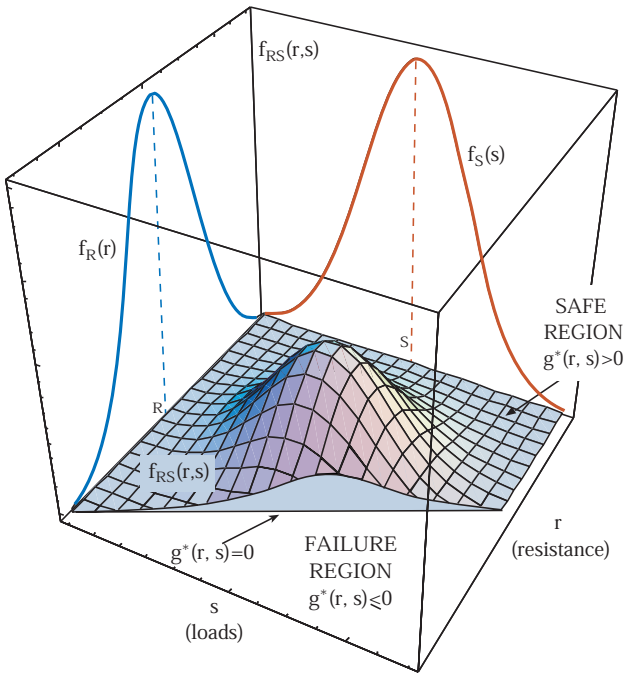

Figure 1.4: Graphical representation of the joint distribution $f_{RS}(r, s)$, marginals $f_R(r)$ and $f_S(s)$ and failure region.

$$p_f = \int_{-\infty}^{\infty} [1 - F_S(r)] f_R(r) dr. \tag{1.8}$$

In Figure 1.5 those concepts already mathematically expressed are shown. In 1.5(a), R and S marginal distributions are plotted. In 1.5(b) and (d), the failure probability, expressed as function of S and R, respectively, is represented; while in 1.5(c) the graphical interpretation of the failure probability is shown.

Nevertheless, in most of the structural design cases it is not possible to reduce the problem to only two variables R and S. Because of this, R and S must be replaced by $h_R(x_1, x_2, ..., x_n)$ and $h_S(x_1, x_2, ..., x_n)$ respectively. This way the joint density function is given by:

$$f(\mathrm{x}) = f_{X_1, X_2, ..., X_n}(x_1, x_2, ..., x_n; \Theta), \tag{1.9}$$

where Θ is a parametric vector defining variable distributions. With that, failure probability may be obtained with:

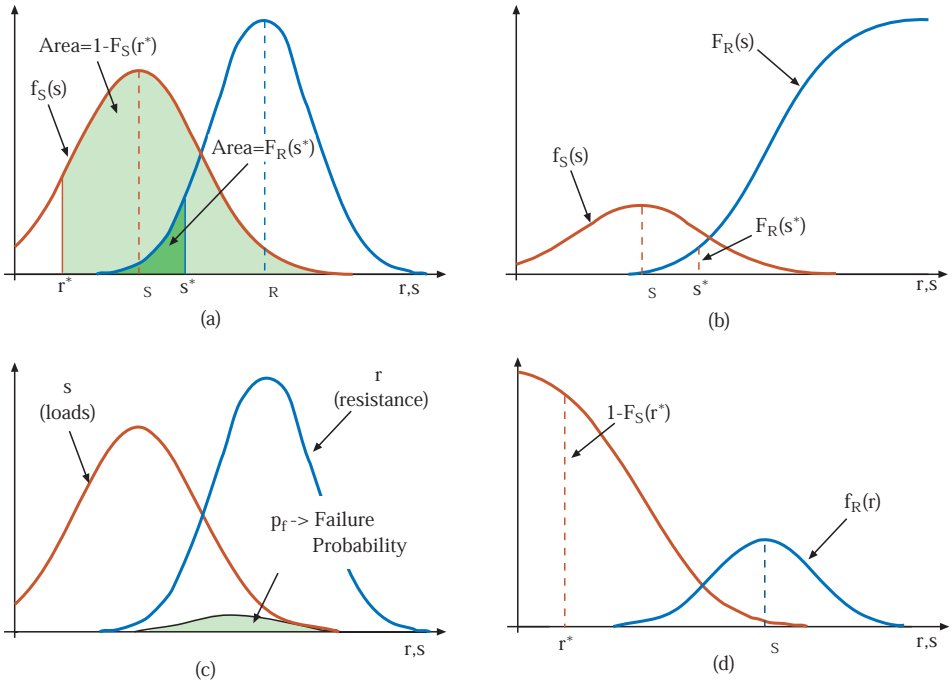

Figure 1.5: Graphical representation of marginal density distributions $f_R(r)$ and $f_S(s)$, distribution and probability of failure.

$$fp_f(\Theta) = \int_{g^*(x_1,x_2,...,x_n)\leq 0} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n; \Theta)dx_1dx_2...dx_n, \qquad (1.10)$$

with $g^*(x_1, x_2, ..., x_n)$ being the equation that defines the limit state.

In this way, a conditional probability is obtained; being a punctual estimation of the failure probability for a determined value of Θ. But this can be generalized if the expected value of the failure probability is considered ([40]):

$$p_f(\Theta) = E[p_f(\Theta)] = \int_{\Theta} p_f(\Theta) f_{\Xi}(\Theta) d\Theta, \qquad (1.11)$$

with $E[\cdot]$ being the expected value and $f_{\Xi}(\Theta)$ the joint density function of Θ.

3 Methods of reliability estimation

In the maritime structure design field in Spain, the R.O.M. (Recomendaciones de Obras Martimas) of Puertos del Estado ([139]) provides a catalogue of rules, regulations and technical criteria to be applied to the four phases into which the service life of infrastructures can be divided:

1. Planning and design.
2. Construction.
3. Operating and maintenance.
4. Recycling and dismantling.

The R.O.M. is divided into different sections, with the *ROM 0.0, Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias*, being the one that specifically refers to the different verification and design methods to be applied in the maritime structure field. In this document, reliability methods are divided into Level I, Level II and Level III methods. By its application, a design that ensures safety, service and operation over the structure's lifetime can be formulated.

3.1 Level I

Level I methods are those based on safety factors estimation. There are two main branches within these methods: methods based on global safety factors, and those based on partial safety factors for all the involved variables. With the first set, a minimum safety factor that the infrastructure has to ensure over its lifetime is estimated. This coefficient is obtained from the main failure modes and lifetime of the structure, and from the limit states, ultimate and service. On the other hand, methods based on the estimation of partial safety factors disaggregate the safety factor into several partial factors. Each of them is obtained through the quantiles of the different involved variables. To these obtained values, some corrections of weighting and compatibility are applied.

Level I methods represent an easy and traditional approach to structural design. They are very useful and easily applied methods, but they are not able to inform about the possibility of not satisfying the requirements.

3.2 Level II

Level II and III methods are based on the failure probability estimation, p_f. The analytical solution of p_f, where all the involved variables are taken into account, can be expressed as:

$$p_f = \int_{g(x_1,x_2,...,x_n)\leq 0} f\mathrm{x_1,x_2,...,x_n}(x_1,x_2,...,x_n)dx_1dx_2...dx_n \tag{1.12}$$

Usually, the achievement of an analytical solution of this integration is not possible. There are different approaches to estimate the failure probability. In the case where the involved variables are not correlated it is possible to estimate the two first moments of the joint distribution and proceed with them to the design ([54], [103], [144], [8], [33], etc.). But, in the case of correlated variables, they must first be transformed into independent variables.

3.2.1 FOSM. First Order Second Moment

Level II methods are also known as FOSM (First Order Second Moment) methods. FOSM methods simplify the ultimate limit state equation to a straight line, and the point on this line for which the distance is the minimum distance will be the design point (Figure 1.6).

The distance β in Figure 1.6 represents the reliability index proposed by Hasofer and Lind as a constant measurement, to determine whether the system is safe or not. This index can be estimated by:

$$\beta = \underset{\mathbf{x}}{\text{Minimum}} \quad \sqrt{(\mathbf{x}-\boldsymbol{\mu}_X)^T\boldsymbol{\sigma}_X^{-1}(\mathbf{x}-\boldsymbol{\mu}_X)^T} \tag{1.13}$$

subject to $g_{\mathbf{X}}(\mathbf{x}) = 0$

Unlike Level I methods, there is no weighting or correction to be applied to the parameters obtained using Level II and III methods.

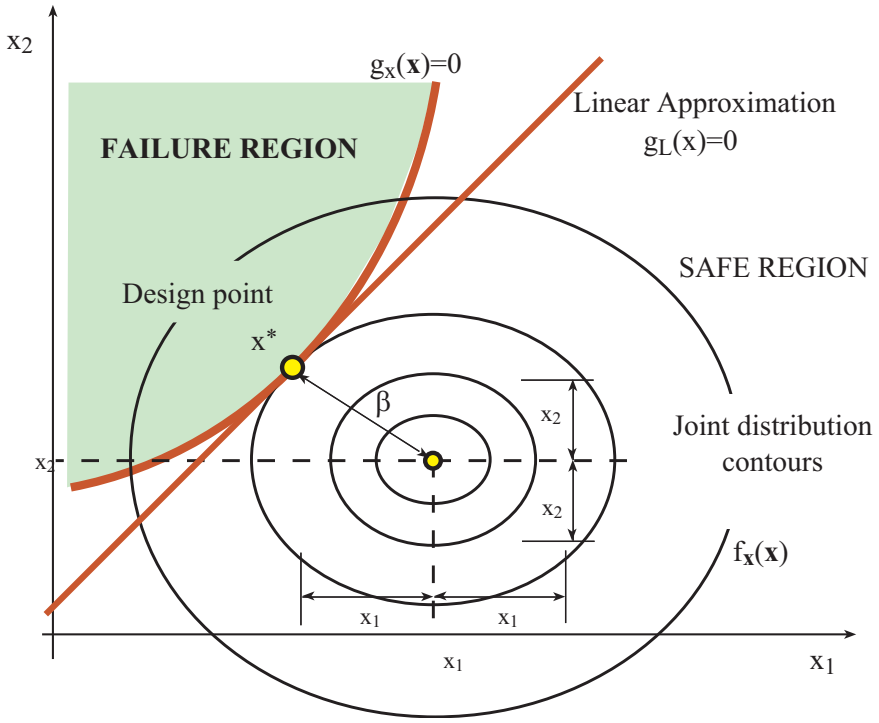

Figure 1.6: Limit state surface graphical representation $G_{\mathbf{X}}(\mathbf{X}) = 0$ and its linear approximation for a 2-D case.

3.2.2 **Hasofer-Lind transformation**

Among FOSM methods, Hasofer and Lind ([74]) proposed a transformation of the multivariate standard variables into independent normal variables $N(0,1)$. In the case of having independent variables this transformation is simple:

$$Z_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}; \forall i = 1, ..., n \tag{1.14}$$

μ_{X_i} and σ_{X_i} each being the i variable's mean and variance respectively. But, in the case of the variables being correlated an orthogonal transformation is required to obtain random normal variables

Orthogonal transformation for random normal variables.

Let $X = (X_1, X_2, ..., X_n)$ be a vector containing the random correlated variables involved in the system, and its mean and variance-covariance values be

$$\boldsymbol{\mu}_X = E(\mathbf{X}) = (E(X_1), E(X_2), \ldots, E(X_n)) = (\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}),$$

and

$$\boldsymbol{\sigma}_X = \text{cov}(X_i, X_j) = \sigma_X^2(i,j)$$

If the variables are independent, the variance-covariance matrix is strictly diagonal.

The purpose of this transformation is the obtention of a linearly independent vector $\mathbf{U}$, and a transformation matrix $\mathbf{B}$ which satisfies:

$$\mathbf{U} = \mathbf{BX} \tag{1.15}$$

The transformation must be orthogonal to keep constant the distance between both spaces $\mathbf{X}$ and $\mathbf{Z}$. Vector $\mathbf{U}$ will not be correlated if the variance-covariance transformed matrix $\boldsymbol{\sigma}_U$ is diagonal.

$$\begin{aligned} \boldsymbol{\sigma}_U &= \text{cov}(\mathbf{U}, \mathbf{U}^T) = \text{cov}(\mathbf{BX}, \mathbf{X}^T\mathbf{B}^T) = & (1.16)\\ &= \mathbf{B}\ \text{cov}(\mathbf{X}, \mathbf{X}^T)\ \mathbf{B}^T = \mathbf{B}\ \boldsymbol{\sigma}_X\ \mathbf{B}^T & (1.17) \end{aligned}$$

The variance-covariance being a symmetric definite positive matrix, it can be decomposed by Cholesky decomposition:

$$\boldsymbol{\sigma}_X = \mathbf{L}\mathbf{L}^T \tag{1.18}$$

where $\mathbf{L}$ is a lower triangular matrix such that its' inverse $\mathbf{B} = \mathbf{L}^{-1}$ will also be lower triangular and easy to achieve. So:

$$\mathbf{B}\ \boldsymbol{\sigma}_X\ \mathbf{B}^T = (\mathbf{B}\ \mathbf{L})\left(\mathbf{L}^T\ \mathbf{B}^T\right) = \mathbf{I} \tag{1.19}$$

where $\mathbf{I}$ is the identity matrix.

Substituting 1.19 in 1.17 results in:

$$\boldsymbol{\sigma}_U = \mathbf{I}. \tag{1.20}$$

Finally, variables $\mathbf{U} \sim N(\boldsymbol{\mu}_U, \mathbf{I})$ are transformed into $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I_n})$. To do this:

$$\mathbf{Z} = \frac{\mathbf{Z} - \boldsymbol{\mu}_Z}{\mathbf{I}} = \mathbf{B}(\mathbf{X} - \boldsymbol{\mu}_X) \tag{1.21}$$

This final expression 1.21 allows changing from space $\mathbf{X}$ to $\mathbf{Z}$.

3.3 Level III

Finally, Level III methods are those that try to solve equation 1.12 by integration techniques (analytically or numerically), or by using simulation techniques. Only in a few cases does an analytical solution exist, so normally simulation techniques are used. Level III methods based on the estimation of the failure region can be split into FORM (First Order Reliability Methods), and SORM (Second Order Reliability Methods). On the other hand, Level III methods that consist of the synthetic generation of involved variables are generally based on simulation techniques such as Monte Carlo ([112]).

3.3.1 FORM. First Order Reliability Methods

While FOSM estimates the failure probability considering two first moments of the random variables, FORM methods are a linear approximation of the problem taking into account the actual distributions of the variables. These methods were firstly proposed by Freudenthal [54] in the field of structural reliability and then developed by [74], [134], [79], [42], etc. The process undertaken is similar to FOSM methods, the main difference is related to the transformation applied to the random variables to transfer them to the multivariate normal space.

A random variable X can be transformed into a normally distributed variable Z by:

$$F_X(x) = \Phi(z) \quad \text{or} \quad z = \Phi^{-1}(F_X(x)) \tag{1.22}$$

where $F_X(x)$ is the distribution function of X, and Φ is the distribution function of the normal variable $Z \sim N(0,1)$..

When the variables are not independent, the Rosenblatt transformation ([140]) may be used.

3.3.2 Rosenblatt transformation.

With X representing random variables with an unknown distribution, they can be transformed into independent uniform variables $U(0,1)$ by:

$$\begin{array}{rcl} u_1 & = & F_1(x_1) \\ u_2 & = & F_2(x_2|x_1) \\ \vdots & \vdots & \vdots \\ u_n & = & F_n(x_n|x_1, x_2, \ldots, x_{n-1}), \end{array} \tag{1.23}$$

Then, these uniform variables can be transformed into normally distributed variables Z by:

$$\begin{array}{rcl} z_1 & = & \Phi^{-1}\left(F_1(x_1)\right) \\ z_2 & = & \Phi^{-1}\left(F_2(x_2|x_1)\right) \\ \vdots & \vdots & \vdots \\ z_n & = & \Phi^{-1}\left(F_n(x_n|x_1, x_2, \ldots, x_{n-1})\right). \end{array} \tag{1.24}$$

where $F_1(x_1)$, $F_2(x_2|x_1), \ldots, F_n(x_n|x_1, x_2, \ldots, x_{n-1})$ are the marginal distribution functions of X_1 and the conditional variables, respectively.

Once the variables have been transformed to normally distributed ones, it is necessary to transform the state limit equation from $g_{\mathbf{X}}(\mathbf{x})$ to $g_{\mathbf{Z}}(\mathbf{z})$. In order to do this, the jacobian J of the transformation has to be estimated. When the joint density function of the variables is not available, the transformation proposed by ([123]) can be used.

3.3.3 SORM. Second Order Reliability Methods

There are cases where it is not feasible to approximate the state limit equation to a straight line. This occurs when the state limit equation is a sharp curve, or when this curve is not so sharp in the original space but is sharp in the transformed space. When this occurs, the problem can be solved by using SORM (Second Order Reliability Methods). Among these methods can be listed the approaches proposed by [15], [26], [27], [28], [73], [159], [41], [94], [125], etc.

These second order methods are based on an approximation of the state limit equations by parabolic or spheric approximations in the surroundings of the maximum verisimilitude design point. The results obtained by these methods are very accurate and can be more efficient than Monte Carlo simulation techniques at estimating extreme percentiles.

3.3.4 Simulation techniques

Failure probability estimation (1.12) through simulation techniques belongs to Level III methods. These techniques are based on the synthetic generation of a high number of vectors containing the involved variables ($\hat{\mathrm{x}} = \hat{x}_1, \hat{x}_2, \hat{x}_n$), and in a later check to determine whether or not the limit state is exceeded with all of them. In this way the failure probability can be estimated as:

$$p_f \approx \frac{n(g(\hat{\mathrm{x}}) \leq 0)}{N} \tag{1.25}$$

This procedure is known as Monte Carlo. There are many variation of it which may be used to improve the results and/or adapt them to specific cases. Two examples of these variations are a weighted simulation and a directional simulation.

The first example is a modification of Monte Carlo proposed by [71], [147], [141], [167] or [60]. It provides efficiently and simultaneously an estimation of the failure probability and the error. Moreover, it is able to deal with limit state functions which are non–differentiable. On the other hand, directional simulation is an adaption to polar coordinates proposed by [46], [49] or [44].

In order to apply these simulation techniques to structural reliability problems it is necessary to identify the variables involved and to be able to systematically generate random vectors from them.

4 Objectives

The general objective of this thesis is the generation of synthetic time series to be applied in the design of maritime structures. This design would be set within the Level III methods. For a correct simulation of met-ocean variables it is necessary to understand their behaviour and variability. Moreover, an efficient treatment of long term multivariate time series requires the use of different statistical techniques.

This general objective can be divided into two partial objectives which are:

1. To be able to generate synthetic multivariate time series of the variables that mainly define wave climate and atmospheric conditions. To do this, different

techniques are explored, allowing for the consideration of different temporal and spatial scales.

2. To transform the simulated time series of met-ocean variables into time series of the parameters, on which the main failure modes of the structures depend.

The achievement of the proposed objective requires a deep exploration and development of different techniques and algorithms, and its subsequent validation.

5 Methodology

The general methodology to generate synthetic time series to be applied to maritime structure design is depicted in Figure 1.7. Taking into account the historical data of wave climate and atmospheric pressure fields, a large number of lifetimes are simulated. This simulation involves some steps in order to consider the different temporal and spatial scales involved. The availability of synthetic time series enables its transference to the structure in terms of the parameters related to its failure mode (loads, momentums, scour,...). This transference is done by a clusterization and a later interpolation to reconstruct time series of those parameters. The availability of these long term series enables a structural integrity analysis of the element or the estimation of the loads return periods. In Figure 1.7 the scope of each partial objective previously defined in order to achieve the general objective is depicted.

5.1 Simulation

The achievement of the proposed methodology requires the use of different techniques and algorithms. To properly simulate multivariate sea states it is necessary to know the associated atmospheric conditions, as they constitute the main force which generates waves. Moreover, it is necessary to include seasonality and the autoregressive character inherent to waves and atmospheric conditions. For these reasons the simulation is undertaken in three steps, schematized in Figure 1.8

1. Simulation of the atmospheric pressure fields. In this way the spatial scale of influence in the wave climate is taken into account.

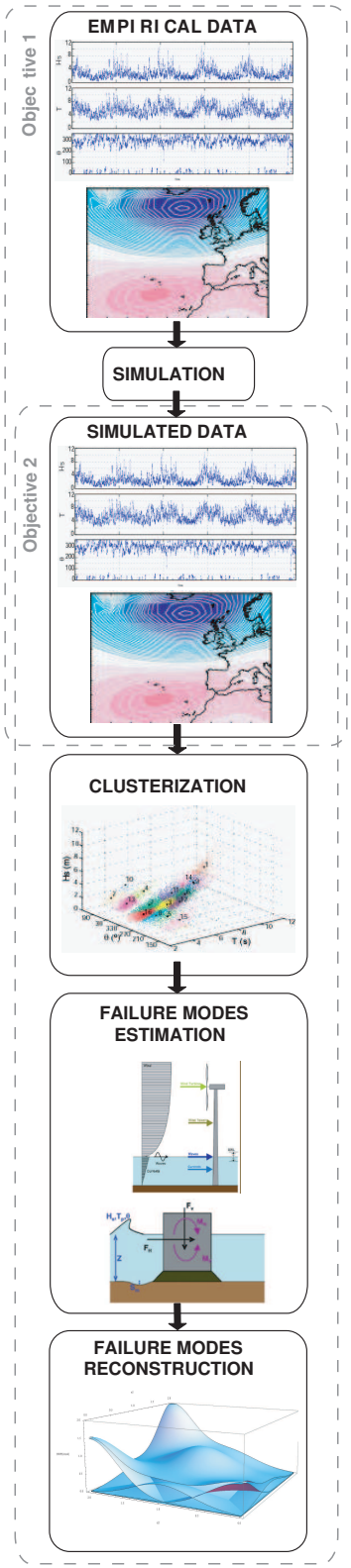

Figure 1.7: Diagram of the methodology.

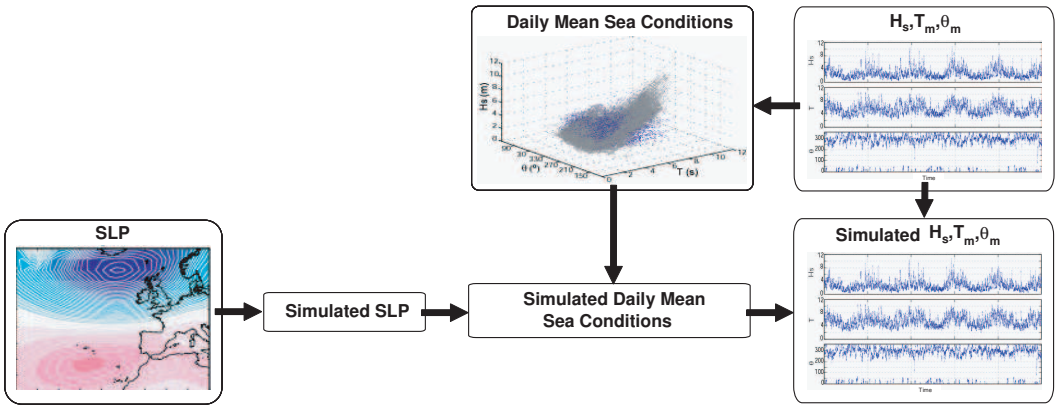

Figure 1.8: Diagram of the simulation process.

2. Simulation of mean sea conditions at the structure location. In this step the spatial scale is reduced to match that of the actual location of the structure.

3. Simulation of hourly sea states at the structure location. This third step is made to preserve the initial temporal resolution.

5.2 Transference

Alternatively, once the synthetic time series at the structure location are available, loads and other parameters can be estimated. The parameters to be estimated are those on which the principal failure modes depend. This process is undertaken using the following steps:

1. Sea state clusterization in a certain number of representative patterns.

2. Loads/scour,... produced by the representative cases are estimated. This estimation can be made by using semi–empirical formulation, numerical models or even physical models.

3. Reconstruction of the loads/scour,... time series by using an interpolation technique.

The stated methodology is developed considering the empirical distributions of the variables. This may be a limitation because this way does not take into account

the different behaviours of mean and extreme values. Moreover, the simulation is constrained to the historical events recorded. Regarding these issues, a univariate technique to solve them has been developed. This technique is able to consider both regimes simultaneously, reproducing both the extreme events and mean conditions accurately. Moreover, the upper tail is adjusted to an extreme model, allowing for the simulation of events not constrained to the historical maxima.

In this thesis work, each of the steps outlined in the proposed methodology will be explained in detail, allowing the achievement of the formulated objectives.

6 Organization of the Thesis

The Thesis is composed of 7 chapters where the previously stated objectives are deeply explored. The main body of this document, Chapters 2–6, each correspond to one paper that has been already published or is under consideration for acceptance. Each paper has only been changed in terms of format to adapt it to the present document.

The papers follows the same outline of the previous section. Chapters 2 and 3 solve the problem of multivariate simulation of met-ocean variables. In order to achieve this a logistic model has been developed to deal with aggregated multivariate systems (Chapter 2). Then, based on the developed model and combining it with another technique existent in the literature ([120]) a Monte Carlo simulation method applied to trivariate sea states is developed (Chapter 3). This method takes into account the influence of the atmospheric conditions. Chapters 4 and 5 present a hybrid model to estimate design parameters from met-ocean time series efficiently and saving computational effort. This model is applied to two different kinds of structures: a vertical breakwater (Chapter 4) and an offshore wind turbine (Chapter 5). Finally, Chapter 6 explores the possibility of dealing simultaneously with extreme and mean regimes. In Chapter 7 a summary of the conclusions obtained from the work is given and some future research topics are proposed.

Chapter 2. *Autoregressive logistic regression applied to atmospheric circula-*

tion patterns.[1]

Based on the atmospheric pressure fields, a logistic model to simulate circulation patterns is developed. In order to do this, pressure fields are classified into synoptic patterns by a clusterization technique, such as k-means. This model allows the use of nominal variables (in this case, the pressure fields previously classified) and the inclusion of previous states (autoregressive behaviour, Markov Chain), seasonality, trends, and other covariates of influence in the process.

Chapter 3. *Climate-based Monte Carlo simulation of trivariate sea states.*[2]

Combining the autoregressive model explained in Chapter 2 and a technique of multivariate simulation proposed by [120], this chapter develops a methodology to simulate hourly trivariate sea states. This method can be divided into three steps: i) simulation of the pressure fields, ii) simulation of the mean sea conditions and iii) simulation of the hourly sea states. The process undertaken enables the consideration of different temporal and spatial scales.

Chapter 4. *A simplified method to downscale wave dynamics on vertical breakwaters.*[3]

By means of the application of a hybrid method wave climate, time series are transformed into loads and scour time series in front of a vertical breakwater. This hybrid method combines a clusterization technique of maximum dissimilarity (MDA) and a later interpolation algorithm (RBF). Time series of the parameters involved in structure design are thereby, obtained.

Chapter 5. *A multivariate approach to estimate design loads for offshore wind*

[1] Guanche,Y., Mínguez, R. and Méndez, F.J. (2013). Autoregressive logistic regression applied to atmospheric circulation patterns. *Clymate Dynamics.* doi: 10.1007/s00382-013-1690-3

[2] Guanche,Y., Mínguez, R. and Méndez, F.J. (2013). Climate-based Monte Carlo simulation of trivariate sea states. *Coastal Engineering, under revision*

[3] Guanche,Y., Camus, P., Guanche, R., Méndez, F.J. and Medina R. (2013). A simplified method to downscale wave dynamics on vertical breakwaters. *Coastal Engineering.* **71** 68–77

turbines.[1]

This chapter, as an extension of the previous one, shows the application of the same methodology to an offshore wind turbine. In this case wave conditions as well as wind and current conditions are relevant. With the application of the methodology, loads and momentum time series acting upon the structure over its lifetime are obtained.

Chapter 6. *Point-in-time and extreme-value probability simulation technique for engineering design.*[2]

As stated before, the previous chapters' simulations were based on the use of the empirical distributions involved in each case. Nevertheless, it is true that there is a discrepancy between the extreme events distribution and the mean regime. Because of this discrepancy, this chapter explores the possibility of jointing both regimes and taking them into account simultaneously during the simulation. In order to demonstrate this simulation technique, an application has been added to the already published paper; the environmental conditions for offshore wind turbines have been analyzed and the 50–year return period environmental contours have been evaluated.[3]

Chapter 7. *Conclusions and Future Research.*

This last chapter summarizes all the conclusions from the presented work. Moreover, some future research topics suggested by this study are presented.

[1] Guanche,Y., Guanche, R., Camus, P., Méndez, F.J. and Medina R. (2012). A multivariate approach to estimate design loads for offshore wind turbines. *Wind Energy.* doi: 10.1002/we.1542

[2] Mínguez, R., Guanche,Y., and Méndez, F.J. (2012). Point-in-time and extreme-value probability simulation technique for engineering design. *Structural Safety.* **41** 29–36

[3] Mínguez, R., Guanche,Y., Jaime, F.F., Méndez, F.J. and Tomás, A. (2013). Filling the gap between point-in-time and extreme value distributions. *to be presented in ICCOSAR '13, New York.*

CHAPTER

2

Autoregressive logistic regression applied to atmospheric circulation patterns

1 Abstract

Autoregressive logistic regression (ALR) models have been successfully applied in medical and pharmacology research fields, and in simple models to analyze weather types. The main purpose of this chapter is to introduce a general framework to study atmospheric circulation patterns capable of dealing simultaneously with: seasonality, interannual variability, long-term trends, and autocorrelation of different orders. To show its effectiveness on modeling performance, daily atmospheric circulation patterns identified from observed sea level pressure (DSLP) fields over the Northeastern Atlantic, have been analyzed using this framework. Model predictions are compared with probabilities from the historical database, showing very good fitting diagnostics. In addition, the fitted model is used to simulate the evolution over time of atmospheric circulation patterns using Monte Carlo method.

Simulation results are statistically consistent with respect to the historical sequence in terms of i) probability of occurrence of the different weather types, ii) transition probabilities and iii) persistence. The proposed model constitutes an easy-to-use and powerful tool for a better understanding of the climate system.

2 Introduction

The study of atmospheric patterns, weather types or circulation patterns, is a topic deeply studied by climatologists, and it is widely accepted to disaggregate the atmospheric conditions over regions in a certain number of representative states. This consensus allows simplifying the study of climate conditions to improve weather predictions and a better knowledge of the influence produced by anthropogenic activities on the climate system [83, 84, 85, 129].

The atmospheric pattern classification can be achieved by using either manual or automated methods. Some authors prefer to distinguish between subjective and objective methods. Strictly speaking, both classifications are not equivalent because, although automated methods could be regarded as objective, they always include subjective decisions. Among subjective classification methods and based on their expertise about the effect of certain circulation patterns, [77] identify up to 29 different large scale weather types for Europe. Based on their study, different classifications have been developed, for instance, [58], [59] and [163] among others. To avoid the possible bias induced by subjective classification methods, and supported by the increment of computational resources, several automated classification (clusterization) methods have been developed, which may be divided into 4 main groups according to their mathematical fundamentals: i) threshold based (THR), ii) principal component analysis based (PCA), iii) methods based on leader algorithms (LDR), and iv) optimization methods (OPT). A detailed description of all these methods and their use with European circulation patterns can be found in [129].

Once the atmospheric conditions have been reduced to a catalogue of representative states, the next step is to develop numerical models for a better understanding of the weather dynamics. An appropriate modeling of weather dynamics is very useful for weather predictions, to study the possible influence of well-known

synoptic patterns such as East Atlantic (EA), North Atlantic Oscillation (NAO), Southern Oscillation Index (SOI), etc., as well as to analyze climate change studying trends in the probability of occurrence of weather types, and so on. For example, [153] investigated long term trends in annual frequencies associated with weather types, demonstrating the utility of weather classification for climate change detection beyond its short-term prognosis capabilities. [124] studied the dynamics of weather types using 1st order Markovian and non-Markovian models, however seasonality is not considered. [89] introduced a seasonal Markov chain model to analyze the weather in the central Alps considering three weather types. The transition probabilities are determined using a linear logit regression model. [126] implemented a cyclic Markov chain to introduce the influence of the El Niño-Southern Oscillation (ENSO).

Generalized linear regression, and especially autoregressive logistic regression, has proved to be a promising framework for dealing with seasonal Markovian models, and not only for atmospheric conditions. Similar models have been applied successfully in medical and pharmacological research fields [10, 39, 130]. The main advantages of autoregressive logistic regression (ALR) are that i) it can be used to model polytomous outcome variables, such as weather types, and ii) standard statistical software can be used for fitting purposes.

The aim of this chapter is twofold; firstly, to introduce autoregressive logistic regression models in order to deal with weather types analysis including: seasonality, interannual variability in the form of covariates, long-term trends, and Markov chains; and secondly, to apply this model to the Northeastern Atlantic in order to show its potential for analyzing atmospheric conditions and dynamics over this area. Results obtained show how the model is capable of dealing simultaneously with predictors related to different time scales, which can be used to predict the behaviour of circulation patterns. This may constitute a very powerful and easy-to-use tool for climate research.

3 Autoregressive Logistic Model

Traditional uni- or multivariate linear regression models assume that responses (dependent variables or outcomes) are normally distributed and centered at a linear

function of the predictors (independent variables or covariates). For some regression scenarios, such as the case considered here, this model is not adequate because the response variable Y is categorical and its possible outcomes are associated with each weather type ($Y \in \{1, 2, \ldots, n_{wt}\}$ being n_{wt} the number of weather types), which are not normally distributed. Thus the necessity to dispose of alternative regression models.

Logistic regression was originally defined as a technique to model dependent binary responses ([12, 35]). The likelihood of the binary dependent outcome is expressed as the product of logistic conditional probabilities. [121] introduced the capability of dealing with transition probabilities using Markov chains, which was further explored by [39] to predict the outcome of the supervised exercise for intermittent claudication, extending the model to polytomous outcomes.

Let $Y_t;\ t = 1, \ldots, n$ be the observation weather type at time t, with the following possible outcomes $Y_t \in \{1, \ldots, n_{wt}\}$ related to each weather type. Considering $\boldsymbol{X}_t;\ t = 1, \ldots, n$ to be a time-dependent row vector of covariates with dimensions $(1 \times n_c)$, i.e. seasonal cycle, NAO, SOI, principal components of synoptic circulation, long-term trend, etc., the autoregressive logistic model is stated as follows:

$$\ln\left(\frac{\text{Prob}(Y_t = i|Y_{t-1}, \ldots, Y_{t-d}, \boldsymbol{X}_t)}{\text{Prob}(Y_t = i^*|Y_{t-1}, \ldots, Y_{t-d}, \boldsymbol{X}_t)}\right) = \alpha_i + \boldsymbol{X}_t\boldsymbol{\beta}_i + \textstyle\sum_{j=1}^{d} Y_{t-j}\gamma_{ij}; \quad \forall i = 1, \ldots, n_{wt}|i \neq i*, \tag{2.1}$$

where $\boldsymbol{\alpha}_i$ is a constant term and $\boldsymbol{\beta}_i$ $(n_c \times 1)$ and γ_{ij} $(j = 1, \ldots, d)$ correspond, for each possible weather type i, to the parameter vectors associated with covariates and d-previous weather states, respectively. Note that d corresponds to the order of the Markov model. The model synthesized in equation 2.1 provides the natural logarithm of the probability ratio between weather type i and the reference weather type i^*, conditional on covariates $\boldsymbol{X}_t$ and the d previous weather states, i.e. the odds. The left hand side of equation 2.1 is also known as *logit*. According to this

expression, the conditional probability for any weather type is given by:

$$\text{Prob}(Y_t = i|Y_{t-1},\ldots,Y_{t-d},\boldsymbol{X}_t) = \frac{\exp\left(\alpha_i + \boldsymbol{X}_t\boldsymbol{\beta}_i + \sum_{j=1}^{d} Y_{t-j}\gamma_{ij}\right)}{\sum_{k=1}^{n_{wt}} \exp\left(\alpha_k + \boldsymbol{X}_t\boldsymbol{\beta}_k + \sum_{j=1}^{d} Y_{t-j}\gamma_{kj}\right)};$$

$$\forall i = 1,\ldots,n_{wt}. \tag{2.2}$$

Note that in order to make parameters unique we impose an additional condition, which fixes the parameter values related to the reference weather i^* (arbitrary chosen) to zero.

3.1 Description of the parameters

Since the purpose of this chapter is to present a unique model able to reproduce different weather dynamic characteristics, including: seasonality, covariates influence, long-term trends, and Markov chains; the inclusion of these features in the model (2.1) will be briefly described in this subsection:

- **Seasonality:** It is known that there is a strong seasonality on weather type frequencies, for example, [89] modeled this effect for the weather in the central Alps. In their work the seasonality is introduced in the model as an autoregressive term but it could be also introduced by adding harmonic factors. Here, the seasonality is introduced in the model using harmonics as follows:

$$\pi^S = \beta_0^S + \beta_1^S \cos(wt) + \beta_2^S \sin(wt), \tag{2.3}$$

  where π^S represents the seasonality effect on the *logit*, t is given in years, β_0^S correspond to annual mean values, and β_1^S and β_2^S are the amplitudes of harmonics, $w = 2\pi/T$ is the angular frequency. Since β_0^S is a constant term, it replaces the independent term $\boldsymbol{\alpha}_i$ in 2.1. For this particular case, we choose T to be defined in years, and thus $T = 1$ and t is in annual scale. This means, for instance, that the time associated with day 45 within year 2000 is equal to $2000 + 45/365.25 = 2000.1232$. However, according to the definition of the harmonic argument ($wt = \frac{2\pi t}{T}$), t could be given in days, then T must be equal to 365.25.

Analogously to Autoregressive Moving Average (ARMA) models [14], seasonality can also be incorporated trough an autoregressive term at lag 365. Details about how to incorporate this autoregressive component are given in the autoregressive or Markov chain parameters description below.

- **Covariates:** To introduce the effect of different covariates, the model is stated as follows:

$$\pi^C = \boldsymbol{X}\boldsymbol{\beta}^C = (X_1, \ldots, X_{n_c}) \begin{pmatrix} \beta_1^C \\ \vdots \\ \beta_{n_c}^C \end{pmatrix} = \sum_{i=1}^{n_c} X_i \beta_i^C, \tag{2.4}$$

where π^C is the covariates effect on the *logit*, $\boldsymbol{X}$ is a row vector including the values of different n_c covariates considered (SOI, NAO, monthly mean sea level pressure anomalies principal components, etc.), and $\boldsymbol{\beta}^C$ is the corresponding parameter vector.

- **Long-term trends:** The long-term trend is a very important issue because many authors, such as [16, 64, 85] , perform a linear regression analysis using as predictand the probabilities of each weather type, and the time as predictor. However, mathematically speaking, this may conduct to inconsistencies, such as probabilities outside the range 0 and 1, which is not possible. To avoid this shortcoming, we use a linear regression model but for the logits, being considered as a particular case of covariate:

$$\pi^{LT} = \beta^{LT} t, \tag{2.5}$$

where π^{LT} represents the long-term trend effect on the *logit*, and t is given in years. The parameter represents the annual rate of change associated with the logarithm of the probability for each weather type, divided by the probability of the reference weather type, i.e. $\Delta \log \frac{p_i}{p_i^*}$. The regression coefficient β^{LT} is a dimensionless parameter, which for small values of the coefficient may be interpreted as the relative change in the odds $\frac{\delta p_i}{p_i^*}$ due to a small change in time δt. Note that 2.5 does not correspond to the typical trend analysis because trends are analyzed on logits. However, as numerical results show, this codification provides consistent results on long-term changes of the weather type probabilities.

- **Autoregressive or Markov chain:** The sequence of atmospheric circulation patterns can be described as a Markov chain. [89] proved that a first order autoregressive logistic model is appropriate for reproducing the weather types in the central Alps. This effect can be included in the model using the following parameterization:

$$\pi^{AR_d} = \sum_{j=1}^{d} Y_{t-j}\gamma_j, \tag{2.6}$$

where π^{AR_d} represents the autoregressive effect of order d on the *logit*. The order d corresponds to the number of previous states which are considered to influence the actual weather type, Y_{t-j} is the weather type on previous j-states, and γ_j is the parameter associated with previous j-state.

Note that each $Y_{t-j};\ j = 1, \ldots, d$ in 2.6 corresponds to a different weather type, according to the polytomous character of the variable. In order to facilitate parameter estimation using standard logistic regression techniques, the autoregressive parts must be transformed using a contrast matrix, such as the Helmert matrix [39] so that each covariate Y_{t-j} transforms into $n_{wt} - 1$ dummy variables $\boldsymbol{Z}^{t-j}$. The Helmer contrast matrix for transforming outcome Y_t into the dummy variable row vector $\boldsymbol{Z}^t$ is provided in Table 2.1. According to this transformation matrix, equation 2.6 becomes:

Y_t	$\boldsymbol{Z}^t\ (1 \times (n_{wt} - 1))$					
1	-1	-1	-1	$\ldots$	-1	-1
2	-1	-1	-1	$\ldots$	-1	1
3	-1	-1	-1	$\ldots$	2	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$n_{wt} - 2$	-1	-1	$n_{wt} - 3$	$\ldots$	0	0
$n_{wt} - 1$	-1	$n_{wt} - 2$	0	$\ldots$	0	0
n_{wt}	$n_{wt} - 1$	0	0	$\ldots$	0	0

Table 2.1: Helmert Contrast Matrix

$$\pi^{AR_d} = \sum_{j=1}^{d} Y_{t-j}\gamma_j = \sum_{j=1}^{d}\sum_{k=1}^{n_{wt}-1} Z_k^{t-j}\gamma_{jk}. \tag{2.7}$$

Regarding seasonality, and according to expression 2.7, it can be included in the model as follows:

$$\pi^{AR_{365}} = Y_{t-365}\gamma_{365} = \sum_{k=1}^{n_{wt}-1} Z_k^{t-365}\gamma_{365,k}, \tag{2.8}$$

which corresponds to an autoregressive component at lag 365.

Note that the prize for using standard logistic regression fitting is an increment on the number of parameters, i.e. from d to $d \times (n_{wt} - 1)$.

The model can include all these effects adding the logits, i.e. $\pi = \pi^S + \pi^C + \pi^{LT} + \pi^{AR_d}$. Thus, expression 2.2 can be expressed as follows:

$$\text{Prob}(Y_t = i|Y_{t-1}, \ldots, Y_{t-d}, \boldsymbol{X}_t) = \frac{\exp\left(\pi_i^S + \pi_i^C + \pi_i^{LT} + \pi_i^{AR}\right)}{\sum_{k=1}^{n_{wt}} \exp\left(\pi_k^S + \pi_k^C + \pi_k^{LT} + \pi_k^{AR}\right)}; \tag{2.9}$$

$\forall i = 1, \ldots, n_{wt}$.

In order to deal with different time-scales within the model: annual, monthly and daily; all the parameters to be included are transformed to the lowest scale considered, i.e. daily. Thus, we require a covariate value for each day. This value may be chosen assuming a piecewise constant function over the data period (a month for monthly data, a year for yearly data, and so on), which is the one considered here, or using interpolation and/or smoothing techniques, such as splines. Note that in our case, the same covariate value keeps constant for the entire month (during 30-31 days).

3.2 Data set-up

Once the mathematical modeling is defined, this section describes the data set-up from the practical perspective. Let $\boldsymbol{Y}$ correspond to the vector of weather types

		$\boldsymbol{x}$					
					Autocorrelation		
t	$\boldsymbol{Y}_t$	$\boldsymbol{y}$	Seasonality	Trend	Covariates	Lag 1	...	Lag d
t_1	Y_1	$y_{1,1} \dots y_{1,n_{wt}}$	$\cos(wt_1)$ $\sin(wt_1)$	t_1	$X_{1,1} \dots X_{1,n_c}$	$Z_{1,1}^{t-1} \dots Z_{1,n_{wt}-1}^{t-1}$	...	$Z_{1,1}^{t-d} \dots Z_{1,n_{wt}-1}^{t-d}$
t_2	Y_2	$y_{2,1} \dots y_{2,n_{wt}}$	$\cos(wt_2)$ $\sin(wt_2)$	t_2	$X_{2,1} \dots X_{2,n_c}$	$Z_{2,1}^{t-1} \dots Z_{2,n_{wt}-1}^{t-1}$	...	$Z_{2,1}^{t-d} \dots Z_{2,n_{wt}-1}^{t-d}$
t_3	Y_3	$y_{3,1} \dots y_{3,n_{wt}}$	$\cos(wt_3)$ $\sin(wt_3)$	t_3	$X_{3,1} \dots X_{3,n_c}$	$Z_{3,1}^{t-1} \dots Z_{3,n_{wt}-1}^{t-1}$	...	$Z_{3,1}^{t-d} \dots Z_{3,n_{wt}-1}^{t-d}$
$\vdots$	$\vdots$	$\vdots$ $\vdots$ $\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$ $\vdots$ $\vdots$	$\vdots$ $\vdots$ $\vdots$	$\vdots$	$\vdots$ $\vdots$ $\vdots$
t_n	Y_n	$y_{n,1} \dots y_{n,n_{wt}}$	$\cos(wt_n)$ $\sin(wt_n)$	t_n	$X_{n,1} \dots X_{n,n_c}$	$Z_{n,1}^{t-1} \dots Z_{n,n_{wt}-1}^{t-1}$	...	$Z_{n,1}^{t-d} \dots Z_{n,n_{wt}-1}^{t-d}$

Table 2.2: Data Setup for the Autoregressive Logistic Regression applied to Weather Types

at different times of dimensions $(n \times 1)$, so that $Y_t \in \{1, \dots, n_{wt}\}$. To deal with polytomous variables a matrix $\boldsymbol{y}$ of dimensions $(n \times n_{wt})$ is constructed as:

$$y_{tj} = \begin{cases} 0 & \text{if} \quad j \neq Y_t \\ 1 & \text{if} \quad j = Y_t \end{cases} ; \; \forall j = 1, \dots, n_{wt}; \; \forall t = 1, \dots, n. \tag{2.10}$$

Note that since only one weather type at a time is possible, $\sum_{j=1}^{n_{wt}} y_{tj} = 1; \; \forall t$. The matrix $\boldsymbol{x}$ of dimensions $n \times (3 + n_c + 1 + d \times (n_{wt} - 1))$ includes all predictors at each of the n observations. Three parameters for seasonality (2.3), n_c parameters for covariates (2.4), one parameter for the long term trend (2.5), and $d \times (n_{wt} - 1)$ parameters for the autocorrelation (2.7). The general data setup for the autoregressive logistic regression applied to weather types is provided in Table 2.2.

Note that the column associated with the seasonality constant term β_0 in 2.3, which corresponds to a column vector $(1, 1, \dots, 1)^T$, must be included in matrix $\boldsymbol{x}$ depending on the standard logistic model used. While some of those models automatically include this constant, others do not.

3.3 Parameter estimation

Parameter estimation is performed using the maximum likelihood estimator, which requires the definition of the likelihood function. For a given sequence of n weather types $\boldsymbol{Y}$, the likelihood function becomes:

$$\ell(\boldsymbol{\Theta}, \boldsymbol{Y}, \boldsymbol{X}_t) = \prod_{t=1}^{n} \prod_{i=1}^{n_{wt}} \text{Prob}(Y_t = i | Y_{t-1}, \dots, Y_{t-d}, \boldsymbol{X}_t)^{u_{ti}}, \tag{2.11}$$

where $\boldsymbol{\Theta}$ is the parameter matrix, and the auxiliary variable u_{ti} is equal to:

$$u_{ti} = \begin{cases} 0 & \text{if} \quad y_t \neq i \\ 1 & \text{if} \quad y_t = i \end{cases} ; \forall i = 1, \ldots, n_{wt}; \ \forall t = 1, \ldots, n. \tag{2.12}$$

Note that the likelihood function (2.11) is the product of univariate logistic functions.

An important issue for the appropriate modeling of weather types, is to decide whether the inclusion of a covariate is relevant or not. There are several tests and methods to deal with this problem, such as Akaike's information criteria or Wald's test. Further information related to logistic regression parameterization and fitting can be found in [47, 138, 162].

There are several statistical software packages which are able to solve a polytomous logistic regression fitting (e.g. `SYSTAT`, `NONMEM`), but for this particular case, the function `mnrfit` in MATLAB has been used. This function estimates the coefficients for the multinomial logistic regression problem taking as input arguments matrices $\boldsymbol{x}$ and $\boldsymbol{y}$ from Table 2.2.

4 Case study: Weather types in the Northeastern Atlantic

In the last decade, the availability of long term databases (reanalysis, in situ measurements, satellite) allows a detailed description of the atmospheric and ocean variability all over the globe, which include the analysis and study of atmospheric patterns. To show the performance of the proposed model, Daily Sea Level Pressure (DSLP) data from NCEP-NCAR database [90] have been used. The area under study corresponds to the Northeastern Atlantic covering latitudes from $25°$ to $65°$N and longitudes from $52.5°$W to $15°$E. The data record covers 55 years, from 1957 up to 2011. Note that NCEP-NCAR data records start in 1948, however it is accepted by the scientific community that recorded data up to 1957 is less reliable [93].

The first step to apply the proposed method is data clustering. However, in order to avoid spatially correlated variables that may disturb the clusterization, a principal components analysis is applied to the daily mean sea level pressures (DSLP). From

this analysis, it turns out that 11 linearly independent components represent 95% of the variability.

As proposed by several authors, such as [34, 51] and [96] among others, the non-hierarchical K-means algorithm is able to classify multivariate patterns into a previously determined number of groups, eliminating any subjectivity in the classification. To reduce the likelihood of reaching local minima with the algorithm, clusterization is repeated a hundred times, each with a new set of initial cluster centroid positions. The algorithm returns the solution with the lowest value for the objective function. In this application, the daily mean sea level pressures corresponding to the 55 years of data ($n = 20088$ days), represented by 11 principal components, are classified into $n_{wt} = 9$ groups.

Note that in this particular case we select 9 weather types for the sake of simplicity, to facilitate the implementation, fit and interpretation of the model results. However, the selection of the appropriate number of clusters is an open issue not solved yet. There are authors, such as [34, 89, 160], that defend the use of a maximum of 10 weather types, others ([16, 87, 101]) claim that a higher number of weather types is required to represent the intrannual/interannual variations and seasonality appropriately. Being more specific, [34] uses only 4 weather types to represent daily precipitation scenarios, [51] classifies into 20 weather types the daily atmospheric circulation patterns, or for example, [87] uses 64 weather types to study the extreme wave height variability. This methodology does not solve the problem of establishing the appropriate number of weather types, which must be decided by the user according to his/her experience. But due to the facility to implement, fit and interpret model results might help establishing a rationale criteria for solving this problem.

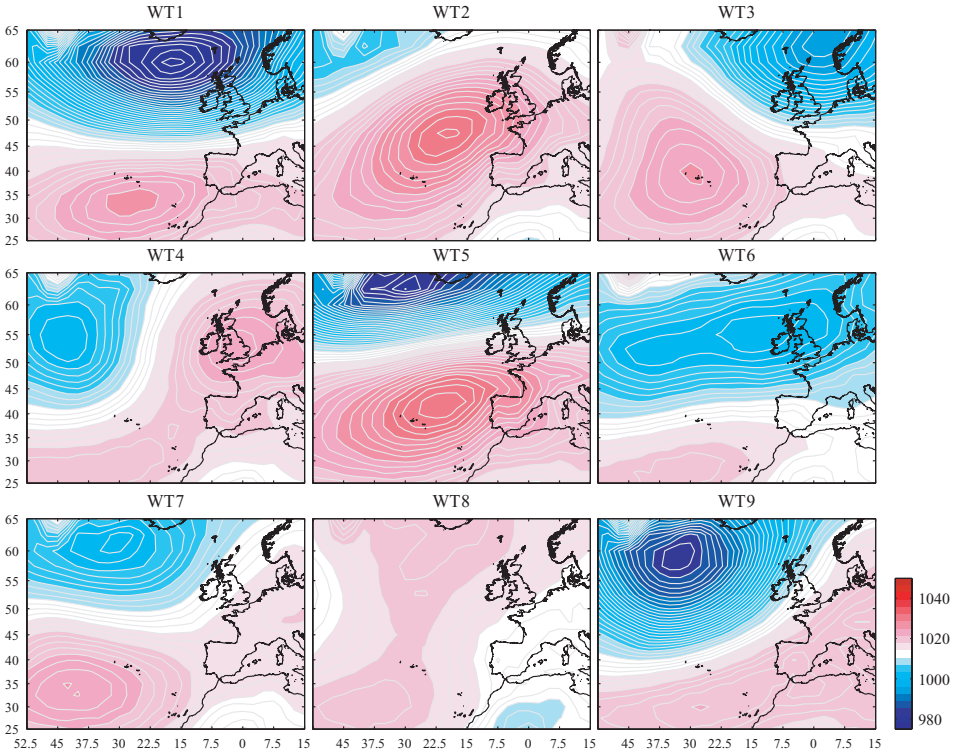

Figure 2.1: DSLP synoptical patterns associated with the clusterization.

Figure 2.1 shows the 9 representative weather types obtained from the clusterization. For instance, the upper left subplot represents a synoptical circulation pattern with a low pressure center above the Britannic Islands while the Azores High remains southwestern the Iberian Peninsula, whereas the upper central subplot shows the Azores High with its center southwest of the United Kingdom.

Assigning arbitrarily an integer value between 1 and $n_{wt} = 9$, for each weather type in Figure 2.1, we get the time series of weather types $\boldsymbol{Y}$, which is the input for the model.

To fit the data and according to the parameterizations given in (2.3)-(2.7), long-term trend, seasonality, covariates and a first order autoregressive Markov chain are included. Each study and location may require a pre-process to select the parameters to be included according to their influence. Related to covariates, it is worth to mention that Monthly Sea Level Pressure Anomalies fluctuations (MSLPA) have been considered. These anomalies correspond to monthly deviations from the 55-

year monthly averages, which allows obtaining interannual variations. This interannual modulation can be related to well known synoptic patterns, such as EA, NAO, SOI, etc. [82], but we preferred to use the principal components of the anomalies to avoid discrepancies about what predictors should be used instead. Nevertheless, we could have used those indices within the analysis. In this case, the first 9 principal components of the monthly sea level pressure anomalies (MSLPA) that explain more than 96% of the variability are included as covariates. Figure 2.2 shows the spatial modes related to those principal components. Note, for instance, that the correlation between the first mode and NAO index is $r = -0.618$ and the correlation between the second mode and EA synoptic pattern is $r = 0.482$.

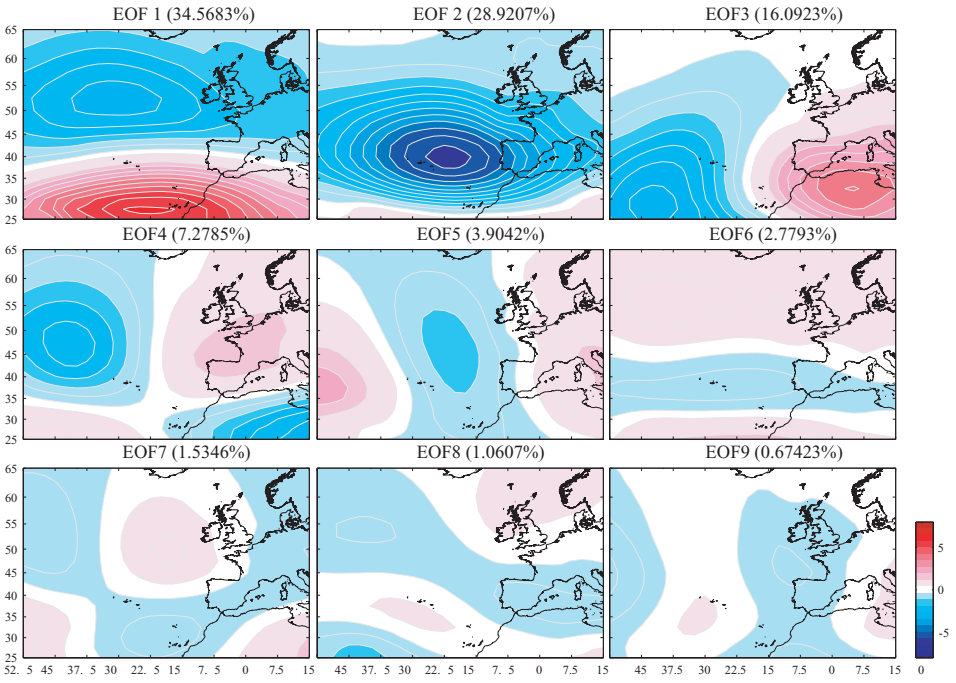

Figure 2.2: MSLPA spatial modes related to the Principal Components included as covariates in the model.

4.1 Model Fitting

Results obtained from the application of the proposed model to the Northeastern Atlantic are described in detail. The output given by function `mnrfit` is a matrix

$\hat{\boldsymbol{p}}$ of dimensions $(n_p \times (n_{wt} - 1))$ including parameter estimates by the maximum likelihood method, where n_p is the number of parameters in the model and n_{wt} is the number of weather types considered. Note that each weather type has an associated parameter except for the reference weather type, whose parameters are set to zero.

The criteria to choose the final model, i.e. the order d of the auto-regressive component, seasonality, covariates, etc. is based on statistical significance, in particular, using the likelihood ratio (LR) statistic. This statistical method is appropriate to compare nested models by comparing the deviance ratio ΔDev., which measures the change of fitting quality for two different parameterizations, and the chi-square distribution with $\Delta df = \Delta n_p \times (n_{wt} - 1)$ degrees of freedom. Note that Δn_p is the difference in terms of number of parameters for both parameterizations. Basically, it tries to check if the increment of fitting quality induced by increasing the number of parameters is justified, i.e. does the increment on fitted parameters conduct to a better model? For instance, assuming a confidence level $\alpha = 95\%$, if $\Delta\text{Dev.} > \chi^2_{0.95,\Delta df}$, the improvement achieved by adding n_p additional parameters is significant. This test allows to analyze which parameters or covariates are relevant to represent climate dynamics in a particular location.

In order to evaluate the goodness-of-fit related to the predictors, several different fits are considered. In Table 2.3, up to 7 nested models are compared depending on the predictors involved. In this table, the number of parameters (n_p), the deviance of the fitting (Dev.), the degrees of freedom (df) and the rate of change on deviance (ΔDev.) are provided. Model 0 is the so-called $Null$ model that only takes into account an independent term (β_0). Model I adds the possible influence of seasonality(π^S), which according to the increment on deviance with respect to the null model $\Delta\text{Dev.} = 7417 > \chi^2{}_{95\%,16}$ is significant, confirming the hypothesis that there is a seasonality pattern in the occurrence of the different weather types. Model II includes seasonality and MSLPA covariates ($\pi^S + \pi^C$), which also provide significant information. Model III is fitted accounting for seasonality, MSLPA covariates and long-term trend ($\pi^S + \pi^C + \pi^{LT}$). In this particular case, the increment on quality fit induced by the inclusion of an additional parameter, related to long-term trend, is not significant, i.e. $\Delta\text{Dev.} = 9 < \chi^2{}_{95\%,8}$. Models IV and V include the influence of autoregressive terms (Markov Chains, MC) with

orders $d = 1$ and $d = 2$, respectively ($\pi^S + \pi^C + \pi^{LT} + \pi^{AR_d}$). Note that both autoregressive components are significant. Additionally, due to the importance of long-term changes in the probabilities of occurrence of the different weather types, a model that only takes the long term trend into account has also been fitted, model VI (π^{LT}). This additional factor is statistically significant $\Delta\text{Dev.} = 69 > {\chi^2}_{95\%,8}$, which means that there is a long-term evolution on the probability of occurrence related to each weather type. However, there is an inconsistency with respect to model III, where this factor is not statistically significant. The reason for this behaviour is simple, when using covariates, the long-term effects are implicitly included in the covariates and there is no reason to include additional effects not explained by those covariates.

It is important to point out that deciding which model is more appropriate for each case depends on weather dynamics knowledge of the user, and its ability to confront or contrast its feeling about which physical phenomena is more relevant, with respect to the statistical significance of the corresponding fitted model. The main advantage of the proposed method is that it provides an statistical and objective tool for deciding what information is more relevant to explain climate variability.

Model	Predictors	n_p	df	Dev.	Δ Dev.	${\chi^2}_{95\%,\Delta df}$
0	β_0	1	160696	85736		
					7417	26.9
I	π^S	3	160680	78319		
					10214	92.8
II	$\pi^S + \pi^C$	12	160608	68105		
					9	15.5
III	$\pi^S + \pi^C + \pi^{LT}$	13	160600	68096		
					22159	83.7
IV	$\pi^S + \pi^C + \pi^{LT} + \pi^{AR_1}$	21	160536	45937		
					327	83.7
V	$\pi^S + \pi^C + \pi^{LT} + \pi^{AR_2}$	29	160472	45610		
0	β_0	1	160696	85736		
					69	15.5
VI	π^{LT}	2	160688	85667		

Table 2.3: Fitting diagnostics for different model parameterizations, including number of parameters (n_p), the deviance of the fitting (Dev.), the degrees of freedom (df) and the rate of change on deviance (ΔDev.)

Note that as said in Section 3.1 of this chapter, the seasonality constant term β_0, which corresponds to a column vector $(1, 1, \ldots, 1)^T$ is automatically included in the model depending on the standard logistic model used. Using the function `mnrfit` this constant is automatically added, thus the null model (a) has $n_p = 1$ and the model fitted only with the trend (g) has $n_p = 2$.

If we consider model IV, which accounts for seasonality, MSLPA covariates, long-term trend and a first order autoregressive component as predictors ($\pi^S + \pi^C + \pi^{LT} + \pi^{AR_1}$), the model has 21 parameters, $n_p = 21 = 3 + n_c + 1 + d \times (n_{wt} - 1) = 3 + 9 + 1 + 1 \times 8$: i) three for seasonality π^S, ii) nine for the MSLPA principal components π^C, iii) one for the long-term trend π^{LT}, and eight for the dummy variables of the first autoregressive component π^{AR_1}.

Once the parameter estimates for the models $\hat{\boldsymbol{\Theta}}$ are known, the predicted probabilities $\hat{\boldsymbol{p}}$ for the multinomial logistic regression model associated with given predictors $\tilde{\boldsymbol{x}}$ can be easily calculated. This task can be performed using the MATLAB function `mnrval`, which receives as arguments the estimated parameters $\hat{\boldsymbol{\Theta}}$ and the covariate values $\tilde{\boldsymbol{x}}$. In addition, confidence bounds for the predicted probabilities related to a given confidence level ($\alpha = 0.99, 0.95, 0.90$) can be computed under the assumption of normally distributed uncertainty. Note that these probabilities $\hat{\boldsymbol{p}}$ correspond to the probability of occurrence for each weather type according to the predictor values $\tilde{\boldsymbol{x}}$.

These probabilities allow direct comparison with the empirical probabilities from the data, and the possibility to simulate random sequences of weather types. The graphical comparison between fitted model and observed data can be done in different time scales, aggregating the probabilities of occurrence within a year, year-to-year or for different values of the covariates (MSLPA).

- **Seasonality** To analyze the importance of seasonality, Figure 2.3 shows the comparison of the probabilities of occurrence for each weather type within a year. Color bars represent cumulative empirical probabilities, and black lines represent the same values but given by the fitted model I, which only accounts for seasonality using harmonics (panel above in Figure 2.3), and also using an autoregressive term at lag 365 (panel below in Figure 2.3). For each day within a year the bars represent cumulative probabilities of

occurrence of all the 9 weather types, which are calculated for each day using the 55 data associated with each year.

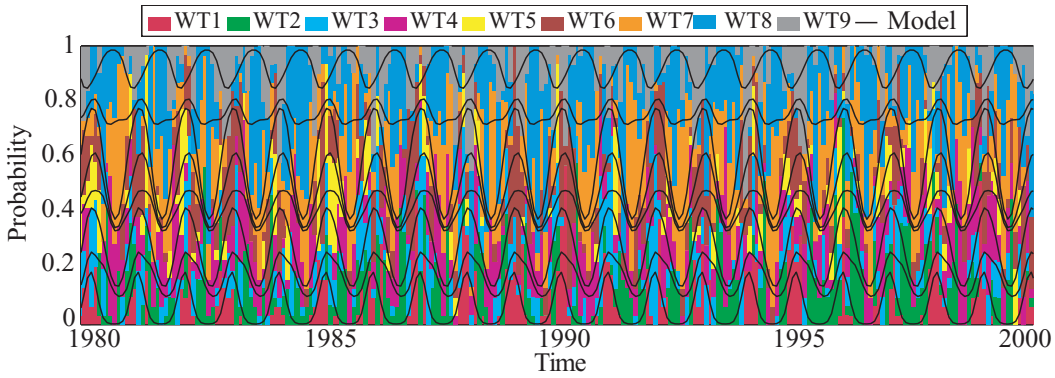

Figure 2.4: Evolution of the monthly probabilities of occurrence during 20 years and comparison with the seasonal fitted model I (black line).

Note that there is a clear seasonal pattern which is captured by the model using harmonics, being circulation patterns 4, 7 and 8 the most likely weather types during the summer, while groups 1, 6 and 9 are more relevant during the winter. Comparing both ways of accounting for seasonality, the harmonic (panel above of Figure 2.3) is capable of reproducing the seasonal behavior better than the autocorrelation term at lag 365 (panel below of Figure 2.3).

This seasonal variation through the years is also shown in Figure 2.4. In this particular case color bars represent cumulative monthly probabilities. Note that the model (black line) repeats the same pattern all over the years since we are using fitting results associated with model I. Analogously to the previous Figure 2.3, it is observed a clear seasonal pattern. For example, in the lower part of the graph it is observed how weather types 1 and 2, mostly related to winter and summer, respectively, change the occurrence probability depending on the season within the year. The same behavior is observed in the upper part of the graph related to weather types 3 and 9.

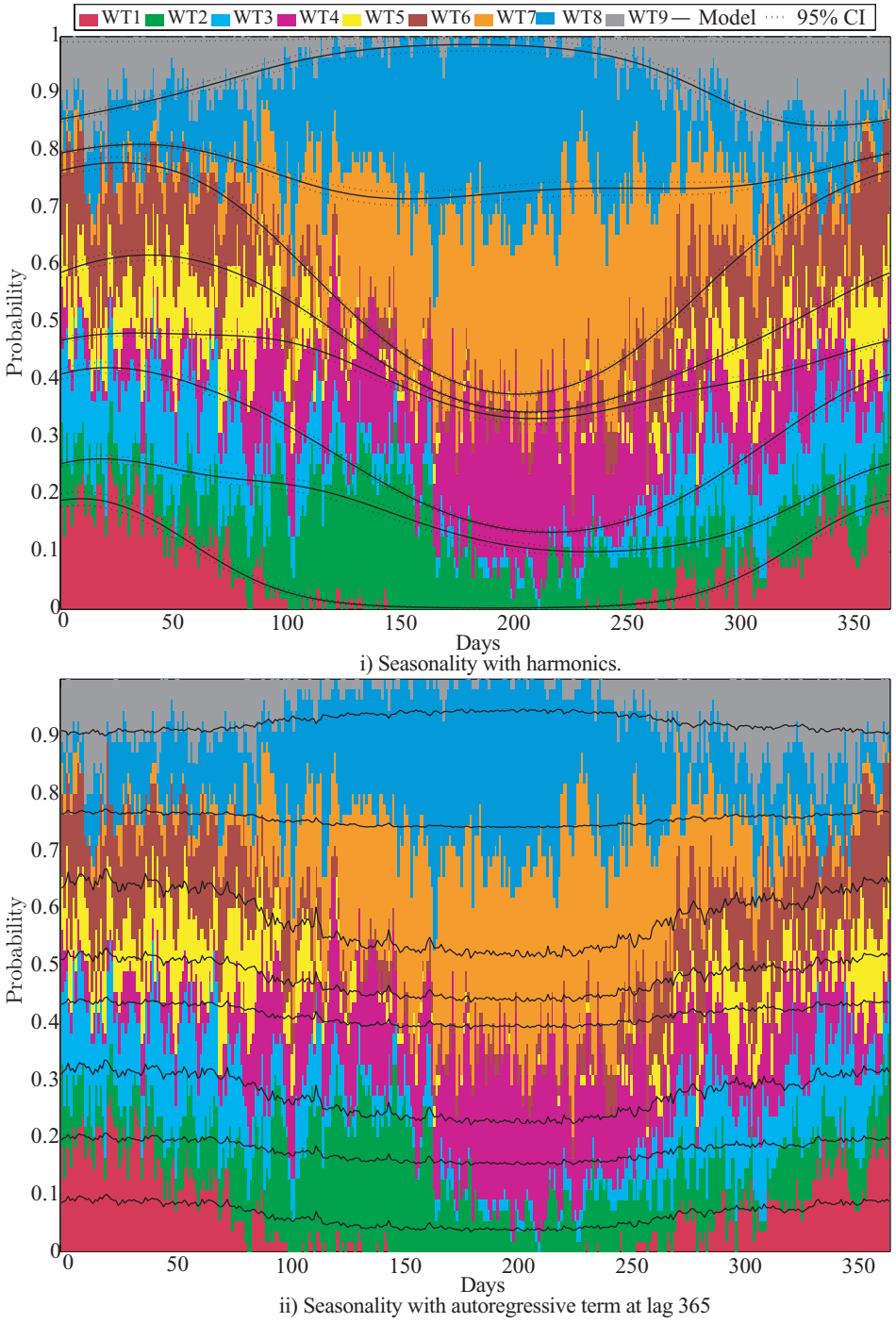

Figure 2.3: Model fitting diagnostic plot considering seasonality: i) using harmonics (Model I), and ii) using an autoregressive term at lag 365.

- **Mean Sea Level Pressure Anomalies (MSLPA)** Although model I reproduces and explains the important seasonality effect, it can be observed in Figures 2.3 and 2.4 that there are important fluctuations and discrepancies between the empirical data and the model on a daily and monthly basis, respectively. If model IV including seasonality, MSLPA covariates, an autoregressive component of order $d = 1$ and long-term trend ($\pi^S + \pi^C + \pi^{AR_1} + \pi^{LT}$) is considered, results are shown in Figures 2.5 and 2.6 . The fitted model now explains all fluctuations both on the daily and monthly scale.

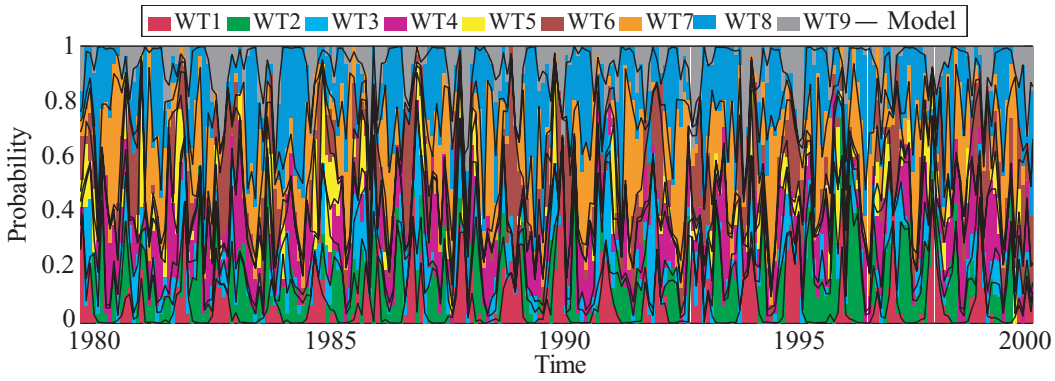

Figure 2.6: Evolution of the monthly probabilities of occurrence during 20 years and comparison with the seasonal fitted model IV (black line).

Note that once the noise on daily and monthly probabilities is explained by those additional factors, the consideration of seasonality through the 365-lag autoregressive model also provides similar diagnostic fitting plots, i.e. model $IV : \pi^S + \pi^C + \pi^{AR_1} + \pi^{LR} \equiv \pi^{AR_{365}} + \pi^C + \pi^{AR_1} + \pi^{LR}$.

It is relevant to point out how the inclusion of MSLPA allows explaining the monthly fluctuations on the probabilities of occurrence associated with the different weather types (see Figure 2.6). These results confirm that model IV is capable of reproducing and explaining the weather dynamics accurately, both on a daily and monthly basis. Using this model we manage to model atmospheric processes on both the short and the long term, using a combination

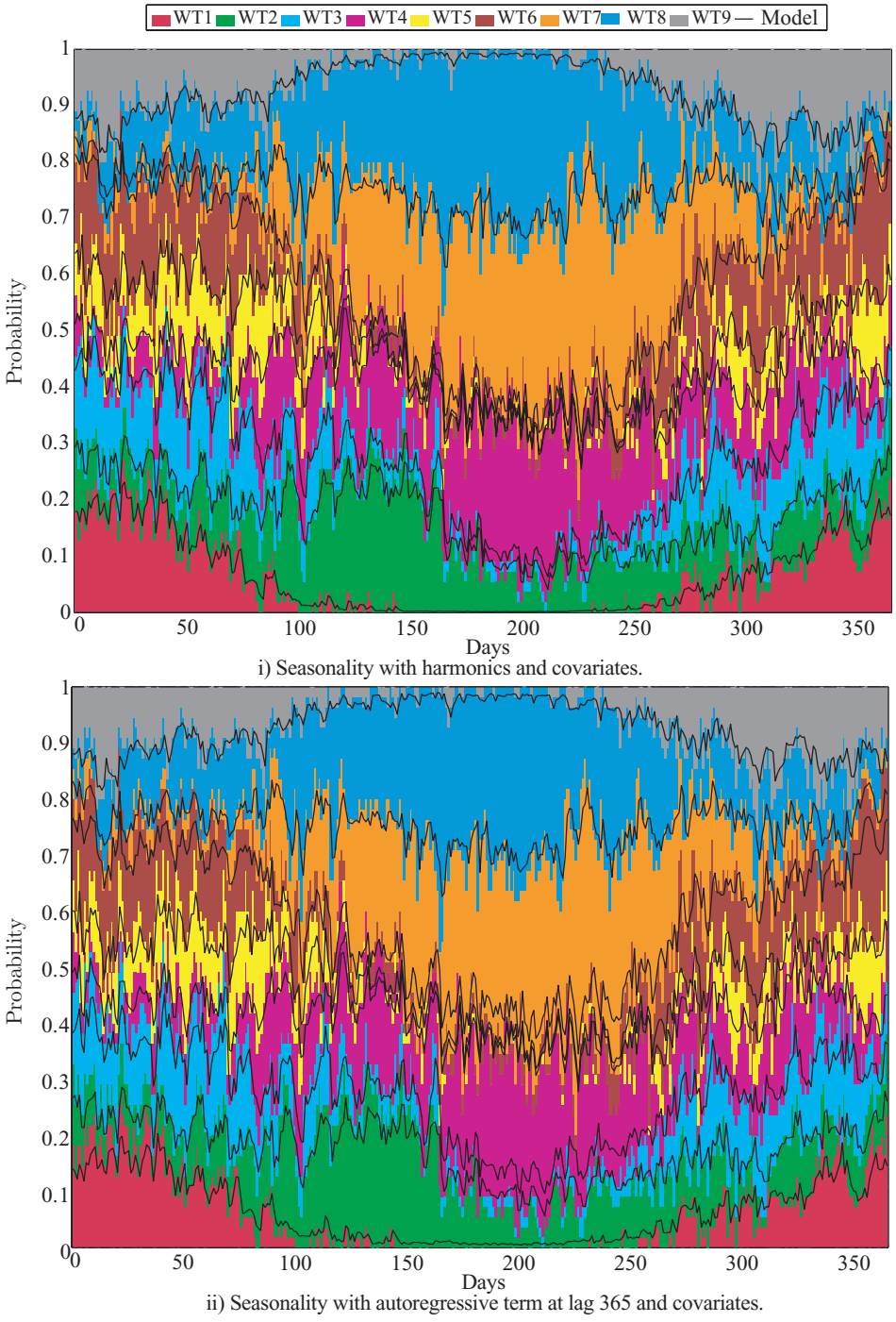

Figure 2.5: Model fitting diagnostic plot considering model IV: i) using harmonics (Model I), and ii) using an autoregressive term at lag 365.

of short-term sequencing through autocorrelation terms and long-term correlations included implicitly through seasonality, covariates and long-term variations.

To further explore the influence of the MSLPA on the occurrence probability for each weather type, Figure 2.7 shows the probability of occurrence of each weather type conditioned to the value of the MSLPA principal components ($PC_i;\ i = 1, \ldots, 9$) included as covariates. Color bars represent the cumulative empirical probabilities from data, and the black lines are fitted model probabilities.

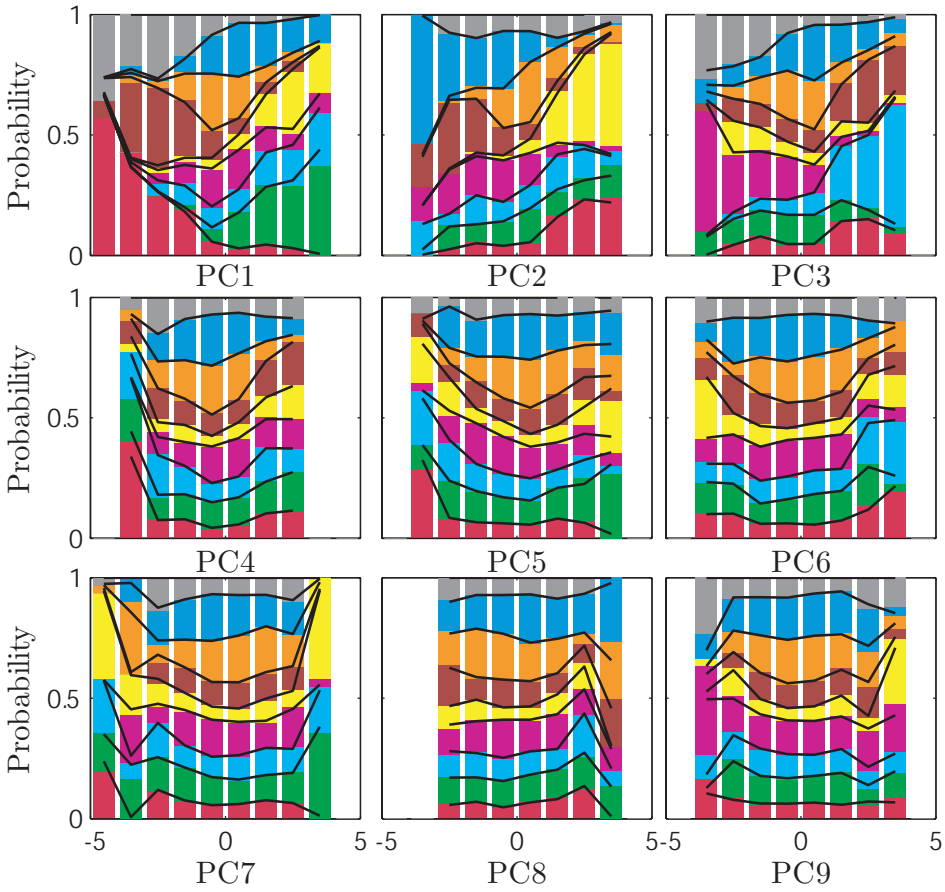

Figure 2.7: Evolution of the probabilities of occurrence of each weather type conditioned to the principal component value associated with fitted model IV (black line).

According to results shown in Figure 2.7, the presence or absence of a weather type may be related with the value of the PC anomaly. For instance, in the subplot associated with the first principal component (upper left subplot), negative values of the principal component imply an increment on the occurrence of weather types 1 (red), 6 (maroon) and 9 (grey); while for positive values the most likely weather types are 2 (green), 3 (light blue) and 5 (yellow). On the other hand, for negative values of the second principal component, the dominant weather type is the blue one (8), prevailing weather types 1 and 5 for positive values of the PC. Finally, for the third principal component, the behavior is different; the lowest values of this principal component indicate a higher likelihood of weather types 4 and 9, while higher values increase the probability of occurrence of weather type 3.

Note that according to the low variance explained by principal component from 4 to 9, we could be tempted to omit them from the analysis. To check whether these covariates improve significantly the quality of the fit, we have included the principal components one at a time, and check the likelihood ratio (LR) statistic. Table 2.4 provides the results from the analysis. Note that although it is clear that the most relevant information is given by the first three principal components, which represent important increments on deviance, the remainder covariates also improve the quality of the model from an statistical viewpoint. For this particular case, all principal components are statistically significant on a 95% confidence level.

- **Trends** Finally, in order to show the possible influence of a long-term trends, results associated with model VI, which only accounts for long term trends, are shown in Figure 2.8. Color bars represent the annual probability of occurrence for each year (55 data record) associated with the 9 established weather types. The black line represents the model fitting (model VI in Table 2.3). Note that we do not present results associated with model IV because the long term trend is not statisticcally significant in that model, because long-term effects are implicitly accounted for through the covariates.

Model	df	Dev.	Δ Dev.	$\chi^2_{95\%,8}$
0	160696	85736		
			4428	15.5
PC_1	160688	81308		
			3879	15.5
PC_2	160680	77429		
			2292	15.5
PC_3	160672	75137		
			225	15.5
PC_4	160664	74912		
			122	15.5
PC_5	160656	74790		
			61	15.5
PC_6	160648	74729		
			79	15.5
PC_7	160640	74650		
			99	15.5
PC_8	160632	74551		
			18	15.5
PC_9	160624	74533		

Table 2.4: Fitting diagnostics related to the principal components associated with MSLPA, including the deviance of the fitting (Dev.), the degrees of freedom (df) and the rate of change on deviance (ΔDev.)

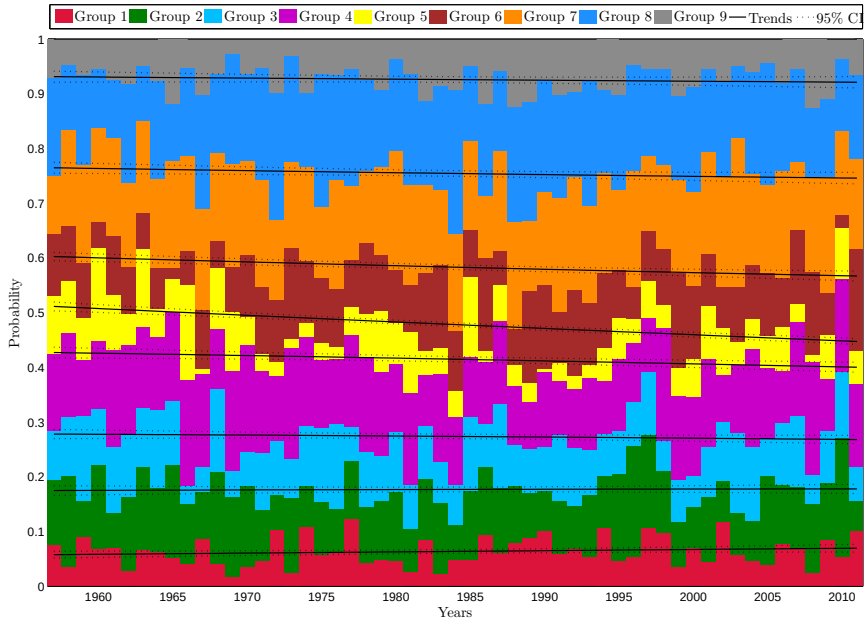

Figure 2.8: Annual probabilities of occurrence for each weather type and comparison with model VI fitting results (black line) in the period $1957 - 2011$.

	WT_1	WT_2	WT_3	WT_4	WT_5	WT_6	WT_7	WT_8
Trend$(\times 10^{-2})$	0.09	*-0.4*	**-0.51**	**-0.48**	**-1.33**	0.25	-0.07	-0.16
$\sigma_{Trend}(\times 10^{-2})$	0.24	0.21	0.22	0.20	0.24	0.21	0.19	0.19

Table 2.5: Fitting parameters associated with model VI including long-term trends, and their corresponding standard error. Values in bold are statistically significant at 95% confidence level and values in cursive are significant at 90% confidence level.

The parameters for the trends and their corresponding standard errors are provided in Table 2.5. Note that statistically significant trends at 95% confidence levels are boldfaced, while trends which are statistically significant at 90% confidence level are in italics. According to results given in this table the following observations are pertinent:

- The reference weather type is weather type number 9. That is the reason why there is no parameter related to this case. Note that it is a typical winter weather type.
- The coefficients may be interpreted as the relative change in the odds due to a small change in time δt, i.e. the percentage of change in odds between weather types 5 and 9 during one year is approximately equal to -1.33%.
- Weather types 4, 7 and 8, which represent summer weather types, decrease with respect to type 9. This means that weather types related to winter are increasing its occurrence probability. This result is consistent with recent studies about the increment of wave climate severity, which is linked to weather types during the winter season.
- Note that weather type 1, also typical during winter, slightly increases the odds with respect to type 9. Confirming the increment of occurrence related to winter weather types.

4.2 Monte Carlo Simulations

Once the model has been fitted and the $\hat{\boldsymbol{p}}$ matrix is obtained, synthetic sequences of weather types can be generated through Monte Carlo method. In this particular case, since we require the knowledge of the covariate values during the simulation period, 55 years of daily data series ($n = 20088$) are sampled using the original covariates. In order to obtain statistically sound conclusions according to the stochastic nature of the process, the simulation is repeated 100 times. The results obtained are validated with a threefold comparison against the original sequence of weather types: i) occurrence probabilities of WT, ii) transition probability matrix between WT and iii) persistence analysis of WT.

- **Occurrence Probabilities**

  The probabilities of occurrence of the 9 groups for the 100 simulations, against the empirical probability of occurrence from the 55-year sample data, are shown in Figure 2.9. Note that results are close to the diagonal, which demonstrates that the model simulations are capable of reproducing the probability of occurrence associated with weather types appropriately.

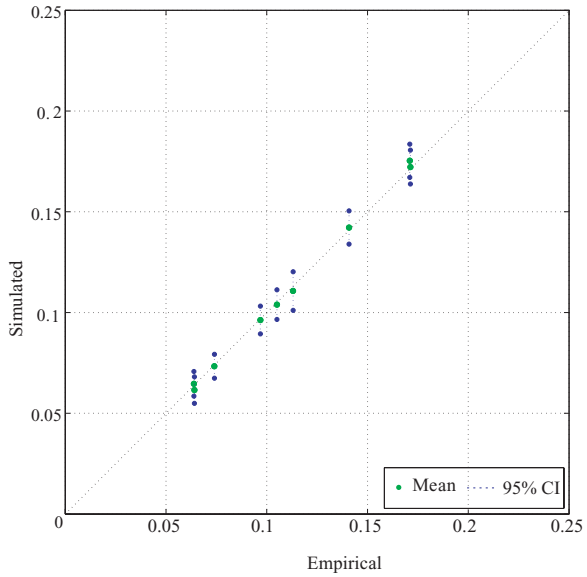

Figure 2.9: Scatter plot of the empirical occurrence probabilities associated with the weather types versus Monte Carlo simulation results.

- **Transition Probabilities Matrix**

  The transition probabilities express the probability of changing from group i to group j between consecutive days. Thus, in the case of having 9 weather types, the transition matrix (T) has dimensions 9×9, and each cell $T_{i,j}$ is the probability of changing from weather type i to weather type j ([132]). The diagonal of the transition matrix T corresponds to the probability of staying in the same group. The transition matrix is calculated for each of the 100 simulated samples. Figure 2.10 shows the scatter plot related to the $9 \times 9 = 81$ elements of transition matrix, including its uncertainty due to the simulation procedure, against the empirical transition probabilities obtained from the initial data set. The model is able to reproduce correctly the transitions between circulation patterns within the sequence. In this particular case, the points with probabilities in the range $0.6 - 0.8$ are those representing the probability of staying in the same group (diagonal of the transition matrix).

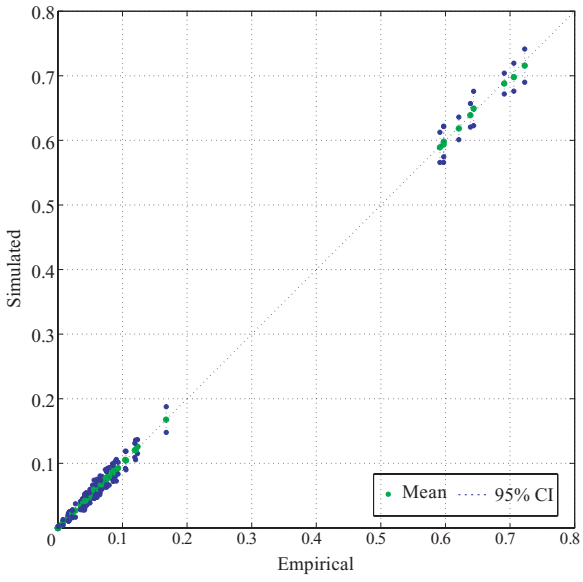

Figure 2.10: Scatter plot of the empirical transition probabilities between weather types versus Monte Carlo simulation results.

- **Persistence Analysis**

  Finally, a persistence analysis is performed over the simulated samples in order to check the ability of the model to reproduce weather dynamics. The correct reproduction of the weather types persistence is very important for many climate related studies, because it may be related to length of droughts, heat waves, etc. Figure 2.11 shows the empirical cumulative distributions of the persistence associated with each weather type. Note that the average empirical distribution (green line) is very close to the one related to the historical sample data (blue line) for all cases. This blue line stays between the 95% confidence intervals (red dotted line) related to the 100 simulations. To further analyze the performance on persistence from an statistical viewpoint, we perform a two-sample Kolmogorov-Smirnov ([102]) goodness-of-fit hypothesis test between the original data and each sampled data. This test allows determining if two different samples come from the same distribution without specifying what that common distribution is. In Figure 2.12 the box plots associated with the p-values from the 100 tests for each weather type are shown. Note that if the p-value is higher than the significance level (5%) the null hypothesis that both samples come from the same distribution is accepted. Results shown in Figure 2.12 prove that for most of the cases the persistence distributions from the Monte Carlo simulation procedure come from the same distribution as the persistence distribution from the historical data. For all the weather types the interquartile range (blue box) is above the 5% significance level (red dotted line). These results confirm the capability of the model to reproduce synthetic sequences of weather types coherent in term of persistence.

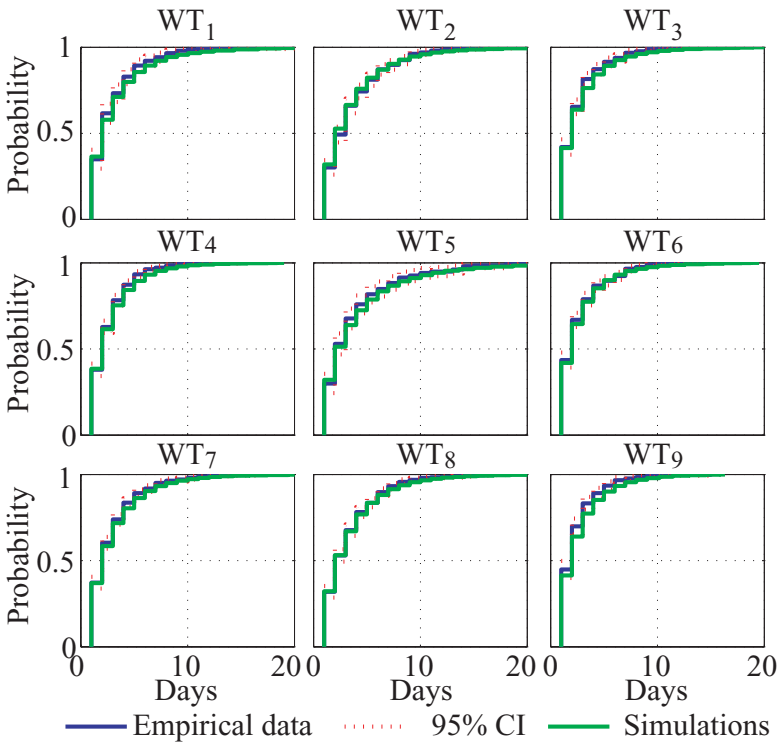

Figure 2.11: Empirical cumulative distribution of the persistence for the 9 groups related to: i) historical data and ii) sampled data using Monte Carlo method.

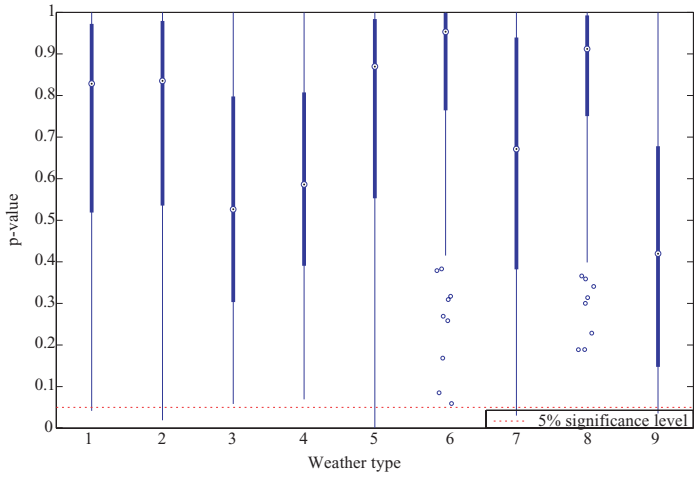

Figure 2.12: Box plot associated with the p-values from the 100 tests for each weather type.

5 Conclusions

This work presents an autoregressive logistic model which is able to reproduce weather dynamics in terms of weather types. The method provides new insights on the relation between the classification of circulation patterns and the predictors implied. The advances with respect to the state-of-the-art can be summarized as follows:

- The availability of the model to include autoregressive components allows the consideration of previous time steps and its influence in the present.
- The models allows including long-term trends which are mathematically consistent, so that the probabilities associated with each weather type always range between 0 and 1.
- The proposed model allows to take into account simultaneously covariates of different nature, such as MSLPA or autoregressive influence, where the time scales are completely different.
- The capability of the model to deal with nominal classifications enhances the physical point of view of the problem.
- The flexibility of the proposed model allows the study of the influence of any change in the covariates due to long-term climate variability.

On the other hand, the proposed methodology presents a weakness in relation with the data required for fitting purposes, because a long-term data base is needed to correctly study the dynamics of the weather types.

Although further research must be done on the application of the proposed model to study processes that are directly related with weather types, such as marine dynamics (wave height, storm surge, etc.) or rainfall, this method provides the appropriate framework to analyze the variability of circulation patterns for different climate change scenarios ([127]).

CHAPTER

3

Climate-based Monte Carlo simulation of trivariate sea states

1 Abstract

Accurate wave climate characterization, which is vital to understand wave-driven coastal processes and to design coastal and offshore structures, requires the availability of long term data series. Where existing data are sparse, synthetically generated time series offer a practical alternative. The main purpose of this chapter is to propose a methodology to simulate multivariate hourly sea state time series that preserve the statistical characteristics of the existing empirical data. This methodology combines different techniques such as univariate ARMAs, autoregressive logistic regression and K-means clusterization algorithms, and is able to take into account different time and space scales. The proposed methodology can be broken down into three interrelated steps: i) simulation of sea level pressure fields, ii) simulation of daily mean sea conditions time series and iii) simulation of hourly sea state time series. Its effectiveness is demonstrated by synthetically generating multivariate hourly sea states from a specific location near the Spanish Coast. The

direct comparison between simulated and empirical time series confirms the ability of the developed methodology to generate multivariate hourly time series of sea states.

2 Introduction

Sea condition data is required for the long-term analysis of wave-driven coastal areas and the design of coastal structures and related risk assessments. Depending on the kind of structure to be designed, the extreme value distribution, the long-term distribution or even both are required. There are several different data sources available for designers such as: buoy, satellite, reanalysis (or hindcast) data and visual observation records. To obtain an accurate characterization of the extreme conditions at a specific location, a long term data set is required, which is rarely available. Reanalysis data usually provide longer records and avoid missing data and sparse spatial resolution. In the last decade, long-term databases from numerical models have been developed improving the knowledge of deep water wave climate ([136], [30], [135]) and its propagation to obtain nearshore conditions ([19]). It is of note however, the length of these hindcast databases is limited, typically up to 60 years. Because of limited record lengths, synthetic stochastically similar time series may be required for structural design and flood and erosion risk analysis, for example.

Simulating synthetic time series in order to represent sea conditions is not new. In 1952, Longuett-Higgins gave an approach for the statistical distribution of wave heights based on a short term model ([99]). Since then, many authors have made contributions to achieve improved models representing time series of significant wave heights. A revision and an application to the Portuguese Coast can be found in [69]. More recently, [151] proposed a unified distribution model that mixes different fits for central, minimum and maximum regimes, objectively determining the thresholds within the different regimes.

Despite being one of the most important variables to define a sea state, the wave height (H_s) alone is not sufficient to fully characterise the prevailing wave conditions. As a minimum, the mean period (or peak period) associated with the

significant wave height is required. It is well known that wave period can be an influencing variable in many situations, such as overtopping or transmission through permeable breakwaters, for example. There exist in the technical literature different analyses that explore the joint distribution of $H_s - T_m$. For example, in [68], a bivariate autoregressive model is described that reproduces time series of significant wave height and mean period. In their work, an ARMA (Autoregressive Moving Average) model is used to represent both variables, significant wave height and mean period. Recently, in [48] a maximum entropy distribution of significant wave heights and peak periods was proposed. One of their approaches uses a univariate maximum entropy distribution, while an alternative approach consists of adjusting the maximum entropy marginal distribution of wave height, followed by conditioning the wave period distribution on the wave height. With an extension from the ARMA models to the VARMA models (Vector Autorregressive Moving Average), [17] proposed a multivariate simulation able to deal simultaneously with more than two variables but with the inherent complexity of the multivariate ARMA parameter estimation.

In [22], long-term statistics of storms are characterized by a set of three variables that represent the maximum significant wave height $H_{s_{\max}}$ during each storm, its maximum wave height $H_{\max}$, and the associated wave period $T_{z_{\max}}$ (that occurring with $H_{\max}$). The joint probability distribution and dependence structure is derived from real data so that once a storm has occurred, its intensity and characteristics can be derived from this joint distribution, i.e., a set of values $(H_{s_{\max}}, H_{\max}, T_{z_{\max}}$ can be drawn at random from a population with the corresponding distribution. Their model defines (1) The marginal distribution of $H_{s_{\max}}$, 2) The conditional distribution of $H_{\max}$ given $H_{s_{\max}}$ and (3) the conditional distribution of $T_{z_{\max}}$ given $H_{\max}$, $H_{s_{\max}}$. Alternatively, if more than two variables are included in the analysis, copula functions may be used instead. This way, [38] presents a multivariate model to study sea storms. In their study, up to four variables (wave height, storm duration, storm direction and time in between storms) are taken into account to develop a model capable of simulating sea storm behaviour.

[120] proposed a methodology to generate statistically dependent wind speed scenarios decoupling the process into univariate ARMA models and their cross-correlations. This allows the reproduction of more than two variables avoiding

the complexity of multivariate ARMA parameter estimation or the use of copula functions. This methodology provides good results not only in the joint distribution but also in the marginals.

Alternatively, to deal efficiently with long time-series, several clustering methods have been developed in the field of data mining. These techniques extract features from the original data, giving a representative subset of selected values.. The K-means algorithm (KMA) and self-organizing maps (SOM) are two of the most popular clustering techniques in this field. These allow the definition of a number of synoptic patterns. These algorithms have been widely used in both: atmospherical and marine climate data. For example, [87] classify SLP (Sea Level Pressure) fields and use them to explain the wave climate and its variability; while [19] use a clusterization of the met-ocean parameters to propose a wave propagation methodology.

Moreover, plausible time series of circulation patterns can be simulated using an autoregressive logistic model as described by [67]. In that approach, a model that takes into account seasonality, interannual variability in terms of sea level pressure anomalies, long-term trends and Markov Chains is developed to accurately reproduce stochastically similar time series of circulation patterns.

The aim of this chapter is to combine the different space and time scales to generate plausible long-term hourly time-series of trivariate (significant wave height, H_s, mean period, T_m and mean direction θ_m) sea state parameters by using some of the aforementioned techniques. To achieve this objective, a methodology has been developed that comprises three interrelated steps. In the first step, synthetic daily sea level pressures fields (DSLP), decomposed into principal components, are simulated by using the multivariate simulation technique proposed by [120]. During the second step, daily mean sea conditions (DMSC), clustered by K-means as proposed by [19], are simulated by applying an autoregressive logistic model and taking into account the previously simulated DSLP as covariates. The third step consists of a modified version of the methodology proposed by [120], to be used with hourly sea state (HSS) parameters and conditioned to the catalogue of synoptic DSS patterns simulated in the previous step.

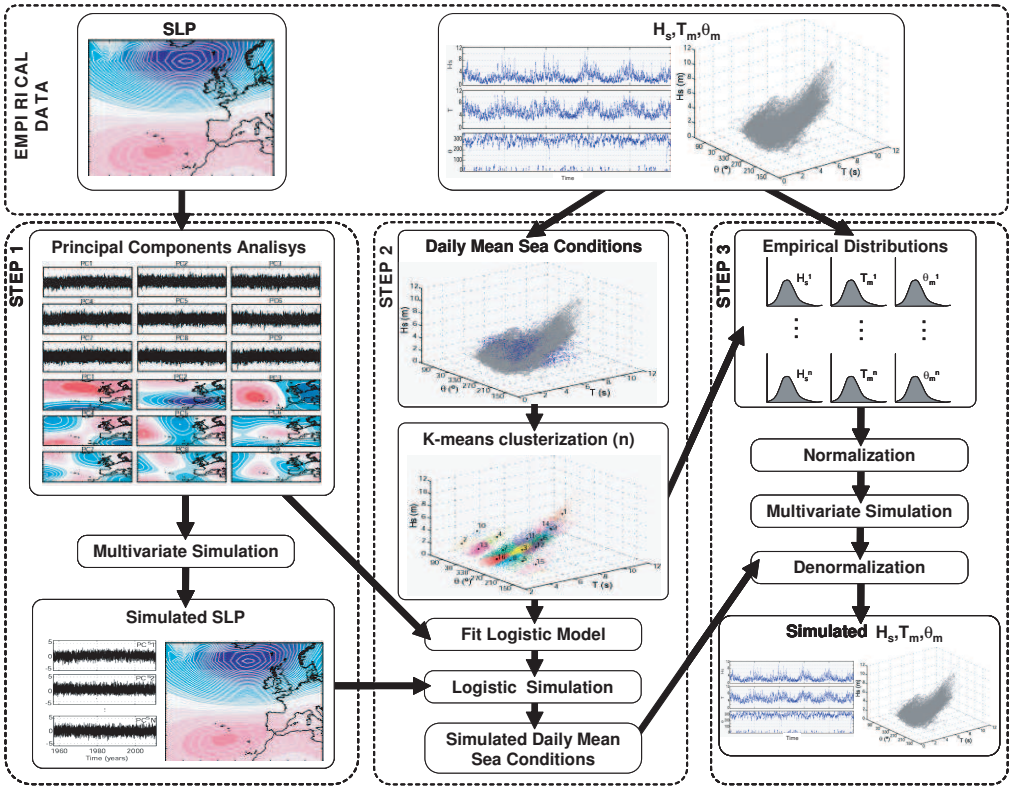

Figure 3.1: Diagram of the methodology.

3 Methodology

In this section, the procedure to generate plausible synthetic multivariate sea state time series is described step-by-step. As mentioned before, the entire methodology can be divided into three interrelated processes: the simulation of DSLP fields, the simulation of daily mean sea conditions (DMSC) and the simulation of hourly sea states (HSS) conditioned to daily mean sea conditions. A general framework of the methodology is shown in Figure 3.1 and is explained in detail below.

Although the final objective is to undertake a stochastic simulation of the wave parameter time series, both SLP and wave climate (H_s, T_m and θ_m) databases are needed as inputs for the algorithm. SLP fields in the wave generation area of the study location are used as a covariate (predictor). In addition, historical time series of the wave climate variables to be simulated (predictand), at the same location,

are also required . The SLP fields are daily averaged (prior to simulation). These fields relate to the required time for wave generation and cyclones evolution in the location where the methodology is to be applied. Within different study areas, the timescales of relevance will vary. Wave parameters data is analysed at two different temporal scales: daily and hourly. In the first instance, the data is daily averaged to establish the relationship with the daily averaged SLP field. This information is the one used in the subsequent stochastic simulation of hourly data, thus preserving the initial time resolution within the raw data. This enables the combination of the different temporal and spatial scales that are required in the subsequent analysis.

The first step of the proposed methodology is the simulation of the daily averaged SLP fields. These simulated fields are later used as an explicative variable when studying the daily mean wave conditions. The DSLP fields, decomposed into PCs, are simulated with a multivariate simulation technique ([120]). This technique enables both the autocorrelation of each variable and the cross-correlations between variables to be accounted for.

The second step consists of the simulation of daily mean sea conditions. In order to simplify this step, and to achieve more accurate results in the third step, the DMSC data is clustered (K-means) into groups, with each group containing data with similar characteristics. Thus, a discrete time series of daily mean conditions is obtained. The simulation of DMSC uses an autoregressive logistic model ([67]). This kind of model enables the consideration of previous states (autoregressive processes) as well as other explicative variables (covariates). In the case presented here, the model is fitted using the DSLP decomposed into PCs as covariates as well as the significant previous states (Autoregressive process). Once the model is fitted using the historical data, the previously simulated DSLP PCs are taken into account for synthetic DMSC time series simulation.

Finally, the third and last step of the proposed methodology consists of simulating the variables defining the wave parameters with an hourly time resolution. Prior to the simulation, the hourly and the daily historical time series are linked, using the original database timestamp, obtaining the empirical distributions of H_s, T_m and θ_m for each cluster. By using these empirical distributions, the three variables are normalized. This enables the use of the same multivariate simulation technique as in step 1. As detailed above, this technique accounts for the existing

autocorrelation in H_s, T_m and θ_m, as well as the cross-correlation between them. After simulating the normalized variables, and considering the empirical distributions, together with the synthetic sequence of DMSC generated before, the three variables are transformed back onto the original scales. The resulting H_s, T_m and θ_m time series show similar marginal and joint distributions as those obtained with the historical data. The disaggregation into daily condition groups improves the simulation results and preserves the seasonality of the variables. Moreover, the use of the synthetic DMSC is an indirect consideration of the Sea Level Pressure in the wave generation area related to the location of study.

3.1 STEP 1. Simulating Daily Sea Level Pressure (DSLP) fields

The simulation of daily sea level pressure fields during Step 1 is carried out by using the multivariate simulation method proposed by [120]. In order to do so, the original DSLP data are monthly standardized, and decomposed into principal components. By this, seasonality is avoided and the dimensionality of the DSLP data is reduced, which may affect the following steps of the methodology. The decomposition into principal component is especially useful for reducing the number of dimensions and to identify patterns in environmental data. This technique removes the data dependency and data redundancy with a minimum loss of variance, which is sometimes required for the assumptions made by many statistical methods. More information related to this analysis can be found in [87] and [133]. Once the data is pre-treated, the simulation is carried out.

The process of the multivariate simulation ([120]), once the DSLP is decomposed into PCs, can be summarized as follows:

1. For the N_{PC} principal components, a normalization is done according to the empirical distributions for each PC, obtaining the transformed variables Z_i, with $i = 1, ...N_{PC}$.

2. A univariate ARMA model is adjusted for each normalized variable (Z_i). This fitting process yields uncorrelated normal residuals (historical errors). Further information related to these models and their performance can be found in [14].

3. Although the historical error series obtained before are not autocorrelated, they should be cross-correlated if the variables involved are interrelated. To compute the cross-correlation between residuals, the variance-covariance matrix (G) is built. As explained in [120], the size of G is $(K + N_{PC}) \times N_s$, K being the number of lags (positive or negative) taken into account and N_s the length of the following simulations to be done.

4. Once the variance-covariance matrix is built, Cholesky decomposition is done (i.e., computing L such $G = LL^T$) obtaining the orthogonal transformation needed to cross-correlate the independent errors.

5. $\xi = N_s$ independent standard normal errors are simulated.

6. The generated errors are cross-correlated by using the orthogonal transformation obtained earlier.

7. For the N_{PC} components involved, using the ARMA model fitted previously and the N_s cross-correlated residuals corresponding to each variable, Zs_i transformed variables are simulated.

8. At this point, N_{PC} autocorrelated and cross-correlated series have been simulated with a length of N_s. The multivariate simulation ends after the denormalization of Zs_i, which is done using the empirical distribution of each PC.

A flowchart of Step 1 is depicted in Figure 3.2. A more detailed explanation of the multivariate simulation process is provided by [120].

3.2 STEP 2. Simulating Daily Mean Sea Conditions (DMSC) time series

In order to generate plausible sequences of DMSC an autoregressive logistic regression model is applied to the met-ocean variables data. [67] provides a detailed description of the model and one application to the simulation of synthetic series of circulation patterns of mean daily sea level pressures can be found. The aim of

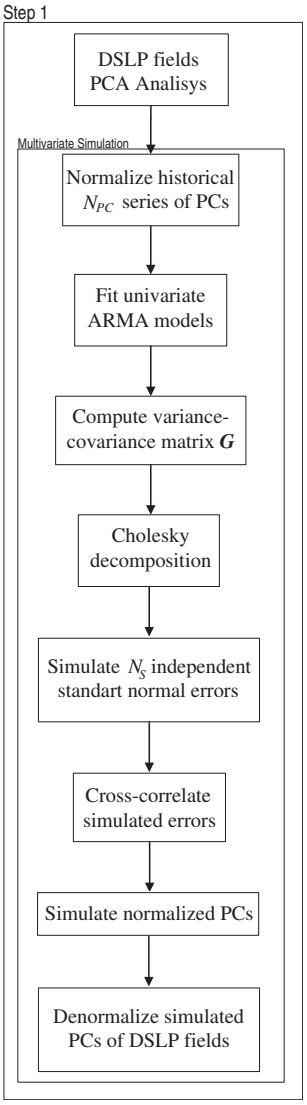

Figure 3.2: Flow diagram of the Step 1.

Step 2 is to simulate statistically similar time series of DMSC, taking into account the DSLP fields from Step 1. The sequence of Step 2 is:

1. Classification of the multivariate met-ocean variables of the DMSC into n groups by using a K-means classification technique. This technique splits the multivariate data into different clusters, each one defined by a centroid and formed by the data for which the centroid is the nearest.

2. Fitting the obtained DMSC sequence by using a logistic model. Seasonality, DSLP and Markov Chains of sea states and DSLP have been used as covariates. The DSLP, decomposed into PCs, are introduced to the model, while the seasonality is introduced by adding intra-annual harmonics. The fitting process allows the selection of the significant covariates to be included in the model and is unique for each study location. The PCs of the DSLP used to fit the model are obtained from the historical data, not the simulated.

3. Once the model is fitted, synthetic sequences of DMSC can be easily simulated. Taking into account the significant covariates established in the fitting process and the simulated PCs of DSLP obtained in Step 1, synthetic sequences of DMSC are then sampled.

Figure 3.3 shows the flow diagram for the sequence during Step 2 of the simulation process. The model parametrization, the theoretical foundation of the logistic models and a detailed example simulating weather type sequences, can be found in [67].

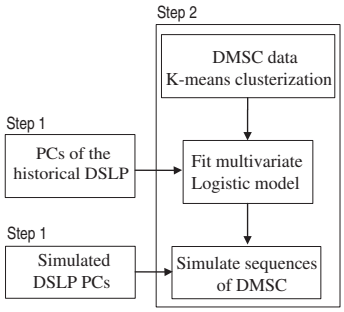

Figure 3.3: Flow diagram of the Step 2.

3.3 STEP 3. Simulating Hourly Sea States (HSS) time series

Once the sequence of DMSC is simulated taking into account the DSLP fields, the third and last step of the presented method is to transfer the daily data of sea mean conditions to an hourly scale. This is done by a modified version of the methodology proposed by [120]. The modification consists of a disaggregation of the entire

process into as many DMSC groups as defined in Step 2. This process ensures we capture the dominant properties of the data and thus improves in accuracy during the simulation process. The development of this step can be summarized as follows:

1. Transference of the DMSC original sequence to the empirical hourly sea state time series. This way, nxN groups are obtained, where n is the number of DMSC previously determined and N the number of variables taken into account. This disaggregation incorporates the historical sequence of DMSC.

2. All the n groups of N variables are normalized considering their empirical distribution(Z_i, with $i = 1, ...N$).

3. Fitting a univariate ARMA model for each Z_i.

4. Computing the variance-covariance matrix (G).

5. Cholesky decomposition ($G = LL^T$) .

6. Simulation of independent standard normal errors ($\xi = N_s$).

7. Cross-correlation of the independent errors.

8. Simulation of the transformed variables (Zs_i) by using the fitted ARMA models.

9. Denormalization of the N variables taking into account the n DMSC empirical distributions and the simulated DMSC temporal sequence.

Except for the initial decomposition of the hourly database into the DMSC groups, the process is analogous to the one in Step 1, and is represented in Figure 3.4.

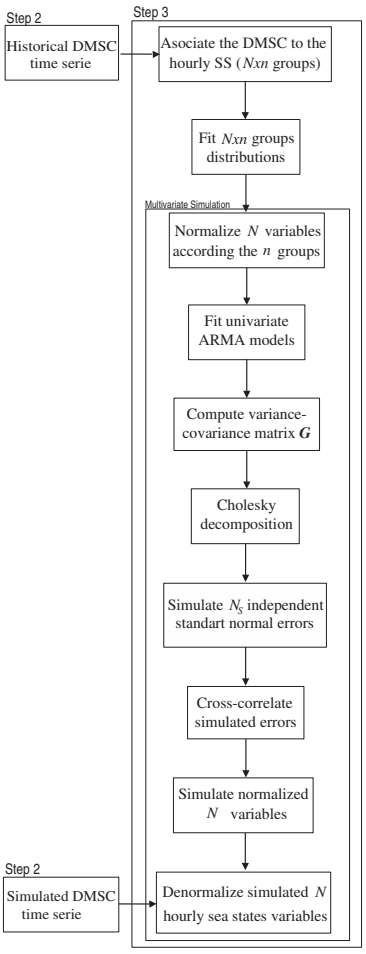

Figure 3.4: Flow diagram of the Step 3.

4 Case study: Hourly Sea States time series simulation in a location in NW Spanish coast.

An application to demonstrate the methodology has been undertaken on the North West Coast of Spain. The area is located close to Langosteira's harbour, in Galicia. In this section, the methodology is implemented and the results are verified.

4.1 Data

To show the performance of the proposed methodology, long-term data sets of waves and Sea Level Pressure fields are required. Data from reanalysis databases are used because of their length and consistency.

The SLP data is extracted from the NCEP-NCAR database, [90]. The area under study corresponds to the Northeastern Atlantic covering latitudes from $25°$ to $65°$N and longitudes from $52.5°$W to $15°$E. This area covers the wave generation area of the waves arriving to the Northwestern coast of Spain. The data record covers 55 years, from 1957 up to 2011. Note that NCEP-NCAR data records start in 1948, however it is accepted by the scientific community that recorded data up to 1957 is less reliable [93]. This database has a 6-hour temporal resolution, thus, in order to obtain the Daily Sea Level Pressure (DSLP), the four values of each day are averaged.

The hindcast wave data used in this work is: DOW 1.1 (Downscaled Ocean Waves, [19]) developed by IH Cantabria. These data comprises an hourly regional wave reanalysis for the period 1948-2008 with spatial resolution of $\sim$200 m along the Spanish coast. A hybrid method is used to downscale the GOW 1.1 reanalysis ([136]) to coastal areas, based on a propagation catalog using the SWAN model ([13]) and statistical techniques: MDA in the selection process and RBF in the time series reconstruction; and calibrated using instrumental data (see [50], [115] and [118]). The forcing inputs come from the SeaWind NCEP/NCAR dataset ([111]).

Both data sets comprise a common period of time from 1957 to 2008. This 52 year period of data constitutes the basic information for the implementation of the methodology.

4.2 Step 1

In order to avoid spatially correlated variables and to reduce dimensionality, a principal components analysis is applied to the daily mean sea level pressures (DSLP). Prior to the principal component analysis, the data seasonality is removed by undertaking a monthly standarization. This will help the simulation avoiding inaccurate results because the multivariate simulation is not able by itself to reproduce

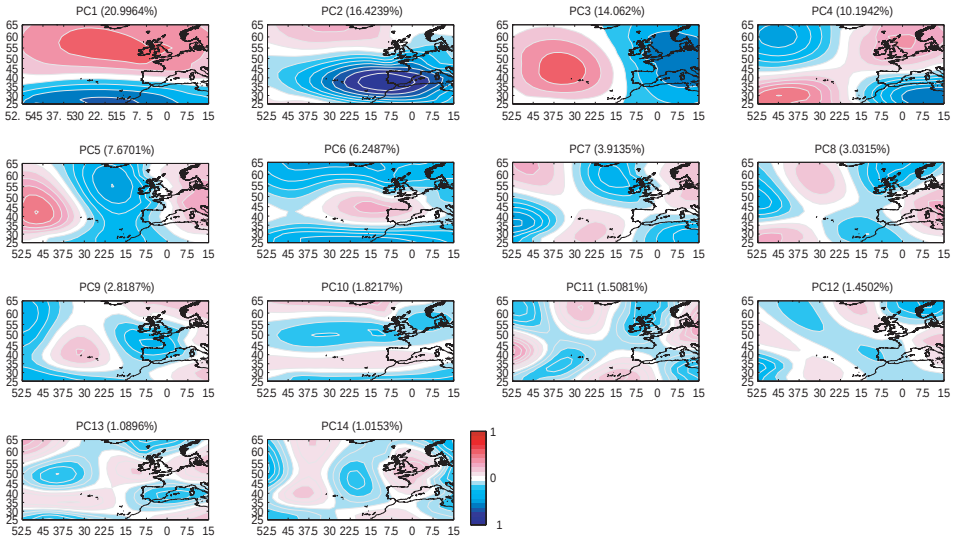

Figure 3.5: Spatial modes related to the DSLP Principal Components.

seasonality accurately. The PCs analysis shows that 14 ($N_{PC} = 14$) linearly independent components represent more than 92% of the variability. These 14 PCs are simulated by using the multivariate simulation technique proposed by [120]. Figure 3.5 shows the spatial modes related to those principal components.

Taking the empirical distributions into account the 14 PCs are standardized, obtaining the transformed variables Z_i, with $i = 1, ...14$. These transformed variables are fitted to univariate ARMA(p, q) models. In this particular case the use of an ARMA$(2, 1)$ model for all the time series provides appropriated results. The fitting process of these kind of models is a well-known process and further information can be found in [14].

The residual errors given in the comparison between the standardized PCs and the adjusted ARMAs are not autocorrelated, but they are cross-correlated if there exists temporal correlation between the different PCs. In Figure 3.6, the cross-correlation of residuals between some PCs are plotted. Due to the high amount of possible combinations of PCs, only those with higher correlation coefficients are shown. From the results obtained from the cross-correlation diagrams the number of lags (K) to be considered is determined, in this case $K = 4$.

The variance-covariance matrix G is built from the cross-correlations coeffi-

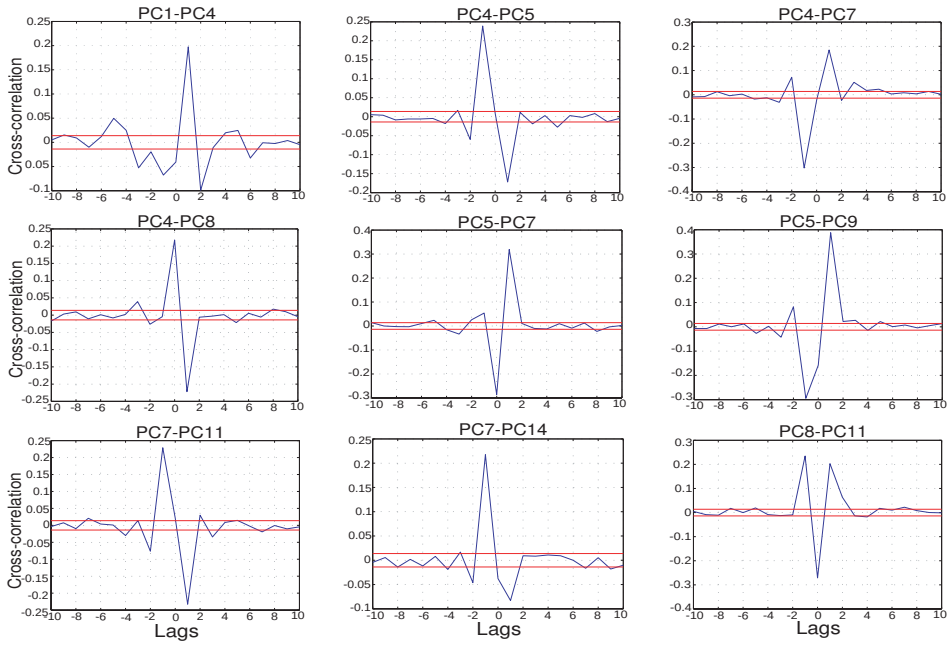

Figure 3.6: Cross-correlations of residuals obtained after the ARMA fits of different PC combinations .

cients. For lately comparative purposes, the simulations are going to be the same size of the original time series; thus the N_S taken was the same as the original time series length.

With the G matrix built, the Cholesky decomposition is constructed. Then the N_S independent standard normal errors are generated and cross-correlated. When they are introduced within the ARMA models, the standardized PCs are simulated. Finally, the variables are denormalized.

Figure 3.7 shows the comparison between the original and simulated empirical probability density functions for the 14 PCs. Visual inspection shows that the simulation accurately reproduces the marginal distribution of the DSLP PCs. Due to the high number of joint distributions within the 14 PCs, a graphical representation for comparison purposes is not feasible. Thus, to compare between the original and simulated joint distributions, the Kullback-Leibler divergence measurement [95] has been calculated for all the possible combinations. The Kullback-Leibler divergence measurement, also known as Relative Entropy, is given by the following

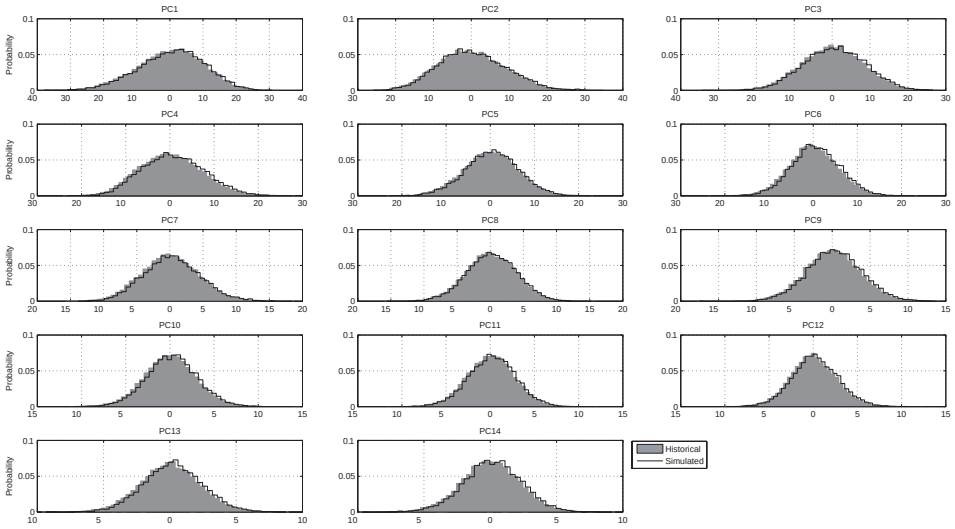

Figure 3.7: Empirical probability density function of each PC related to: i) historical data (grey bars) and ii) simulated data (black line).

expression:

$$D_{KL}(P \mid Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx \tag{3.1}$$

Which estimates the difference between two probability distributions P and Q. This measurement is always positive and only would be 0 if the two distributions compared were the same.

The estimation of D_{KL}, assuming that P is the empirical bidimensional distribution of each $PC_i - PC_j$ (with $i, j = 1, ...14$) and Q the bidimensional distribution of the simulated data for all the $PC\mathrm{sim}_i - PC\mathrm{sim}_j$ (with $i, j = 1, ...14$), results in the symmetric matrix shown in Table 3.1. In this table, each value (i, j) represent the D_{KL} of the comparison between the 2-D distribution of $PC_i - PC_j$ against the 2-D distribution of $PC\mathrm{sim}_i - PC\mathrm{sim}_j$. The values of the diagonal are the comparison between marginals, PC_i against $PC\mathrm{sim}_i$. Note that all the values from the table are close to zero, i.e. there is a good agreement between the empirical and the simulated 2-D distributions.

	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8	PC_9	PC_{10}	PC_{11}	PC_{12}	PC_{13}	PC_{14}
PC_1	0.0035	0.0221	0.0198	0.0179	0.0257	0.0151	0.0455	0.0149	0.0148	0.0257	0.0173	0.0123	0.0164	0.0226
PC_2		0.0009	0.0173	0.0135	0.0161	0.0142	0.0350	0.0114	0.0109	0.0169	0.0168	0.0127	0.0107	0.0116
PC_3			0.0026	0.0132	0.0172	0.0149	0.0390	0.0139	0.0155	0.0147	0.0166	0.0122	0.0120	0.0117
PC_4				0.0009	0.0119	0.0131	0.0345	0.0117	0.0130	0.0183	0.0105	0.0104	0.0105	0.0108
PC_5					0.0005	0.0103	0.0395	0.0096	0.0122	0.0138	0.0144	0.0091	0.0119	0.0108
PC_6						0.0007	0.0353	0.0116	0.0091	0.0161	0.0139	0.0130	0.0107	0.0089
PC_7							0.0257	0.0336	0.0371	0.0409	0.0354	0.0340	0.0377	0.0336
PC_8								0.0006	0.0082	0.0125	0.0132	0.0104	0.0114	0.0086
PC_9									0.0004	0.0141	0.0185	0.0089	0.0104	0.0095
PC_{10}										0.0047	0.0180	0.0171	0.0184	0.0159
PC_{11}											0.0030	0.0114	0.0146	0.0141
PC_{12}												0.0004	0.0095	0.0103
PC_{13}													0.0006	0.0111
PC_{14}														0.0005

Table 3.1: Kullback-Leibler divergence measurement results.

4.3 Step 2

The second step of the process starts with the clusterization of the trivariate Daily Mean Sea Conditions, (H_s, T_m, θ_m). The non-hierarchical K-means algorithm is used to classify the data into a previously determined number of groups represented by a centroid. In this case, the number of groups chosen is $n = 16$. To reduce the likelihood of reaching local minima with the algorithm, clusterization is repeated a hundred times, each with a new set of initial cluster centroid positions. The algorithm returns the solution with the lowest value for the objective function. Note that in this particular case we select 16 clusters for several reasons: i) for the sake of simplicity, ii) to facilitate the implementation, fit and interpretation of the model results and iii) to reduce computational effort. The subset obtained by KMA algorithm applied to the data is shown in Figure 3.8, where the centroid positions are represented by the larger dots.

Assigning arbitrarily an integer value between 1 and n=16 for each DMSC in Figure 3.8, we get the sequence of DMSC, which is the input for the autoregressive logistic model.

The autoregressive logistic model allows the simulation of synthetic sequences of DMSC taking into account different covariates such as seasonality, DSLP and autoregressive terms of both: DMSC and DSLP. It is important because it is known that sea conditions at any given point in time depend on the previous states and may be affected by atmospheric behaviour of present and previous days. Moreover, its nominative nature is useful when working with classified data. Further information related to the theoretical foundation and covariates implementation can be found in [67]. Here, the fitting process is described briefly. The criteria to choose the final model, i.e. the order of the auto-regressive components, seasonality, number of DSLP PCs, etc. is based on statistical significance, in particular, using the likelihood ratio (LR) statistic. This method compares nested models by comparing the deviance ratio ΔDev., which measures the change of fitting quality for two different parameterizations, and the chi-square distribution with $\Delta df = \Delta n_p \times (n-1)$ degrees of freedom, being Δn_p the difference in terms of number of parameters for both parameterizations. Here, a confidence level $\alpha = 0.95$ is assumed, so that if $\Delta\text{Dev.} > \chi^2_{0.95,\Delta df}$ the improvement achieved by adding n_p additional parameters is

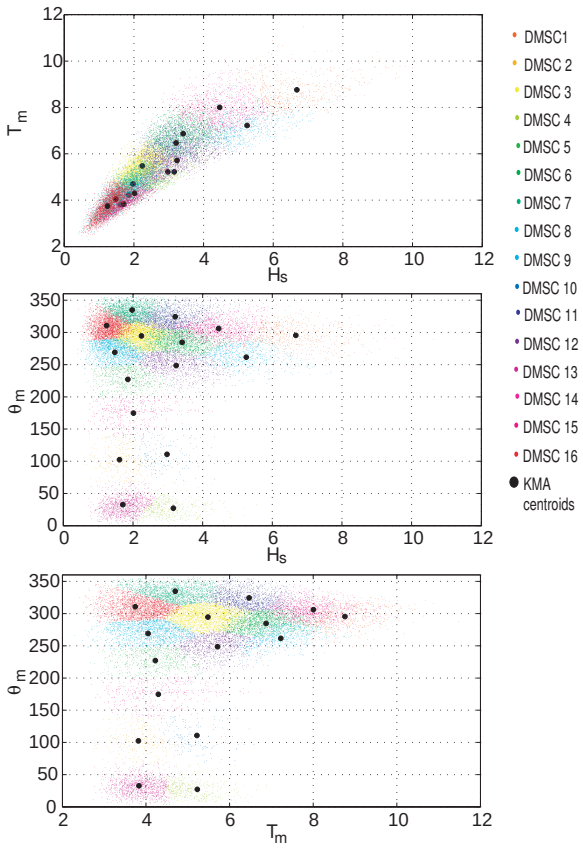

Figure 3.8: KMA classification of the Daily Mean Sea Conditions.

significant. The initial model to compare with is the null model, which has $n_p = 1$ because a constant term is automatically added by the model.

In Table 3.2 the fitting process is presented. In the Model column: X states for the model fitted just taking into account seasonality (by adding 2 harmonics), Z and Z_2 refers to the first and second autoregressive terms of the DMSC, each $Y_i = \sum_{j=1}^{i} PC_j$ refers to the addition of consecutive DSLP PCs to the model and $S_i = \sum_{j=1}^{i} Z_{PC_j}$, $S_{2,i} = \sum_{j=1}^{i} Z_{2,PC_j}$ and $S_{3,i} = \sum_{j=1}^{i} Z_{3,PC_j}$ state for the addition of the first, second and third autoregressive term of consecutive DSLP PCs, respectively. Firstly the addition of the two first autoregressive terms of DMSC is tested. The results show that only the first one is significant. Note that the inclusion of each

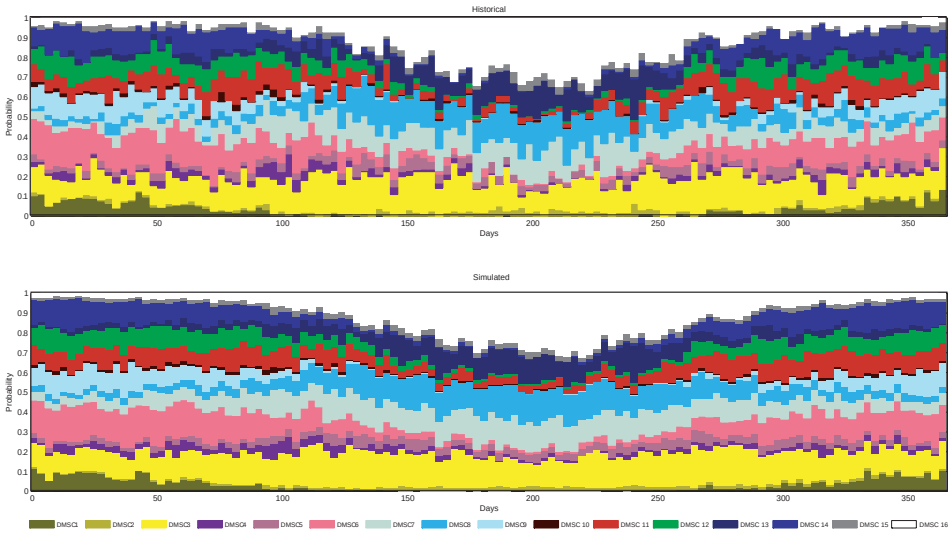

Figure 3.9: Model fitting diagnostic plot.

autoregressive term involves $n-1$ parameters, where n is the number of DMSC groups. Then, the DSLP PCs are included one by one to determine significant ones. As shown in Table 3.2, the 14 PCs of DSLP in the concurrent day, the first 3 PCs of the previous day and the first 5 PCs of two days before are significant. The influence of the third autoregressive term of the PCs is not significant. So, the resulting fitted model takes into account: seasonality, first autoregressive term of the DMSC, 14 PCs of the concurrent day, 3 first PCs of the day before (first autoregressive term of the DSLP PCs) and 5 first PCs of two days before (second autoregressive term of the DSLP PCs). Note that this fitting process has been made taking into account the historical DSLP decomposed into PCs, not the simulated PCs obtained in step 1.

Figure 3.9 shows a comparison between the empirical and the simulated probabilities of the $n = 16$ DMSC groups within a year. The simulation is made considering the significant covariates estimated before. The data are 3-daily aggregated for a better graphical representation.

Once the model is fitted, synthetic sequences of weather types can be generated through a Monte Carlo method, using the simulated PCs as covariates. To check the ability of the model to reproduce statistically similar sequences of DMSC the

Model	df	Δ df	Dev.	Δ Dev.	$\chi^2_{0.95,\Delta df}$
$Null$	284880	30	97409	6287	43.8
X	284850	225	91122	21976	260.9
$X+Z$	284625	225	69146	-6350	260.9
$X+Z+Z_2$	284400		75496		
$X+Z$	284625	15	69146	864	260.9
$X+Z+Y_1$	284610	15	68282	4350	25
$X+Z+Y_2$	284595	15	63932	1817	25
$X+Z+Y_3$	284580	15	62115	263	25
$X+Z+Y_4$	284565	15	61852	774	25
$X+Z+Y_5$	284550	15	61078	1552	25
$X+Z+Y_6$	284535	15	59526	342	25
$X+Z+Y_7$	284520	15	59184	163	25
$X+Z+Y_8$	284505	15	59021	197	25
$X+Z+Y_9$	284490	15	58824	739	25
$X+Z+Y_{10}$	284475	15	58085	65	25
$X+Z+Y_{11}$	284460	15	58020	300	25
$X+Z+Y_{12}$	284445	15	57720	526	25
$X+Z+Y_{13}$	284430	15	57194	811	25
$X+Z+Y_{14}$	284415	15	56383	301	25
$X+Z+Y_{14}+S_1$	284400	15	56082	700	25
$X+Z+Y_{14}+S_2$	284385	15	55382	327	25
$X+Z+Y_{14}+S_3$	284370	15	55145	-100	25
$X+Z+Y_{14}+S_4$	284355		55245		
$X+Z+Y_{14}+S_3$	284370	15	55145	51	25
$X+Z+Y_{14}+S_3+S_{2,1}$	284355	15	55094	90	25
$X+Z+Y_{14}+S_3+S_{2,2}$	284340	15	55004	61	25
$X+Z+Y_{14}+S_3+S_{2,3}$	284325	15	54943	349	25
$X+Z+Y_{14}+S_3+S_{2,4}$	284310	15	54594	250	25
$X+Z+Y_{14}+S_3+S_{2,5}$	284295	15	54344	-47	25
$X+Z+Y_{14}+S_3+S_{2,6}$	284280		54402		
$\boldsymbol{X+Z+Y_{14}+S_3+S_{2,5}}$	284295	15	54344	-40	25
$X+Z+Y_{14}+S_3+S_{2,5}+S_{3,1}$	284280		54384		

Table 3.2: Fitting process results.

simulation has been repeated $n_{sims} = 100$ times. The results obtained have been validated with a threefold comparison against the original sequence of DMSC: i) occurrence probabilities of DMSC, ii) transition probability matrix between DMSC and iii) persistence analysis of DMSC.

Firstly, the probabilities of occurrence of the n groups for the n_{sims} simulations, against the empirical probability of occurrence from the 52-year sample data, are shown in Figure 3.10. The results are close to the diagonal, demonstrating the model capability to reproduce the probability of occurrence appropriately.

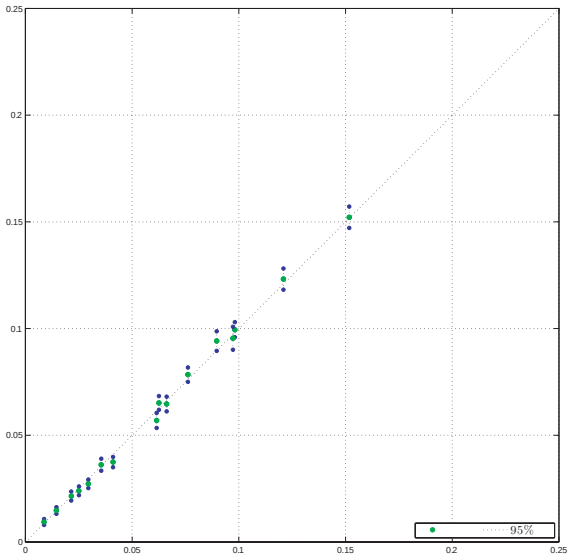

Figure 3.10: Scatter plot of the empirical occurrence probabilities associated with the DMSC versus 100 Monte Carlo simulation results.

The transition probabilities express the probability of changing from group i to group j between consecutive days. Thus, in the case of having 16 wave conditions groups, the transition matrix (T) has dimensions 16×16, and each cell $T_{i,j}$ is the probability of changing from DMSC group i to DMSC group j. The diagonal of the transition matrix T corresponds to the probability of staying in the same group. The transition matrix is calculated for each of the 100 simulated samples and Figure 3.11 shows the scatter plot related to the $16 \times 16 = 256$ elements of transition matrix, including its uncertainty due to the simulation procedure, against the em-

pirical transition probabilities obtained from the historical data set. As shown, the model is able to reproduce correctly the transitions between DMSC.

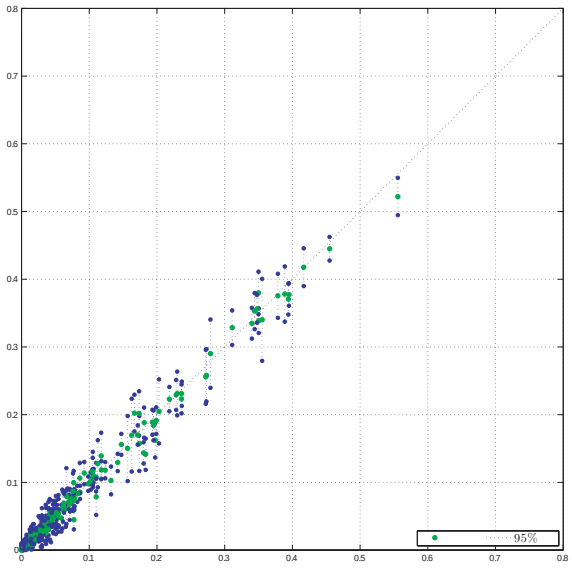

Figure 3.11: Scatter plot of the empirical transition probabilities associated with the DMSC versus 100 Monte Carlo simulation results.

Thirdly, a persistence analysis is performed over the simulated samples in order to check the ability of the model to reproduce wave climate dynamics. Figure 3.12 shows the empirical cumulative distributions of the persistence associated with each DMSC group. The green line represents the average empirical distribution while the one related to the historical sample data is coloured in blue for all cases. The red dotted line represents the 95% confidence intervals related to the 100 simulations. To further analyze the performance of persistence from an statistical view-point, a two-sample Kolmogorov-Smirnov ([102]) goodness-of-fit hypothesis test between the original data and each sampled data is performed. This test determines if two different samples come from the same distribution without specifying what that common distribution is. In Figure 3.13 the box plots associated with the p-values from the 100 tests for each DMSC group are shown. Note that if the p-value is higher than the significance level (5%) the null hypothesis that both samples come from the same distribution is accepted. Results shown in Figure 3.12

prove that for most of the cases the persistence distributions from the Monte Carlo simulation procedure come from the same distribution as the persistence distribution from the historical data. For all the DMSC groups the interquartile range (blue box) is above the 5% significance level (red dotted line).

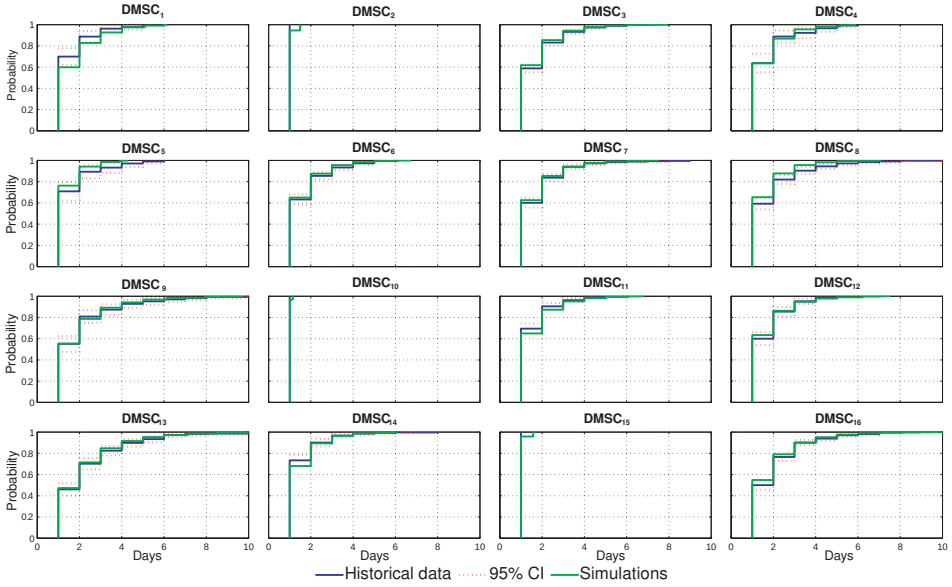

Figure 3.12: Empirical cumulative distribution of the persistence for the 16 groups related to: i) historical data and ii) sampled data using Monte Carlo method.

4.4 Step 3

In the two previous steps the simulations are performed with a daily temporal resolution and the main objective is to reproduce the daily mean wave characteristics taking into account the influence of the atmospheric conditions. This third step of the process involves a change in the temporal scale, transferring the daily information obtained in the previous steps into hourly sea states simulation.

To accomplish this third step and thus the entire process, a modified version of the multivariate simulation technique proposed by [120] is used. Initially, the trivariate hourly data are split into the daily mean sea conditions obtaining $n = 16$ groups for each of the $N = 3$ variables: H_s, T_m and θ_m.

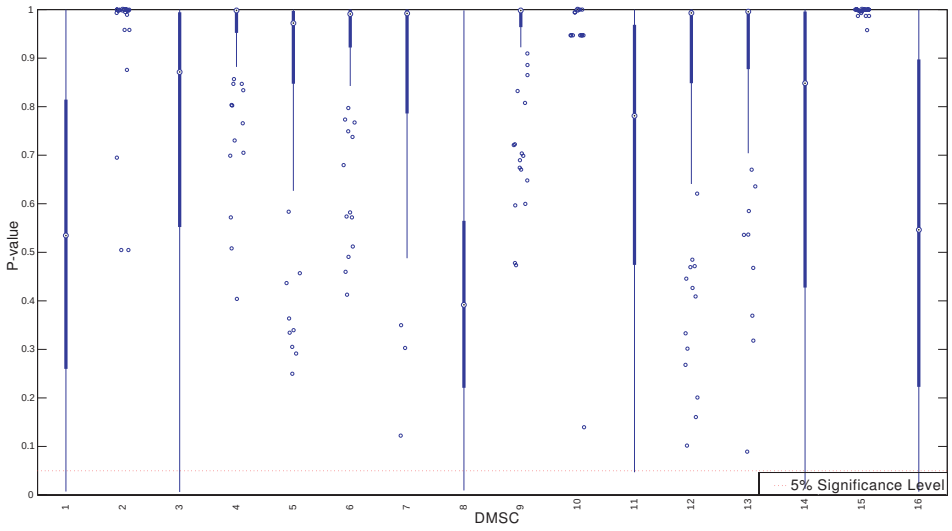

Figure 3.13: Box plot associated with the p-values from the 100 tests for each DMSC.

For all the N variables a normalization is made taking into account the empirical distribution of the n groups and the historical sequence of DMSC. Thus, Z_i (with $i = 1, ...3$) transformed variables are obtained. These transformed variables are fitted into ARMA(p, q) models, in this case ARMA$(2, 1)$ models for the three variables provides appropriated results. The cross-correlation that exists between the residual errors of the models is shown in Figure 3.14. As seen, there is a strong concurrent cross-correlation between H_s and T, while the correlations with the wave direction present lower values.

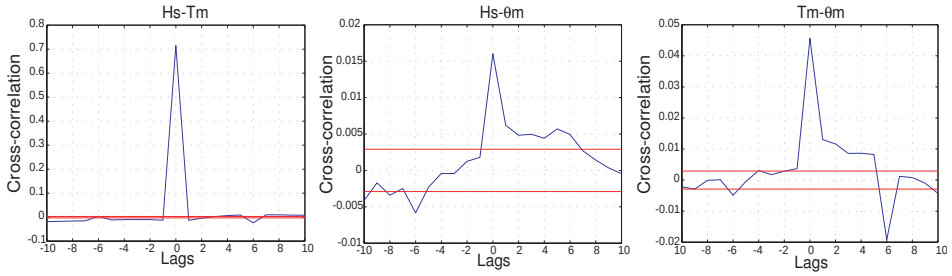

Figure 3.14: Cross-correlations of residuals obtained after the ARMA fits of the three variables.

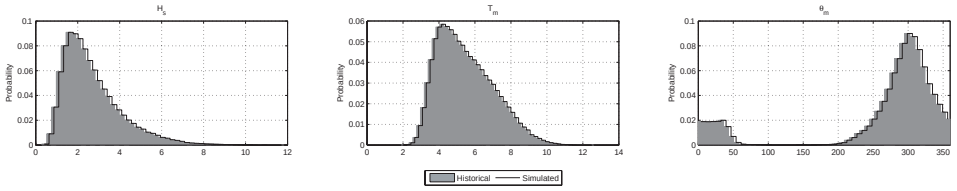

Figure 3.15: Empirical probability density function of H_s, T_m and θ_m related to: i) historical data (grey bars) and ii) simulated data (black line).

From these results, up to 10 lags ($K = 10$) are considered to be appropriate to build the variance-covariance matrix G. Once the matrix G is built, the simulation process is similar at the one explained in step 1, but the main difference comes in the denormalization of the simulated variables. This denormalization is made taking into account the empirical distribution of the $n = 16$ groups in which each variable is split but considering the simulated sequences of DMSC. This way, synthetic hourly trivariate sea states are obtained.

Figures 3.15 and 3.16 show the ability of the process to provide good results not only in the joint distribution but also in the marginals. To quantify the similarity between the historical data and the simulated ones, the Kullback-Leibler measurement is estimated for all the combinations. Results of this measurement are shown in Table 3.3, each element of diagonal represents the D_{KL} measurement for comparison between marginal distributions of the three variables while the non-diagonal elements are the estimations for the 2D joint distributions within variables and its comparison between empirical and simulated data. Note that all the estimated values are close to 0, meaning that both data (historical and simulated) present similar marginal and joint distributions.

	H_s	T_m	θ_m
H_s	0.0003	0.0345	0.0068
T_m		0.0003	0.0100
θ_m			0.0001

Table 3.3: Kullback-Leibler divergence measurement results.

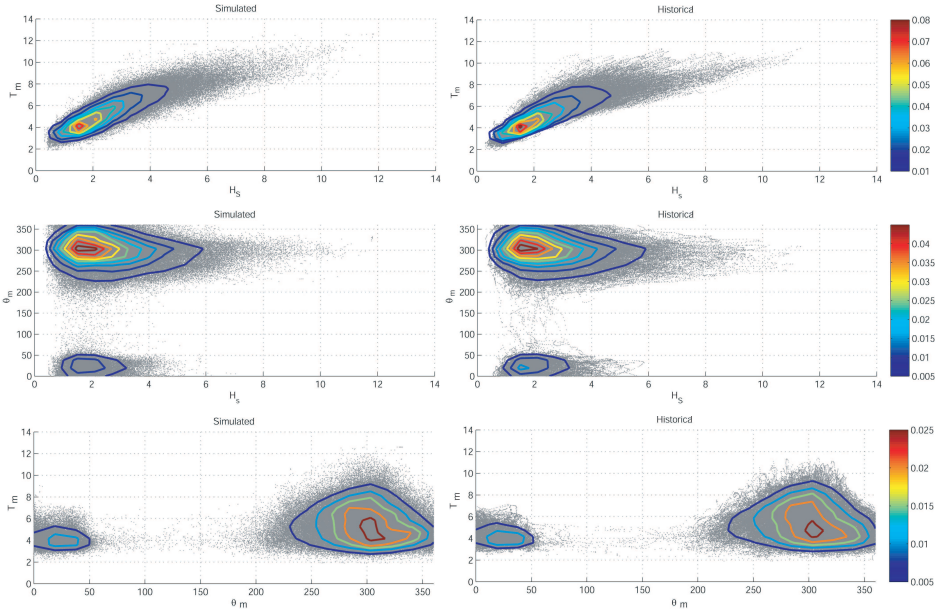

Figure 3.16: Comparison between historical and simulated joint distributions. Contour lines represent the empirical joint density distribution while dots are hourly data.

In terms of simulating correctly the seasonality present in the parameters distribution, Figure 3.17 shows a comparison between the original and simulated data. In the upper subplots the three variables aggregated in a year while in the lower ones the simulated data are presented.

The methodology allows the generation of synthetic time series with statistically similar behaviour. Because of the use of the empirical distributions when denormalizating the simulated time series, the maximal value of H_s and T are constrained to be the historical maxima of the original data. This could be avoided by using a fitted distribution (i.e. Pareto) to describe the extreme values, but this involves the definition of a threshold and the choice of the best fit. This is out of the scope of the present work and it is a subject for further research.

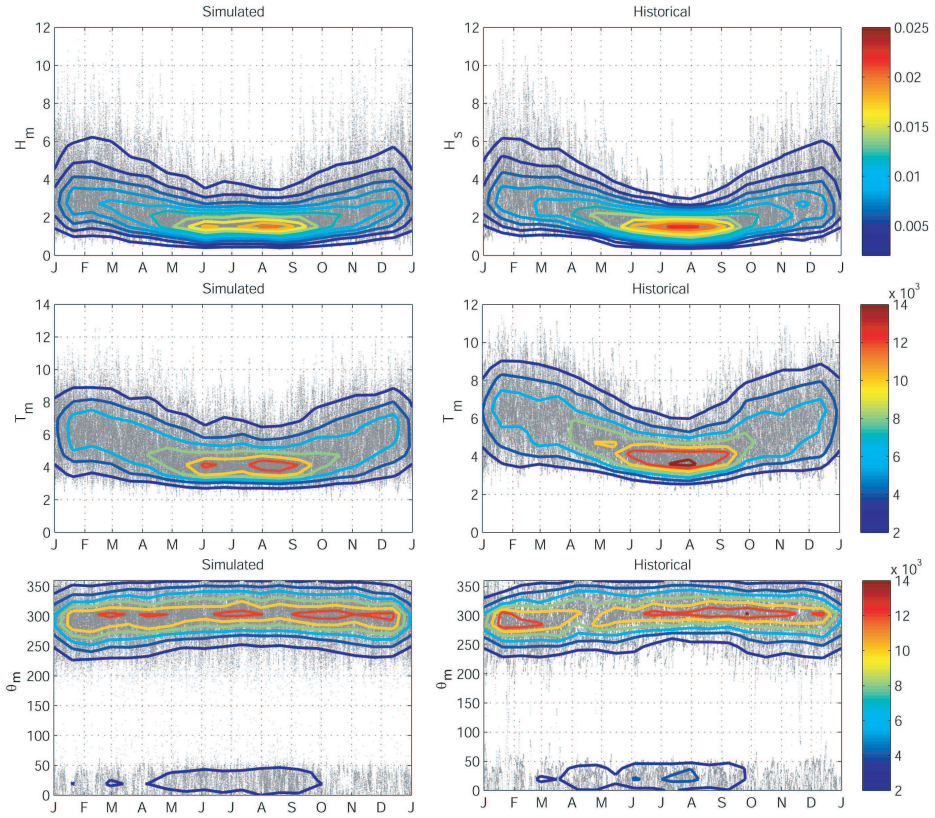

Figure 3.17: Comparison of the seasonality of historical and simulated. Contour lines represent the empirical joint density distribution while dots are hourly data.

5 Conclusions

This work presents a methodology to reproduce hourly trivariate sea state time series. The method combines the use of univariate ARMA models cross-correlated with an autoregressive logistic regression model. This combination of techniques allows the simulation of wave climate time series taking into account the different temporal and spatial scales involved. The advances with respect to the state-of-the-art can be summarized as follows:

- The possibility to simulate daily Sea Level Pressure fields decomposed into PCs allows the generation of different atmospheric scenarios.
- The autoregressive logistic model takes into account simultaneously cova-

riates of different nature, such as DSLP, seasonality or autoregressive influence, where the time and space scales are completely different.

- The methodology is able to reproduce multivariate time series of interrelated variables.
- The developed model is DSLP-driven, facilitating the understanding of local wave climate as a function of given synoptic circulation patterns.

Further research is required on the correct characterization of extreme events. Although the use of the empirical distributions gives accurate results, a better definition of extreme values would provide the possibility of simulating different severe events from those actually recorded. This definition could be done by fitting a Pareto distribution over a predetermined threshold.

CHAPTER

4

A simplified method to downscale wave dynamics on vertical breakwaters

1 Abstract

A coastal structure is usually designed with the final objective to guarantee its functionality and stability throughout its life cycle. Regarding stability, the three main failure modes are sliding, overturning and failure of the foundations. To accomplish the design objectives, a design sea state is usually used when calculating the loads and scour around the structure. This design sea state corresponds to a certain sea state with specific return period values of a significant wave height. However, the combination of different simultaneous sea state parameters can produce other critical situations compromising the stability of the structure which then require the calculation of long time series of wave forces corresponding to long-term historical wave situations. Moreover, a design force associated to a certain return period can be defined from the time series of the stability parameters. The most accurate techniques which can be used to estimate structure stability are based on numerical and physical models, but these are very time consuming and the calculation of

long time series is therefore unfeasible. Here, we propose a hybrid methodology to transform wave conditions into wave forces acting upon vertical structures and scour around it. The methodology consists of a selection of a subset of sea states representative of wave climate at the structure location, using a maximum dissimilarity algorithm. The wave forces acting upon the structure and scour around it, for the wave situations selected, are then estimated. Finally, a reconstruction of the calculated parameters corresponding to historical sea states using an interpolation technique is done based on radial basis function. The validation of the results, through a direct comparison between reconstructed series and analytically (semi-empirical formulations) calculated ones, confirms the ability of the developed methodology to reconstruct time series of stability parameters on vertical breakwaters. This methodology allows its application to numerical and physical models.

2 Introduction

The design of a vertical breakwater requires information on several parameters that influence its behavior during its life cycle. The three main major failures can be synthesized in the following: sliding, overturning and failure of foundation. In order to achieve an accurate design, in terms of structural stability, it is crucial to have the data to estimate loads (sliding and overturning) and erosion (foundation failure).

The determination of wave loads has been a challenge since the beginnings of coastal structure engineering. The early stages of the definition of the pressure law over submerged walls took place in field tests conducted on the breakwaters of the Great Lakes, by Gaillard at the beginning of the 20th century. This study laid the foundations for many subsequent studies.

During the past 100 years, several authors have studied and proposed different methods to determine wave loads on a vertical wall ([78], [145], [113], [119], [143], [86],[61],[62], [157]). The Goda-Takahashi method (Goda, 1974 with the modifications of Takahashi and Tanimoto, 1994) is commonly used, as a semi-empirical formula, as it allows estimating both the wave load acting upon a vertical breakwater and the buoyancy and uplift pressures, allowing the design of the caisson.

Apart from semi-empirical methods, there are two other ways of estimating the forces acting upon a caisson: numerical models and physical models. The former have been developed over the last decades, due to the improvement in computational resources, allowing us to determine the forces acting upon a caisson ([65]). As for physical models, several studies have been conducted to determine wave loads based on laboratory tests ([70], [31], [36], [37]).

In terms of erosion around breakwaters, many studies have tried to investigate the influence of the complicated flow originated by the interaction between the structure and the seabed. Every structure in the sea will change the flow patterns in its neighborhood which can then result in scour at the breakwater toe and a possible total failure of the breakwater.

Throughout the last decades, many authors have studied and proposed several estimations of the scour depending on the type of structure and the sea conditions it is exposed to ([166], [81], [152], [154], [155], [156], [97], [122]). In the majority of these studies, semi empirical formulations are established taking into account the results obtained with experimental tests or with numerical models.

In the design of vertical breakwaters, the critical stability parameters (wave loads and scour) are usually calculated for the sea state with a significant wave height associated with a specific return period (design sea state) and the other corresponding sea state parameters which influence the structure stability obtained from the joint probability function between the significant wave height and the rest of the parameters. However, we consider that calculating the wave loads and the maximal scour produced on the breakwater as a result of all the possible sea states occurring at this location provides more realistic results. Consequently, each sea state is defined by means of the simultaneous significant wave height and the other sea state parameters considered (e.g., wave peak period, mean sea level, mean wave direction), which can produce worse combination of sea state parameters for the stability of the structure. A design force with a certain return period can be defined by means of stability parameter time series and can be used to optimize the breakwater design. Consequently, long-term wave data are required at the breakwater location to carry out a proper statistical definition. Instrumental data are rarely available for the required position, making reanalysis databases with high spatial and hourly resolutions a good choice. This can be used to define wave climate at

the breakwater location. The transformation of wave reanalysis sea state time series (usually consisting of over 50 years of hourly data) to the parameters involved in the breakwater stability (loads and scour) is unfeasible due to the high amount of time required if numerical or physical models are used to calculate wave loads. Therefore, the objective of this work is to propose a methodology to transform wave time series at the breakwater location into stability parameter time series, in order to be able to use numerical or physical models with reasonable computational time demands.

In meteorology, the concept downscaling refers to several methods proposed in the literature to gain subgrid detail of the outputs of the numerical atmospheric global circulation models (GCMs) or to simulate a local variable not generated in the GCM ([164]). On the one hand, the dynamic downscaling methods use the ACM-integrated gridded fields as boundary conditions for a new higher resolution limited-area model, including parameterizations adapted for the region of interests. On the other hand, the statistical downscaling methods define a statistical model which relates the gridded atmospheric patterns to historic local observations. Based on the same idea, different downscaling methods have been developed to transfer wave climate from deep water to coastal areas at higher resolutions while still simulating shallow water transformations. Another group of methodologies which combines numerical models (dynamic downscaling) and mathematical tools (statistical downscaling) to reduce the computational effort ([76], [20]) are also proposed. Therefore, to achieve the previously stated objective, a hybrid methodology is adopted, based on the selection of a number of sea states, the calculation of the stability parameters mentioned previously (loads and scour) and their reconstruction using a statistical technique.

This methodology is mainly based on the maximum dissimilarity algorithm (MDA) and the radial basis function (RBF) as the hybrid approach to downscale wave climate to coastal areas as proposed in [19], [20]. In the present application, the term downscale means the transformation of the sea state time series at the structure location to wave load parameters on the breakwater. The hybrid methodology proposed selects a subset of representative sea states of historical wave data at the location of coastal structures using MDA, the calculation of the stability parameters with a numerical or physical model corresponding to selected wave

conditions (dynamical downscaling), and an interpolation technique based on radial basis function (RBF) to reconstruct long time series of parameters involved in the stability of coastal structures.

A rigorous validation of the methodology would imply calculating wave forces using numerical or physical models for each hourly sea state condition over a period of several years, which supposes an unfeasible computational time, in order to compare these time series with the reconstructed ones obtained applying the proposed methodology (based only on several numerical or physical simulations). Therefore, a semi-empirical method is considered to explain the methodology. In the present application, the Goda-Takahashi method was applied to estimate the wave forces corresponding to each sea state of the wave reanalysis database while Hughes and Fowler (1991), [81], was used to estimate the scour around the breakwater, this method being more efficient both economically and computationally. The obtained time series were compared to those reached when using the proposed methodology. Each of the steps of the proposed methodology is explained in detail in the following sections.

3 Proposed methodology

The methodology proposed to downscale stability parameters calculating these parameters for a representative subset of wave conditions extracting them from a historical database and performing a statistical reconstruction of the time series created for these parameters. The steps involved are: (a) selection of sea states; (b) calculation of wave loads acting upon the vertical breakwater and scour around it; and (c) reconstruction of the complete series of sea states using an interpolation scheme. To validate this methodology, two semi-empirical methods (Goda-Takahashi method for wave loads and Hughes and Fowles formulation for scour) were applied to calculate wave forces and scour allowing us to obtain the analytical time series on the breakwater which corresponds to each sea state of the historical database. The reconstructed time series of each load parameter obtained with the proposed methodology could then be directly compared with the analytically calculated time series. The different steps are explained in the following sections.

A double application is considered to explain the proposed methodology. The breakwater under study is located on the north coast of Spain, at Gijon's harbor. The hindcast wave data used in this work is: DOW 1.1 (Downscaled Ocean Waves, [19]) obtained by IH Cantabria, an hourly regional wave reanalysis for the period 1948-2008 with coverage of $\sim$200 m along the Spanish coast. A hybrid method is used to downscale the GOW 1.1 reanalysis ([136]) to coastal areas, based on a propagation catalog using the SWAN model and statistical techniques: MDA in the selection process and RBF in the time series reconstruction. The forcing inputs come from the SeaWind NCEP/ NCAR dataset ([111]).

The sea state (considered to be 1 h) parameters required by the GodaTakahashi formula to define the wave loads on the breakwater are: significant wave height (m) H_s, peak period (s) T_p, mean wave direction $(^\circ)$ θ_m and hourly mean sea level (m) Z_m. On the other hand, the variables involved in the Hughes and Fowler formulation of scour are: significant wave height (m) H_s, peak period (s) T_p, and hourly mean sea level (m) Z_m. The methodology hereafter was established as a function of the four parameters used by the GodaTakahashi method; similarly, this could also be established for the Hughes and Fowler formulation if the wave direction was discarded.

4 Selection of sea states

The aim of the selection process was to extract a subset of wave situations which were representative of those conditions available at the breakwater location through the DOW 1.1 reanalysis database. There are several selection algorithms to extract a subset of data from a database (i.e., KMA (K-Means Algorithm), SOM (Self Organizing Maps), MDA (Maximum Dissimilarity Algorithm)). Out of these options, the MDA algorithm was chosen due to its ability to distribute the selected data fairly evenly throughout the space, with some points selected along the borders of the data space, therefore guaranteeing the most representative subset in comparison with the original sample ([20]). Consequently, the MDA algorithm allows an automatic selection of a subset of sea states which are representative of wave climate at the breakwater location. This algorithm was then applied to reconstruct wave load and scour time series on a marine structure. Multivariate data were defined as:

$X_i = H_{s,i}, T_{p,i}, \theta_{m,i}, Z_{m,i}$; $i = 1, , N$, where N amounted to 534,000 hourly sea states, corresponding to the period between 1948 and 2009. This reanalysis data is defined by scalar and directional parameters of different magnitudes which require normalization and an implementation of the distance in the circle for the directional parameter on the MDA algorithm.

The scalar variables are normalized by scaling the variable values between 0 and 1 with a simple linear transformation which requires two parameters, the minimum and maximum values of the two scalar variables. In the case of circular variables (defined in radians or in sexagesimal degrees using the scaling factor $\pi/180$), and considering that the maximum difference between two directions over the circle is equal to π and the minimum difference is equal to 0, normalization was achieved by dividing the direction values between p. Similarly, the circular distance was rescaled between 0 and 1. The dimensionless input data were expressed as $X_i = H_i, T_i, \theta_i, Z_i$; $i = 1, , N$, after these transformations.

Therefore, given a sample data $X_i = H_i, T_i, \theta_i, Z_i$; $i = 1, ..., N$ consisting of N n-dimensional vectors, a subset of M vectors $D_1, ..., D_M$ is selected by the MDA algorithm. The initial data of the subset D_1 is considered to be the sea state with the highest significant wave height. The rest of the $M - 1$ elements are selected iteratively, transferring the most dissimilar one to the subset established by the MaxMin version of the algorithm ([165]). For example, if the subset is formed by $R(R = M)$ vectors, the dissimilarity between the vector i of the data sample N-R and the j vectors belonging to the R subset is calculated as:

$$d_{ij} = \|X_i - D_j\|; i = 1, ..., N - R; j = 1, ..., R \tag{4.1}$$

Subsequently, the dissimilarity $d_{i,subset}$ between vector i and subset R, is calculated as:

$$d_{i,subset} = min\{\|X_i - D_j\|\}; i = 1, ..., N - R; j = 1, ..., R \tag{4.2}$$

Once the $N - R$ dissimilarities have been calculated, the next selected data is that with the maximum $d_{i,subset}$. We used the algorithm developed by Polinsky et al. (1996).

A Euclidean-Circular distance (E for the Euclidean distance in scalar parameters and C for the circular distance in directional parameters) was implemented in the MDA algorithm:

$$\|X_i - D_j\| = \sqrt{(H_i - {H_j}^D)^2 + (T_i - {T_j}^D)^2 + (Z_i - {Z_j}^D)^2 + (min\{|\theta_i - {\theta_j}^D|, 2 - |\theta_i - {\theta_j}^D|\})} \tag{4.3}$$

Finally, we de-normalized the subset applying the opposite transformation used during normalization. The MDA subset was then defined as ${D_j}^* = \{{H_{s,j}}^D, {T_{s,j}}^D, {Z_{s,j}}^D, {\theta_{s,j}}^D\}; j = 1, ...M$

The MDA was applied to the 60 year time series of the four parameters considered in the definition of wave conditions at the breakwater location. We used different subset sizes (from 10 to 1000 sea states) to analyze the influence of representative case numbers in the calculation of the parameters. Figure 4.1 shows a zoom over the time series of the four parameters $\{H_{s,i}, T_{p,i}, Z_{m,i}, \theta_{m,i}\}$ and MDA subsets of different size (M being the number of selected cases). The first R selected data using MDA of different subset sizes are the same, meaning that for a selection of M=200 cases and another of M=100 cases, the first 100 cases of the first selection are exactly the same cases as those belonging to the second selection.

Figure 4.2 shows the distribution of the sample data for different 2-dimensional combinations of the 4 parameters, both for the 60 year data series and for the M=1000 selected data. As seen in both figures (Figs. 4.1 and 4.2), the selected cases spread all over the data range.

5 Stability parameter calculation

5.1 Dynamic loads

Wave forces were calculated for each set of selected (Section 5) wave conditions at the structure location according to the previously defined sea state parameters. The dynamic loads (F) can be expressed as a function of wave parameters:

$$F_i = f_F(H_{s,i}, T_{p,i}, \theta_{m,i}, Z_i) \tag{4.4}$$

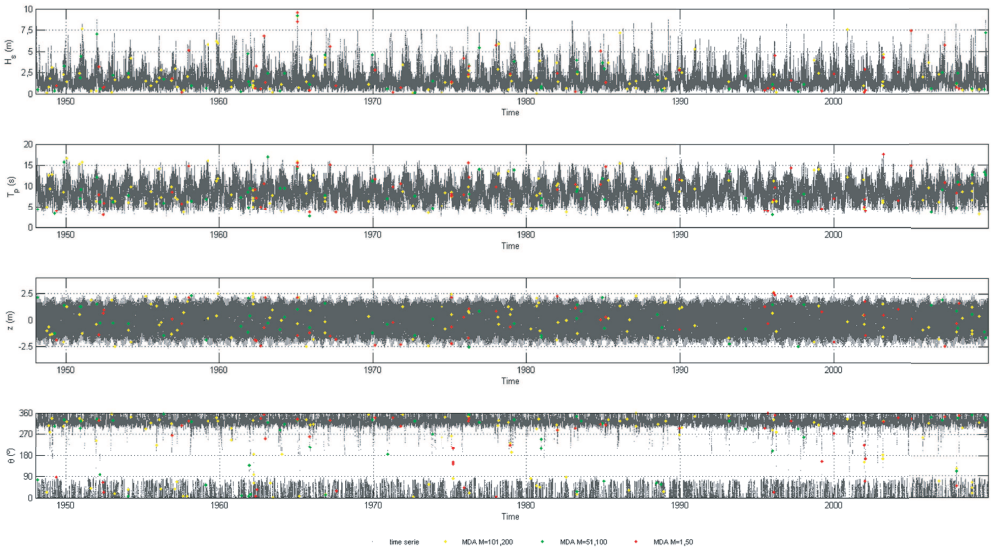

Figure 4.1: Time series of H_s, T_p, θ_m and Z_m in front of the breakwater (grey dots), the selected cases by MDA algorithm, $M = 1, 50$ red points, $M = 51, 100$ green points, $M = 101, 200$ yellow points.

The transformation of a sea state to wave forces should be estimated by a numerical or physical model. However, a semi-empirical formula (Goda-Takahashi) was considered because it allowed the direct transformation of the 60 year hourly reanalysis time series, making the validation of the proposed methodology possible.

The Goda-Takahashi method is explained in section 5.1.1 of this chapter. Pressures exerted upon the breakwater, as estimated by this empirical method, are a function of the parameters H_s, T_p, θ_m, Z and the breakwater geometry. Once the pressure distribution and the uplift pressure distribution have been calculated, they can be integrated. Then the horizontal F_h, and vertical, F_v forces, as well as their moment around the bottom of an upright section: the horizontal M_h, and vertical force M_v momentums, are obtained. Finally, the safety factor against sliding, SSC and overturning, OSC are defined as follows:

$$SSC = \frac{\mu(W - F_v)}{F_h} \tag{4.5}$$

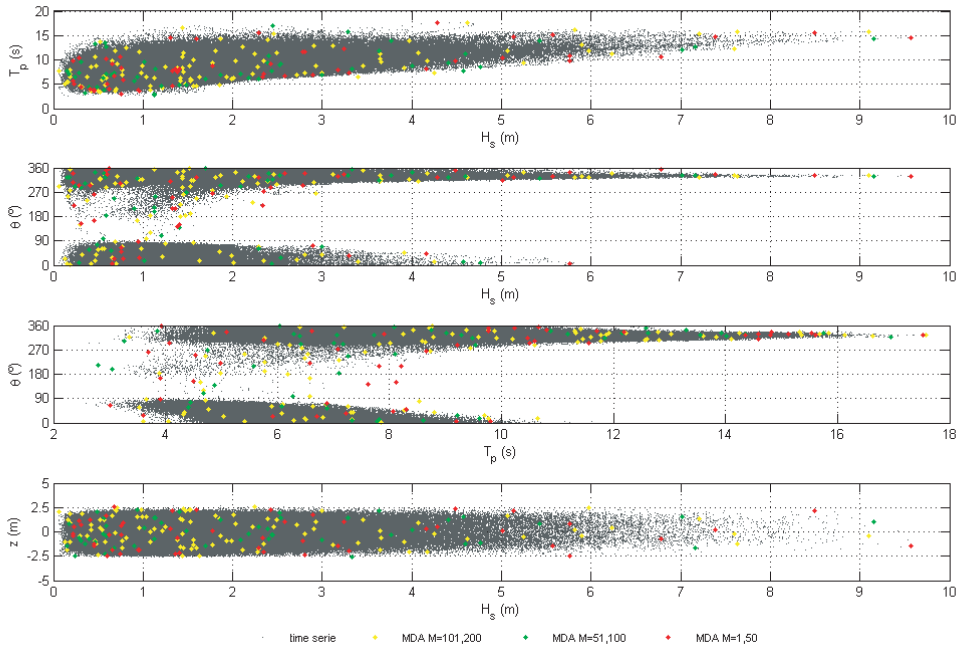

Figure 4.2: Distribution of the selected data ($M = 1, 50$ red points, $M = 51, 100$ green points, $M = 101, 200$ yellow points) obtained by MDA algorithm in the sample time series (grey points).

$$OSC = \frac{M_w - M_v}{Mh} \tag{4.6}$$

where W stands for the breakwater weight, M_w for its momentum and μ is a friction coefficient (0.65 in this application). The critical stability of the structure can be analyzed by means of the sliding and overturning factor time series.

This procedure has been applied to each sea state selected by the MDA algorithm and to the complete wave reanalysis time series at the vertical breakwater location, at the Gijon's harbor (Fig. 4.3).

5.1.1 **Goda-Takahashi method**

Since 1974, Goda's formulas have been widely used to predict wave loads acting on vertical breakwaters for irregular waves, both for breaking and non breaking

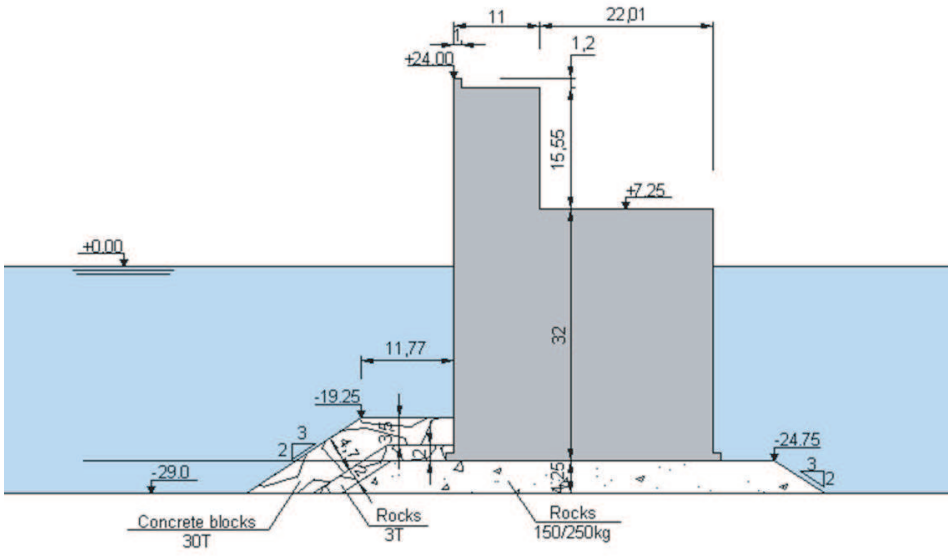

Figure 4.3: Vertical breakwater cross section at the harbor of Gijon extension.

waves. In 1994, Takahashi implemented a coefficient taking into consideration an impulsive breaking wave on a vertical wall.

This method assumes the existence of a trapezoidal pressure distribution along the vertical wall, as shown in Figure 4.4. In this figure, h represents the sea level at the breakwater, i.e.: the depth considering the mean sea level and the tidal elevation Z, d is the depth above the armor layer of the rubble foundation, h' the distance from the design water level to the bottom of the upright section, and hc the crest elevation of the breakwater above the design water level. The highest wave in the design sea state should be employed.

The elevation to which the wave pressure is exerted is shown as η^*:

$$\eta^* = 0.75(1 + \cos(\theta_m))(1.8H_s) \tag{4.7}$$

where θ_m denotes the angle between the direction of wave approach and a line normal to the breakwater. Goda proposed rotating this wave direction by an amount of up to $15°$ toward this line. The wave pressure on the front of a vertical wall can be calculated by:

$$p_1 = 0.5(1 + \cos(\theta_m))(\alpha_1 + \alpha^* \cos^2(\theta_m))\rho g(1.8H_s) \tag{4.8}$$

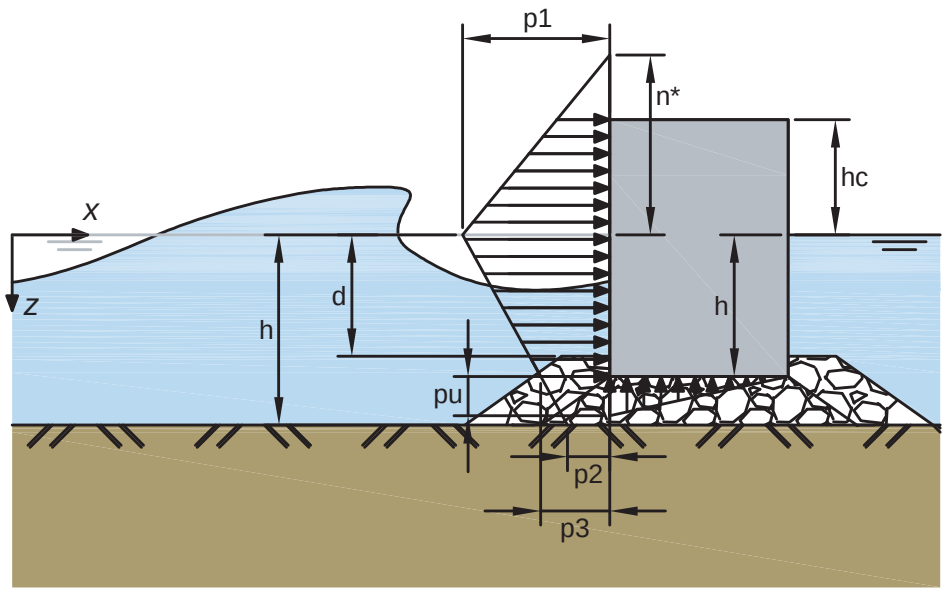

Figure 4.4: Distribution of wave pressure on an upright section of a vertical breakwater. Goda-Takahashi (1994) method.

$$p_2 = \begin{cases} \left(1 - \frac{h_c}{\eta^*}\right) p_1 & \text{if} \quad \eta^* > h_c \\ 0 & \text{if} \quad \eta^* < h_c \end{cases} \tag{4.9}$$

$$p_3 = \alpha_3 p_1 \tag{4.10}$$

where the coefficients α_1, α_2 and α_3 are a function of H_s, T_p, θ_m, Z and the geometry of the breakwater (h_s, h, d):

$$L = \frac{gT_p^2}{2\pi} \tanh\left(\frac{2\pi h}{L}\right) \tag{4.11}$$

$$\alpha_1 = 0.6 + 0.5 \left[\frac{4\pi h_s/L}{\sinh(4\pi h_s/L)}\right] \tag{4.12}$$

$$\alpha_* = \max\{\alpha_1, \alpha_2\} \tag{4.13}$$

$$\alpha_3 = 1 - \frac{h}{h_s}\left[1 - \frac{1}{\cosh(2\pi h_s/L)}\right] \tag{4.14}$$

where:

$$\alpha_2 = \min\left\{\frac{h_b - d}{3h_b}\left(\frac{1.8H_s}{d}\right)^2, \frac{2d}{1.8H_s}\right\} \tag{4.15}$$

$$\alpha_I = \alpha_{I0}\alpha_{I1} \tag{4.16}$$

$$\alpha_{I0} = \begin{cases} 1.8H_s/d & \text{if} \quad 1.8H_s/2 \leq 2 \\ 2 & \text{if} \quad 1.8H_s/d > 2 \end{cases} \tag{4.17}$$

$$\alpha_{I1} = \begin{cases} \frac{\cosh\delta_2}{\cosh\delta_1} & \text{if} \quad \delta_2 \leq 0 \\ \frac{\cosh\delta_2}{\cosh\delta_1(\cosh\delta_2)^{0.5}} & \text{if} \quad \delta_2 > 0 \end{cases} \tag{4.18}$$

$$\delta_1 = \begin{cases} 20\delta_{11} & \text{if} \quad \delta_{11} \leq 0 \\ 15\delta_{11} & \text{if} \quad \delta_{11} > 0 \end{cases} \tag{4.19}$$

$$\delta_2 = \begin{cases} 4.9\delta_{22} & \text{if} \quad \delta_{22} \leq 0 \\ 3\delta_{22} & \text{if} \quad \delta_2 > 0 \end{cases} \tag{4.20}$$

$$\delta_{11} = 0.93\left(\frac{B}{L} - 0.12\right) + 0.36\left(\frac{h_s - d}{h_s} - 0.6\right) \tag{4.21}$$

$$\delta_{22} = -0.36\left(\frac{B}{L} - 0.12\right) + 0.93\left(\frac{h_s - d}{h_s} - 0.6\right) \tag{4.22}$$

where B is the toe berm width.

The uplift pressure acting on the bottom of the upright section is assumed to have a triangular distribution where toe pressure pu is given by the equation below:

$$p_u = 0.5(1 + \cos(\theta_m))\alpha_1\alpha_3\rho g(1.8H_s) \tag{4.23}$$

Once the pressure distribution and the uplift pressure distribution are calculated, they can be integrated obtaining the horizontal force, F_h, and the vertical force, F_v and its moment around the bottom of an upright section: the horizontal force momentum M_h, the vertical force momentum M_v.

5.2 Scour around the breakwater

For each of the selected cases defined by the MDA algorithm, the maximum scour depth at the foot of the breakwater was calculated. In 1991, Hughes and Fowler performed several laboratory tests, and from these results obtained the following empirical equation:

$$\frac{S_m}{(U_{rms})_m \cdot T_p} = \frac{0.05}{(\sinh(k_p \cdot h))^{0.35}} \tag{4.24}$$

where k_p is the wave number determined from the dispersion relationship, h is the depth and $(U_{rms})_m$ is given by the equation:

$$\frac{(U_{rms})_m}{g \cdot k_p \cdot T_p \cdot H_s} = \frac{\sqrt{2}}{4\pi \cdot \cosh(k_p \cdot h)} \cdot \left[0.54 \cdot \cosh\left(\frac{1.5 - k_p \cdot h}{2.8}\right)\right] \tag{4.25}$$

with T_p as the peak period, H_s the significant wave and g the gravity acceleration.

6 Time series reconstruction

The reconstruction of the time series of the stability parameters on the vertical breakwater was carried out by means of an interpolation technique based on the radial basis function (RBF), a very convenient method for scattered and multivariate data ([53], [72]). This interpolation method approximates the real-valued function $f = f(x)$ using a weighted sum of radially symmetric basic functions located on the scattered data points $x_1, , x_M$ where the associated real function values $f_1, , f_M$ are available. The approximation function is assumed to be of the form:

$$RBF(x) = p(x) + \sum_{j=1}^{M} a_j \Phi(\|x - x_j\|) \tag{4.26}$$

where Φ is the radial basis function, $\|\|$ being the Euclidian norm; $p(x)$ is a monomial basis $p_0, p_1, , p_n$, formed by a number of monomials of degree=1 equal to the data dimension (n) and a monomial of degree=0 and $b = b_0, b_1, , b_n$ the coefficients of these monomials. The RBF coefficients a_j and the monomial coefficients b are obtained by enforcing the interpolation constraints $RBF(x_i) =$

f_i. Gaussian functions with a shape parameter are used as the radial basis coefficients. The optimal value of the shape parameter is estimated using the algorithm proposed by [137]. To implement the RBF interpolation technique in the pressure distribution time series reconstruction, we have M 4-dimensional points $D_j{}^* = \{H_{s,j}{}^D, T_{s,j}{}^D, Z_{s,j}{}^D, \theta_{s,j}{}^D\}$; $j = 1, ...M$, corresponding to the M cases selected by the MDA algorithm and the associated semi-empirical parameters obtained applying the GodaTakahashi method. In parallel and for the scour distribution time series reconstruction we have M 4-dimensional points $D_j{}^* = \{H_{s,j}{}^D, T_{s,j}{}^D, Z_{s,j}{}^D\}; j = 1, ...M$, for the M cases selected and the associated parameters of the Hughes and Fowler formulation. These semi-empirical parameters are: the horizontal force F_h, the vertical force F_v, the horizontal force momentum M_h, the vertical force momentum M_v and the scour around the breakwater S_m. With the parameters mentioned above, the stability of the vertical breakwater can be analyzed and the safety factor against sliding (SSC) and overturning (OSC) can be estimated. The aim of the RBF application is the evaluation of the interpolation function of each of the following parameters: the horizontal force RBF_{F_h}, the vertical force RBF_{F_v}, the horizontal force momentum RBF_{M_h}, the vertical force momentum RBF_{M_v}, and the scour RBF_{S_m}. To calculate the interpolation functions, scalar variables are normalized with a simple linear transformation which scales the values between 0 and 1. Circular variables are normalized by dividing the direction values by π. Therefore each situation in the time series (60 years) is defined as $X_i = H_{s,i}, T_{p,i}, \theta_{m,i}, Z_{m,i}$; $i = 1, ..., N$, while each selected case is expressed as $D_j = \{H_{s,j}, T_{s,j}, Z_{s,j}, \theta_{s,j}\}; j = 1, ...M$ The interpolation function is calculated using the following expression:

$$RBF(X_i) = p(X_i) + \sum_{j=1}^{M} a_j \Phi(\|X_i - D_j\|) \tag{4.27}$$

where $p(X_i) = b_o + b_1 H_{si} + b_2 T_{pi} + b_3 \theta_{mi} + b_4 Z_i$ and Φ is a Gaussian function with a shape parameter c. The Euclidean distance has been replaced by the distance EC as in the MDA algorithm.

$$\Phi(\|X_i - D_j\|) = \exp(-\frac{\|X_i - D_j\|^2}{2c^2}) \tag{4.28}$$

The optimal shape parameter is estimated by the Rippa algorithm. The coefficients $b_l = [b_0, b_1, b_2, b_3, b_4]^T$ of the monomials and the coefficients $a_j = [a_1, , a_M]^T$ of the radial basis functions are obtained by the interpolation conditions:

$$RBF(D_j) = f_j(D_j); j = 1, ..., M \tag{4.29}$$

where the real functions f_j are defined by the parameters F_h, F_v, M_h or M_v obtained by GodaTakahashi, corresponding to the sea states selected by the MDA algorithm D_j.

Therefore, pressure and scour distributions are calculated for the entire 60 year series by means of the RBF functions obtained for each parameter. These functions are defined as:

$$F_{h,i} = RBF_{F_h}(\{D_j, F_{h,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{4.30}$$

$$F_{v,i} = RBF_{F_v}(\{D_j, F_{v,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{4.31}$$

$$M_{h,i} = RBF_{M_h}(\{D_j, M_{h,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{4.32}$$

$$M_{v,i} = RBF_{M_v}(\{D_j, M_{v,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{4.33}$$

$$S_{m,i} = RBF_{S_m}(\{D_j, S_{m,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{4.34}$$

And the final result is the reconstructed time series of the parameters that define the wave loads and the scour at a vertical breakwater:

$$X_{p,i} = \{F_{h,i}, F_{v,i}, M_{h,i}, M_{v,i}, S_{m,i}\}; i = 1, ..., N \tag{4.35}$$

7 Validation

The proposed methodology was applied to calculate wave loads and scour at a breakwater. The time series of the propagated parameters F_h, F_v, M_h, M_v and S_m were reconstructed considering different numbers of cases selected by the MDA algorithm ($M = 10$, 20, 30, 40, 50, 75, 100, 150, 200, 300, 500, 750 and 1000). It means that, for example with $M = 10$, the RBF is applied to reconstruct the 60 years hourly time series taking into account the 10 sea states selected with the MDA algorithm. That process was then repeated for all the values of M considered and for all the parameters of study.

On the other hand, values for the same four parameters (F_h, F_v, M_h, $M_v$and S_m) for the entire 60 year time series were calculated analytically. Thus we have real values available for the validation of the time series reconstructed by means of the proposed methodology.

To validate the time series, 99% of the percentiles for each time series reconstructed were calculated, varying the number of cases selected by the MDA ($M = 10, 20, 30, 40, 50, 75, 100, 150, 200, 300, 500, 750, 1000$) for all the parameters obtained. Once this was done, the error made by these percentiles was calculated according to the following:

$$E_i = \frac{(X_i - Y_i)}{X_i} \cdot 100 \tag{4.36}$$

where X_i stands for the value of the 99% percentile of the series obtained numerically and Y_i for the 99% percentile of each reconstructed time series. Figure 4.5 shows the evolution of the error of the 99% percentile of the F_h, and F_v series reconstructed by varying the number of selected cases. Fairly good results were obtained for the 99% percentiles of the horizontal and vertical forces and momentums, with maximum differences of $< 5\%$ for the horizontal force and momentum, and 78% for the vertical force and momentum. Note that these values of disagreement correspond to series reconstructed with less than 50 cases from a data base of about 500,000 cases (60 year hourly data). Whenever series were reconstructed with more than 100-200 cases, this error reached values of less than 1%.

The quality of the reconstruction of the F_h and F_v time series is shown in Figure 4.6, which represents the real time series and those reconstructed with $M = 50$ and

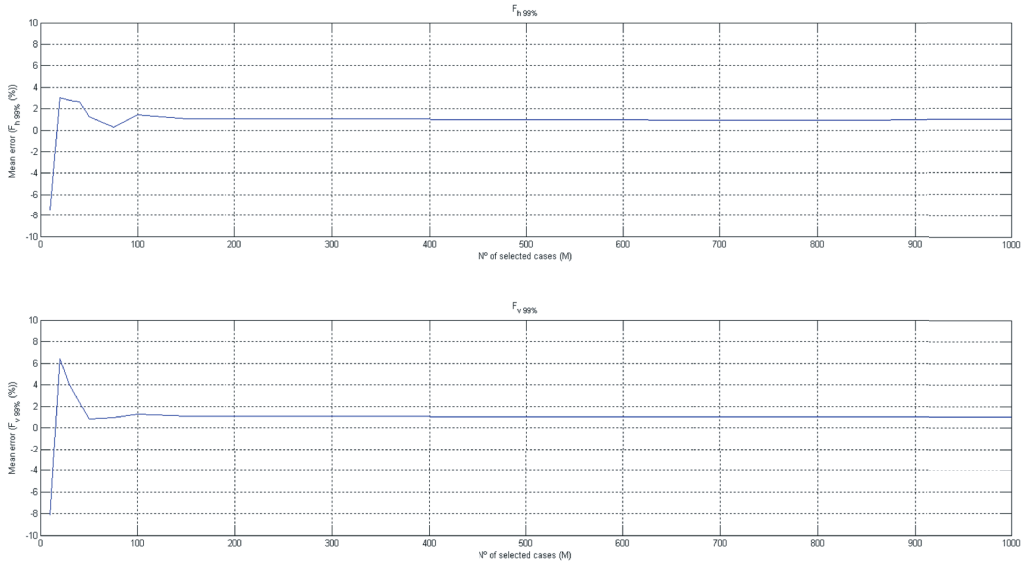

Figure 4.5: Evolution of the error in the 99% percentile of F_h and F_v varying the number of selected cases to reconstruct the time series.

$M = 500$ selected with the MDA algorithm from the entire series, respectively. In this figure, the line describing the F_h and F_v series using 500 cases is hardly different from that of the real time-series. The line corresponding to 50 cases presents larger differences with the real one.

In terms of scour, the results obtained are significantly better than those regarding the forces. Figure 4.7 shows the evolution of a scatter plot for an increasing number of selected cases used to reconstruct the series. As can be seen, the reconstructed time series matches almost perfectly the analytically calculated one.

As a final summary of the proposed methodology, Figure 4.8 shows a diagram of the steps followed throughout the entire process

8 Conclusions

A hybrid methodology was developed to obtain long-time series of wave loads on a breakwater using numerical or physical models with a reasonable computational time. The methodology is based on the selection of a number of representative

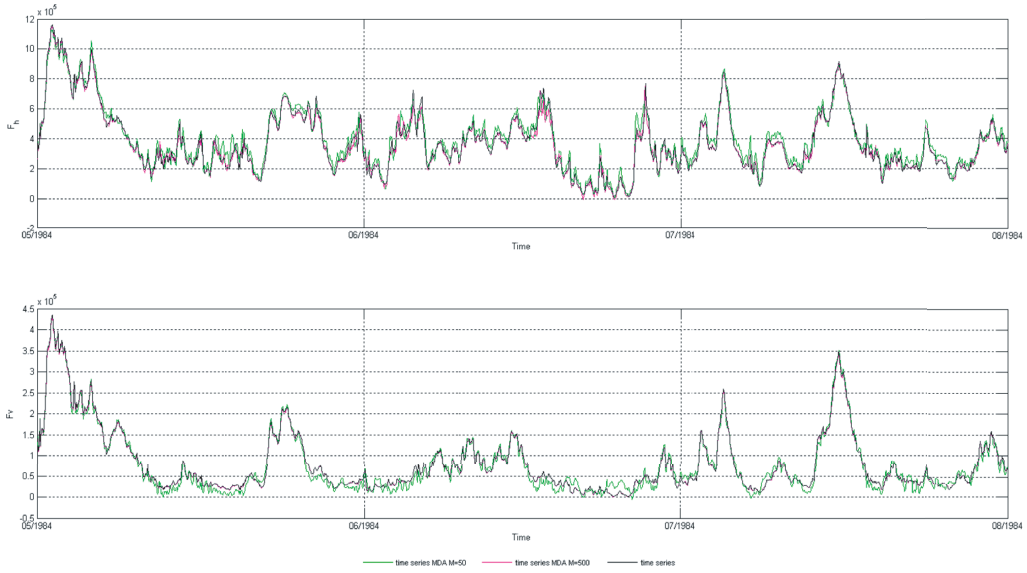

Figure 4.6: Time series calculated (in grey) and reconstructed considering $M = 50$ cases (in green) and $M = 500$ cases (in pink) of the parameters F_h and F_v.

wave climate sea states at the breakwater location, the calculation of the dynamic loads corresponding to these selected wave situations and a multidimensional RBF interpolation to reconstruct the wave load time series.

Although the methodology was developed to be used with numerical or physical models, requiring huge computational times to simulate each sea state, in this study a semi-empirical formula (Goda-Takahashi method) was used to carry out the validation, due to the possibility of calculating load parameters corresponding to every sea state of the reanalysis database. Validation of the results confirmed that the proposed methodology can reproduce the time series of wave loads and scour. The results shown above state that for a selection of less than 500 cases, the error made in the estimated parameters is almost negligible in a reconstruction of a 60-year hourly time series. The proposed methodology therefore provides a tool with accurate results of estimated parameters avoiding many calculations. In terms of computational effort and using a standard PC, it takes less than an hour to reconstruct a time series using RBF and around half an hour to extract a representative group of cases by using MDA algorithm.

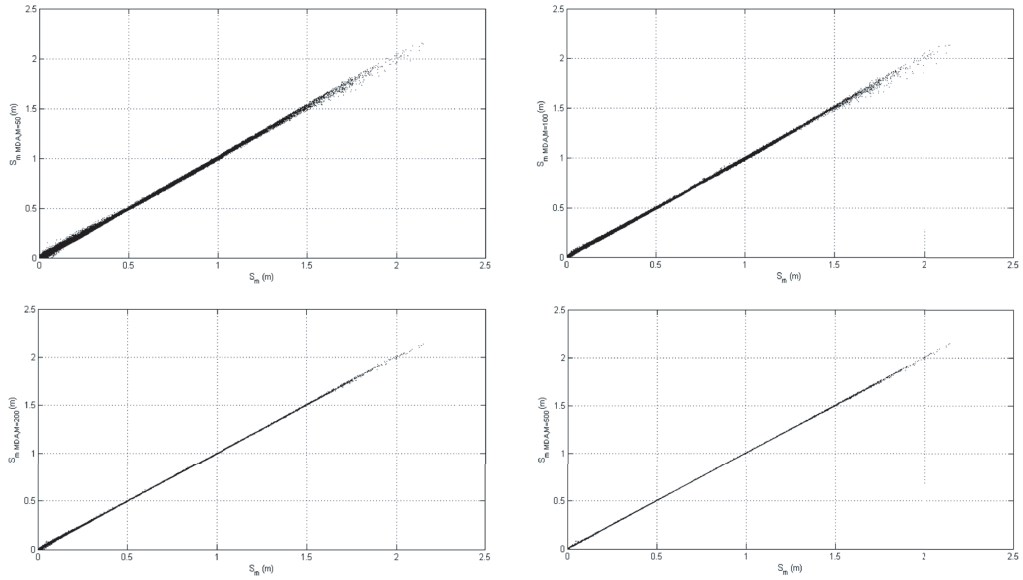

Figure 4.7: Scatter plot of the calculated series of scour and the reconstructed ones. In the upper-left plot with $M = 50$ cases, in the upper-right plot with $M = 100$ cases, in the lower-left with $M = 200$ cases and with $M = 500$ in the lower-right plot.

The MDA algorithm automatically selects a subset of multidimensional sea states evenly covering the diversity of wave situations at the structure location and is very convenient for the subsequent RBF interpolation technique.

The availability of long time series of loads acting on a vertical breakwater and the scour around it allows the analysis of critical stability situations, taking into account the combination of different correlated sea state parameters, and the definition of the design force with a certain return period, which constitutes an improvement in the design of coastal structures.

As shown, the proposed methodology can be applied to reconstruct parameters of different nature such as loads and scour. Both examples explained in this chapter were selected due to their relation to the breakwater stability. The methodology can also be used to transform stability parameters which depend on other sea state parameters with higher dimensionality due to the capability of MDA and RBF to work with high dimensional data.

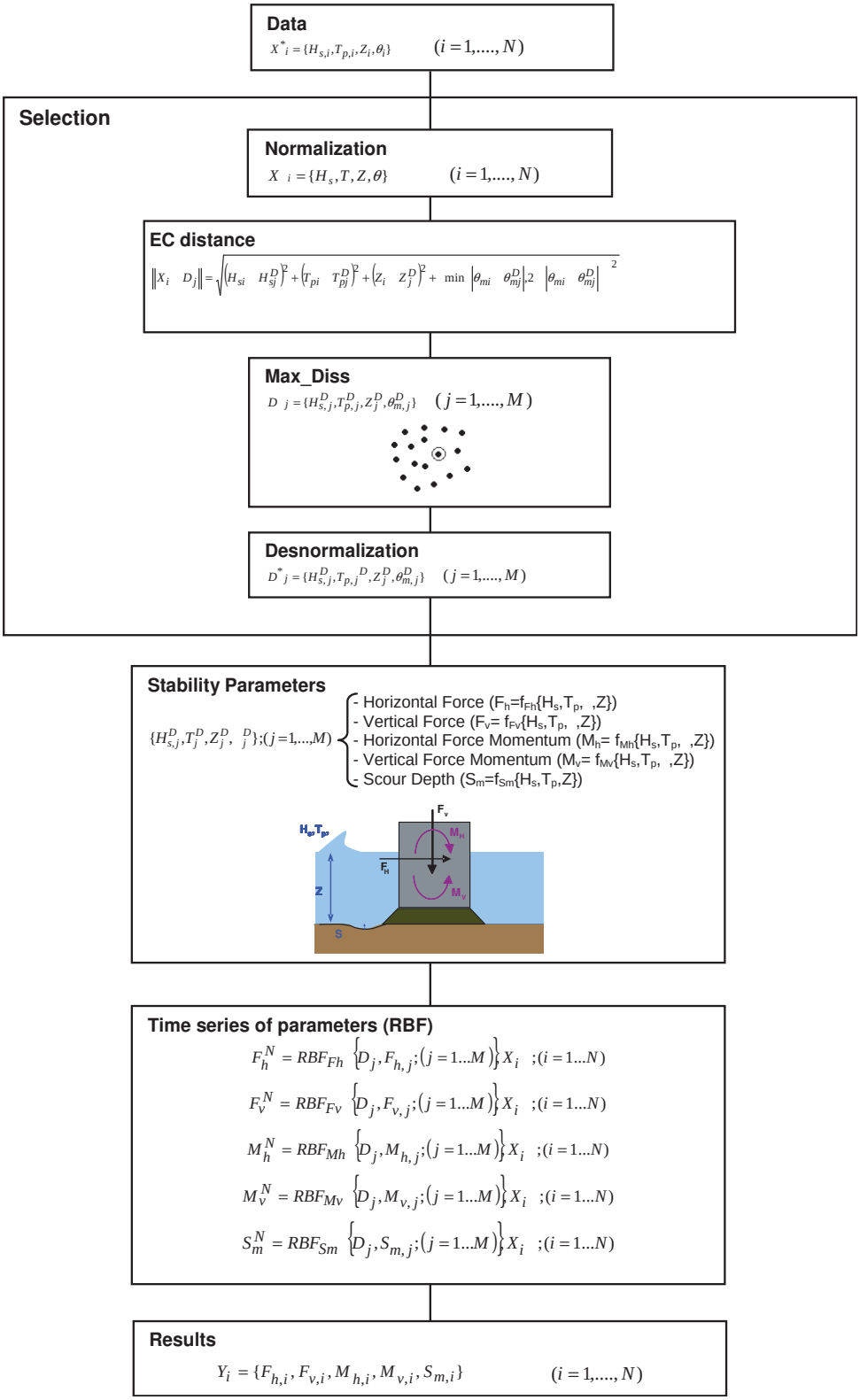

Figure 4.8: Scheme of the methodology for Gijon's breakwater application (example).

CHAPTER

5

A multivariate approach to estimate design loads for offshore wind turbines

1 Abstract

The design of offshore wind farms is a complex process that requires a detailed study of the oceanographic, meteorological and geotechnical conditions at the site. The structure and all structural members shall be designed in a way that they can be resistant against different kinds of loads: permanent, variable, environmental, accidental and deformations. This chapter is focused on those called environmental loads. The main environmental conditions that may contribute to structural damage, operational disturbances or other failures are wind, waves, currents and sea ice. Thus, the combination of the different parameters may produce many different critical situations for the integrity of the structure, requiring the calculation of long time series corresponding to long-term historical data situations. The most accurate techniques available at the moment to estimate loads acting upon a structure are numerical and physical models; however, they are very time consuming,

and the calculation of long time series of data is unfeasible. Therefore, a new hybrid methodology to select waveswindcurrent representative conditions that allow the interpolation of long time series of forces on a wind turbine is proposed. The methodology consists of a selection of a subset of representative cases of wave-windcurrent climate at the structures location by using a maximum dissimilarity algorithm, then estimating loads acting upon the structure for the seawind states selected and the reconstruction of loads corresponding to historical data using an interpolation technique based on radial basis function. To validate the proposed methodology and because of there is no availability of long time records of loads on wind turbines, the well-known IEC 61400-3 has been applied to estimate the loads for the complete reanalysis time series of waves, winds and currents. The validation of the results confirms the ability of the methodology developed to reconstruct time series of forces on the structure on the basis of the previously selected cases. This methodology permits application of numerical and physical models to offshore wind farm design, considerably reducing the number of tests or simulations.

2 Introduction

The need for clean, renewable and sustainable energy sources has increased in recent decades. Along with this need comes the development of technology and the capabilities of human society. As a result of these two ideas, offshore wind energy appears. Wind energy is one of the most promising options for electricity generation. According to the Europe 2020 Renewable Energy Targets, wind energy-installed production will increase from 2.9GW in 2010 to 40GW in 2020.

Technology today makes offshore structures possible where conditions are favourable. A good example is seen on the North West Europe coast, especially the North Sea, where the 2020 installed production target is 25GW. There, continental shelf conditions allow good foundations for turbines on the sea floor, and the constant winds provide the energy needed to start the process. On the Spanish coast, the Mediterranean would be where it would be possible to instal wind farms moored to the sea bottom. The Atlantic Coast and the Canary Islands Coasts are too deep for that.

The analysis and design of offshore wind farms are a complex process that involves many technical requirements. This design has to guarantee safe and economical operation during the service life of at least 20 years and must take account of the most important external conditions, closely related to the wind farm location.

According to the DNV–OS–J101 Offshore Standards ([5]), when designing wind turbine farms, a high number of load cases should be considered. These load cases fall into different categories: permanent (mass), variable (actuation, operational), environmental (wind, waves, current, ice), accidental (collisions, explosions, fire, etc.) and deformation loads (temperature, built-in deformations and settlement of foundations). This study is focused on environmental applied loads (wind, wave and currents) during the operational life of the wind turbine. Other kinds of loads (e.g. accidental loads, internal loads due to the dynamic response of the structure and loads associated with turbine or grid faults) are not considered.

For offshore structures, environmental load involves loads due to waves, wind, current, earthquakes, ice, snow, tidal, scouring, etc., but the most important is those related to waves, winds and currents. From all the external agents referred to before, the design rules establish multiple combinations of situations that may occur during the life cycle of the structure to be studied and/or simulated, both numerically and physically. In terms of numerical models, several computer programs have been developed to simulate the behaviour of offshore structures([2], [1],[3]). Related to these aero–hydro–servo–elastic time–domain simulations have appeared in the lasts years the contour line or surface methods, which can be very useful in terms of extreme responses for limit state designs,([91],[92]) but the assessment of combined wave, wind and current loads on offshore structures continues to be a time-consuming process. On the other hand, physical models give good results, but the number of cases that can be tested is limited by financial and time factors. Despite their limitations, physical models are very useful for optimizing design and for the verification and validation of numerical models.

This study proposes a hybrid methodology to obtain long time series of environmental loads on an offshore wind turbine by simulating a limited number of seawind states. The methodology is based on that proposed to downscale wave dynamics in coastal structures ([66]) and combines a selection of seawind states

by using data mining algorithms and a later reconstruction interpolation with radial basis functions (RBFs). To validate the proposed methodology, the loads have been estimated applying IEC 61400-3 (2009), but the final objective of the proposed methodology is to apply numerical or physical models. It must be noted that some simplifications have been made when estimating the loads.

By using this methodology, the number of numerical simulations or physical tests needed to reproduce long time series of loads on the wind turbine can be reduced. The availability of long time series of loads on the wind turbine gives access to all the combinations of effects because of external agents that may not be included in the recommended load cases of the standards. This study therefore aims to provide a useful tool in wind turbine structure design.

3 Proposed methodology

The methodology proposed is a transformation to design parameters (loads) of a representative subset of sea and wind conditions from a historical database, and a statistical reconstruction of the time series of the parameters is calculated ([19], [20]). The stages of the methodology are the following: (i) definition of the met-ocean database; (ii) selection of representative wind and sea states using data mining algorithms; (iii) loads on a wind turbine calculation; and (iv) transference of the complete series of wind and sea states using an interpolation scheme. The proposed methodology has been successfully used to study loads acting upon vertical breakwaters ([66]). A diagram of the methodology is shown in Figure 5.1.

- (1) Use of this methodology requires a long-term database (N data) of the different variables involved at the location. For this application, the number of independent met-ocean parameters is $n = 8$. These variables define the characteristics of waves (significant wave height (H_s), wave period (T_m), wave direction (θ_{Waves})), wind (magnitude and direction (V_{1-hour}, θ_{Wind})), currents (the speed and direction of the tidal current (U_{Tidal}, θ_{Tidal})) and sea level (SWL).

- (2) The aim of the selection process was to extract a representative subset of windwave conditions of the long time series from reanalysis databases.

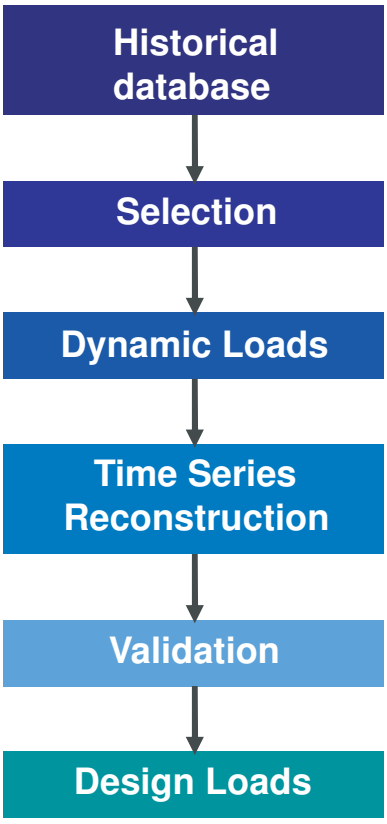

Figure 5.1: Diagram of the methodology.

The maximum dissimilarity algorithm (MDA) allows an automatic selection of a subset of sea states that are representative of met-ocean climate at the location useful to be combined with an interpolation technique ([19]).

These reanalysis data are defined by scalar and directional parameters of different magnitudes that require normalization and an implementation of the distance in the circle for the directional parameter in the MDA algorithm. The scalar variables are normalized by scaling the variable values between 0 and 1 with a simple linear transformation requiring two parameters, the minimum and maximum values

of the two scalar variables. In the case of circular variables (defined in radians or in sexagesimal degrees using the scaling factor $\pi/180$) and considering that that the maximum difference between two directions over the circle equals π and the minimum difference equals 0, normalization was achieved by dividing the direction values by π. In this way, the circular distance was rescaled between 0 and 1. The initial data of the subset D_1 were considered to be the sea state with the highest significant wave height. The rest of the $M-1$ elements were selected iteratively, transferring the most dissimilar ones to the subset established by the MaxMin version of the algorithm ([165]). This dissimilarity was determined using the Euclidean circular distance. Finally, we de-normalized the subset applying the opposite transformation used during normalization.

- (3) In this phase of the methodology, loads were calculated for each of the selected seawind conditions at the structures location. A simplified method (based on IEC 61400-3) is considered because it allows the direct transformation of the long time hourly reanalysis series (20 years for this application), which made the validation of the proposed methodology possible.

- (4) The reconstruction of the time series of loads on the wind turbine used an interpolation technique based on RBF, a very handy method for scattered and multivariate data ([53], [72]).

Figure 5.2 shows a detailed schematic diagram of the methodology.

An application is considered to explain the proposed methodology. The offshore wind turbine under study is located on the northeast Mediterranean coast of Spain between the city of Tarragona and the Ebro River Delta. This place was chosen because the continental shelf in the area would, as already pointed out, allows an offshore wind farm to be located there.

Figure 5.3 shows the location of the case study.

4 Database

To apply the methodology described, long-term database of waves, currents and winds from the location is needed. The parameters were taken from reanalysis databases because of its adequate length, covering 20 years (1989-2009). The wave

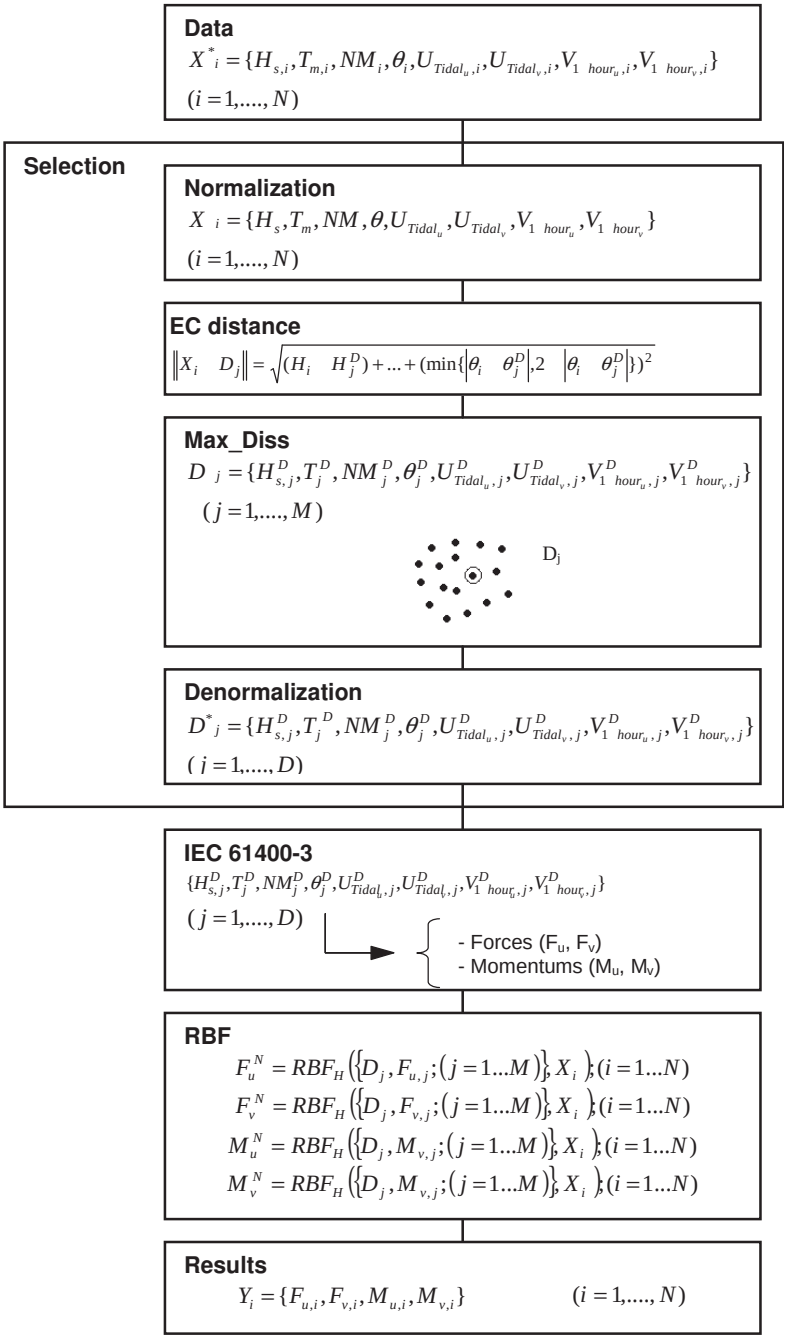

Figure 5.2: Schematic diagram of the methodology.

Figure 5.3: Location of the case study.

data were taken from Downscaled Ocean Waves 2.1, a coastal hourly wave reanalysis database with a spatial of resolution of 0.125° in the western Mediterranean Sea. This database has been obtained downscaling the Global Ocean Waves 2.1 database applying a hybrid methodology ([19]) that combines the SWAN wave model (dynamical downscaling) with statistical tools (MDA algorithm for the selection of 500 representative deep water conditions and RBF interpolation technique for the 20 year hourly time series reconstruction). Global Ocean Waves 2.1 (IH Cantabria) is a wave reanalysis database generated usingWW3 model and forced by 15 km resolution wind fields from a dynamic downscaling (WRF model) nested to ERA-Interim (1989-2009) atmospheric reanalysis.

The wind data were extracted from the SeaWind-ERA-Interim database (IH Cantabria)([111]). SeaWind constitutes an hourly wind reanalysis database on a 15 km spatial resolution grid, which provides surface winds at a height of 10m. It was generated by dynamic downscaling, using the WRF model from a global reanalysis (ERA-Interim, developed by the European Centre for Medium-Range Weather Forecasts) over the South Atlantic-European region and the Mediterranean basin.

The storm surge data used were extracted from the Global Ocean Surges 2.1

database (IH Cantabria) ([6]), a high-resolution hourly storm surge reanalysis database covering 20 years (1998-2009) in Southern Europe, with a spatial resolution of 1/8° (∼13 km). It was generated with the three-dimensional Regional Ocean Modelling System model developed by the Rutgers Ocean Modelling Group and forced with high-resolution wind and pressure data (∼15 km) from a dynamic downscaling of ERA-INTERIM.

5 Selection of wind and sea states

The subset is selected using the MDA. Figure 5.4 shows the resulting clusters or selected data when applying three different data mining techniques: K-means algorithm, self-organizing maps and MDA, in a data sample located in a circle domain ([20]). It can be observed that MDA distributes selected data rather evenly across the space, with some points selected in the outline of the data space guarantee the most representative subset from the original sample.

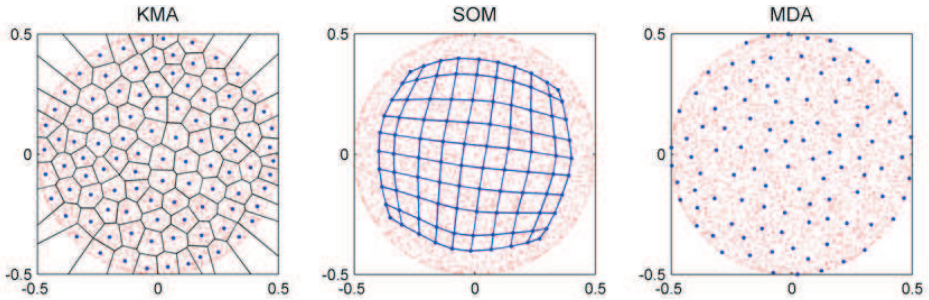

Figure 5.4: Comparison of K-means (KMA), self-organizing maps (SOM) and maximum dissimilarity (MDA) algorithm.

Therefore, selection was carried out using this technique because of its ability to fairly represent the space with some points selected in the outline of the data set. An explanation of the algorithm and the required data pre-processes is given in section 5.1 of this chapter. Different selections have been performed using MDA for different subset sizes. Figure 5.5 shows the long time series of the eight independent variables with the data selected by MDA algorithm.

Figure 5.6 shows different scatter plots of the hourly time series and the data selected in different colours (M=1, 50 in red, M= 51, 100 in green and M= 101,

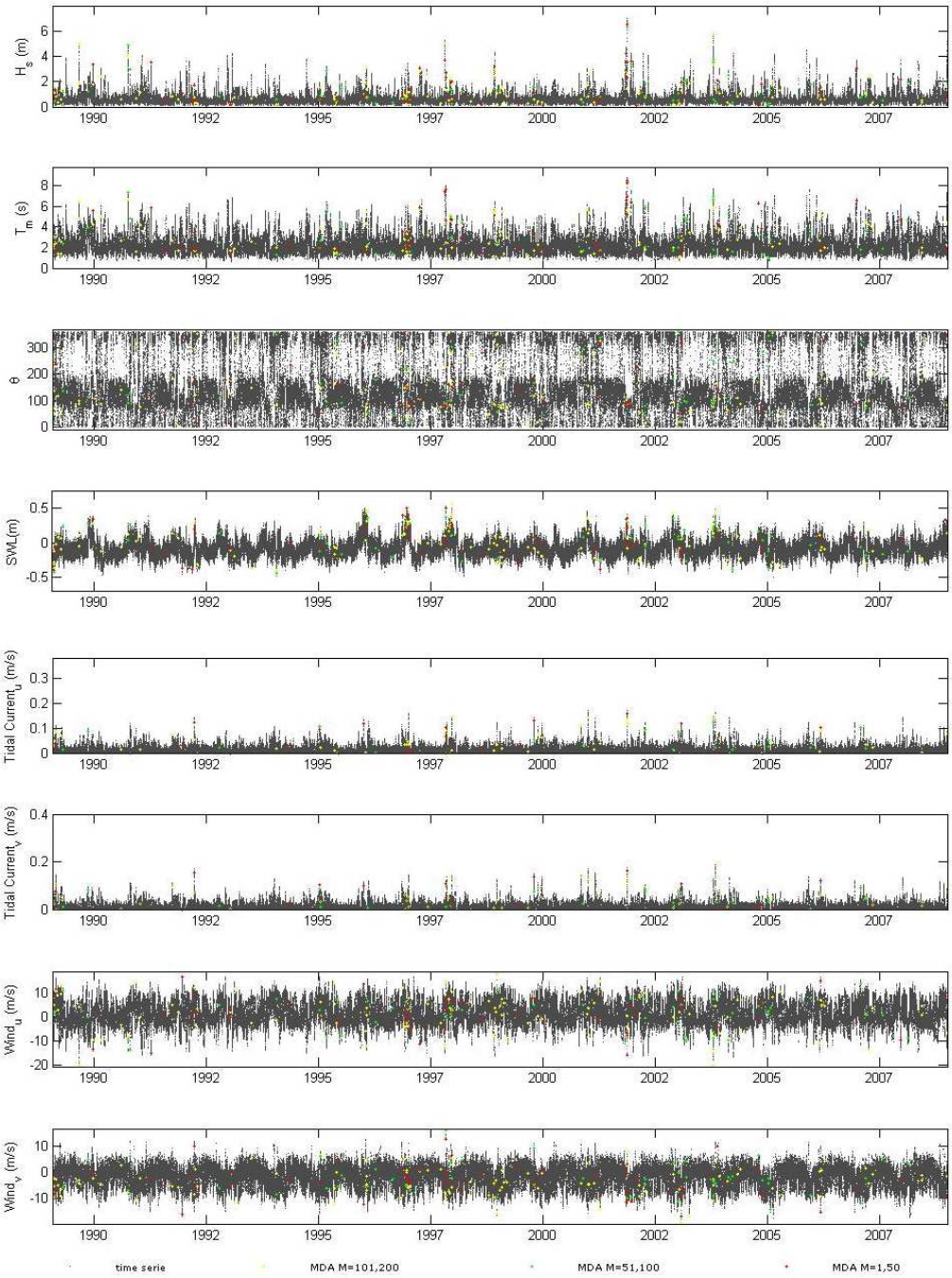

Figure 5.5: Time series of H_s, T_p, θ(wave direction), SWL, tidal current (in its components X and Y) and wind (in its components X and Y) at the wind turbine location (grey points), the cases selected by MDA algorithm, $M = 1, 50$ red points, $M = 51, 100$ green points and $M = 101, 200$ yellow points .

200 in yellow). The cases selected span the space of the input data, trying to cover it evenly and fill it uniformly. It must be mentioned that the MDA algorithm is a deterministic method, so for example in a subset of M= 200 cases, the first 100 will be the same as those for a subset of M= 100 cases.

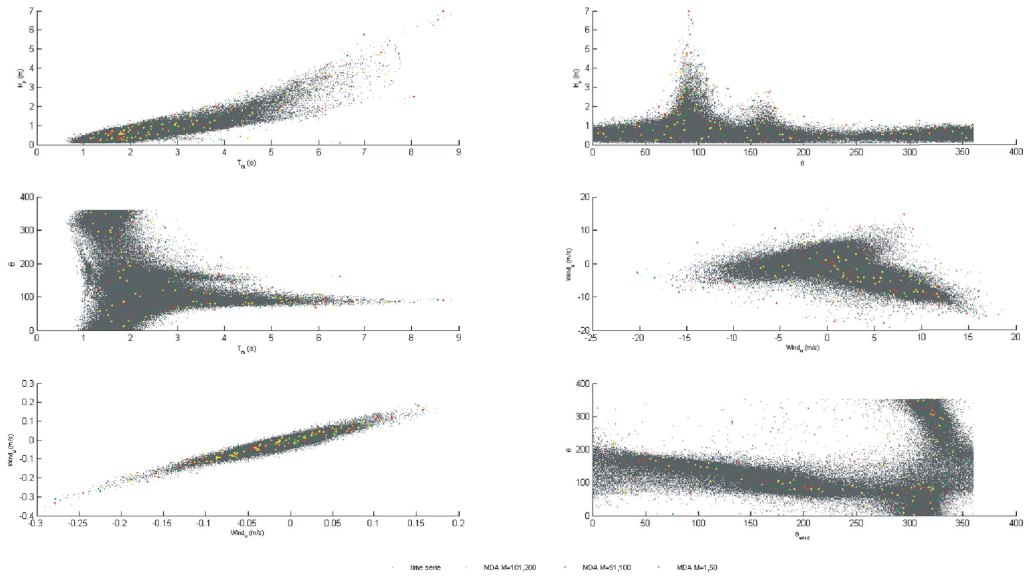

Figure 5.6: Distribution of cases selected by MDA algorithm ($M = 1, 50$ red points, $M = 51, 100$ green points and $M = 101, 200$ yellow points).

5.1 Maximum dissimilarity algorithm

In the development of computer-based methods for selecting sets of structurally diverse compounds from chemical databases, dissimilarity-based compound selection has been suggested as an effective method to identify a subset comprising the most dissimilar data in a database ([148]). The maximum dissimilarity algorithm (MDA) is one subclass of these selection techniques. The selection process starts initializing the subset by transferring one vector from the data sample. The remaining elements are selected iteratively, calculating the dissimilarity between each remaining data in the database and the elements of the subset and transferring the most dissimilar one to the subset.

In this work, sample $X_i = \{H_{si}, T_{mi}, SWL_i, \theta_{Wavesi}, U_{Tidali}, \theta_{Tidali}, V_{1-houri}, \theta_{Windi}\}$; $i = 1, ..., N$ are defined by scalar and directional variables of different magnitudes. The vector components require to be normalized to be equally weighted in the similarity criterion. The scalar variables are normalized by scaling the variables values between [0,1] with a simple linear transformation, which requires two parameters, the minimum and maximum value of the two scalar variables. The circular variables are normalized by dividing the direction values between π, therefore rescaling the circular distance, which could be maximum equal to π, between [0,1]. The Euclidean circular distance is implemented in the similarity criterion.

Between all the algorithm variants ([165]), the MaxMin version has been considered to obtain the most representative subset of the diversity of the data. For example, if the subset is formed by R $(R \leq M)$ vectors, the dissimilarity between the vector i of the data sample $N - R$ and the j vectors belonging to the R subset is calculated as

$$d_{ij} = \|X_i - D_j\|; i = 1, ..., N - R; j = 1, ..., R \tag{5.1}$$

Subsequently, the dissimilarity $d_{i,subset}$ between the vector i and the subset R is calculated as

$$d_{i,subset} = min\{\|X_i - D_j\|\}; i = 1, ..., N - R; j = 1, ..., R \tag{5.2}$$

Once the $N - R$ dissimilarities are calculated, the next selected data is the one with the largest value of $d_{i,subset}$. Moreover, the efficient algorithm developed by [131] has been implemented, which implies not calculating the distance between the different vectors d_{ij} in the definition of the distance $d_{i,subset}$. For example, in the selection of the R vector, the distance $d_{i,subset}$ is defined as the minimum distance between the vector i of the data sample (consisting of $N - (R - 1)$ vectors at this cycle) and the last vector transferred to the subset R, and the minimum distance between the vector i and the $R - 1$ vectors of the subset determined in the previous cycle:

$$d^{min}_{i,subset} = min\{d_{i,R}, d^{min}_{1,subset(R-1)}\} \tag{5.3}$$

After finishing the selection process, a denormalization of the subset has to be carried out. Finally, the MDA subset is defined by $D_i = \{H_{si}, T_{mi}, SWL_i, \theta_{Wavesi}, U_{Tidali}, \theta_{Tidali}, V_{1-houri}, \theta_{Windi}\}$; $i = 1, ..., M$.

6 Calculation of wind turbines loads

IEC 61400 [4] is a class of IEC international standards for wind turbines. Subclass 614003 refers to design requirements for offshore wind turbines, explaining how to study the structural components to provide an appropriate level of protection against damage from all hazards during the planned lifetime. The most important environmental loads on a monopile offshore wind turbine can be divided into the forces because of wind, waves and currents as shown in Figure 5.7. These forces arise as a function of different parameters. The momentum calculation follows the same scheme. All moments are referred to the bottom of the sea:

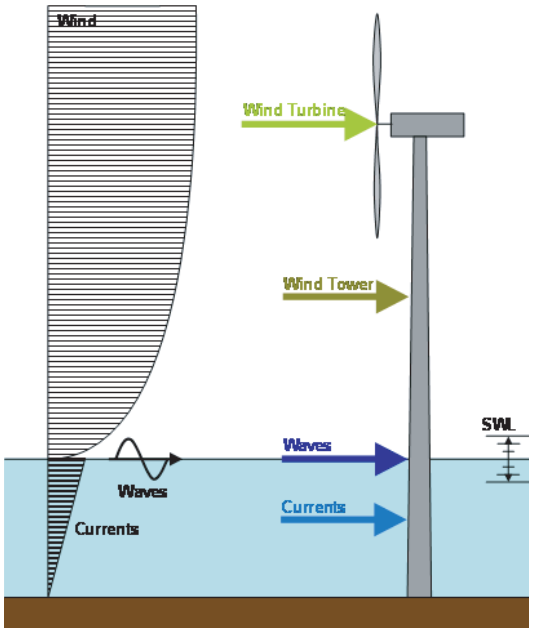

Figure 5.7: Loads on a wind turbine.

$$\overrightarrow{F} = \overrightarrow{F}_{Waves-Current} + \overrightarrow{F}_{Wind} \tag{5.4}$$

$$\overrightarrow{M} = \overrightarrow{M}_{Waves-Current} + \overrightarrow{M}_{Wind} \tag{5.5}$$

6.1 Wave-current forces

Particle velocities due to waves depend on H_s, T_p, θ_{Waves} and SWL. There are many ways to attempt to describe the velocity and acceleration profile. For this study in the interest of simplicity, linear wave theory was used despite not being the most accurate method taking into account the largest waves in water depths typical for offshore wind farms.

$$\overrightarrow{W}(z,t) = \pm\frac{H}{2}\frac{2\pi}{T_m}\frac{\cosh(k(h+z))}{\sinh(kh)}\cos\left(\frac{2\pi}{T_m}t\right) \tag{5.6}$$

$$\overrightarrow{\dot{W}}(z,t) = \pm\frac{H}{2}\left(\frac{2\pi}{T_m}\right)^2\frac{\cosh(k(h+z))}{\sinh(kh)}\cos\left(\frac{2\pi}{T_m}t\right) \tag{5.7}$$

As seen, because of water oscillation movement, velocities are time dependent on the wave period. Therefore, there are both a positive and a negative component inside the same wave. In this way, the highest wave ($H_{max} = 1.8H_s$) with the mean period (T_m) of each sea state was chosen, and maximum velocity instant was taken from its period.

When there is a current, the velocity profile given by it can be assumed to be described by:

$$\overrightarrow{U}(z)_{Tidal} = |\overrightarrow{U}(z)_{Tidal}|\left[\frac{z+h}{h}\right]^{1/7}(\sin(\theta_{Tide})\vec{i} + \cos(\theta_{Tide})\vec{j}) \tag{5.8}$$

$$\overrightarrow{U}(z)_{Wind} = 0.01\overrightarrow{V}(z=10m)_{1-hour}\left[\frac{z+20}{20}\right](\sin(\theta_{Wind})\vec{i} + \cos(\theta_{Wind})\vec{j}) \tag{5.9}$$

For collinear waves and currents, velocity profiles (waves and current in the wave direction component) should be added before estimating forces to take into account wavecurrent interaction, so the hydrodynamic forces (wave-current) are calculated with

$$\overrightarrow{F}(t)_{WC_{\theta w}} = \int_{z=-h}^{z=0} \left\{ \left[\frac{1}{2} C_d \rho_{Water} D |\overrightarrow{W}(z,t) + \overrightarrow{U}(z)_{\theta w}| \left(\overrightarrow{W}(z,t) + \overrightarrow{U}(z)_{\theta w} \right) \right] + \left[\frac{\pi}{4} C_M \rho_{Water} D^2 \left(\dot{\overrightarrow{W}}(z,t) + \dot{\overrightarrow{U}}(z)_{\theta w} \right) \right] \right\} dz \tag{5.10}$$

$$\overrightarrow{F}_{WC_{\theta w}} = \max \left(|\overrightarrow{F}_{WC_{\theta w}}| \right) (\sin(\theta_{Waves})\vec{i} + \cos(\theta_{Waves})\vec{j}) \tag{5.11}$$

$$\overrightarrow{M}(t)_{WC_{\theta w}} = \int_{z=-h}^{z=0} \left\{ \left[\frac{1}{2} C_d \rho_{Water} D |\overrightarrow{W}(z,t) + \overrightarrow{U}(z)_{\theta w}| \left(\overrightarrow{W}(z,t) + \overrightarrow{U}(z)_{\theta w} \right) \right] + \left[\frac{\pi}{4} C_M \rho_{Water} D^2 \left(\dot{\overrightarrow{W}}(z,t) + \dot{\overrightarrow{U}}(z)_{\theta w} \right) \right] \right\} (z+h) dz \tag{5.12}$$

$$\overrightarrow{M}_{WC_{\theta w}} = \max \left(|\overrightarrow{M}_{WC_{\theta w}}| \right) (\sin(\theta_{Waves})\vec{i} + \cos(\theta_{Waves})\vec{j}) \tag{5.13}$$

The non-collinear component of the current with the wave direction (current orthogonal to the wave direction) contributes with a drag force calculated by

$$\overrightarrow{F}(t)_{WC_{\perp\theta w}} = \int_{z=-h}^{z=0} \left[\frac{1}{2} C_d \rho_{Water} D |\overrightarrow{U}(z)_{\perp\theta w_{Tidal}}| \overrightarrow{U}(z)_{\perp\theta w_{Tidal}} \right] dz \tag{5.14}$$

$$\overrightarrow{M}(t)_{WC_{\perp\theta w}} = \int_{z=-h}^{z=0} \left[\frac{1}{2} C_d \rho_{Water} D |\overrightarrow{U}(z)_{\perp\theta w_{Tidal}}| \overrightarrow{U}(z)_{\perp\theta w_{Tidal}} \right] (z+h) dz \tag{5.15}$$

6.2 **Wind forces**

Finally, forces due to wind are a function of V_{1-hour}, θ_{Wind} and SWL and can be divided into the forces acting on the tower of the turbine and those acting on the

wind turbine. For this particular application, forces and moments on the tower are calculated using the IEC standards [4], and those on the turbine are calculated on the assumption of a 5 MW Reference turbine as defined by the National Renewable Energy Laboratory ([88]). It must be said that the wind forces considered are steady state wind forces. The cut-off velocity was set at 25 m/s, so wind forces over the turbine due to higher wind speeds have been neglected. However, these forces should be considered for a better prediction of extreme values. Nevertheless, the methodology proposed allows use other sophisticated numerical models that can consider additional non-linearities. Figure 5.8 shows the thrust force applied on the wind turbine related to wind speed.

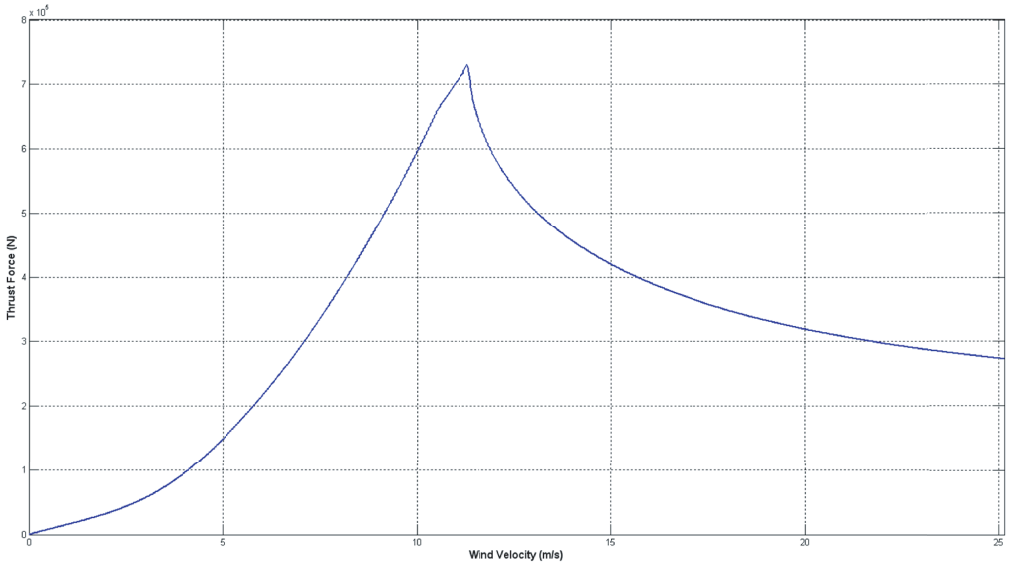

Figure 5.8: Thrust force on the turbine.

$$\overrightarrow{F}_{Wind} = \overrightarrow{F}_{\overrightarrow{V}_{Tower}} + \overrightarrow{F}_{\overrightarrow{V}_{Turbine}} \tag{5.16}$$

with:

$$\overrightarrow{F}_{\overrightarrow{V}_{Tower}} = \int_{z=0-SWL}^{z=nacelle} \frac{1}{2} C_d \rho_{Air} D |\overrightarrow{V}(z)_{1-hour}| \overrightarrow{V}(z)_{1-hour} dz \tag{5.17}$$

$$\overrightarrow{F}_{\overrightarrow{V}Turbine} \rightarrow \text{NREL/TP-500-38060} \tag{5.18}$$

and:

$$\overrightarrow{M}_{Wind} = \overrightarrow{M}_{\overrightarrow{V}Tower} + \overrightarrow{M}_{\overrightarrow{V}Turbine} \tag{5.19}$$

with:

$$\overrightarrow{M}_{\overrightarrow{V}Tower} = \int_{z=0-SWL}^{z=nacelle} \frac{1}{2} C_d \rho_{Air} D |\overrightarrow{V}(z)_{1-hour}| \overrightarrow{V}(z)_{1-hour}(z+h)dz \tag{5.20}$$

$$\overrightarrow{M}_{\overrightarrow{V}Turbine} = \overrightarrow{F}_{\overrightarrow{V}Turbine} z \tag{5.21}$$

where:

$$\overrightarrow{V}(z)_{1-hour} = \overrightarrow{V}(z=10m)_{1-hour} \left(\frac{z}{10}\right)^{0.14} (\sin(\theta_{Wind})\vec{i} + \cos(\theta_{Wind})\vec{j}) \tag{5.22}$$

7 Time series reconstruction

The reconstruction of the time series of loads on the wind turbine used an interpolation technique based on RBF, a very useful scheme for scattered and multivariate data. The RBF approximation has been successfully applied in many fields, usually with better results than other interpolation methods.

This interpolation method consists of approximating the real-valued function $f = f(x)$ with a weighted sum of radially symmetric basic function located at the scattered data points $x_1, ..., x_M$ where the associated real function values $f_1, ..., f_M$ are available. Therefore, after calculating the forces indicated in Section 6 for the M select waves-wind-current conditions, functions for the total horizontal and vertical force (F_u, F_v) and for the horizontal and vertical momentum (M_u, M_v) are determined using RBF method. The 20 year time series of these forces and momentums can be reconstructed applying the corresponding RBF functions in the rest of the environmental conditions. A detail explanation can be found in 7.1.

7.1 Radial Basis Function interpolation technique

Suppose that f=f(x) is the real-valued function that we want to approximate.We are given M scattered data points $x_1, ..., x_M$ of dimension n and the associated real function values $f_1, ..., f_M$, being $f_i = f(x_j), j = 1, ..., M$. The radial basis function (RBF) interpolation method consists of a weighted sum of radially symmetric basic function located in the data points. The approximation function is assumed to be of the form:

$$RBF(x) = p(x) + \sum_{j=1}^{M} a_j \Phi(\|x - x_j\|) \tag{5.23}$$

where Φ is the radial basis function, being $\|\|$ the Euclidian norm; $p(x)$ is a monomial basis $p_0, p_1, ..., p_n$, formed of a number of monomials of degree of 1 equal to the data dimension (n) and a monomial of degree of 0, being $b = b_0, b_1, ..., b_d$ the coefficients of these monomials.

The RBF coefficients a_j and the monomial coefficients b are obtained by enforcing the interpolation constraints $RBF(x_i) = f_i$.

There are several expressions for radial basis functions (linear, cubic, Gaussian, multiquadric), some of them containing a shape parameter that plays an important role for the accuracy of the interpolation method. [137] has proposed an algorithm for choosing an optimal value of the shape parameter by minimizing a cost function that imitates the error between the radial interpolant and the unknown function $f(x)$. This cost function collects the errors for a sequence of partial fits to the data: $E = (E_1, ..., E_M)^T$, where E_k is defined as the error between the function f_k in the point x_k and the estimated value by the RBF function calculated removing the point x_k from the original data set.

In the implementation of the RBF interpolation technique in the wind turbine load series reconstruction, we have M points eight-dimensional $D_i = \{H_{si}, T_{mi}, SWL_i, \theta_{Wavesi}, U_{Tidali}, \theta_{Tidali}, V_{1-houri}, \theta_{Windi}\}$; $i = 1, ..., M$, corresponding to the M cases selected by MDA algorithm and the associated (real) forces obtained by the application of the specific formulation to evaluate the environmental forces acting on the wind turbine (i.e. the total horizontal and vertical forces $F_{u,j}$ and $F_{v,j}$, and the total horizontal and vertical momentum $M_{u,j}$ and $M_{v,j}$).

Therefore, the aim of the RBF application is the evaluation of the interpolation function of the forces acting on the wind turbine decomposed in the components $x-$ and y, RBF_{F_u} and RBF_{F_v}, respectively.

To calculate the interpolation functions, the variables that define the process are normalized by means of a lineal transformation that scales the values between 0 and 1. Therefore, each initial situation is defined as $X_i = \{H_{si}, T_{mi}, SWL_i, \theta_{Wavesi}, U_{Tidali}, \theta_{Tidali}, V_{1-houri}, \theta_{Windi}\}$; $i = 1, ..., N$, whereas each selected case, where the real forces are available, is expressed as $D_i = \{H_{si}, T_{mi}, SWL_i, \theta_{Wavesi}, U_{Tidali}, \theta_{Tidali}, V_{1-houri}, \theta_{Windi}\}$; $i = 1, ..., M$

The interpolation function is calculated by means this expression:

$$RBF(X_i) = p(X_i) + \sum_{j=1}^{M} a_j \Phi(\|X_i - D_j\|) \tag{5.24}$$

where $p(X_i) = b_0 + b_1 H_i + b_2 T_{mi} + b_3 SWL_i + b_4 \theta_{wi} + b_5 U_{ti} + b_6 \theta_{Uti} + b_7 V_{1hi} + b_8 \theta_{windi}$ and Φ is a Gaussian function with a shape parameter c.

$$\Phi(\|X_i - D_j\|) = \exp(-\frac{\|X_i - D_j\|^2}{2c^2}) \tag{5.25}$$

The optimal shape parameter is estimated by the Rippa algorithm. The coefficients $b_l = [b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8]T$ of the monomials and the coefficients $a_j = [a_1, ..., a_M]^T$ of the radial basis functions are obtained by the interpolation conditions:

$$RBF(D_j) = f_j(D_j); j = 1, ..., M \tag{5.26}$$

where the real functions f_j are defined by the parameters F_h, F_v, M_u or M_v obtained, corresponding to the selected sea states by MDA algorithm D_j.

Therefore, loads over the wind turbine are reconstructed to the entire period of data by means the RBF functions calculated for each calculated parameter. These functions are defined as

$$F_{u,i} = RBF_{F_u}(\{D_j, F_{u,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{5.27}$$

$$F_{v,i} = RBF_{F_v}(\{D_j, F_{v,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{5.28}$$

$$M_{u,i} = RBF_{M_u}(\{D_j, M_{u,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{5.29}$$

$$M_{v,i} = RBF_{M_v}(\{D_j, M_{v,j}(j = 1, ..., M)\}, X_i); i = 1, ..., N \tag{5.30}$$

8 Validation

To evaluate the progression of the accuracy in the reconstructed time series of the parameters calculated, F_u, F_v, M_u and M_v (horizontal and vertical forces and momentums, respectively), they were reconstructed considering different numbers of cases selected by the MDA algorithm (M=10, 20, 30, 40, 50, 75, 100, 150, 200, 300, 500, 750 and 1000).

On the other hand, the values for the same four parameters (F_u, F_v, M_u and M_v) for the entire 20 year time series were calculated analytically. Thus in this application, we have the real values available for the validation of the reconstructed time series by using the proposed methodology.

The scatter plots for the propagated time series and the reconstructed time series for F_u and F_v are shown in Figure 5.9. The reconstructed time series with M=75, 200, 500 and 1000 selected cases are shown in the subplots. The normalized root mean square error (NRMSE) and the correlation coefficient (r) were computed for F_u and F_v. These statistics are given in Table I.

It can be seen that, with the proposed methodology, the reconstructed time series are more accurate when using more cases in the calculation of the corresponding RBF function. The differences between the calculated and the reconstructed time series are more significant using M=75 cases. However, the increased accuracy of reconstruction is not so great when using M=500 cases or M= 1000 cases.

The reconstructed time series of F_u and F_v using M=200 cases (in orange), M=1000 cases (in red) and the real-time series (in black) are shown in Figure 5.10,

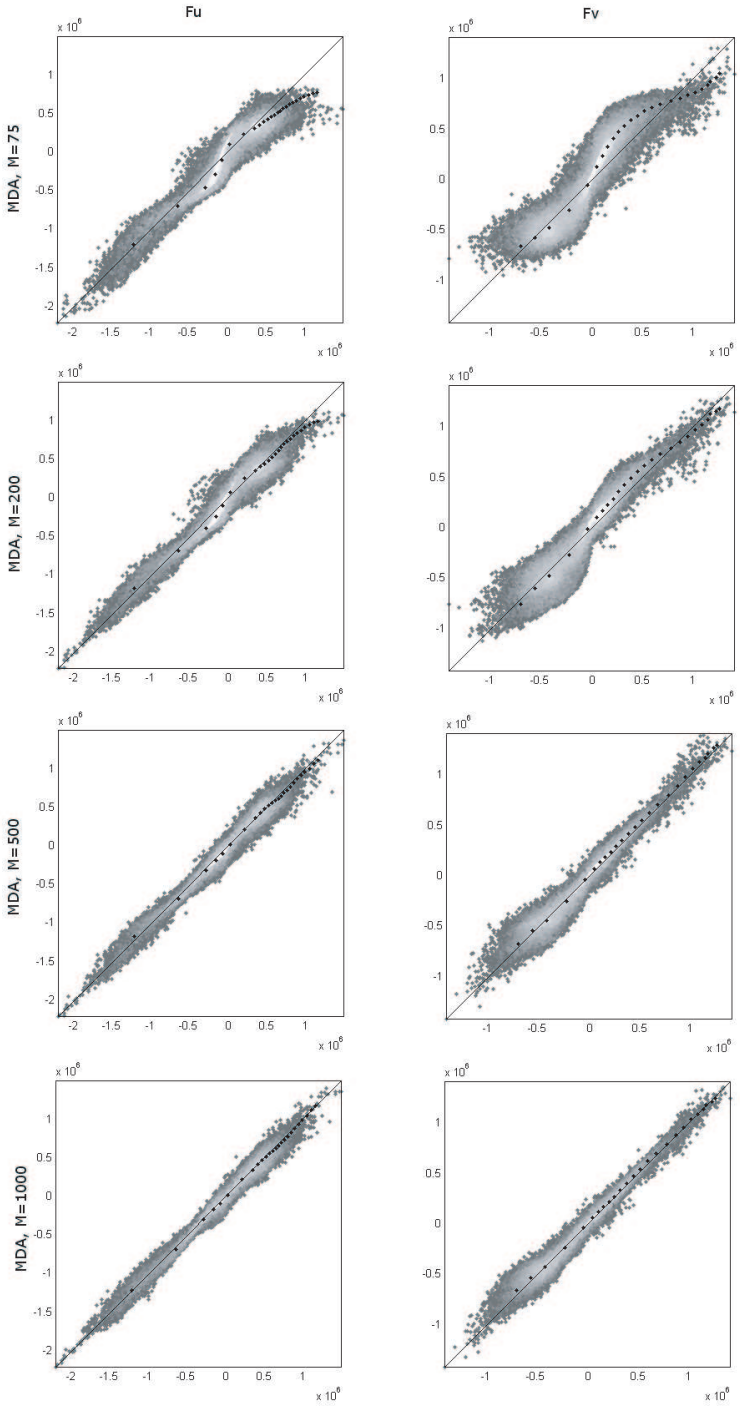

Figure 5.9: Scatter plot of the time series (calculated vs. reconstructed) of F_u (X component) and F_v (Y component) considering $M = 75$, $M = 200$, $M = 500$ and $M = 1000$.

demonstrating that the proposed methodology is able to reproduce the structure of the time series

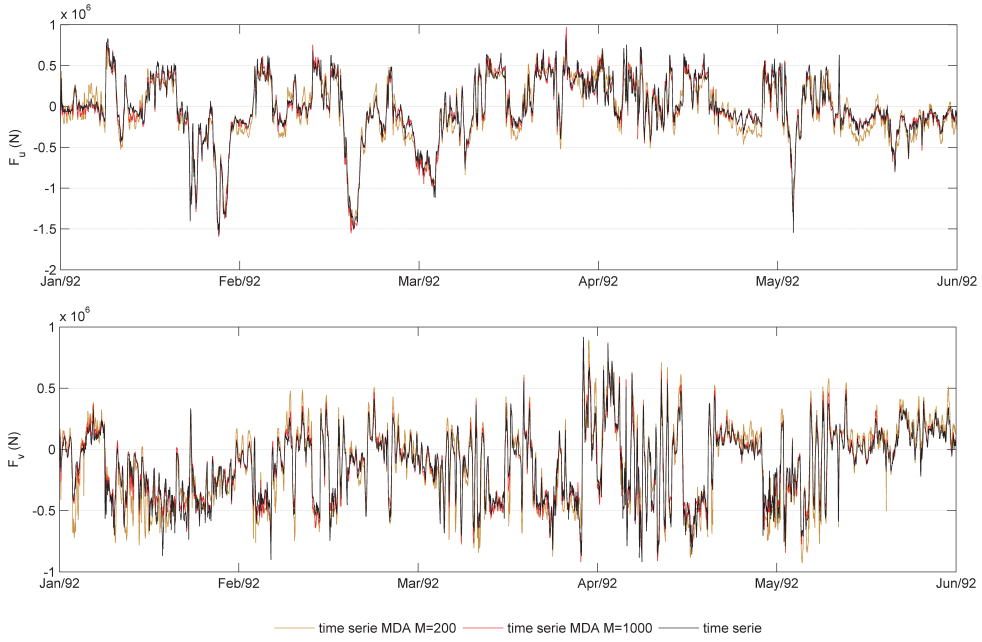

Figure 5.10: Time series calculated (in black) and the time series reconstructed considering $M = 200$ cases (in orange) and $M = 1000$ cases (in red) of parameters F_u (X component) and F_v (Y component).

	F_u					F_v				
MDA	75	200	500	750	1000	75	200	500	750	1000
$NRMSE$	0.0367	0.0276	0.0165	0.015	0.0138	0.0531	0.0375	0.0220	0.0183	0.0163
ρ	0.9389	0.9635	0.9865	0.9887	0.9905	0.8875	0.9381	0.9773	0.9840	0.9873

Table 5.1: The normalized root mean square error and the correlation coefficient of F_u (X component) and F_v (Y component)

9 Design Tools

The availability of long time series of environmental forces acting on a wind turbine could become a useful tool in offshore wind turbine design and would allow analysis of critical structural integrity situations representing an improvement in the design of offshore wind turbine farms. Moreover, this study might be understood from different points of view. The data used are the reconstructed series based on MDA= 1000 selected cases.

In terms of extreme situations, Figure 5.11 shows a Peak Over Threshold (POT) distribution of the environmental forces acting upon the wind turbine, with a threshold at the 99.5 percentile.

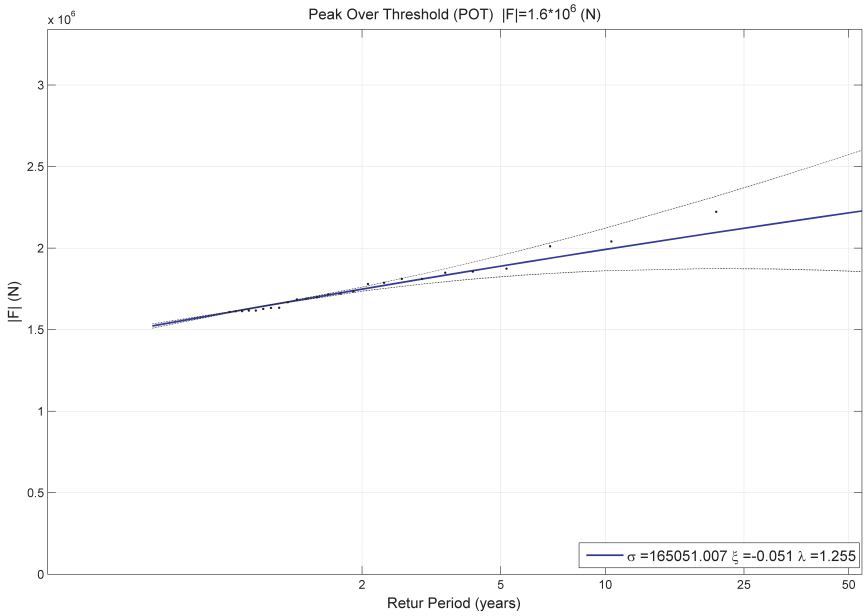

Figure 5.11: Peak over threshold (POT) distribution od the module of forces at the wind turbine (95% confidence intervals in dashed lines).

Figure 5.12 shows the rose directional distribution of probabilities for the wind (upper left rose), wave (upper central rose), currents (upper right rose) and total force (lower rose). However, Figure 5.13 shows wind, wave and current contributions only during 1992. Moreover, Figure 5.14 shows the spatial total force

distribution during 1999 (left side rose), 2000 (central rose) and 2001 (right side rose).

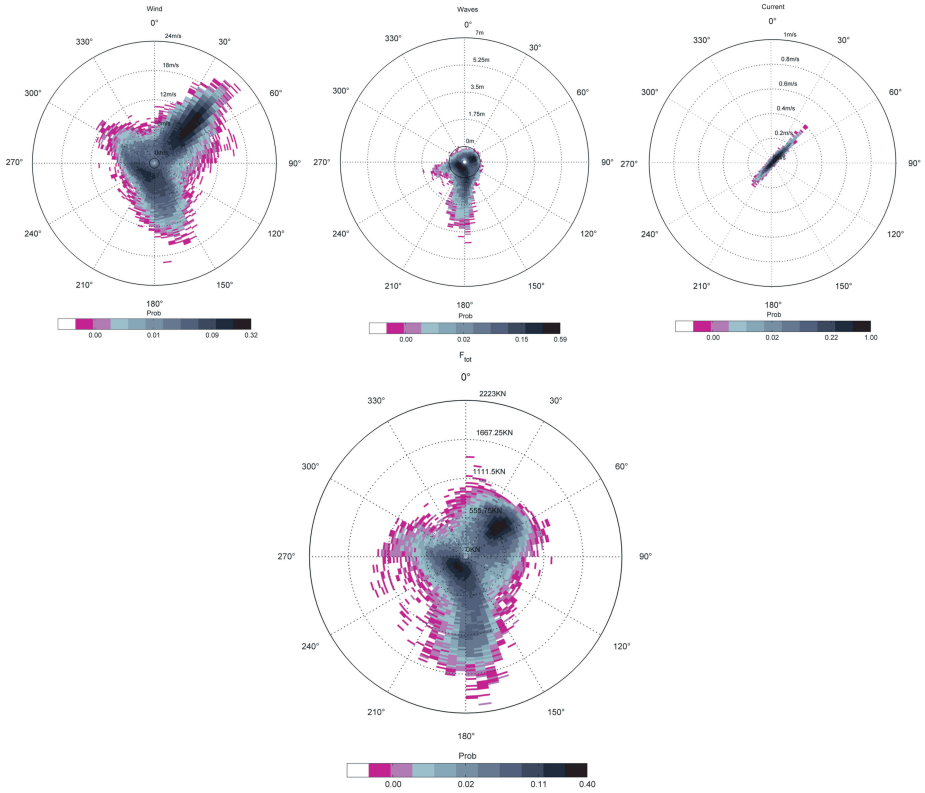

Figure 5.12: Roses of spatial distribution of wind (upper left rose), wave (upper central rose), current (upper right rose) and total force magnitude (lower rose).

The methodology proposed also allows the availability of long-term series for the different components of the environmental forces on the monopile and so may be made a spatial and/or temporal study depending on the components. Figure 5.13 shows the probabilistic roses of the three main components (waves, wind and currents) and the total force for a period of only 1 year (1992). It can be seen that in that year (1992), the worst states are due to waves, but that the most common state arises from the combination of waves and the main wind direction. Notice that the forces are misaligned, so that induced movements should be carefully studied to avoid fatigue problems. At this case study location, current forces are negligible compared with waves or wind, but at other location, they may become a significant

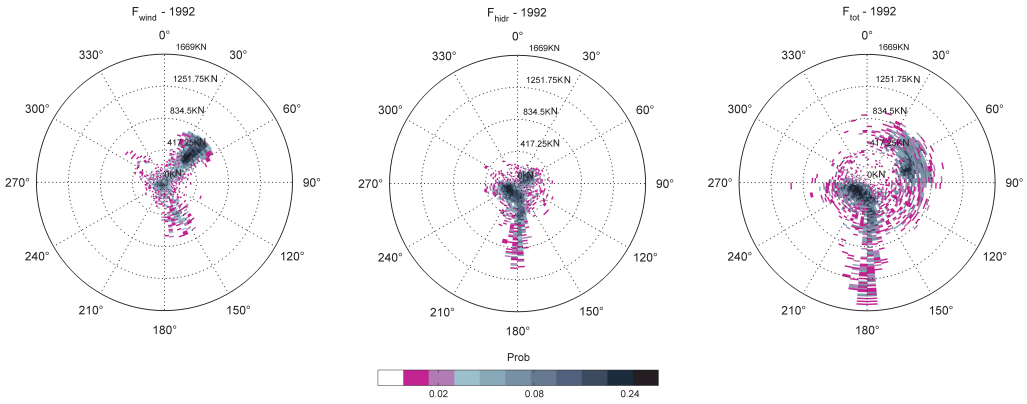

Figure 5.13: Probabilistic roses of the different components of forces acting on a wind turbine, 1992 (wind component: left panel, hydraulic component(wave–current):central panel and total force(sum of both):right panel).

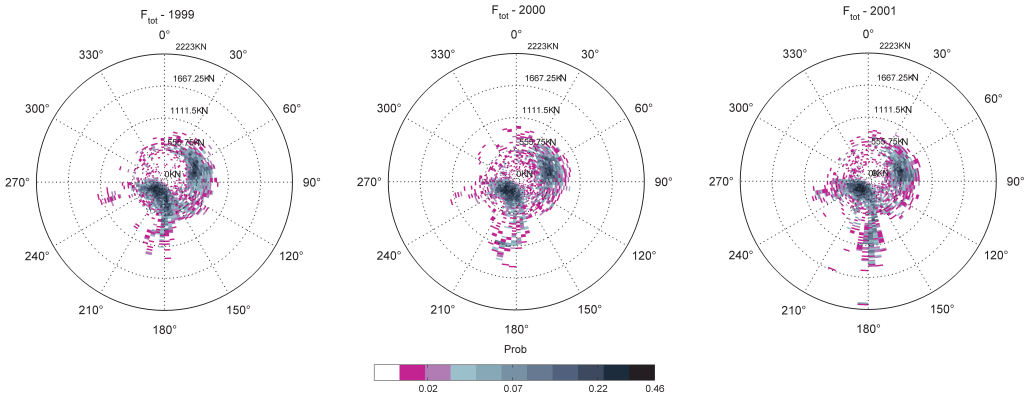

Figure 5.14: Probabilistic roses for the total force acting upon a wind turbine, 1999 (left panel), 2000 (central panel) and 2001 (right panel).

force. The inclusion of current forces in the proposed methodology makes it more versatile.

But not only the spatial distributions of force components are important when designing wind turbine farms. Temporary variations of the spatial distributions of total forces are also very important, and their study calls for long-term series. Figure 5.14 shows the spatial total forces distribution acting on the wind turbine over 3 consecutive years (1999 left side rose, 2000 central rose and 2001 right side

rose). The variation in the shape and magnitude of the cluster of dots is clear.

Figures 5.11-5.14 serve as examples of the analysis that can be made with the availability of long time series of forces and also the combination of their different components. Thus, the availability of long time load series for a wind turbine seems to be useful, taking account of some combinations of external agents (waves, wind and currents) that may not be included in the load cases proposed by the standards.

10 Conclusions

In this study, a hybrid methodology has been developed to obtain long time series of the environmental loads on an offshore wind turbine, using some simplifications that drastically reduce the computational effort. These loads can be determined by numerical models or semi-empirical formulations. The methodology is based on a selection of representative seawind states at the turbine location, calculation of the dynamic loads corresponding to these selected seawind states and a multidimensional RBF interpolation to reconstruct the long time series of dynamic loads.

The availability of long time series of the loads acting on an offshore wind turbine allows the analysis of critical stability situations, representing an improvement in their design. Nevertheless, this methodology can be also applied to reconstruct other parameters involved in offshore structure design and also allows the inclusion of more variables that may influence the process.

To analyse possible application of the proposed methodology to real cases, an application at a specific location is presented. Moreover, some examples of design tools extracted from the results obtained are described.

CHAPTER

6

Point-in-time and extreme-value probability simulation technique for engineering design

1 Abstract

Engineering design of structural elements entails the satisfaction of different requirements during each of the phases that the structure undergoes: construction, service life and dismantling. Those requirements are settled in form of limit states, each of them with an associated probability of failure. Depending on the consequences of each failure, the acceptable probability varies and also the denomination of the limit state: ultimate, damage, serviceability, or operating stop. This distinction between limit states forces engineers to: i) characterize both the point-in-time and extreme probability distributions of the random variables involved (agents), which are characterized independently, and ii) use the appropriate distribution for each limit state depending on the failure consequences. This chapter proposes a

Monte Carlo simulation technique, which allows the generation of possible outcomes for agents holding the following conditions: i) both the point-in-time and the extreme value distributions are appropriately reproduced within the simulation procedure, and ii) it maintains the temporal dependence structure of the stochastic process. In addition, a graphical representation of both distributions on a compatible scale is given, this graph clarifies the link between point-in-time and extreme regimes and helps quantifying the degree of accuracy of the simulation results. In addition, new insights for the development of First-Order-Reliability methods (FORM) combining point-in-time and extreme distributions simultaneously are provided. The method is illustrated through several simulation examples from well-known distributions, whereas its skill over real data is shown using the significant wave height data record from a buoy located on the Northern Spanish coast.

2 Introduction

Engineering structures undergo different phases during their lifetime: construction, service life and dismantling. During each of these phases, the structure must satisfy different requirements, which from the engineering design point of view, are defined as limit states. The objective of the design is to verify that the structure satisfies those project requirements in terms of acceptable failure rates and costs (see [100] and [139]).

Acceptable failure rates are established by codes and expert committees [7, 80, 98, 139] on the basis of the consequences of failure for each limit state, and trying to counter-balance safety and costs (direct, societal and environmental). Since the consequences of failure might be very different depending on the limit state considered, these limit conditions are classified in different categories: ultimate, damage, serviceability, or operating stop. The acceptable probabilities of failure for each category depends on the type of structure and environmental conditions, but in all cases it increases from the minimum acceptable probability of failure related to the ultimate limite state, up to the maximum acceptable probability associated with the operating stop limit state.

From the design perspective, these different probability thresholds encompass the consideration of different probability distributions for agents. Serviceability

or operating stop limit conditions depend on regular, central or mean values of those agents, whereas damage and ultimate limit states require extreme conditions, i.e. to pay attention to singular values. The statistical theory for dealing with mean values (point-in-time distribution) is different from the theory for extreme values [23, 24, 25, 32]. Traditionally, both problems are treated independently, which makes difficult to understand the link between point-in-time and extreme distributions and their implications from the practical point of view.

There are several attempts in the literature to incorporate both the point-in-time (central) and extreme information in the same probability distribution model (mixture models), see for instance, [9, 18, 55, 56, 149, 158, 161]. The common characteristic of these works is that all are applied to specific distributions, and the parameter estimation is fuzzy, not providing a general framework to deal with the problem. This work is intended to fill this niche.

As previously mentioned, safety of structures is the fundamental criterion for design, and once limit states and required probabilities are defined, engineering design must ensure satisfaction of the safety constraints. There are several methods to check the satisfaction of the safety requirements which can be classified in two main groups: (a) the classical safety factor approach, and (b) the probability based approach. The latter deals with probabilities of failure, which are difficult to deal with because (a) it requires the definition of the joint probability of all variables involved, and (b) the evaluation of the failure probability is not an easy task. The problem becomes even more difficult if several failure modes are analyzed, because the failure region is the union of the different failure mode regions, and regions defined as unions are difficult to work with because of their irregular and non-differentiable boundaries [106]. A method widely used by engineers to overcome these difficulties is Monte Carlo simulation technique. Once the probability distributions are defined, long records of the random variables involved may be sampled [11, 141, 142] and used to check whether the safety constraints are satisfied in terms of probabilities of failure. The simplicity on its implementation has increased the development of different methods for structural reliability analysis [26, 71, 146], such as directional simulation techniques [11, 46], importance sampling [44, 104, 105], or techniques which allows reproducing on multidimensional

settings, not only the marginal distributions but the temporal dependence of the stochastic processes involved as well [68, 120, 150].

The aim of this chapter is threefold: i) to develop a Monte Carlo simulation method for reproducing both the point-in-time (mean values) and extreme value distributions of random variables, while keeping the temporal dependence structure of the stochastic process involved, and valid for any kind of probability distribution function, ii) to present a graphical interpretation of simulation results to merge both distributions on a compatible scale, and iii) to provide new insights for the use of the point-in-time and extreme regimes simultaneously within First-Order-Reliability methods (FORM). The theoretical and practical material developed in this chapter is intended to support engineers on the design process and help understanding the relationship between both distributions, freeing engineers of deciding which conditions, average or extreme, must be used for each failure mode, because both conditions are considered into the proposed distribution.

3 Order Statistics and Extremes

Let consider the point-in-time probability density and distribution functions of a random variable X, i.e. $f_X(x)$ and $F_X(x)$. If we draw from this distribution a sample $x_1, x_2, \ldots, x_n$ of size n, and arrange it in increasing order $x_{1:n}, x_{2:n}, \ldots, x_n$, we could obtain the probability distribution of the rth element of this sequence, $X_{r:n}$, so-called the rth order statistic of a sample of size n. The first and last order statistics are the minimum $X_{1:n}$ and maximum $X_{n:n}$ respectively, and are called extremes [25, 57].

This maximum and minimum are very important for the design considering ultimate and damage limit states, and assuming that the point-in-time distribution of the variable of interest $F_X(x)$ (loads, significant wave height, strength, etc.) is known, the cumulative distribution functions of the maximum and minimum order statistics of a sample of size n are, respectively:

$$F_X^{\max}(x) = \left[F_X(x)^n\right], \tag{6.1}$$

and

$$F_X^{\min}(x) = 1 - \left[1 - F_X(x)\right]^n. \tag{6.2}$$

When n tends to infinity, distributions (6.1) and (6.2) are degenerate, only taking values equal to 0 or 1. For these cases linear transformations of x, consisting on location and scale changes, are looked for to avoid degeneracy. Note that when this is possible, $F_X(x)$ is considered to belong to the domain of attraction of the limit distribution.

[52] proved that there is only one parametric family for each of the limit distributions of maxima and minima, which correspond to the Generalized Extreme Value Distributions for maxima (GEV) and minima (GEVm), respectively. For instance, the cumulative distribution function (CDF) for maxima is given by:

$$F_X^{\max}(x;\mu,\psi,\xi) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\psi}\right)\right]_+^{-\frac{1}{\xi}}\right\}; \xi \neq 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\psi}\right)\right]\right\}; \xi = 0, \end{cases}$$

where μ, ψ, and ξ are the location, scale and shape parameter, $[a]_+ = \max(0, a)$, and the support is $x \leq \mu - \psi/\xi$, if $\xi < 0$, or $x \geq \mu - \psi/\xi$, if $\xi > 0$. The GEV family includes three distributions corresponding to the different types of tail behavior: Gumbel ($\xi = 0$) with a light tail decaying exponentially; Fréchet distribution ($\xi > 0$) with a heavy tail decaying polinomially; and Weibull ($\xi < 0$) with a bounded tail.

Note that this result has two very important practical implications:

1. The complexity to characterize the point-in-time regime $F_X(x)$ of a given random variable X, which allows using multiple distributions as possible candidates, contrasts with respect to the apparent simplicity to characterize the probability distributions for maxima and/or minima, which only requires the estimation of the three parameters μ, ψ, and ξ from the corresponding limit distribution family.

2. Since different point-in-time distributions may have the same domain of attraction, the best way to characterize the tail (upper/lower) of the distribution is using data belonging to the corresponding tail (maxima/minima) and estimate the parameters of the corresponding limit distribution.

From the practical point of view, the use of the GEV distribution for maxima is not appropriate in many cases because it uses small samples for the fitting process. For those cases, it is preferable to use the Pareto-Poisson model, which is valid for independent and identically distributed processes, or the Peaks Over Threshold (POT) method, suitable for dependent and identically distributed processes. The method presented is valid for those distributions or any other distribution for maxima.

Traditionally, engineers treat both point-in-time and extreme value distributions independently depending on the kind of limit state under consideration. The method proposed facilitates the engineering task as follows:

1. By presenting a graphical interpretation which makes easier to check if the right-tail of the distribution is appropriately reproduced or fitted by the point-in-time distribution, and then decide if an additional analysis of those extremes is required.

2. For those cases where both analysis are relevant and required, we present the methodology to link both distributions and use them simultaneously. Thus avoiding the decision to choose one or the other depending on the limit state considered.

4 Relationship between point-in-time and extreme value distributions: Graphical representation

From the practical point of view, it would be very useful for engineers to establish the relationship between the point-in-time and the extreme value distributions for random agents, or even to have a graphical visualization of this relationship, which would allow them to quantify the skill of any Monte Carlo simulation technique to deal with both central and extreme conditions at the same time.

The aim of this section is to present a graphical representation to accomplish the aforementioned task. Let assume an stochastic process X_t with associated sampling or occurrence frequency $f = 1/T_x$ (T_x is the sampling period, for instance 1 hour, 2 hours, etc.) and whose point-in-time distribution is $F^{\mathrm{PT}}(x)$. If we simulate samples of size n from the stochastic process X_t and calculate their maximum

values, this maximum is a random variable X_M with probability distribution function $F^{\mathrm{EV}}(x)$. Both distributions may be plotted on the same return period graph as follows:

1. Calculate the "equivalent return period" from the point-in-time distribution, i.e. $T^{\mathrm{PT}} = \dfrac{1}{1 - F^{\mathrm{PT}}(x)}$.
2. Plot T^{PT} versus x.
3. Calculate the "return period" from the extreme value distribution, i.e. $T^{\mathrm{EV}} = \dfrac{1}{1 - F^{\mathrm{EV}}(x)}$.
4. Plot the re-scaled return period using the sample size n, i.e. $T_r^{\mathrm{EV}} = n \cdot T^{\mathrm{EV}}$ versus x.

Figure 6.1 (a) shows the proposed graphical interpretation associated with an hourly stochastic process ($T_x = 1$ hour) without temporal correlation, and whose marginal (point-in-time) distribution corresponds to the standard normal ($X_t \sim N(0, 1^2)$). Dark gray line corresponds to (T^{PT}, x). We sample $n_y = 1000$ years of data and look for the annual maximum $x^{\max}$. Black dots correspond to $(T_r^{\mathrm{EV}}, x^{\max})$. The re-scaled return period is calculated as $T_r^{EV} = \frac{n}{1-\hat{F}^{\mathrm{EV}}(x)}$, where $\hat{F}^{\mathrm{EV}}(x_i) = \frac{i}{n_y+1};\ \forall i = 1, \ldots, n_y$ is the empirical annual maxima probability distribution for the sample, and $n = 8766$. Finally, the light gray line represents $(T_r^{\mathrm{EV}}, x^{\max})$, where T_{EV_r} has been calculated using the GEV fitted distribution to annual maxima. Note that both the point-in-time and the extreme regimes converge on the tail of interest, however, there are slight differences between the point-in-time and the maxima fitted distribution due to the simulation and fitting process uncertainty. These differences tend to zero as the sample size tends to infinity. Note that the true abscissas axis units in Figure 6.1 are hours, however, we have re-scaled the ticks to years to facilitate the interpretation.

Analogous results are shown in Figure 6.1 (b) for a gamma distributed stochastic process with scale and shape parameters $\theta = 5$ and $\kappa = 10$, respectively. Note that as in the previous case, both the point-in-time and extreme-value probability distributions converge on the tail of interest.

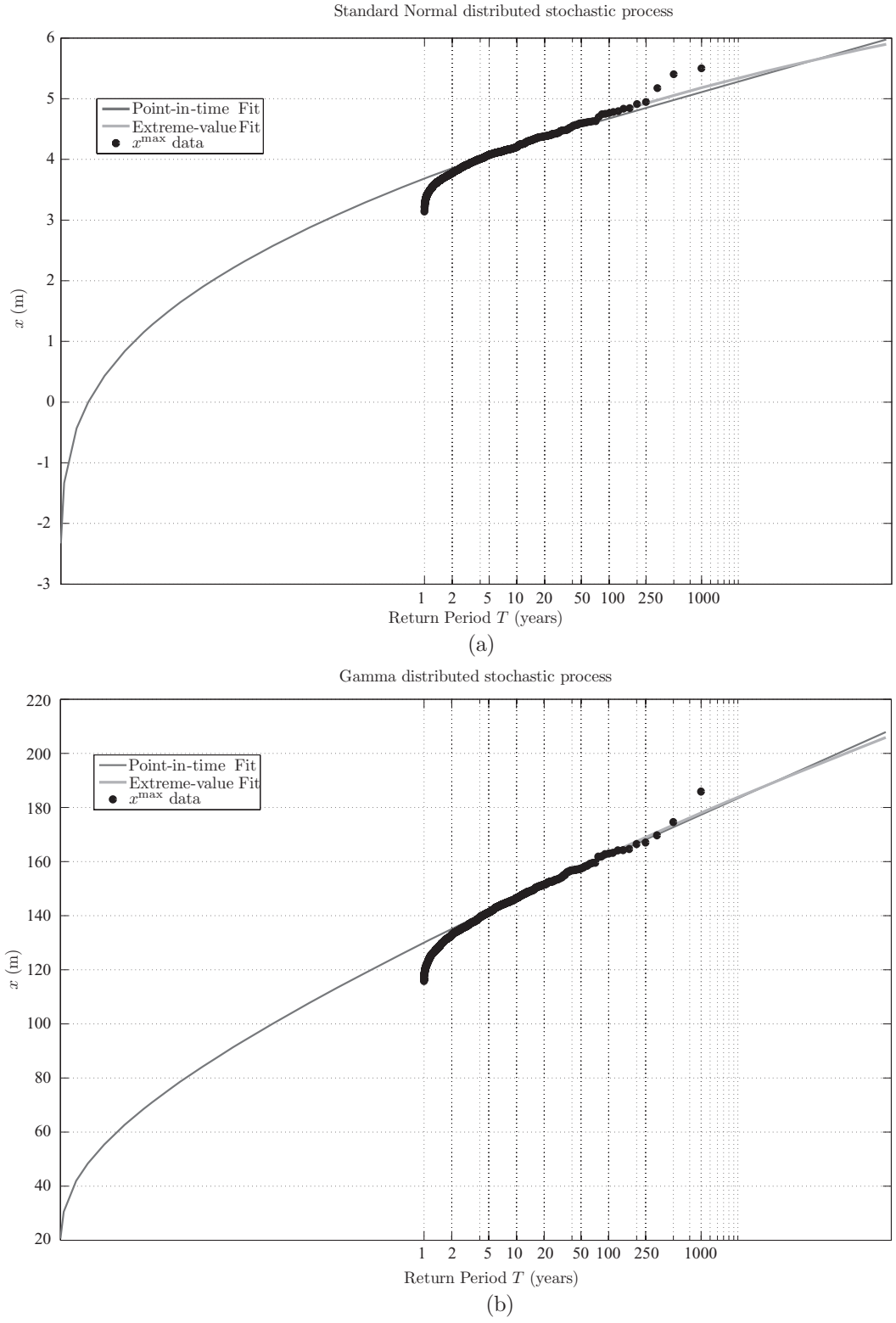

Figure 6.1: Graphical representation of the point-in-time and extreme regimes for: a) an standard normal and b) a gamma distributed stochastic processes.

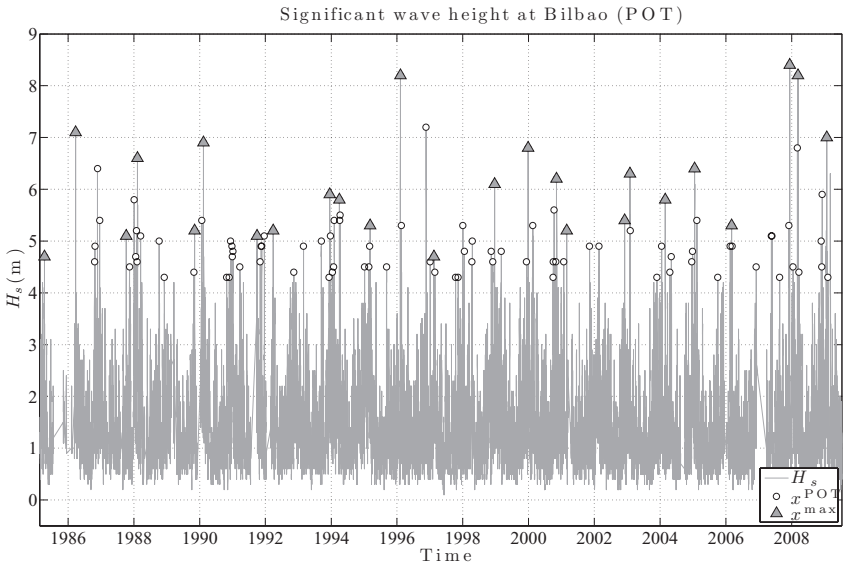

Figure 6.2: Hourly significant wave height record at Bilbao buoy from February 21, 1985 to July 13, 2009, annual maxima and peaks over the threshold $u = 4.2$.

These results are not surprising, since we are sampling from given point-in-time distributions, and thus the sampled data reproduce appropriately the tail of interest, especially if large samples are used. However, when dealing with real data sets, the point-in-time distribution does not usually fit appropriately the tail of interest. This is the case for the significant wave height instrumental record (gray line) associated with Bilbao buoy, shown in Figure 6.2. Their corresponding annual maxima (triangle dots) and peaks over the threshold $u = 4.2$ m (circle dots) are also shown. Note that the latter correspond to maximum values during independent storms. The independence assumption is considered assuming that the minimum distance in time between peaks must be 3 days. This data set consists of an hourly time series of significant wave height in meters from February 21, 1985 to July 13, 2009.

Significant wave height is a very important parameter for harbor design. Average conditions of significant wave height are relevant to analyze operating conditions for ships, whereas extreme significant wave heights are used for the stability design of protection structures, such as, vertical breakwaters, dikes, etc.. Thus the importance of characterizing both distributions.

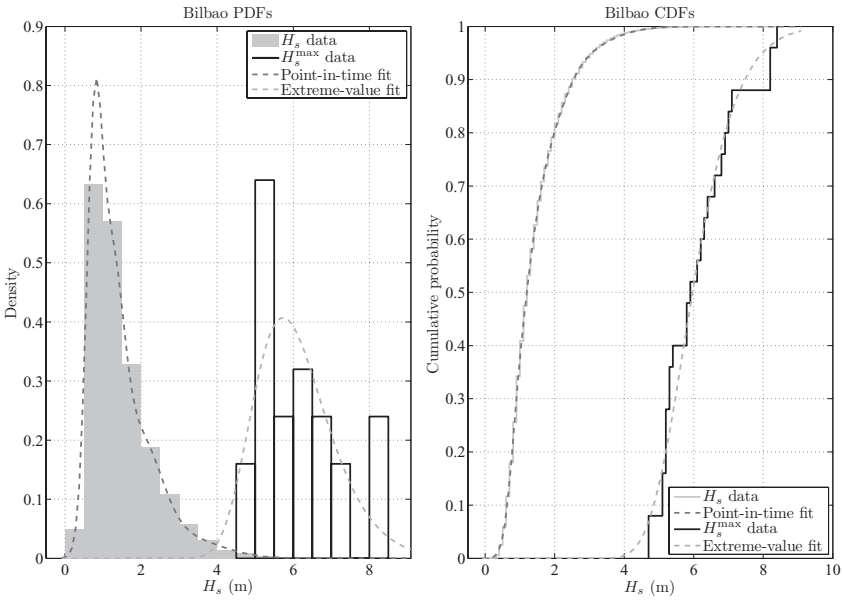

Figure 6.3: Graphical representation of the point-in-time and extreme (annual maxima) distributions for the significant wave height record at Bilbao buoy.

We fit both the significant wave height record and the corresponding peaks over the selected threshold to different parametric distributions: i) a Gaussian Mixture with 4 components for the point-in-time distribution, and ii) a POT model for the annual maxima (extreme-value) distribution, it is possible to plot i) the histograms, ii) the fitted densities, iii) the empirical cumulative and iv) fitted cumulative distributions (see Figure 6.3). Note that they all present very good fitting diagnostic plots. However, it is difficult to establish whether the fitted point-in-time distribution is capable of reproducing the tail of interest.

If data and fits from Bilbao buoy are plotted using the proposed graphical representation, results shown in Figure 6.4 are obtained. Note that this representation allows establishing the range of validity of the fitted point-in-time distribution, which starts distorting results above 4.8 meters of significant wave height approximately. The hourly probability of not exceeding this value within the year is 0.996. Above these quantile and probability thresholds, the point-in-time distribution is no longer valid. It can be observed that the extreme value fit allows reproducing appropriately the tail of the distribution, especially for long equivalent return periods.

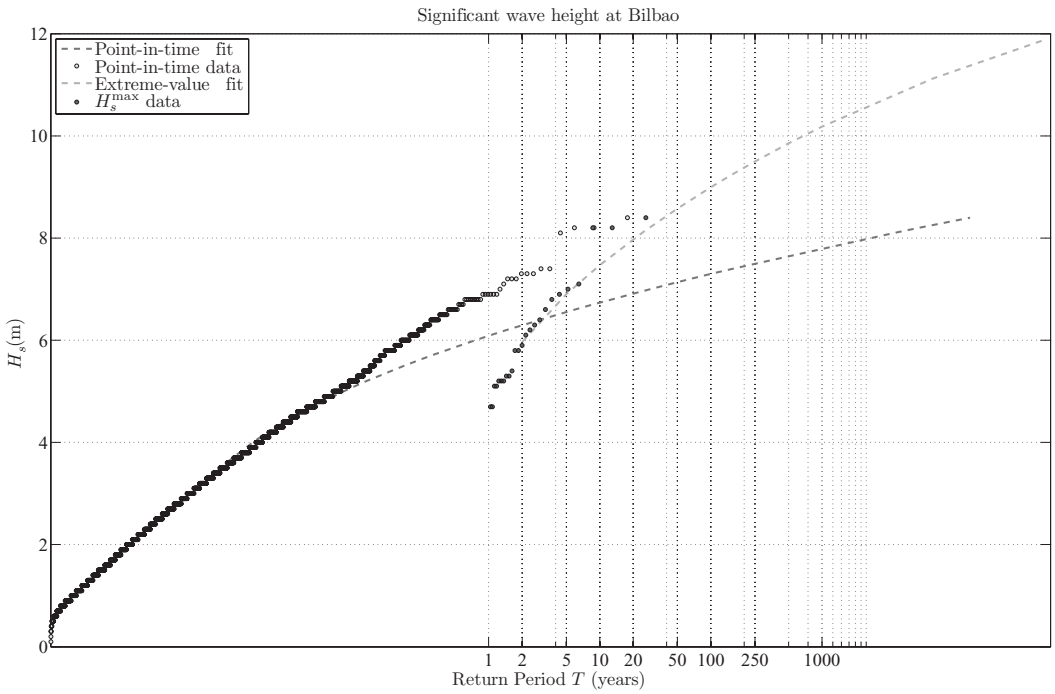

Figure 6.4: Graphical representation of the point-in-time and extreme distributions for the significant wave height record at Bilbao buoy using the proposed method.

These results confirm the appropriateness of using the graphical representation to help understanding the relationship between both the point-in-time and extreme regimes, posing a new challenge for LEVEL III reliability methods based on Monte Carlo simulation techniques: is it possible to simulate, from given point-in-time and extreme-value fitted distributions, samples reproducing both regimes simultaneously? The answer to this question is given in the next section.

5 Point-in-time and extreme-value simultaneous Monte Carlo simulation technique

Consider the stochastic process X_t, whose point-in-time and extreme-value probability distributions are $F^{\mathrm{PT}}(x)$ and $F^{\mathrm{EV}}(x)$, respectively. In Figure 6.5 the PDFs and CDFs of both distributions in case of maxima are plotted. The first important issue in order to reproduce both distributions is to select the threshold x_{lim},

this limit corresponds to the maximum value which is governed by the point-in-time distribution. From the practical point of view, it is established based on the proposed graphical representation, as shown in panel left-below from Figure 6.5, being the x-value whose associated return periods T^{PT} and T_r^{EV} are closer. This condition can be mathematically defined as:

$$\underset{x}{\text{Minimize}} \quad \left(T^{\mathrm{PT}} - T_r^{\mathrm{EV}}\right)^2, \tag{6.3}$$

which in case of dealing with maxima becomes:

$$\underset{x}{\text{Minimize}} \quad \left(\frac{1}{1 - F^{\mathrm{PT}}(x)} - \frac{n}{1 - F^{\mathrm{EV}}(x)}\right)^2. \tag{6.4}$$

Note that in case both regimes intersect, as it is shown in left-bottom panel from Figure 6.5, the optimal solution from problem (6.3) corresponds to zero, i.e. x_{lim} is the solution of the implicit equation $T^{\mathrm{PT}} = T_r^{\mathrm{EV}}$. Nevertheless, we advocate this approach to overcome the difficulties of solving the implicit equation for those cases where there is no solution (no intersection of regimes). In case of multiple solutions, we take the minimum solution if we are dealing with maxima.

The probability of not exceeding the maximum value x_{lim} within the point-in-time distribution is equal to $p_{\mathrm{lim}}^{\mathrm{PT}} = F^{\mathrm{PT}}(x_{\mathrm{lim}})$, thus the simulation technique uses $F^{\mathrm{PT}}(x)$ for probabilities lower than or equal to $p_{\mathrm{lim}}^{\mathrm{PT}}$ (which is equivalent to x lower than x_{lim}), and $F^{\mathrm{EV}}(x)$ otherwise. However, for the extreme regime the probability must be re-scaled considering:

1. The extreme distribution is related to the maximum of n elements from the point-in-time distribution.

2. There is a probability $p_{\mathrm{lim}}^{\mathrm{EV}} = F^{\mathrm{EV}}(x_{\mathrm{lim}})$ of not exceeding the x_{lim}-value within the extreme distribution, which is usually different from zero. This is the case shown in Figure 6.5. Thus, those values are not sampled again because they are already considered within the point-in-time distribution.

Finally, when dealing with maxima, and for given uniformly distributed random number u^{PT} representing a probability, the corresponding simulated value is obtained as follows:

$$x = \begin{cases} F^{\mathrm{PT}^{-1}}\left(u^{\mathrm{PT}}\right) & \text{if} \quad u^{\mathrm{PT}} \leq p_{\mathrm{lim}}^{\mathrm{PT}} \; (x \leq x_{\mathrm{lim}}) \\ F^{\mathrm{EV}^{-1}}\left(u^{\mathrm{EV}}\right) & \text{if} \quad u^{\mathrm{PT}} > p_{\mathrm{lim}}^{\mathrm{PT}} \; (x > x_{\mathrm{lim}}), \end{cases} \tag{6.5}$$

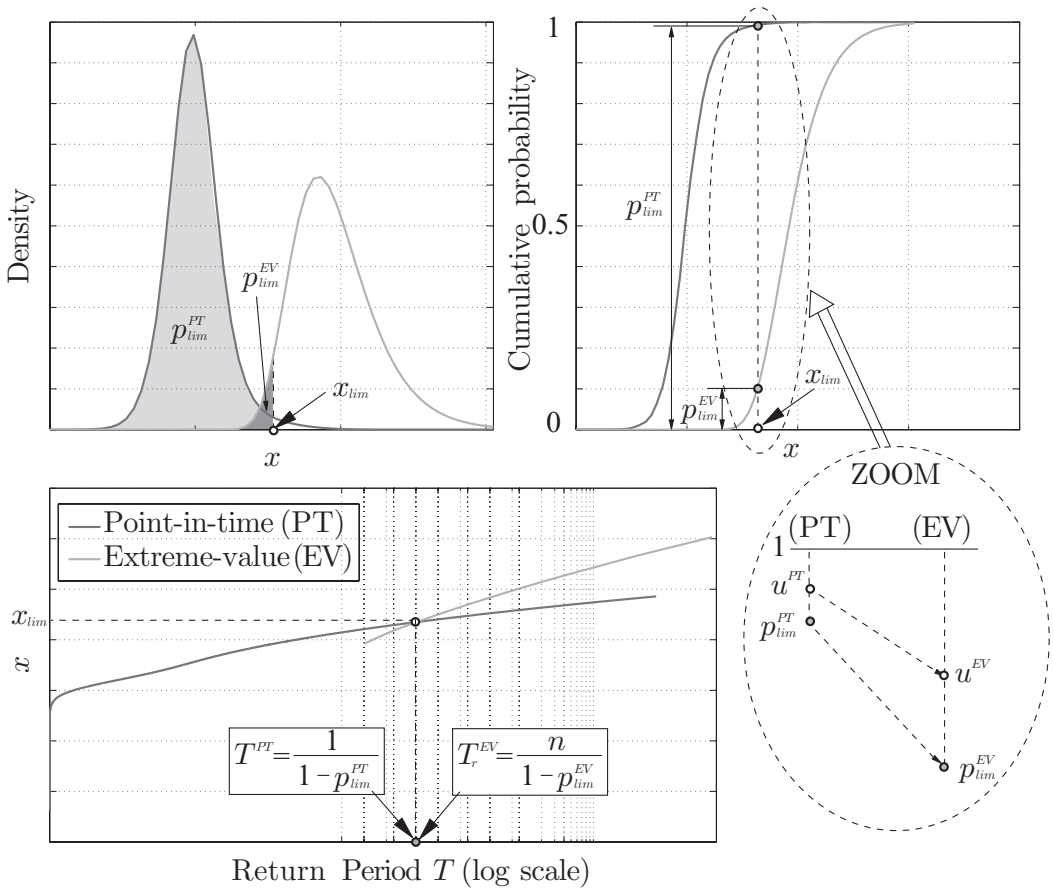

Figure 6.5: Graphical illustration of the simulation process.

where the re-scaled probability, is equal to:

$$u^{\mathrm{EV}} = p_{\mathrm{lim}}^{\mathrm{EV}} + \frac{u^{\mathrm{PT}} - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}). \tag{6.6}$$

The bottom-right panel of Figure 6.5 shows the graphical interpretation of the probability re-scaling, which constitutes a distorted zoom of the panel above. Note that expression (6.5) allows reproducing both the point-in-time and extreme-value distributions simultaneously.

To shown the functioning of the proposed simulation technique, 1000 years of hourly significant wave height data has been sampled using (6.5) and the fitted distributions at Bilbao buoy location. For this particular case $n = 8766$ corresponds to the mean number of hours per year used to evaluate the annual maxima.

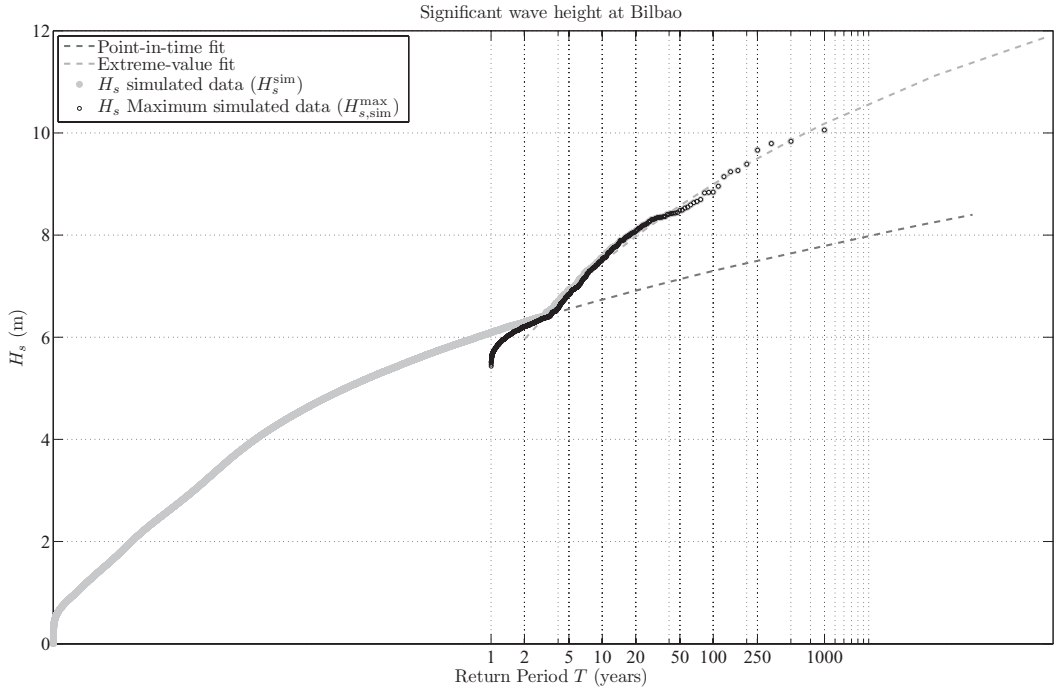

Figure 6.6: Graphical representation of the point-in-time and extreme simulation results for the significant wave height at Bilbao buoy location using the proposed method.

For the significant wave record, the solution of equation (6.4) is $x_{\mathrm{lim}} = 6.404$, and the associated probabilities are $p_{\mathrm{lim}}^{\mathrm{PT}} = 0.99996$ and $p_{\mathrm{lim}}^{\mathrm{EV}} = 0.66$. These values correspond to return periods $T_{\mathrm{lim}} \approx 25817$ hours and $T_{\mathrm{lim}} = 2.94$ years, respectively, which are equivalent.

Results from the simulation process are shown in Figure 6.6. Note that the sample fits appropriately the point-in-time distribution up to the probability related to $T_{\mathrm{lim}} = 2.94$ years return period, and finally the data fits to the extreme distribution for larger return periods. In addition, results related to the annual maxima are also shown. Note also the good fitting shown with respect to the theoretical extreme value distribution above $T_{\mathrm{lim}} = 2.94$ years return period.

These results confirm the validity and good performance of the proposed procedure.

6 Autocorrelation

The Monte Carlo simulation method provided in the previous section focusses on the marginal distribution associated with an stochastic process. However, it has been recognized by different authors the importance of the temporal correlation of any stochastic process, or even the cross correlation between different stochastic processes (see, for instance, [18, 29, 120]).

An appropriate description of any stochastic process requires recognizing its time dependent nature. For this particular case, the proposed method given in (6.5) is combined with results from [68, 120]. Basically, the method encompasses the following sequential procedure:

1. Using the point-in-time marginal distribution function, transform the time series of historical values $\boldsymbol{x}_t$ into a normalized Gaussian time series using the following transformation [123]:

$$\Phi(\boldsymbol{z}_t) = F^{\mathrm{PT}}(\boldsymbol{x}_t). \tag{6.7}$$

   Transformation (6.7) allows preserving the marginal distribution of the random variables involved.

2. Fitting of a time series model (e.g., an ARMA process) to the transformed historical values obtained in step 1 above. The obtained model allows taking into account temporal correlations.

The time series theory based on *autoregressive moving average* (ARMA) models allows incorporating the temporal structure. An ARMA(p, q) process $\boldsymbol{Z}$ is mathematically expressed as

$$z_t = \sum_{j=1}^{p} \phi_j z_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \tag{6.8}$$

where $\phi_i;\ i = 1, \ldots, p$ are the autoregressive parameters, and $\theta_j;\ j = 1, \ldots, q$ are the moving average parameters. The term ε_t stands for an uncorrelated normal stochastic process with mean zero and variance σ_ε^2, and it is also uncorrelated with $z_{t-1}, z_{t-2}, \ldots, z_{t-p}$. This process is so-called *white noise*, *innovation term*, or *error term*.

Observe in (6.8) that z_t boils down to a linear combination of white noises, and as such, the marginal distribution associated with the stochastic process $\boldsymbol{Z}$ is necessarily normal, which is in accordance with the first assumption (6.7).

Note that a stationary process is assumed. In case of dealing with seasonal behaviors, which could wreck the stationarity, non-stationary probability distributions could be used instead [107, 108, 110, 116] to easily overcome this difficulty.

It is important to point out that only the data belonging to the point-in-time distribution is used to characterize the autocorrelation structure of the stochastic process, because the extreme data is by definition independent, and has no information about autocorrelations. Note also that using ARMA models, only second-moment properties are preserved by the autocorrelation, and this might not provide a complete description in the case of a non-Gaussian process. That is the reason why non-gaussian processes are transformed into gaussian processes using (6.7). References [68, 120] proved that this approach reproduces autocorrelations in the original domain with a high degree of accuracy.

6.1 The algorithm

Once the parameters of the ARMA model are estimated from the transformed time series, it is very simple to incorporate the autocorrelation structure to the final series. The overall method consists of the following sequential procedure:

- *Step 1:* Estimate the parameters of the probability distributions that best fits both the point-in-time and the extreme-value regimes. This is done using the available historical data.

- *Step 2:* Apply transformation (6.7) to the historical time series using the point-in-time marginal cumulative distribution function. This way, a *transformed* series is obtained with an associated standard normal marginal distribution.

- *Step 3:* Adjust a univariate ARMA model to the corresponding *transformed* series (obtained in Step 2 above). The fitting process to be performed in this step is well known (see, e.g., [128]) and yields uncorrelated normal residuals (historical errors) with zero mean and constant variance σ_{ε}^2.

- *Step 4:* Simulate independent normal errors with zero mean and variance σ_ε^2, i.e. $\boldsymbol{\varepsilon}_t^{\text{sim}}$.
- *Step 5:* Introduce the simulated error series $\boldsymbol{\varepsilon}_t^{\text{sim}}$ into the ARMA model fitted in Step 3, obtaining $\boldsymbol{z}_t^{\text{sim}}$.
- *Step 6:* Calculate the time series $\boldsymbol{u}_t^{\text{sim}} = \Phi(\boldsymbol{z}_t^{\text{sim}})$, which is uniformly distributed.
- *Step 7:* In this step, the inverse transformation (6.5) is applied to this serie in order to enforce the actual marginal distribution that has been estimated in Step 1.

Note that the method proposed has the following advantages with respect to existing Monte Carlo simulation methods:

1. It reproduces the autocorrelation function as in [21] or [22].
2. It preserves the statistical properties of the stochastic process in terms of the marginal distribution, reproducing appropriately not only the central part of the distribution (point-in-time) but also the right-tail (extremes).

In addition, as proposed in [120], the method could be easily extended to simulate different stochastic processes at the same time. This would allow replicating the main cross-correlations coefficients characterizing those stochastic processes, and not just the contemporaneous. However, since it is not clear how the cross correlation in the point-in-time and extreme distributions behaves, this is a subject for further research.

6.2 Illustrative example

To show the functioning of the proposed algorithm, the following ARMA process $(1,1)$ with parameters $\phi_1 = -0.8$, $\theta_1 = 0.3$, and variance $\sigma_\varepsilon^2 = 1$ is considered. According to [14], the variance of the process is $\sigma_Z^2 = \dfrac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma_\varepsilon^2 \approx 4.3611$ ($\sigma_Z \approx 2.0883$). One hundred years of hourly data $\boldsymbol{x}^{\text{sim}}$ ($n = 100 \times 24 \times 365.25 = 876600$) is sampled from this stochastic process. This sample is considered as our

initial data set. The idea is to use this sample and the algorithm presented in the previous subsection to generate one thousand years ($n_y = 1000$) of hourly data $\boldsymbol{y}^{\text{sim}}$ considering the autocorrelation, the fitted point-in-time and extreme-value distributions, and compare results with the reference values from the original stochastic process and the initial sample data $\boldsymbol{x}^{\text{sim}}$.

The results of the application of the algorithm are the following:

- *Step 1:* The point-in-time distribution of $\boldsymbol{x}^{\text{sim}}$ is normal, and its estimated parameters and 95% confidence bands are:

$$\begin{array}{rcl} \hat{\mu} & = & 2.7840 \times 10^{-4}\ (-0.0041, 0.0047) \\ \hat{\sigma} & = & 2.0910\ (2.0879, 2.0941), \end{array} \tag{6.9}$$

  which both contain the true values 0 and 2.0883, respectively. For the extreme value distribution, the annual maxima from the sample $\boldsymbol{x}^{\text{sim}}$, i.e. $\boldsymbol{x}^{\max}_{\text{sim}}$, follows a GEV distribution with estimated parameters and 95% confidence bounds:

$$\begin{array}{rcl} \hat{\mu}_e & = & 7.7050\ (7.7007, 7.7093) \\ \hat{\psi} & = & 0.6020\ (0.5989, 0.6051) \\ \hat{\xi} & = & -0.0178\ (-0.0227, -0.0129). \end{array} \tag{6.10}$$

- *Step 2:* Apply transformation (6.7) to the historical time series (x^{sim}) using the normally distributed point-in-time marginal distribution.

- *Step 3:* Adjust an univariate ARMA model to the corresponding *transformed* series $\boldsymbol{z}^{\text{sim}}$, obtaining the following parameter estimates: $\hat{\phi} = -0.8011$ and $\hat{\theta} = 0.2984$. The residuals standard deviation is $\hat{\sigma}_\varepsilon = 0.4781$.

Figure 6.7 shows the proposed graphical interpretation applied to the sample data $\boldsymbol{x}^{\text{sim}}$ and $\boldsymbol{x}^{\max}_{\text{sim}}$. Dark gray line corresponds to (T^{PT}, x) for the point-in-time fitted distribution. Light gray line corresponds to (T_r^{EV}, x) associated with the GEV fitted distribution for annual maxima. Note that both fitted distributions differ at the right tail of the distribution (see the corresponding zoom in the figure, where data has been remove to ease visualization), which is usually the case when fitting real data. Light gray circle dots correspond to the sample data values, and black dots are related to sample annual maxima.

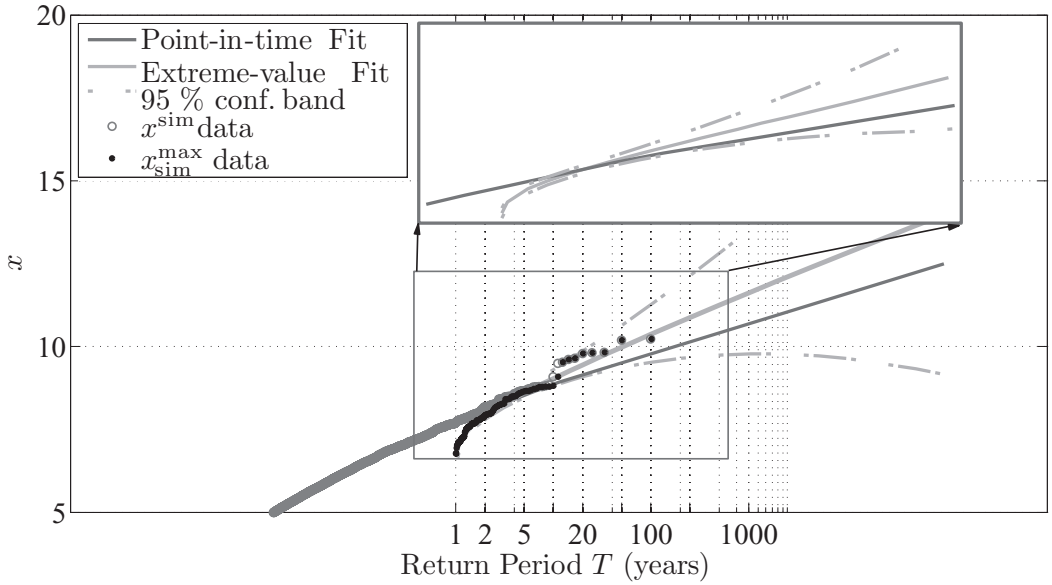

Figure 6.7: Graphical representation of the point-in-time and extreme regimes for the illustrative autocorrelated normal stochastic process.

- *Step 4:* In order to obtain $n_y = 1000$ years of hourly data, $n = 1000 \times 365.25 \times 24$ independent normal errors are sampled with standard deviation $\hat{\sigma}_\varepsilon = 0.4781$, i.e. $\boldsymbol{\varepsilon}_y^{\mathrm{sim}}$.

- *Step 5:* Introduce the sampled error time series into the ARMA model fitted in Step 3, obtaining $\boldsymbol{z}_y^{\mathrm{sim}}$.

- *Step 6:* Calculate the uniformly distributed time series of probabilities $\boldsymbol{u}^{\mathrm{sim}}$.

- *Step 7:* In this step, the inverse transformation (6.5) is applied to get $\boldsymbol{y}^{\mathrm{sim}}$. Note that for this particular example, the solution of equation (6.4) is $x_{\mathrm{lim}} = 8.3396$, and the associated probabilities are $p_{\mathrm{lim}}^{\mathrm{PT}} = 0.9999667$ and $p_{\mathrm{lim}}^{\mathrm{EV}} = 0.7082156$. These values correspond to return periods $T_{\mathrm{lim}} = 30042.7$ hours and $T_{\mathrm{lim}} = 3.427$ years, respectively, which are equivalent.

The graphical illustration of the 1000 years simulated sample is given in Figure 6.8. Note that it shows the same results as Figure 6.7 but replacing the sample data $\boldsymbol{x}^{\mathrm{sim}}$ and $\boldsymbol{x}^{\max}_{\mathrm{sim}}$ used to fit the distributions, by the 1000 years simulated samples

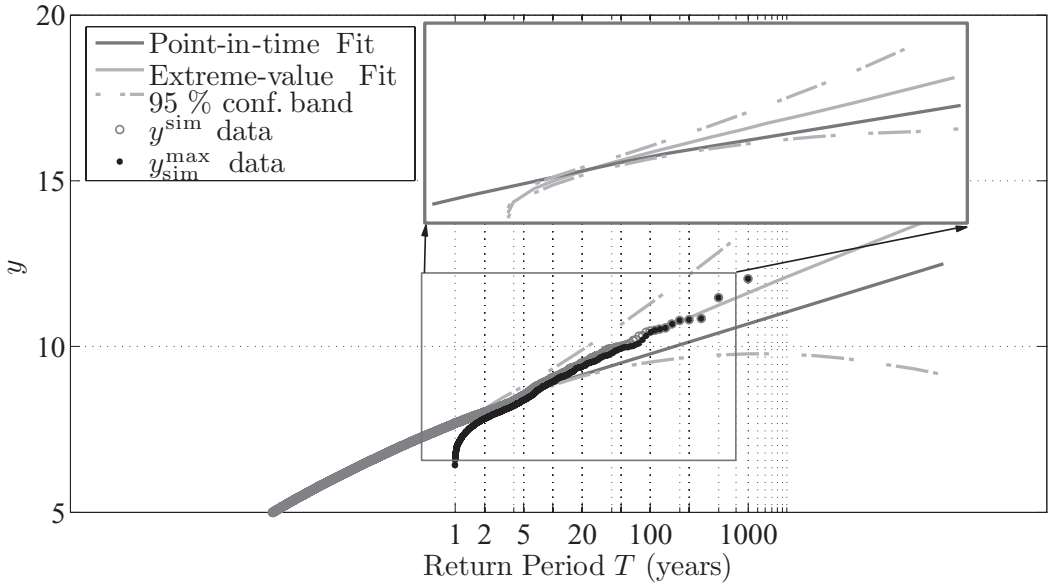

Figure 6.8: Graphical representation of the point-in-time and extreme regimes for the illustrative autocorrelated normal stochastic process.

$\boldsymbol{y}^{\mathrm{sim}}$ and $\boldsymbol{y}^{\max}_{\mathrm{sim}}$ using the proposed procedure. Note the accuracy of the method to reproduce both the point-in-time distribution up to $x_{\mathrm{lim}} = 8.3396$, where the simulated sample starts following the extreme-value distribution. In addition, several tests have been performed to check simulation results :

1. For the point-in-time distribution, a two-sample Kolmogorov-Smirnov test with 0.05 significance level is performed to compare the distributions of the initial sample data ($\boldsymbol{x}^{\mathrm{sim}}$) and the simulation results ($\boldsymbol{y}^{\mathrm{sim}}$). Note that the p-value obtained is 0.5966, which is higher than the significance level, i.e. the null hypothesis that both samples come from the same continuous distribution is accepted.

2. Analogously, the two-sample Kolmogorov-Smirnov test is applied to compare the samples related to annual maxima, i.e. $\boldsymbol{x}^{\max}_{\mathrm{sim}}$ versus $\boldsymbol{y}^{\max}_{\mathrm{sim}}$. Note that the p-value obtained is 0.1183, so that the null hypothesis that both samples come from the same extreme-value continuous distribution is accepted.

3. Finally, an ARMA model is fitted to the simulated sample, obtaining the

following parameter estimates: $\hat{\phi} = -0.8002$ and $\hat{\theta} = 0.2998$. The corresponding residuals standard deviation is $\hat{\sigma}_{\varepsilon} = 1.00015$, which almost coincide with the one from the initial ARMA process.

These results confirm the appropriate performance of the proposed procedure to reproduce i) the point-in-time and extreme-value distributions and ii) the temporal dependence structure of any stochastic process.

7 New insights into structural reliability methods

Besides providing a new Monte Carlo simulation method for dealing with point-in-time and extreme-value distributions, new insights about how to incorporate this methodology within alternative reliability analysis methods, such as *First Order Reliability Methods* (FORM), are also given. Note that we assume that the reader is familiar with LEVEL III methods ([43, 54, 74, 79, 134]) for evaluating the reliability index associated with any mode of failure:

$$\beta = \underset{\boldsymbol{z}}{\text{Minimum}} \quad \sqrt{\sum_{\forall i} z_i^2} \tag{6.11}$$

subject to

$$g(\boldsymbol{x}, \boldsymbol{\eta}) = 0, \tag{6.12}$$

$$\mathbf{T}(\boldsymbol{x}, \boldsymbol{\eta}) = \boldsymbol{z}, \tag{6.13}$$

where $g(\boldsymbol{x}, \boldsymbol{\eta}) = 0$ is the failure condition, and $\mathbf{T}(\boldsymbol{x}, \boldsymbol{\eta})$ is the transformation ([140]) giving the values of the standard and independent normal variables $\boldsymbol{z}$ as a function of the values of the random $\boldsymbol{x}$ and design $\boldsymbol{\eta}$ variables. The probability of failure p_f is related to the reliability index by the approximate relation $p_f = \Phi(-\beta)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

The key issue when dealing with structural risk problems where the point-in-time and extreme-value distributions may coexist, is to decide which one is more relevant for the corresponding limit state. The Monte Carlo method proposed in this chapter deals with the simulation process giving more importance to the point-in-time probability distribution, and it uses the re-scaled extreme-value regime to

improve accuracy in the right tail of the distribution. According to (6.5)-(6.6) and considering $u = \Phi(z^{\mathrm{PT}})$, the Rosenblatt transformation (6.13) becomes:

$$\begin{aligned} \Phi(z^{\mathrm{PT}}) &= F^{\mathrm{PT}}(x) \quad \text{if} \quad x \leq x_{\mathrm{lim}} \text{ or } z^{\mathrm{PT}} \leq z_{\mathrm{lim}} \\ p_{\mathrm{lim}}^{\mathrm{EV}} + \frac{\Phi(z^{\mathrm{PT}}) - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}) &= F^{\mathrm{EV}}(x) \quad \text{if} \quad x > x_{\mathrm{lim}} \text{ or } z^{\mathrm{PT}} > z_{\mathrm{lim}}, \end{aligned} \tag{6.14}$$

where $z_{\mathrm{lim}} = \Phi^{-1}\left(p_{\mathrm{lim}}^{\mathrm{PT}}\right)$. It is important to point out that transformation (6.14) takes into account the point-in-time distribution, but improving accuracy on the upper tail by using the re-scaled extreme-value distribution. Probabilities of failure obtained from this approach are related to the point-in-time frequency sampling, i.e. hours.

Alternatively, if only the extreme-value distribution is considered, transformation (6.13) becomes:

$$\Phi(z^{\mathrm{EV}}) = F^{\mathrm{EV}}(x)\,. \tag{6.15}$$

In this case, probabilities of failure are associated with the extreme-value frequency sampling, i.e. years.

From the practical point of view, we advocate the use of transformation (6.14) and consider probabilities related to the point-in-time frequency sampling, because it allows the consideration of any kind of limit state equation. However, it is important to define the maximum probabilities of failure for each failure mode in terms of the point-in-time frequency sampling. For instance, if an inner harbor must be designed so that ships might not maneuver during no more than 1000 hours per year, then the acceptable probability of failure must be equal to $p_f = 1000/(365.25 \times 24)$. Besides, if the off-shore breakwater of the same harbor must be designed to fail on average once every 25 years, the acceptable probability of failure must be equal to $p_f = 1/(25 \times 365.25 \times 24)$. Considering those probability values, transformation (6.14) may be used for both operating and ultimate limit states without any further consideration.

8 Case study: Environmental conditions for offshore wind turbine design

8.1 IEC61400-3 Standards for off-shore wind turbine design

To show the importance of considering both the point-in-time and right tail distributions in engineering design, herein we present an example from the IEC61400-3 Standards, which establishes the set of design requirements made to ensure that off-shore wind turbines are appropriately engineered against damage from hazards. This code divide external marine conditions related to agents into normal and extreme categories.

In particular, and for the case of waves, it proposes the consideration of severe sea states, which shall be considered in combination with normal wind conditions for calculation of the ultimate loading of an offshore wind turbine during power production. The model should associate a severe sea state with each wind speed in the range corresponding to power production. The significant wave height for each severe sea state shall be determined by extrapolation of appropriate site-specific met-ocean data such that the combination of the significant wave height and the wind speed has a recurrence period of 50 years. For all wind speeds, the unconditional extreme significant wave height with a recurrence period of 50 years may be used as a conservative value for H_s.

It is recommended by this guide to extrapolate met-ocean data using the so-called Inverse First Order Reliability Method (IFORM). This method produces an environmental contour defining, in a certain sense, 50-year recurrence period combinations of mean wind speeds, v, and significant wave heights, H_s. A common way to construct this transformation is to apply the so called Rosenblatt ([140]) transformation as follows:

$$\begin{array}{rcl} \Phi(z_1) & = & F_V\left(v\right) \\ \Phi(z_2) & = & F_{H_s|V}\left(H_s\right), \end{array} \tag{6.16}$$

where $F_V(v)$ is the marginal distribution of mean wind speed, and $F_{H_s|V}\left(H_s\right)$ is the conditional distribution of significant wave heights for given values of the mean wind speed. Using First Order Reliability Methods [45, 74] the points satisfying the equation $z_1^2 + z_2^2 = \beta^2$, is transformed into a curve in the $v - H_s$ plane, which

constitutes the environmental contour. β is the reliability index, whose required value is obtained from the following equation:

$$\Phi(\beta) = 1 - \frac{1}{N}, \tag{6.17}$$

where N is the number of independent sea states in 50 years.

In our particular case, and according to the IEC 61400-3, for a sea state of 1 hour $N = 50 \times 365 \times 24$ and the required reliability index from expression (6.17) is equal to $\beta \approx 4.6$.

8.2 Application of the evaluation of 50-year recurrence period environmental contours

Let consider a specific location in the Northern Spanish coast as possible candidate for an off-shore wind farm. We have at our disposal two times series of hourly significant wave heights and hourly mean wind speeds at 10 meters height. Both data sets come from reanalysis data bases GOW ([136]), DOW ([21]) and SeaWind [109], respectively, calibrated using instrumental data (see [**?**], [114], [117]).

First of all, marginal and conditional distributions of H_s and V given in expression (6.16) must be defined. The best diagnostic fit for the wind speed data corresponds to the Generalized Extreme Value (GEV) distribution. The maximum likelihood estimates are $\hat{\mu} = 6.0019$, $\hat{\sigma} = 3.5812$ and $\hat{\psi} = 0.0236$ for the location, scale and shape parameters, respectively. The histogram and fitted probability density functions shown in Figure 6.9 (upper panel) apparently present good fitting diagnostics, however, if we analyze in detail the right tail in terms of equivalent return periods (lower panel of Figure 6.9), the GEV distribution does not appropriately reproduce extreme winds (gray dashed line) with respect to data. This problem is solved by fitting the Pareto distribution above threshold $V_{\lim} = 20$ m/s. The maximum likelihood estimates for the Pareto distribution are $\hat{\theta} = 3.1004$ and $\hat{\epsilon} = -0.1467$ for the scale and shape parameters, respectively. Note that the tail behavior is completely different for the GEV and Pareto fits, while the one related to GEV defines a heavy tail (Frechet), the one associated with Pareto exhibits a bounded tail (Weibull).

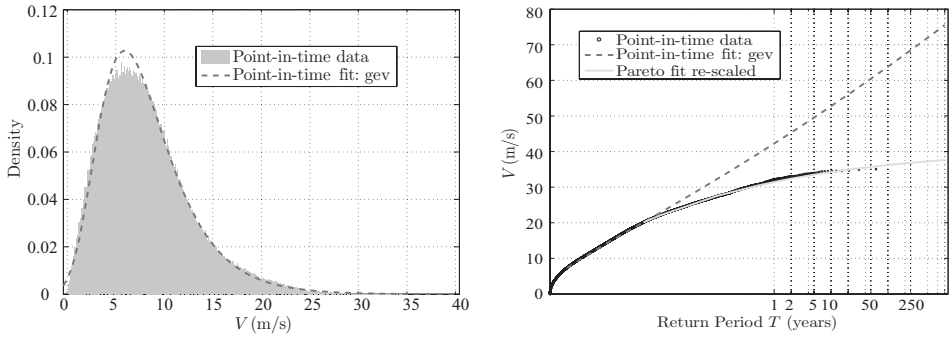

Figure 6.9: Histogram and fitted probability density function of wind speed V, and graphical representation in terms of equivalent return periods using the method proposed by Mínguez, Guanche and Méndez (2012).

The conditional distribution of significant wave height for given values of the wind speed has been fitted using 27 different GEV fits. Each data set $\Omega_i; i = 1, \ldots, 27$ is conformed choosing H_{s_i}-values so that their corresponding wind speeds V_i hold the following conditions: $j - 1 \leq V_i < j;\ \forall j = 1, \ldots, 26$ and $V_i > j - 1;\ j = 27$. Figure 6.10 (upper panel) shows the histograms and fitted probability density functions for each significant wave height data set. Note that they present good fitting diagnostics for the bulk of data.

However, if we take a closer look at the right tail of the distributions (Figure 6.10, lower panel), it is clear that the GEV distribution does not appropriately reproduce extreme significant wave heights (gray dashed line) with respect to data for medium-low values of wind speed. Analogously to the previous case, this problem is solved by fitting the Pareto distribution above the threshold associated with the 95% percentile. Lower panel of Figure 6.10 shows how the Pareto fit reproduces the right tail of the distribution. Note that the GEV distribution is not valid for significant wave height values associated with wind speeds below 16m/s. Above this threshold value, the GEV may be considered appropriate although the use Pareto fit on the tail is more convenient.

To get an smooth transition for the GEV parameters of the probability density function $H_s|V$, the location, scale and shape parameters are fitted to a third order

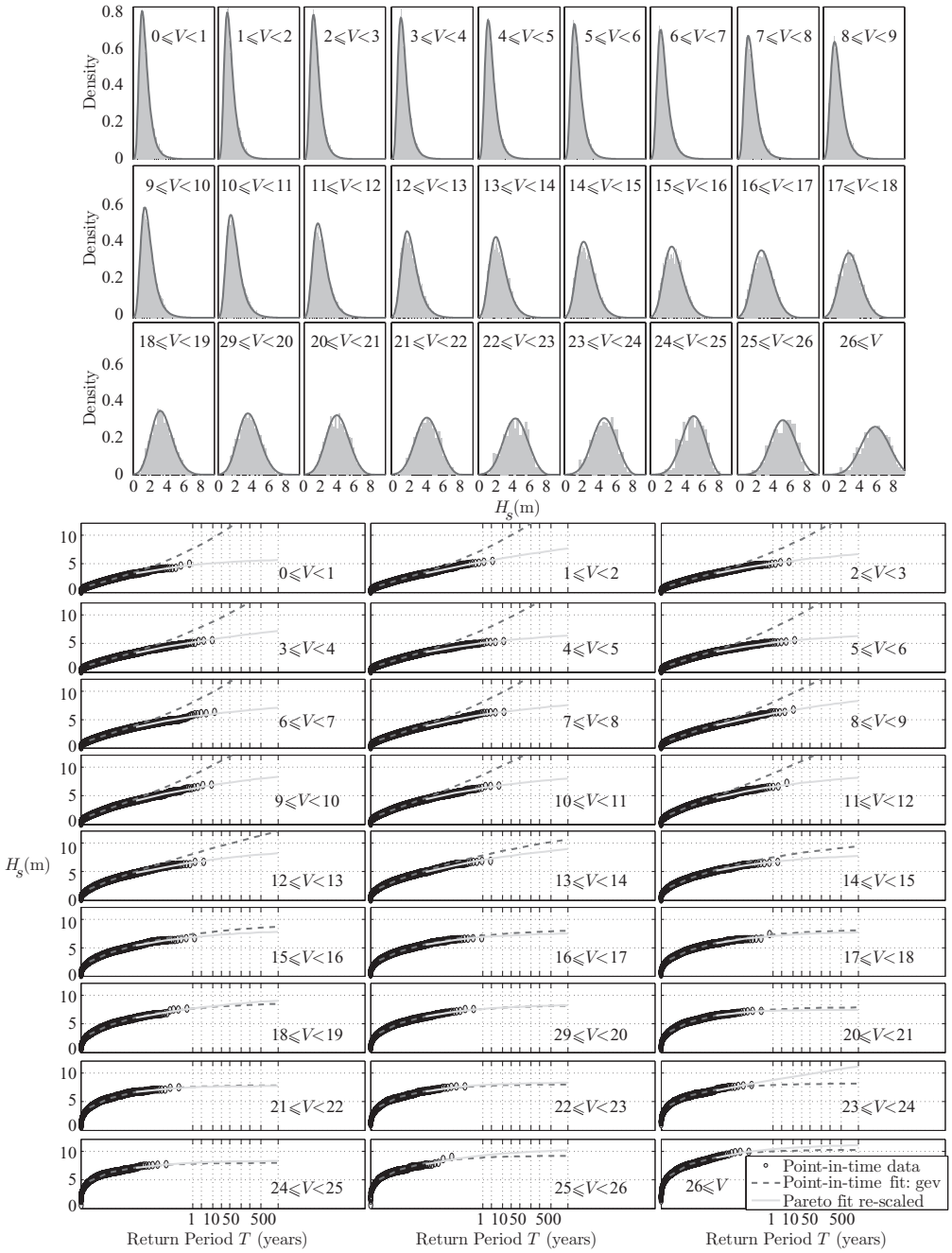

Figure 6.10: Histograms and fitted probability density functions of significant wave height H_s for given values of the wind speed V.

	$\hat{p}_1$ ($\times 10^{-5}$)	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$
$\mu(V)$	-26.150	0.01603	-0.1179	1.191
$\sigma(V)$	-7.240	0.003438	-0.004429	0.4573
$\psi(V)$	6.429	-0.003046	0.02069	0.05389
$\theta(V)$	4.362	-0.002161	0.03037	0.4693
$\epsilon(V)$	3.545	-0.001666	0.01482	-0.1192
$H_s^{\lim}(V)$	-22.120	0.01277	-0.01653	2.568

Table 6.1: Optimal parameter estimates of the regression models given in (6.18) and (6.19).

polynomial:

$$\begin{aligned} \mu(V) &= p_1^{\mu} V^3 + p_2^{\mu} V^2 + p_3^{\mu} V + p_4^{\mu} \\ \sigma(V) &= p_1^{\sigma} V^3 + p_2^{\sigma} V^2 + p_3^{\sigma} V + p_4^{\sigma} \\ \psi(V) &= p_1^{\psi} V^3 + p_2^{\psi} V^2 + p_3^{\psi} V + p_4^{\psi}. \end{aligned} \tag{6.18}$$

Analogously, Pareto distribution parameters (scale and shape) and threshold $H_s^{\lim}$ are also smoothed as follows:

$$\begin{aligned} \theta(V) &= p_1^{\theta} V^3 + p_2^{\theta} V^2 + p_3^{\theta} V + p_4^{\theta} \\ \epsilon(V) &= p_1^{\epsilon} V^3 + p_2^{\epsilon} V^2 + p_3^{\epsilon} V + p_4^{\epsilon} \\ H_s^{\lim}(V) &= p_1^{H_s} V^3 + p_2^{H_s} V^2 + p_3^{H_s} V + p_4^{H_s}. \end{aligned} \tag{6.19}$$

Maximum likelihood estimates of these parameters are given in Table 6.1.

The 50-year environmental contour for a 1-hour sea state duration, using expressions (6.16) and (6.18), and only considering the GEV fittings related to both the wind speed and significant wave height, are shown in Figure 6.11. Contours are associated with equivalent return periods of 100 hours, 1, 5, 50, 100 and 500 years, respectively. 50-year environmental contour is in black. This result confirms that the 50-year environmental contour overestimate in excess significant wave heights for wind speeds lower than ≈ 20 m/s, which is precisely the interval where the GEV fit does not appropriately reproduce the tail of the distribution. In contrast, for wind speeds above ≈ 20 m/s results are in accordance with data.

Alternatively, we could calculate the 50-year environmental contour using the

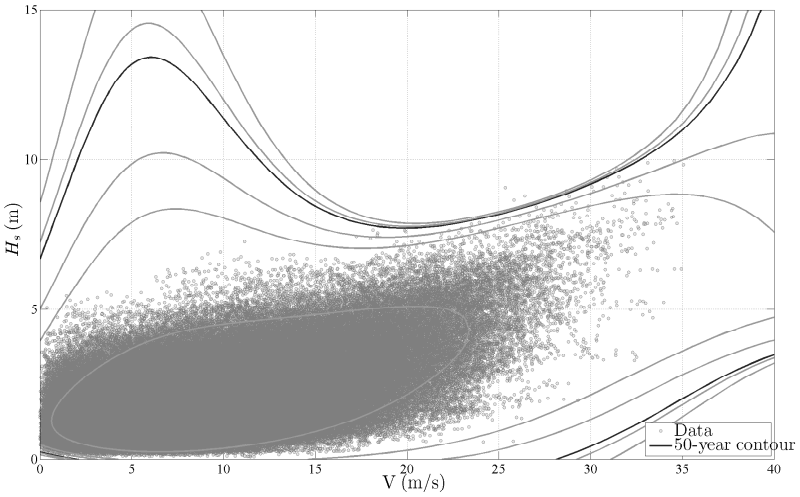

Figure 6.11: 50-year environmental contour plot according to IEC 61400-3 standards, without specific tail fitting.

following Rosenblatt transformation:

$$\Phi(z_1)=\begin{cases} F_V^{\mathrm{PT}}(V) \text{ if } V \leq V_{\mathrm{lim}}, \\ p_{\mathrm{V,lim}}^{\mathrm{PT}} + \frac{F_V^{\mathrm{EV}}(V)-p_{\mathrm{V,lim}}^{\mathrm{EV}}}{1-p_{\mathrm{V,lim}}^{\mathrm{EV}}}(1-p_{\mathrm{V,lim}}^{\mathrm{PT}}) \\ \text{if } V > V_{\mathrm{lim}}, \end{cases}$$
$$\Phi(z_2)=\begin{cases} F_{H_s|V}^{\mathrm{PT}}(H_s) \text{ if } H_s \leq H_{\mathrm{s,lim}}, \\ p_{\mathrm{Hs,lim}}^{\mathrm{PT}} + \frac{F_{Hs|V}^{\mathrm{EV}}(H_s)-p_{\mathrm{Hs,lim}}^{\mathrm{EV}}}{1-p_{\mathrm{Hs,lim}}^{\mathrm{EV}}}(1-p_{\mathrm{Hs,lim}}^{\mathrm{PT}}) \\ \text{if } H_s > H_{\mathrm{s,lim}}. \end{cases} \tag{6.20}$$

In this particular case, we made distinction between the point-in-time distribution, that represents the probabilistic behavior of all data range but the right tail, and the extreme-value distribution, that characterizes the right tail above the selected threshold. Note that the point-in-time distributions F_V^{PT} and $F_{H_s|V}^{\mathrm{PT}}$ correspond, respectively, to $F_V(v)$ and $F_{H_s|V}(H_s)$ from expression (6.16). The new 50-year environmental contour for a 1-hour sea state duration improving the probability density functions at the right tails is shown in Figure 6.12. The improvement in the region of interest, i.e. the one associated with high values of the significant wave height and wind speeds in the range corresponding to power production, is significant. Note that with previous approach significant wave heights related to

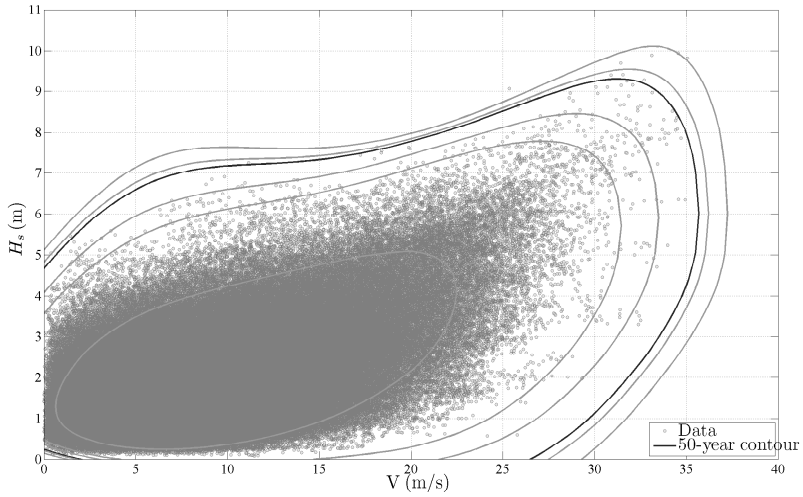

Figure 6.12: 50-year environmental contour plot according to IEC 61400-3 standards, but including specific tail fitting.

the 50-year environmental contour for low wind speed values are above 10 meters, even higher than those related to high wind speeds, which is physically unlikely.

The IEC 61400-3 standards also recommend to use for all wind speeds of the unconditional extreme significant wave height with a recurrence period of 50 years as a conservative value for H_s. Note that according to the fitting shown in Figure 6.13, this conservative value is equal to 9.6483 m (white circle marker specifier). Since this value is above the 50-year environmental contour (see Figure 6.12), it is the one designer must take if the second analysis is performed. In contrast, in case using the first analysis, designer would use significant wave heights considerable above this threshold for low wind speeds, leading to an excessive conservationist.

These results clearly demonstrated the importance of considering both the point-in-time and right-tail distributions.

It is important to clarify that in both cases, the left tail of the distributions is reproduced inappropriately. Contour plot intercepts negative wind speed and significant wave height values, which is impossible. However, this does not distort the relevant results from the engineering perspective, i.e. those in the upper tail.

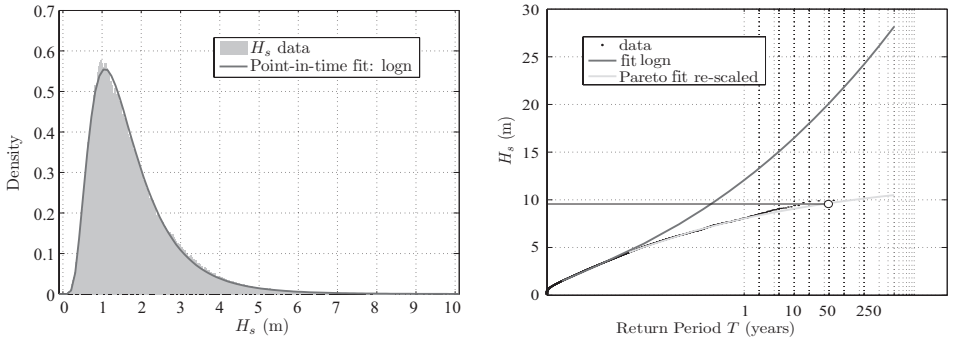

Figure 6.13: Histogram and fitted probability density function of unconditional significant wave height H_s, and graphical representation in terms of equivalent return periods using the method proposed by Mínguez, Guanche and Méndez (2012).

Alternatively, an specific distribution for the left tail could be used instead.

9 Conclusions

The method proposed in this chapter provides new insights on the relationship between the point-in-time and extreme-value distributions associated with any stochastic process, and a possible way to deal with both distributions at the same time. The advances with respect to the state-of-the-art can be summarized as follows:

1. A new graphical representation to help understanding the relationship between both distributions is proposed.

2. A new Monte Carlo simulation technique holding the following requirements is provided:

   - It is able to reproduce both the point-in-time and extreme-value regimes.
   - It maintains the temporal dependence structure of the stochastic process through ARMA models.

3. In addition, some hints about extending the method into FORM techniques are given. In this case, the method frees the engineer to decide about what regime should be used within the design process.

All the methods have been tested using different synthetically generated samples and an example based on real data. Results confirm the good behavior of the proposed methods, and their suitability to i) support engineers on the design process and ii) help understanding the relationship between both point-in-time and extreme-value regime.

To further reinforce our arguments, a practical example about how to construct environmental contours for the definition of design requirements for offshore wind turbines (IEC 61400-3) is given, emphasizing possible problems which may lead to unsafe or excessively conservatism designs.

Nevertheless, further work is required to improve the methodology in terms of:

1. Which is the appropriate threshold value for the definition of the right-tail distribution?

2. How to transform equivalent return periods into real return periods? This issue is related to the dependence assumption of extremes, and could be solved introducing the extremal index concept.

Further research must be also done on cross-correlations between different stochastic processes, however, the results showed in this chapter constitute a clear advance on the knowledge of point-in-time and extreme-value distributions.

Note that although all the material developed in this chapter is related to the upper tail of the point-in-time distribution (maxima), alternative formulations can be straightforwardly obtained for dealing with minima.

CHAPTER

7

Conclusions and Future Research

1 Summary of contributions

In this section a summary of the most relevant contributions presented in this thesis is presented:

1. An autoregressive logistic model has been presented to study atmospheric conditions in terms of synoptical circulation patterns. The nominal character of the model allows the inclusion of autoregressive terms as well as other covariates of different nature such as seasonality, interannual variability and long term trends.

2. A methodology has been developed to simulate hourly trivariate sea state time series. This methodology combines the logistic model mentioned before and an existent technique of multivariate simulation. This methodology takes into account various natural processes with different temporal and spatial scales.

3. A hybrid method of clusterization and interpolation has been used to estimate loads in maritime structures. This hybrid method has already been used to propagate waves from deep seas to shallows seas. Here, the methodology has been successfully applied to two different structures, a vertical breakwater and an offshore wind turbine.

4. Finally, a methodology of simulation has been developed which is able to join both the extreme and mean regimes of single variables.

2 Conclusions

After a revision of the contributions contained within the presented work, the following conclusions can be made

1. An accurate simulation of the variables involved in maritime structure design requires a comprehensive knowledge of the relevant processes and the engaged forces. Similarly, a wave climate study entails the detailed study of the associated atmospheric conditions. Additionally, the winds and pressure fields in the region of influence of the structure being designed must be properly considered.

2. In addition to properly accounting for the spatial scales of the process, a thorough knowledge and consideration of the different temporal scales involved in wave climates is also required. It is obvious that waves present a strong autocorrelation, but there are also other longer term variations that must be taken into account such as seasonality, interannual variability and long term trends.

3. The use of statistical techniques provides for the efficient management of long term time series, enabling its treatment, and allows estimates of the relevant parameters to be derived from them. The hybrid technique proposed serves to objectively select cases to be modeled numerically or in the laboratory.

4. The availability of long term time series allows the simulation to be based on empirical distributions. But to correctly characterize and simulate extreme events, they have to be appropriately modeled. In this way, the simulation is not constrained to the historical records.

2.1 On the autoregressive logistic model

- Previous states of the process can be considered in the model, and this is relevant in atmospheric studies.
- The nominal nature of the model enhances the physical sense of circulation pattern clusterization.
- The model allows for the inclusion of long-term trends which are mathematically consistent, so that the probabilities associated with each weather type always range between 0 and 1.
- It is possible to simultaneously take into account the influence of different variables.
- Synoptical circulation patterns are defined by clusterization techniques such as k-means. This technique is appropriated to define pressure fields.
- Long-term climate variability resulting from any change in the covariates can be studied due to the flexibility of the proposed model.

2.2 On the methodology used to simulate trivariate sea states

- The combination of different statistical techniques allows for the consideration of different temporal and spatial scales.
- The procedure developed enables the reproduction of trivariate sea states, while considering the correlation between the variables.
- The hourly time series simulated are conditioned to daily mean conditions. This is an easy way to facilitate the study of the relationship between the local wave climate and the governing atmospheric conditions.

- Daily mean sea conditions are classified using k-means. This technique creates groups of similar characteristic states, which is convenient to enable the definition of synoptic patterns.

- Simulating daily Sea Level Pressure fields decomposed into PCs allows for the generation of different atmospheric scenarios.

2.3 On the procedure to transfer sea states to design parameters

- Multivariate long term time series management can be efficiently accomplished by combining statistical techniques.

- The MDA algorithm allows the selection of representative sea states. Unlike other techniques this one can include extreme events within the selected states.

- The RBF interpolation technique enables the generation of long term time series of design parameters.

- To estimate the design parameters, semi-empirical formulation has been used, but the methodology would be similar to the cases using numerical or physical models.

2.4 On the simulation technique used to consider mean and extreme regimes simultaneously

- A new way of graphical representation that facilitates the understanding of the relation between both regimes has been presented.

- The simulation technique proposed is able to reproduce both the point–in–time and extreme value regimes, and maintains the temporal dependence structure of the stochastic process using ARMA models.

- Simulating both regimes simultaneously eliminates having to decide which regime to apply in each case.

Nevertheless, further work is required to improve the methodology. Both the appropriate selection of the threshold value for the definition of the upper–tail distribution, as well as how to transform equivalent return periods into real return periods, require additional study. These issues are related to the dependence assumption of extremes, and may be solved by introducing the extremal index concept.

3 Future research topics

This PhD work constitutes a step towards further research related to different issues. Here, some of these challenges are proposed.

Regarding the development of logistic models to analyze atmospheric or coastal conditions:

- The determination of the optimal number of synoptical patterns to describe the atmospheric processes and/or marine dynamics. In the literature there exists discrepancies between authors, related to the optimal number of weather types to consider. Further research must be done on this issue in order to be able to establish objective criteria to determine the correct optimal number of synoptical patterns that should be taken into account.

- Comparison of the different climate change scenarios. By means of the logistic model proposed, the different climate change scenarios could be simulated and comparisons of the different trends obtained.

Concerning the climate–based simulation of sea states technique:

- Inclusion of extreme regimes models to the simulation technique. All the work shown related to the multivariate simulation has been done based on the empirical distribution of the variables from the historical databases. Because of this, the obvious subsequent path for future research would be to develop the methodology of univariate simulation that takes into account both regimes for multivariate cases. This extension is not trivial and requires a detailed study of the limits between both regimes for each variable and an exploration of the correlations between extreme events.

- Addition of more variables of interest to the sea state simulation procedure. Depending on the structure to be designed, it may be of interest to have the availability of wind or currents time series. In order to accomplish this, correlations between variables and autocorrelations within them should be analyzed. Moreover, the addition of other covariates that explain the processes besides seasonality, trends and interannual variability may be required.

- Generalization of the proposed method. Previous to the lines mentioned before, it would be necessary to generalize proposed methodologies. Its applicability in different locations where wave climate characteristics and involved processes are different must be tested.

- Climate-based simulation technique of extreme events. In the literature there are works to generate synthetic time series of extreme events [75] that also distinguish between two regimes when simulating. It would be of considerable interest in maritime design fields to be able to combine the existent extreme simulation technique with the climate–based technique proposed here. In this way, the influence of different natural variables with their own temporal and space scales could be taken into account.

With regard to the transference of sea states to design parameters:

- Application of the MDARBF methodology to laboratory tests planning. The methodology has been proven with semi–empirical formulation, but its application to laboratory tests plan designs would be of considerable interest. In order to do this, some analysis should be done to establish the minimum number of tests to be taken into account and to determine the repeatability achievable in the laboratory.

- Determination of the weaknesses of the process. Further research must be done in order to determine which part of the methodology entails higher uncertainties: the number of the selected cases by the MDA or the interpolation technique. Once this is established, the process could be improved by estimating the uncertainty associated with a certain number of selected cases.

In the area of simultaneous extreme and mean regimes simulation:

- Inclusion of an extremal index. Regarding the simulation technique to consider simultaneously extreme and mean regimes, the inclusion of an extremal index to undertake the transition between mean and extreme regimes has to be thoroughly investigated. In the methodology proposed here, this transition does not consider this index, resulting in a sharp transition that may produce inaccurate results for lower return periods.

- Threshold selection. Concerning the pointintime and extreme distributions used in the simulation technique proposed, the determination of an objective criteria to appropriately define the threshold between both distributions is a subject for additional research

References

[1] *ANSYS Incorporation. ANSYS ASA-OFFSHORE version 14.04*. USA. 153

[2] *Engineering Dynamics Incorporation (EDI). Structural Analysis Computer System.* USA. 153

[3] *STAAD. Offshore Release*. Research Engineers International, Bentley Solution Center, 2005. 153

[4] *International Electrotechnical Commission. IEC 61400–3. Wind Turbines. Design requirements for offshore wind turbines*. 2009. 30, 163, 166

[5] *DNV–OS–J101. Design of offshore wind turbines. Offshore Standard.* Det Norske Veritas, 2011. 153

[6] ABASCAL, A. J., CASTANEDO, S., AND MEDINA, R. Gos, un reanálisis de marea meterorológica de 60 años de alta resolución para el sur de europa. In *Encuentro de Oceanografía Física Española* (Barcelona, Spain, 2010). 29, 159

[7] BAKER, M. J. Evaluation of partial safety factors for level i codes. Example of application of methods to reinforced concrete beams. In *Bulletin d'Information No. 112, Comite Européen due Béton* (Paris, 1976), vol. 112, pp. 190–211. 178

[8] BASLER, B. Untersuchungen über den sicherheitsbegriff von bauwerken. *Schwez. Arch. 27*, 4 (1961), 133–160. 57

[9] BEHRENS, B., FREITAS LOPES, H., AND GAMERMAN. Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling 4* (2004), 227–244. 179

[10] BIZZOTTO, R., ZAMUNER, S., DE NICOLAO, G., KARLSSON, M. O., AND GOMENI, R. Multinomial logistic estimation of Markov-chain models for modelling sleep architecture in primary insomnia patients. *Pharmacokinet Pharmacodyn 37* (2010), 137–155. 71

[11] BJERAGER, P. Probability integration by directional simulation. *J. Engineering Mechanics, ASCE 114*, 8 (1988), 1285–1302. 179

[12] BONNEY, G. E. Logistic regression for dependent binary data. *Biometrics 43*, 4 (1987), 951–973. 72

[13] BOOIJ, N., RIS, R., AND HOLTHUIJSEN, L. A third generation wave model for coastal region: 1. model description and validation. *Journal of Geophisical Research 104* (1999), 7649–7666. 111

[14] BOX, G. E. P., JENKINS, G. M., AND REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. Prentice-Hall International, New Jersey,NJ, 1994. 74, 105, 112, 193

[15] BREITUNG, K. Asymptotic approximations for multinormal integrals. *J. Engrg. Mech. 100*, 3 (1984), 357–366. 61

[16] CAHYNOVA, M., AND HUTH, R. Enhanced lifetime of atmospheric circulation types over europe: fact of fiction? *Tellus 61A* (2009), 407–416. 74, 79

[17] CAI, Y. Multi-variate time-series simulation. *Journal of Time Series Analysis 32* (2011), 566–579. 101

[18] CAI, Y., B., G., HAWKES, P., AND DUNNING, P. Statistical simulation of flood variables: incorporating short-term sequencing. *Journal of Flood Risk Management 1* (2008), 3–12. 179, 191

[19] CAMUS, P., MÉNDEZ, F. J., AND MEDINA, R. A hybrid efcient method to downscale wave climate to coastal areas. *Coastal Engineering 58* (2011), 851–862. 22, 29, 100, 102, 111, 132, 134, 154, 155, 158

[20] CAMUS, P., MÉNDEZ, F. J., MEDINA, R., AND COFIÑO, A. S. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering 58*, 6 (2011), 452–462. 132, 134, 154, 159

[21] CAMUS, P., MÉNDEZ, F. J., MEDINA, R., TOMÁS, A., AND IZAGUIRRE, C. High resolution downscaled ocean waves (dow) reanalysis in coastal areas. *Coastal Engineering 72* (2013), 56–68. 200

[22] CASTILLO, C., MÍNGUEZ, R., CASTILLO, E., AND LOSADA, M. An optimal engineering design method with failure rate constraints and sensitivity analysis. Application to composite breakwaters. *Coastal Engineering 53*, 1 (2006), 1–25. 101

[23] CASTILLO, E. *Extreme Value Theory in Engineering*. Academic Press, New York, 1988. 179

[24] CASTILLO, E., CASTILLO, C., AND MÍNGUEZ, R. Use of extreme value theory in engineering design. In *Proceedings of the European Safety and Reliability Conference 2008 (ESREL 2008), Safety, Reliability and Risk Analysis: Theory, Methods and Applications*, S. Martorel, C. Guedes Solares, and J. Barnett, Eds., vol. 3. Taylor & Francis Group, Valencia, 22-25 September 2008, pp. 2473–2488. 179

[25] CASTILLO, E., HADI, A. S., BALAKRISHNAN, N., AND SARABIA, J. M. *Extreme Value and Related Models in Engineering and Science Applications*. John Wiley & Sons, New York, 2005. 179, 180

[26] CASTILLO, E., SOLARES, C., AND GÓMEZ, P. Estimating extreme probabilities using tail simulated data. *International Journal of Approximate Reasoning: official publication of the North American Fuzzy Information Processing Society 17*, 2-3 (1996), 163–190. 61, 179

[27] CASTILLO, E., SOLARES, C., AND GÓMEZ, P. Tail sensitivity analysis in bayesian networks. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96)* (Portland (Oregon), 1996), Morgan Kaufmann Publishers, San Francisco, California, pp. 133–140. 61

[28] CASTILLO, E., SOLARES, C., AND GÓMEZ, P. High probability one-sided confidence intervals in reliability models. *Nuclear Science and Engineering 126* (1997), 158–167. 61

[29] CHANG, C. H., TUNG, Y. K., AND YANG, J. C. Monte carlo simulation for correlated variables with marginal distributions. *Journal of Hydraulic Engineering 120*, 3 (1994), 313–331. 191

[30] CHAWLA, A., SPINDLER, D. M., AND TOLMAN, H. L. Validation of a thirty year wave hindcast using the climate forecast system reanalysis winds. *Ocean Modelling* (2012). 100

[31] CHIU, Y., LIN, J., CHANG, S., LIN, Y., AND CHEN, C. An experimental studyof wave forces on vertical breakwater. *Journal of Marine Science and Technology 15*, 3 (2007), 158–170. 131

[32] COLES, S. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics, 2001. 179

[33] CORNELL, C. A. A probability based structural code. *J. Amer.Concrete Inst. 66*, 12 (1969), 974–985. 57

[34] CORTE-REAL, J., XU, H., AND QIAN, B. A weather generator for obtaining daily precipitation scenarios based on circulation patterns. *Climate Research 13* (1999), 61–75. 79

[35] COX, D. R. The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 21*, 2 (1972), 113–120. 72

[36] CUOMO, G., ALLSOP, W., BRUCE, T., AND PEARSON, J. Breaking wave loads at vertical seawalls and breakwaters. *Coastal Engineering 57*, 4 (2010), 424–439. 131

[37] CUOMO, G., ALLSOP, W., AND TAKAHASHI, S. Scaling wave impact pressures on vertical walls. *Coastal Engineering 57*, 6 (2010), 604–609. 131

[38] DE MICHELE, C., SALVADORE, G., PASSONI, G., AND VEZZOLI, R. A multivariate model of sea storms using copulas. *Coastal Engineering 54* (2007), 734–751. 101

[39] DE VRIES, S. O., FIEDLER, V., KIUPERS, W. D., AND HUNINK, M. G. M. Fitting multistate transition models with autoregressive logistic regression: Supervised exercise in intermittend claudication. *Medical Decision Making 18* (1998), 52–60. 71, 72, 75

[40] DER KIUREGHIAN, A. Bayesian analysis of model uncertainty in structural reliability. In *Proc. 3rd IFIP WG7.5 Conf. Reliability and Optimization of Structural Systems* (Berlin, 1990), Springer, pp. 211–221. 55

[41] DER KIUREGHIAN, A., AND DE STAFENO, M. Efficient algorithm for second-order reliability analysis. *J. Engineering Mechanics, ASCE 117*, 12 (1991), 2904–2923. 61

[42] DITLEVSEN, O. Structural reliability and the invariance problem. Solid Mechanics Report 22, University of Waterloo, 1973. 60

[43] DITLEVSEN, O. Principle of normal tail approximation. *J. Engineering Mechanics Div., ASCE 107*, 6 (1981), 1191–1208. 197

[44] DITLEVSEN, O., AND BJERAGER, P. Plastic reliability analysis by directional simulation. *J. Engineering Mechanics Div., ASCE 115*, 6 (1989), 1347–1362. 62, 179

[45] DITLEVSEN, O., AND MADSEN, H. O. *Structural reliability methods*. Wiley, Chichester, New York, 1996. 199

[46] DITLEVSEN, O., OLESEN, R., AND MOHR, G. Solution of a class of load combination problems by directional simulation. *Structural Safety 4* (1987), 95–109. 62, 179

[47] DOBSON, A. J. *An Introduction to Generalized Linear Models*, second ed. Chapman & Hall/CRC, Florida, 2002. 78

[48] DONG, S., WANG, N., LIU, W., AND GUEDES SOARES, C. Bivariate maximum entropy distribution of significant wave height and peak period. *Ocean Engineering 59* (2013), 86–99. 101

[49] ENGELUND, S., AND RACKWITZ, R. A benchmark study on importance sampling techniques in structural reliability. *Structural Safety 12*, 4 (1993), 255–276. 62

[50] ESPEJO, A., MÍNGUEZ, R., TOMÁS, A., MENÉNDEZ, M., MÉNDEZ, F., AND LOSADA, I. J. Directional calibrated wind and wave reanalysis databases using instrumental data for optimal design of off-shore wind farms. In *OCEANS, 2011 IEEE - Spain* (june 2011). 111

[51] ESTEBAN, P., MARTÍN-VIDE, J., AND MASES, M. Daily atmospheric circulation catalogue for Western Europe using multivariate techniques. *International Journal of Climatology 26* (2006), 1501–1515. 79

[52] FISHER, R. A., AND TIPPETT, L. H. C. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Journal of Coastal Research 24* (1928), 180–190. 181

[53] FRANKE, R. Scatter data interpolation: test of some methods. *Mathematics of Computation 38* (1982), 181–200. 142, 156

[54] FREUDENTHAL, A. M. Safety and the probability of structural failure. *Transactions, ASCE 121* (1956), 1337–1397. 57, 60, 197

[55] FRIGESSI, A., HAUG, O., AND RUE, H. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes 5* (2002), 219–235. 179

[56] FURRER, E. M., AND KATZ, R. W. Improving the simulation of extreme precipitation events by stochastic weather generators. *Waters Resources Research 44*, 12 (2001), 1–13. 179

[57] GALAMBOS, J. *The Asymptotic Theory of Extreme Order Statistics*. Robert E. Krieger Publishing Company, Malabar, Florida, 1987. 180

[58] GERSTENGARBE, F. W., AND WERNER, P. *Katalog der Grosswetterlagen Europas(1881-1998) nach Paul Hess und Helmuth Brezowsky*. Postdam Institut für Klimafolgenforschung, Postdam, Germany, 1999. 70

[59] GERSTENGARBE, F. W., AND WERNER, P. *Katalog der Grosswetterlagen Europas(1881-2004) nach Paul Hess und Helmuth Brezowsky*. Postdam Institut für Klimafolgenforschung, Postdam, Germany, 2005. 70

[60] GIVENS, G. H., AND RAFTERY, A. E. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of the American Statistical Association 91*, 433 (1996), 132–141. 62

[61] GODA, Y. New wave pressure formulae for composite breakwaters. In *Proceedings 14th International Conference on Coastal Engineering. ASCE* (1974), pp. 1702–1720. 130

[62] GODA, Y. *Random Seas andDesign of Maritime Structures*. University of Tokyo, Tokio, Japan, 1975. 130

[63] GODA, Y. *Random Seas and Design of Maritime Structures*. Tokyo University Press, Tokyo, 1985. 28

[64] GOODESS, C. M., AND JONES, P. D. Links between circulation and changes in the characteristics of Iberian rainfall. *International Journal of Climatology 22* (2002), 1593–1615. 74

[65] GUANCHE, R., LOSADA, I. J., AND LARA, J. L. Numerical analysis of wave loads for coastal structures stability. *Coastal Engineering 56*, 5-6 (2010), 543–588. 131

[66] GUANCHE, Y., CAMUS, P., GUANCHE, R., MÉNDEZ, F. J., AND MEDINA, R. A simplified method to downscale wave dynamics on vertical breakwaters. *Coastal Engineering 71* (2013), 68–77. 153, 154

[67] GUANCHE, Y., MÍNGUEZ, R., AND MÉNDEZ, F. J. Autoregressive logistic regression applied to atmospheric circulation patterns. *Climate Dynamics* (2013). 102, 104, 106, 108, 116

[68] GUEDES SOARES, C., AND CUNHA, C. Bivariate autoregressive models for the time series of significant wave height and mean period. *Coastal Engineering 40* (2000), 297–311. 101, 180, 191, 192

[69] GUEDES SOARES, C., AND FERREIRA, A. M. Representation of non-stationary time series of significant wave height with autoregressive models. *Probabilistic Engineering Mechanics 11* (1996), 139–148. 100

[70] GURHAN, G., AND UNSALAM, D. A comparative study of the first and second order theories of goda's formula for wave-induced pressure on a vertical breakwaterwith regular waves. *Ocean Engineering 32* (2005), 2182–2194. 131

[71] HARBITZ, A. An efficient sampling method for probability of failure calculation. *Structural Safety 3*, 2 (1986), 109–115. 62, 179

[72] HARDY, R. L. Multiquadratic equations of topography and other irregular surfaces. *Journal of Geophysical Research 76* (1971), 1905–1915. 142, 156

[73] HASKIN, F. E., STAPLE, B. D., AND DING, C. Efficient uncertainty analyses using fast probability integration. *Nuclear Engineering and Design 166*, 2 (1996), 225–248. 61

[74] HASOFER, A. M., AND LIND, N. C. Exact and invariant second moment code format. *J. Engrg. Mech. 100*, EM1 (1974), 111–121. 7, 58, 60, 197, 199

[75] HEFFERNAN, J. E., AND TAWN, J. A. A conditional approach for multivariate extreme values. *journal of the Royal Statistic Society 66* (2004), 497–546. 44, 214

[76] HERMAN, A., KAISER, R., AND NIEMEYER, H. D. Wind-wave variability in shallow tidal sea-spectral modellin combined with neural network methods. *Coastal Engineering 56* (2009), 759–772. 132

[77] HESS, P., AND BREZOWSKY, H. *Katalog der Grosswetterlagen Europas(Catalog of the European Large Scale Weather Types)*. Ber. Dt. Wetterd, in der US-Zone 33, Bad Kissinger, Germany, 1952. 70

[78] HIROI, I. On a method of estimatingthe force of waves. *Tokyo University Engineering Reports X* (1919). 130

[79] HOHENBICHLER, M., AND RACKWITZ, R. Non-normal dependent vectors in structural safety. *J. Engineering Mechanics Div., ASCE 107*, 6 (1981), 1227–1238. 60, 197

[80] HORNE, M. R., AND H., P. P. Commentary on the level 2 procedure, rationalization of safety and serviceability factors in structural codes. Report 63, Construction Industry Research and Information Association, London, 1977. 178

[81] HUGHES, S. A., AND FOWLER, J. E. Wave-induced scour prediction at vertical walls. In *Proceedings Coastal Sediments. ASCE* (1991), vol. 2, pp. 1886–1900. 28, 131, 133

[82] HURRELL, J.AND KUSHNIR, Y., OTTERSEN, G., AND VISBECK, M. *The North Atlantic Oscillation: Climate Significance and Environmental Impact*. American Geophysical Union, Washington, DC, 2003. 81

[83] HUTH, R. A circulation classification scheme applicable in GCM studies. *Theoretical and Applied Climatology 67* (2000), 1–18. 70

[84] HUTH, R. Disaggregating climatic trends by classification of circulation patterns. *International Journal of Climatology 21* (2001), 135–153. 70

[85] HUTH, R., BECK, C., PHILIPP, A., DEMUZERE, M., USTRNUL, Z., CAHYNOVÁ, M., KYSELÝ, K., AND TVEITO, O. E. Classification of atmospheric circulation patterns. *Trends and Directions in Climate Research 1146* (2008), 105–152. 70, 74

[86] ITO, Y., AND TANIMOTO, K. Meandering damages of composite type breakwaters. *Technical Note of Port and Harbour Res. Inst.(in Japanese) 112* (19171). 130

[87] IZAGUIRRE, C., MENÉNDEZ, M., CAMUS, P., MÉNDEZ, F. J., MÍNGUEZ, R., AND LOSADA, I. J. Exploring the interannual variability of extreme wave climate in the northeast atlantic ocean. *Ocean Modelling 59–60* (2012), 31–40. 79, 102, 105

[88] JONKMAN, J., BUTTERFILELD, S., MUSIAL, W., AND SCOTT, G. *Definition of a 5–MW reference wind turbine for offshore system development.* Technical Report, NREL/TP 500–38060, 2009. 166

[89] JORDAN, P., AND TALKNER, P. A seasonal Markov chain model for the weather in the central Alps. *Tellus 52A* (2000), 455–469. 71, 73, 75, 79

[90] KALNAY, E. M., KANAMITSU, R., KISTLER, W., COLLINS, D., DEAVEN, L., GANDIN, M., IREDELL, S., SAHA, G., WHITE, J., WOOLLEN, Y., ZHU, M., CHELLIAH, W., EBISUZAKI, W., HIGGINS, J., JANOWIAK, K. C., MO, C., ROPELEWSKI, J., WANG, A., LEETMAA, R., REYNOLDS, R., JENNE, R., AND JOSEPH, D. The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society 77* (1996), 437–470. 22, 78, 111

[91] KARIMIRAD, M., AND MOAN, T. Extreme dynamic structural response anaysis of catenary moored spar wind turbine in harsh environmental conditions. *Journal of Offshore Mechanics and Artic Engineering 133* (2011). 153

[92] KARIMIRAD, M., AND MOAN, T. Wave - and wind-induced dynamic response of a spar-type offshore wind turbine. *Journal of Waterway, Port, Coastal and Ocean Engineering 138* (2011). 153

[93] KISTLER, R., KALNAY, E., COLLINS, W., SAHA, S., WHITE, G., WOOLLEN, J., CHELLIAH, M., EBISUZAKI, W., KANAMITSU, M., KOUSKY, V., VAN DEN DOOL, H., JENNE, R., AND FIORINO, M. The NCEP-NCAR 50-year reanalysis: Monthly means CD-Rom and documentation. *American Meteorological Society 82* (2001), 247–268. 78, 111

[94] KOYLUOGLU, H. U., AND NIELSEN, R. K. New approximations for sorm integrals. *Structural Safety 13* (1994), 235–246. 61

[95] KULLBACK, S., AND LEIBLER, R. A. On information and sufficiency. *Annals of Mathematical Statistics 22* (1951), 79–86. 113

[96] KYSELÝ, K., AND HUTH, R. Changes in atmospheric circulation over Europe detected by objective and subjective methods. *Theoretical and Applied Climatology 85* (2006), 19–36. 79

[97] LEE, K. H., AND MIZUTAMI, N. Experimental study on scour occurring at a vertical impermeable submerged breakwater. *Applied Ocean Research 30*, 2 (2008), 92–99. 131

[98] LIND, N. C. Application to design of level i codes. In *Bulletin d'Information No. 112, Comite Européen due Béton* (Paris, 1976), vol. 112, pp. 73–89. 178

[99] LONGUETT HIGGINS, M. S. On the statistical distribution of the heights of sea waves. *Journal of Marine Research 11* (1952), 245–266. 100

[100] LOSADA, M. A. Recent development in the design of mound breakwaters. In *Chapter 21 in Handbook of Coastal and Ocean Engineering*, J. Herbich, Ed., vol. I. Gulf Publishing, 1990. 178

[101] MAHERAS, P., TOLIKA, K., ANAGNOSTOPOULOU, C., VAFIADIS, M., PATRIKAS, I., AND FLOCAS, H. On the relationships between circulation types and changes in rainfall variability in Greece. *International Journal of Climatology 24* (2004), 1695–1712. 79

[102] MASSEY, F. J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association 46* (1951), 68–78. 95, 121

[103] MAYER, H. *Die Sicherheit der Bauwerke*. Springer, Berlin, 1926. 57

[104] MELCHERS, R. E. Improved importance sampling for structural reliability calculation. In *Proceedings 5th International Conference on Structural Safety and Reliability* (New York, 1989), ASCE, pp. 1185–1192. 179

[105] MELCHERS, R. E. Simulation in time-invariant and time-variant reliability problems. In *Proceedings 4th IFIP Conference on Reliability and Optimization of Structural Systems* (Berlin, 1991), Springer, pp. 39–82. 179

[106] MELCHERS, R. E. *Structural reliability analysis and prediction*, second ed. John Wiley & Sons, New York, 1999. 49, 179

[107] MÉNDEZ, F. J., MENÉNDEZ, M., LUCEÑO, A., AND LOSADA, I. J. Estimation of the long term variability of extreme significant wave height using a time-dependent POT model. *J. of Geophys. Res. 111* (2006). doi:10.1029/2005JC003344. 192

[108] MÉNDEZ, F. J., MENÉNDEZ, M., LUCEÑO, A., AND LOSADA, I. J. Analyzing monthly extreme sea levels with a time-dependent gev model. *J. Atmos. Ocean. Technol. 24* (2007), 894–911. 192

[109] MENÉNDEZ, M., GARCÍA-DÍEZ, M., FERNÁNDEZ, J., AND MÉNDEZ, F. J. High-resolution sea wind reanalyses over the mediterranean area. *Climate Dynamics to appear*. 200

[110] MENÉNDEZ, M., MÉNDEZ, F. J., IZAGUIRRE, C., AND LOSADA, I. J. The influence of seasonality on estimating return values of significant wave height. *Coastal Engineering 56*, 3 (2009), 211–219. 192

[111] MENÉNDEZ, M., TOMÁS, A., CAMUS, P., GARCÍA-DÍEZ, M., FITA, L., FERNÁNDEZ, J., AND MÉNDEZ, F. A methodology to evaluate regional-scale offshore wind energy resources. In *OCEANS, 2011 IEEE - Spain* (june 2011). 29, 111, 134, 158

[112] METROPOLI, N., AND ULAM, S. The monte carlo method. *Journal of the American Statistical Association 44* (1949), 335–341. 7, 60

[113] MICHE, R. *Mouvements ondulatoires de la mer in profendeur constante ou decroissante*, vol. 114. Annals de Ponts et Chausses, Paris, France, 1944. 130

[114] MÍNGUEZ, R., ESPEJO, A., TOMÁS, A., MÉNDEZ, F. J., AND LOSADA, I. J. Directional calibration of wave reanalysis databases using instrumental data. *Journal of Atmospheric and Oceanic Technology* (2010). sent for publication. 200

[115] MÍNGUEZ, R., ESPEJO, A., TOMÁS, A., MÉNDEZ, F. J., AND LOSADA, I. J. Directional calibration of wave reanalysis databases using instrumental data. *Journal of Atmospherical and Oceanic Technology 28* (2011), 1466–1485. 111

[116] MÍNGUEZ, R., MÉNDEZ, F. J., IZAGUIRRE, C., MENÉNDEZ, M., AND LOSADA, I. J. Pseudo-optimal parameter selection of non-stationary generalized extreme value models for environmental variables. *Environmental Modelling & Software 25* (2010), 1592–1607. 192

[117] MÍNGUEZ, R., REGUERO, B. G., LUCEÑO, A., AND MÉNDEZ, F. J. Regression models for outlier identification (hurricanes and typoons) in wave hindcast databases. *Journal of Atmospheric and Oceanic Technology* (2011). Accepted. 200

[118] MÍNGUEZ, R., REGUERO, B. G., LUCEÑO, A., AND MÉNDEZ, F. J. Regression models for outliers identification (hurricanes and typhoons) in wave hindcast databases. *Journal of Atmospheric and Oceanic Technology 29* (2011), 267–285. 111

[119] MINIKIN, R. R. *Winds, waves and maritime structures*. Charles Griffin, London, England, 1950. 130

[120] MORALES, J. M., MÍNGUEZ, R., AND CONEJO, A. J. A methodology to generate statistically dependent wind speed scenarios. *Applied Energy 87* (2010), 843–855. 13, 20, 22, 66, 67, 101, 102, 104, 105, 106, 108, 112, 122, 180, 191, 192, 193

[121] MUENZ, L. R., AND RUBINSTEIN, L. V. Models for covariate dependence of binary sequences. *Biometrics 41*, 1 (1985), 91–101. 72

[122] MYRHAUG, D., AND CHEK, O. M. Random wave-induced scour at the trunk section of a breakwater. *Coastal Engineering 56* (2009), 688–692. 131

[123] NATAF, A. Détermination des distribution de probabilités dont les marges sont données. *Comptes Rendus de l'Academie des Sciences 225* (1962), 42–43. 61, 191

[124] NICOLIS, C., EBELING, W., AND BARALDI, C. Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes. *Appl. Climatol. 49A* (1997), 108–118. 71

[125] NIELSEN, S. R. K., AND BURCHARTH, H. F. Stochastic design of rubble mound breakwaters. In *Proceedings of the 11th IFIP Conference on System Modelling and Optimization* (Copenhagen, Denmark, 1983). 61

[126] PASMANTER, R. A., AND TIMMERMANN, A. Cyclic Markov chains with an application to an intermediate ENSO model. *Nonlinear Processes in Geophysics 10* (2003), 197–210. 71

[127] PASTOR, M. A., AND CASADO, M. J. Use of circulation types classifications to evaluate ar4 climate models over the euro-atlantic region. *Climate Dynamics* (2012). 97

[128] PEÑA, D., TIAO, G. C., AND TSAY, R. S. *A Course in Time Series Analysis*. Probability and Statistics. Wiley, New York, NY, 2001. 192

[129] PHILIPP, A., BARTHOLY, J., BECK, C., ERPICUM, M., ESTEBAN, P., FETTWEIS, X., HUTH, R., JAMES, P., JOUDAIN, S., KREIENKAMP, F., KRENNERT, T., LYKOUDIS, S., MICHALIDES, S. C., PIANKO-KLUCZYNSKA, K., POST, P., RASILLA, D., SCHIEMANN, R., SPEKAT, A., AND TYMVIOS, F. S. Cost733cat- A database of weather and circulation type classification. *Physics and Chemistry of the Earth 35* (2010), 360–373. 70

[130] PLAN, E., ELSHOFF, J. P., STOCKIS, A., SARGENTINI-MAIER, M. L., AND KARLSSON, M. O. Likert pain score modelling: A Markov model and an autoregressive continuous model. *Clinical Pharmacology and Therapeutics 91* (2012), 820–828. 71

[131] POLINSKY, A., FEINSTEIN., R. D., SHI, S., AND KUKI, A. *Librain: software for automated design exploratory and targeted combinational libraries*. American Chemical Society, Washington, DC, 1996. 162

[132] POLO, I., ULLMANN, A., ROUCOU, P., AND FONTAINE, B. Weather regimes in the euro-atlantic and mediterranean sector, and relationship with west african rainfall over the 1989-2008 period from a self-organizing maps approach. *Journal of Climate 24* (2011), 3423–3432. 94

[133] PREISENDORFER, R. W., AND MOBLEY, C. D. Principal component analysis in meteorology and oceanography. *Developments in Atmospheric Science 17* (1988). 105

[134] RACKWITZ, R., AND FIESSLER, B. Structural reliability under combined load sequences. *Comput. Struct. 9* (1978), 489–494. 60, 197

[135] RASCLE, N., AND ARDHUIN, F. A global wave parameter database for geophysical applications. part 2: Model validation with improved source term parameterization. *Ocean Modelling* (2013). 100

[136] REGUERO, B. G., MENÉNDEZ, M., MÉNDEZ, F. J., MÍNGUEZ, R., AND LOSADA, I. J. A global ocean wave (gow) calibrated reanalysis from 1948 onwards. *Coastal Engineering 65* (2012), 38–55. 22, 28, 100, 111, 134, 200

[137] RIPPA, S. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Advances in Computational Mathematics 11* (1999), 193–210. 143, 168

[138] RODRÍGUEZ, G. *Lecture Notes on Generalized Linear Models*, first ed. Princeton University, New Jersey, 2007. 78

[139] ROM 0.0. Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias. Tech. rep., Puertos del Estado, Madrid, España, Noviembre 2001. pp 245. 56, 178

[140] ROSENBLATT, M. Remarks on a multivariate transformation. *Ann. Math. Stat. 23*, 3 (1952), 470–472. 8, 60, 197, 199

[141] RUBINSTEIN, B. V. *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, 1981. 62, 179

[142] RUBINSTEIN, B. V., AND KROESE, D. P. *Simulation and the Monte Carlo method*, 2nd.edition ed. John Wiley & Sons, New York, 1981. 179

[143] RUNDGREN, L. *Water wave forces*, vol. 54. Bulletins of theRoyal Institute of Technology, Division of Hydraulics, Stockholm,Sweden, 1958. 130

[144] RZHANITZYN, R. It is necesary to improve the standars of design of building structures. a statistical method of design of building structures. Technical Translation, by Allan D. E. 1368, National Research Council of Canada, Ottawa, 1957. 57

[145] SAINFLOU, M. *Essai sur les Digues Maritimes Verticales*, vol. 98. Annales des Ponts et Chausses, Paris, France, 1928. 130

[146] SHINOZUKA, M. Stochastic methods in structural dynamics. Stochastic fields and their digital simulations, Martinus Nijhoff, The Hague, 1987. 179

[147] SIEGMUND, D. Importance sampling in the monte carlo study of sequential tests. *The Annals of Statistics 4* (1976), 673–684. 62

[148] SNAREY, M., TERRET, N. K., WILLETT, P., AND WILTON, D. J. Copmparison of algorithms for dissimilarity based compound selection. *Journal of Molecular Graphics and Modelling 15* (1997), 372–385. 161

[149] SOLARI, S. *Metodologías de simulación de agentes naturales y desarrollo de sistemas. Modelo de verificación y gestión de terminales portuarias*. PhD thesis, University of Granada, Córdoba y Málaga, 2011. 179

[150] SOLARI, S., AND LOSADA, M. A. Non-stationary wave height climate modeling and simulation. *Journal of Geophysical Research 116* (2011), 1–13. 180

[151] SOLARI, S., AND LOSADA, M. A. Unified distribution models for met-oceans variables: Application to series of significant waves. *Coastal Engineering 68* (2012), 67–77. 100

[152] SOULSBY, R. *Dinamics of marine sands. A manual for practical applications.*, vol. 98. Thomas Telford, London, UK, 1997. 131

[153] STEFANICKI, G., TALKNER, P., AND WEBER, R. Frequency changes of weather types in the alpine region since 1945. *Appl. Climatol. 60* (1998), 47–61. 71

[154] SUMER, B. M., AND FREDSØE, J. Scour at the head of a vertical-wall breakwater. *Coastal Engineering 29* (1997), 201–230. 131

[155] SUMER, B. M., AND FREDSØE, J. Experimental study of two-dimensional scour and its protection at rubble-mound breakwaters. *Coastal Engineering 40* (2000), 59–87. 131

[156] SUMER, B. M., AND FREDSØE, J. Wave scour around a large vertical circular cylinder. *Journal of Waterway, Port, Coastal and Ocean Engineering 127* (2001), 125–134. 131

[157] TAKAHASHI, S., AND TANIMOTO, K. Design and construction of caisson breakwaters-the japanese experience. *Coastal Engineering 22* (1997), 57–77. 28, 130

[158] TANCREDI, A., ANDERSON, C., AND OHAGAN, A. Accounting for threshold uncertainty in extreme value estimation. *Extremes 9* (2006), 87–106. 179

[159] TVEDT, L. Two second-order approximations to the failure probability. Veritas rep. rdiv/20-004083, Der norske Veritas, Oslo, 1983. 61

[160] VAN DEN BESSELAAR, E., KLEIN TANK, A., AND VAN DER SCHRIER, G. Influence of circulation types on temperature extremes in Europe. *Theoretical and Applied Climatology 99* (2009), 431–439. 79

[161] VAZ DE MELO, B., AND FREITAS LOPES, H. Data driven estimates for mixtures. *Computational Statistics and Data Analysis 47* (2004), 583–598. 179

[162] VIDAKOVIK, B. *Statistics for Bioengineering Sciences*, first ed. Springer, New York, 2011. 78

[163] WERNER, P., AND GERSTENGARBE, F. W. *Katalog der Grosswetterlagen Europas(1881-2009) nach Paul Hess und Helmuth Brezowsky*. Postdam Institut für Klimafolgenforschung, Postdam, Germany, 2010. 70

[164] WILBY, R. L., AND WIBLEY, T. M. L. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography 21* (1997), 530–548. 132

[165] WILLET, P. Molecular diversity techniques for chemical databases. *Information Research 2* (1996). 135, 156, 162

[166] XIE, S. L. *Scouring patterns in front of vertical breakwaters and their influence on the stability of the foundations of the breakwater. Reports of the Department of Civil Engineering*. Delft University of Technology, Delft, The Netherlands, 1981. 131

[167] ZHANG, P. Nonparametric importance sampling. *Journal of the American Statistical Association 91*, 435 (1996), 1245–1253. 62