
**Desarrollo de una metodología
de predicción a medio-largo plazo
de derrames de hidrocarburos
y su aplicación en la lucha
contra la contaminación marina**

**Ana J. Abascal Santillana
y col.**

Ayudas a la investigación 2011

Investigador Principal

Ana J. Abascal Santillana

Doctora por la Universidad de Cantabria
Investigadora de IH Cantabria

Equipo Investigador

Mar Cárdenas Milán

Licenciada en Ciencias del Mar
Investigadora predoctoral de la Universidad de Cantabria

Sonia Castanedo Bárcena

Doctora Ingeniera de Caminos, Canales y Puertos
Profesora titular de la Universidad de Cantabria

Raúl Medina Santamaría

Doctor Ingeniero de Caminos, Canales y Puertos
Catedrático de la Universidad de Cantabria.

Índice

	Página
RESUMEN	5
1. INTRODUCCIÓN Y ANTECEDENTES	5
1.1. Introducción	5
1.2. Antecedentes	7
1.2.1. Aplicación operacional	7
1.2.2. Aplicación probabilística	8
2. OBJETIVOS	8
3. MATERIALES Y METODOLOGÍA	9
3.1 Bases de datos de reanálisis	9
3.2 Clasificación y selección de patrones atmosféricos	9
3.3. Predicción probabilística de la evolución del viento y corrientes	11
3.3.1. Generación de la matriz de transición	11
3.3.2. Aplicación de cadenas de Markov	11
3.4. Predicción probabilística de la trayectoria de un derrame a medio-largo plazo	12
3.4.1. Modelo numérico de transporte de hidrocarburos	12
3.4.2. Simulación de trayectorias y cálculo de la probabilidad de contaminación	13
3.5. Esquema general de la metodología	13
4. RESULTADOS	15
4.1. Clasificación y selección de patrones atmosféricos	16
4.1.1. Clasificación y selección de los datos de viento	16
4.1.2. Clasificación y selección de los datos de presión	18
4.2 Predicción probabilística de la evolución de viento y corrientes	21
4.2.1. Generación de la matriz de transición	21
4.2.2. Selección del estado inicial de la cadena de Markov	21
4.2.3. Aplicación de la cadena de Markov	22
4.3 Predicción probabilística de las trayectorias del vertido	25
4.3.1. Evolución del derrame condicionado al tipo de tiempo del accidente	25
4.3.2. Evolución del derrame en función de los diferentes tipos de tiempo	29
5. DISCUSIÓN	32
6. CONCLUSIONES	33
7. FUTURAS LÍNEAS DE INVESTIGACIÓN	33

8.	ANEJO I. BASE DE DATOS DE REANÁLISIS	34
	AI.1. INTRODUCCIÓN	34
	AI.2. DATOS DE VIENTO	34
	AI.3. DATOS DE PRESIÓN	35
	AI.4. DATOS DE CORRIENTES DE MAREA METEOROLÓGICA	36
9.	ANEJO II. ANÁLISIS DE COMPONENTES PRINCIPALES	38
	AII.1. ANÁLISIS DE COMPONENTES PRINCIPALES	38
10.	ANEJO III. TÉCNICAS DE SELECCIÓN Y CLASIFICACIÓN	39
	AIII.1. INTRODUCCIÓN	39
	AIII.2 ALGORITMO DE MÁXIMA DISIMILITUD (MAXDISS)	39
	AIII.3 TÉCNICA DE CLASIFICACIÓN DE LAS K-MEDIAS (K-MEANS)	39
11.	REFERENCIAS BIBLIOGRÁFICAS	40

RESUMEN

La contaminación por vertidos de hidrocarburos es un problema creciente que supone la contaminación sistemática del litoral. Para el desarrollo de sistemas de prevención y lucha contra la contaminación, así como de los planes de contingencia, es necesario disponer de herramientas técnicas y metodológicas que ayuden al gestor de la crisis en la toma de decisiones. Una de estas herramientas es la predicción de la evolución de un derrame en el medio marino. Los sistemas operacionales existentes hoy en día permiten predecir la deriva de un vertido con un horizonte de predicción de 3 a 5 días. Esta predicción se vuelve insuficiente si se produce un derrame de hidrocarburos a kilómetros de la costa, y el vertido puede permanecer más de este período de tiempo a la deriva. En este caso sería necesario conocer cómo evolucionará y hacia dónde se dirigirá en un intervalo de tiempo mayor (15 días – 1 mes).

Con base en esta problemática, en este proyecto de investigación se ha desarrollado una metodología de predicción a medio-largo plazo de la evolución de un vertido de hidrocarburos en el medio marino. La metodología desarrollada permite relacionar las condiciones océano-meteorológicas actuales con bases de datos de reanálisis o históricas (viento, presión y corrientes) con el objetivo de simular su evolución a medio-largo plazo en términos estadísticos. Una vez conocida la evolución del campo atmosférico y oceanográfico, esta información es utilizada para predecir la evolución del derrame. La metodología se estructura en las siguientes partes:

- 1) **Clasificación y selección de patrones atmosféricos.** Como resultado se obtienen K estados de viento y M patrones de circulación atmosférica (o tipos de tiempo) representativos de la zona de estudio.
- 2) **Predicción probabilística de la evolución de viento y corrientes mediante cadenas de Markov.** En caso de accidente, se simulará la evolución a medio-largo plazo de las condiciones océano-meteorológicas actuales mediante la aplicación de cadenas de Markov, obteniéndose N simulaciones de la evolución del campo de viento y corrientes durante el período de predicción.
- 3) **Predicción probabilística de la evolución del vertido.** Los resultados obtenidos en el paso 2 se utilizan para simular N trayectorias del vertido durante un período de S días (15 días-1mes), utilizando un modelo de transporte de hidrocarburos. A partir de las N simulaciones se calcula la probabilidad de contaminación originada por el vertido.

La metodología desarrollada se ha aplicado en el Golfo de Vizcaya. Para ello se han utilizado datos de reanálisis de viento y corrientes de marea meteorológica en el período 1989-2009 y datos de presión atmosférica desde el año 1957 al 2011. Dichas bases de datos se han

clasificado, obteniéndose 49 estados de viento y 9 patrones de circulación atmosférica representativos de la zona de estudio. A partir de esta clasificación se ha generado la matriz de transición, de dimensiones $49 \times 49 \times 9$, base de la cadena de Markov. Con el objetivo de comparar los resultados con datos reales, se ha simulado la evolución de una mancha observada en el Golfo de Vizcaya (entre Galicia y Asturias), durante el accidente del *Prestige*. Para ello, se ha simulado durante 30 días su evolución a partir del día 27/11/2002, fecha del avistamiento. Los resultados obtenidos muestran que la mayor parte de las trayectorias se dirigen hacia el este, penetrando en el Golfo de Vizcaya. Las zonas con más probabilidad de contaminación son la costa de Asturias y Cantabria, alcanzando después de 1 mes la costa de Francia, acorde a la evolución del vertido en aquel período. Por otro lado, se ha estudiado como afecta el patrón de circulación atmosférica (relacionado con el paso de borrascas y anticiclones) en la evolución estadística de la trayectoria del vertido, encontrándose que la evolución del vertido y la probabilidad de contaminación depende del patrón de circulación atmosférica inicial.

Los resultados obtenidos muestran la capacidad de la metodología desarrollada y su utilidad para proporcionar información necesaria en la planificación y respuesta ante un derrame de hidrocarburos al medio marino. Una aplicación inmediata, de la que ya hay demanda en el sector, es su integración en los sistemas operacionales de predicción de deriva para dar respuesta en tiempo real. En caso de accidente, el sistema operacional proporcionaría la predicción a corto plazo (2-3 días) basada en las técnicas ya existentes, y la predicción estadística a medio-largo plazo (para los próximos 15 días – 1 mes) con base en la metodología presentada. La incorporación de esta metodología en los sistemas operacionales de predicción supone una innovación y una mejora en la planificación y respuesta frente a la contaminación por hidrocarburos en la costa y en el medio marino.

1. INTRODUCCIÓN Y ANTECEDENTES

1.1. Introducción

En enero de 2012 la Fundación MAPFRE concedió a Ana Julia Abascal Santillana del Instituto de Hidráulica Ambiental de la Universidad de Cantabria (IH Cantabria), la Ayuda a la Investigación 2011 para la realización del proyecto de investigación “Desarrollo de una metodología de predicción a medio-largo plazo de derrames de hidrocarburos y su aplicación en la lucha contra la contaminación marina”. En este documento se presenta el trabajo realizado en el marco de dicha Ayuda a la Investigación a lo largo de los doce meses de duración del proyecto.

El presente estudio está enmarcado en el área de la oceanografía operacional y el modelado numérico del transporte de hidrocarburos, y está enfocado principal-

mente al desarrollo de metodologías para mejorar los sistemas de prevención y respuesta frente a la contaminación por hidrocarburos. La contaminación por derrames de hidrocarburos es un problema creciente y global que supone la llegada al océano de millones de toneladas de estos productos cada año y representan la contaminación sistemática del litoral. En los últimos años, la contaminación se ha visto incrementada debido al aumento del transporte marítimo y de las actividades de extracción, explotación y gestión de hidrocarburos. Su origen procede de diversas fuentes de contaminación, como las operaciones de prospección, la carga y descarga de los hidrocarburos, los derrames durante su transporte e incluso las descargas intencionadas e ilegales en la limpieza de sentinas de los buques.

A lo largo de los años y como respuesta a la presión que ha experimentado y experimenta el medio marino ante este tipo de contaminación se han desarrollado diferentes convenios de ámbito internacional que previenen la contaminación de los océanos (MARPOL) o que se comprometen a tomar las medidas adecuadas para prepararse y luchar contra sucesos de contaminación por hidrocarburos (OPRC). En España, en la lucha contra la contaminación en el mar se dispone, entre otros instrumentos, del Plan Nacional de Contingencias por Contaminación Marina Accidental. En este plan se estipula la organización de los recursos humanos y materiales para dar respuesta a un suceso de contaminación marina, así como recomendaciones a las Comunidades Autónomas y otras entidades para la realización de sus respectivos planes de contingencia.

Para responder con rapidez y éxito a un accidente medioambiental de esas características, es necesario que las personas involucradas en la lucha contra la contaminación dispongan de herramientas, datos y conocimientos del medio que permitan una respuesta adecuada en cada situación. Una de estas herramientas es la predicción de la evolución de un derrame en el medio marino.

Es importante destacar que cuando un hidrocarburo es derramado en el medio marino este se mueve principalmente por efecto del viento, el oleaje y las corrientes, siendo estos los forzamientos océano-meteorológicos que condicionan su transporte o deriva. Por lo tanto, para poder predecir la evolución de un derrame, es necesario conocer previamente la evolución del viento, el oleaje y las corrientes superficiales. Actualmente, existen sistemas operacionales oceanográficos y meteorológicos basados en modelos numéricos que proporcionan predicciones de las condiciones océano-meteorológicas a corto plazo (2-3 días), lo que permite realizar la predicción de la evolución de un vertido como máximo en un período de 3 días. Esta predicción se vuelve insuficiente si se produce un derrame de hidrocarburos a kilómetros de la costa, y el vertido puede permanecer más de ese periodo a la deriva. En este caso sería necesario conocer cómo evolucionará y hacia dónde se dirigirá en un intervalo de tiempo mayor (por ejemplo 1 mes). Hoy en día, los sistemas operacionales no proporcionan predicciones océano-meteorológi-

cas a esa escala temporal, por lo que no se puede realizar operacionalmente la predicción de deriva de un vertido con ese horizonte de predicción.

Para cubrir estas carencias, en este trabajo se está investigando y desarrollando una metodología que permita predecir en términos estadísticos la deriva del derrame a medio-largo plazo. Tal y como se comentó en la propuesta, el trabajo que se está realizando se enmarca en la línea de investigación que el IH Cantabria viene desarrollando en los últimos diez años en relación al desarrollo e implementación de metodologías para ser aplicadas en la lucha contra la contaminación por hidrocarburos en el medio marino. En este campo, IH Cantabria tiene una línea de investigación en la predicción de trayectorias de vertidos a corto plazo y otra línea orientada a la predicción de la trayectoria de vertidos a medio-largo plazo. Esta última está siendo financiada por otros dos proyectos de investigación: el proyecto "PLVMA" del Plan Nacional (TRA2011-28900) y el proyecto "AZIMUT-DERRAMES" desarrollado por IH Cantabria en colaboración con las empresas ACCIONA e IBERDROLA.

En este proyecto de investigación se ha desarrollado una metodología que combinando técnicas de clasificación, métodos de análogos, esquemas de "*Tipos de Tiempo*" (*Weather Typing Schemes*) y procesos estocásticos permite calcular en términos estadísticos la evolución de las condiciones océano-meteorológicas (en un período comprendido entre 15 días y 1 mes) y la evolución de un vertido sometido a estas condiciones ambientales. La evolución a futuro de estas condiciones océano-meteorológicas (viento y corrientes), se calculará con base en el conocimiento que se tiene de cómo se han comportado dichas variables en el pasado.

Para estudiar el comportamiento histórico de las variables océano-meteorológicas, es necesario disponer de series de datos que presenten una larga duración temporal y una elevada cobertura espacial. Actualmente existen redes de medidas que proporcionan información *in-situ* de estas variables. Sin embargo, estos datos instrumentales presentan una serie de limitaciones para el estudio propuesto, como duración temporal corta (estas series no suelen superar los 15 años), presentar huecos (no ser homogéneas) o ser medidas puntuales (especialmente). Para suplir estas carencias, en los últimos años se han utilizado modelos numéricos para generar bases de datos de variables oceanográficas o meteorológicas (Kalnay *et al.*, 1996; Ratsimandresy *et al.*, 2008; Menéndez *et al.*, 2011; Abascal *et al.*, 2011), llamadas bases de datos de reanálisis, que se caracterizan por su amplia cobertura espacial y temporal.

Por lo tanto, las bases de datos de reanálisis se caracterizan por ser series de datos de variables oceanográficas o meteorológicas, que se generan a partir de modelos numéricos capaces de simular los procesos que dan lugar a la generación de dichas variables a partir de forzamientos atmosféricos u oceánicos. De este modo, se dispone de series históricas de las variables de mayor duración temporal que los existentes en las bases instru-

mentales y de una mayor resolución espacial. Lo que permite realizar un análisis homogéneo y una caracterización estadística más fiable.

En este estudio, las bases de datos de reanálisis se utilizan para caracterizar las condiciones océano-meteorológicas a partir del comportamiento y evolución de los patrones meteorológicos y oceanográficos en el pasado. Una vez conocido el comportamiento en el pasado, esta información es utilizada para predecir cómo evolucionará, una situación océano-meteorológica actual, a medio-largo plazo y con ello predecir de forma probabilística la evolución del derrame.

1.2. Antecedentes

En los últimos años, se han producido diversos derrames por hidrocarburos en el litoral español. Estos accidentes medioambientales han puesto de manifiesto la necesidad de que las personas que participan en la lucha contra la contaminación marina dispongan de herramientas, datos y conocimientos del medio, para poder dar una respuesta rápida y adecuada en caso de producirse una situación de este tipo. Una de estas herramientas es la predicción de la evolución de un derrame mediante modelos numéricos que resuelvan el transporte del hidrocarburo y su degradación.

Cuando un hidrocarburo es derramado en el mar, los principales factores ambientales que definen su movimiento son: el arrastre del viento, las corrientes marinas superficiales, el arrastre neto de las partículas debido al movimiento oscilatorio del oleaje y la turbulencia (ASCE, 1996; Reed *et al.*, 1999). La mayor parte de los modelos de transporte de hidrocarburos utilizados hoy en día resuelven la ecuación de transporte mediante métodos lagrangianos. Estos métodos están basados en el seguimiento de partículas que representan la mancha del hidrocarburo y que se mueven por el efecto combinado del viento, oleaje, corrientes y turbulencia, por lo que se acercan a la física del problema de una forma más natural que los métodos eulerianos (Maier-Reimer, 1982; Hunter, 1987; James, 2002).

Actualmente, la simulación de vertidos utilizando la técnica de seguimiento de partículas para la advección y el procedimiento de *random walk* para la difusión es la más considerada para simular el transporte de hidrocarburos (Daniel, 1996; Al-Rabeh *et al.*, 2000; Korotenko *et al.*, 2000; Comerma, 2004; Abascal, 2009, entre otros). Además de ser transportados, los hidrocarburos derramados en el medio marino se ven sometidos a un conjunto de procesos físico-químicos que alteran notablemente sus propiedades (densidad y viscosidad) (ASCE, 1996). La variedad de modelos existentes actualmente varían desde modelos bidimensionales que simulan únicamente el transporte superficial a modelos tridimensionales que simulan el transporte y la degradación en la columna de agua.

En la aplicación de los modelos de transporte y degradación de hidrocarburos en la lucha contra la contamina-

ción marina se han desarrollado dos líneas principales de actuación: (1) la primera consiste en la utilización de los modelos de transporte en tiempo real y de forma operacional, proporcionando información sobre la evolución del contaminante en el transcurso de la crisis; (2) la segunda, consiste en utilizar los modelos numéricos para proporcionar información probabilística y estadística, permitiendo al gestor planificar y tomar decisiones teniendo en cuenta la probabilidad de riesgo de las zonas a proteger.

1.2.1. Aplicación operacional

Entre los modelos del estado del conocimiento que actualmente se utilizan en situación de emergencia y de forma operacional se puede destacar: OILMAP (Spaulding *et al.*, 1992), MEDSLIK (Lardner *et al.*, 1998), GNOME (Beegle-Krause, 1999), MOHID (Miranda *et al.*, 2000), MOTHY (Daniel *et al.*, 1996), GULFSPILL (Al-Rabeh *et al.*, 2000), ARLEQUIN (Comerma, 2004), POSEIDON OSM (Nittis *et al.*, 2006), TESEO (Abascal *et al.*, 2007).

Para la aplicación de los modelos de transporte en situación de emergencia, es necesario disponer de sistemas operacionales de predicción atmosféricos y oceanográficos que generen los forzamientos (viento y corrientes) que utiliza el modelo de transporte para simular la deriva del vertido. Por lo tanto, la aplicación de los modelos de transporte de hidrocarburos en tiempo real y en situación de emergencia ha ido ligada a la evolución de la oceanografía operacional. En el caso particular de España, el desarrollo de la oceanografía operacional experimentó un fuerte impulso después del accidente del *Prestige*. Uno de los proyectos impulsores ha sido el proyecto ESEOO (VEM2003-C14-03) (Álvarez-Fanjul *et al.*, 2007). Como parte de los objetivos alcanzados por el proyecto ESEOO, se desarrolló un sistema operacional de predicción de corrientes a escala regional (Sotillo *et al.*, 2008) y el sistema operacional de predicción y transporte de hidrocarburos TESEO (Abascal *et al.*, 2007). El sistema TESEO ha sido utilizado en tiempo real y de forma operacional en diferentes ejercicios de lucha contra la contaminación realizados por Salvamento Marítimo (Abascal *et al.*, 2007; Sotillo *et al.*, 2008; Abascal *et al.*, 2008), lo que muestra el interés creciente de las instituciones gestoras de la crisis en el desarrollo y aplicación de este tipo de modelos.

Los avances en este campo se han visto fomentados con el desarrollo de otros proyectos, como el Proyecto Singular Estratégico PROMARES (PSE-310000-2009-03) (2011) para la protección medioambiental de la costa y del medio marino. El subproyecto OCTOPOS que se desarrolló en el marco del proyecto PROMARES, cuyo objetivo consistió en la implementación de nuevos desarrollos tecnológicos que contribuyeron a la creación de un sistema integral de detección, seguimiento y predicción de vertidos de hidrocarburos e identificación de posibles infractores en zonas costeras, sistemas portuarios, instalaciones singulares y espacios naturales protegidos.

A nivel europeo también se ha fomentado el desarrollo de sistemas operacionales y de herramientas para la

lucha contra la contaminación. Así lo demuestra la financiación de proyectos como ARCOPOLplus (2012-2013), financiado por el Programa Transnacional Espacio Atlántico, cuyo objetivo es reforzar la capacidad de prevención y respuesta de las regiones atlánticas ante episodios de contaminación costera accidental y el proyecto SPRES (*Oil Spill Prevention and Response at Local Scales*) (2012-2014), proyecto INTERREG del programa Área Atlántica, cuyo objetivo es desarrollar herramientas que ayuden a la planificación y a la respuesta ante derrames de hidrocarburos en la escala de actuación local (estuarios y puertos).

1.2.2. Aplicación probabilística

Con respecto a la aplicación probabilística de los modelos de transporte de hidrocarburos, su uso ha estado orientado principalmente a la planificación de la respuesta ante un derrame. Relacionado con este tema, destacan los trabajos de análisis probabilístico de derrames utilizados por la NOAA que tiene implementada una aplicación de estas características en las Bahías de San Francisco, San Diego y Kaneohe, en los lagos Calcasieu y Sabine y, fuera de Estados Unidos, en el Golfo Pérsico y en el puerto de Inchon, Corea (Barker y Galt, 2000). También en EEUU, el MMS (Mineral Management Service) utiliza esta técnica para estimar probabilidades de contaminación en el NE del Golfo de México (Guillen *et al.*, 2004). Otros trabajos a destacar son Condie *et al.* (2005) donde se analiza estadísticamente la conectividad espacial entre zonas costeras de Australia basándose en las corrientes, y Skognes y Johansen (2004) que desarrollan una herramienta de predicción de contaminación para la costa Noruega. En España la aplicación probabilística de los modelos de hidrocarburos no está muy extendida. Se puede destacar el trabajo desarrollado por Díaz *et al.* (2008), en el que se utiliza un modelo de transporte lagrangiano para simular estadísticamente el accidente del *Prestige* y el trabajo desarrollado por Abascal (2009) y Abascal *et al.* (2010) que desarrolla una metodología para implementar un sistema probabilístico de riesgo contra la contaminación en el Golfo de Vizcaya.

La revisión del estado del conocimiento, muestra que en la respuesta ante un derrame de hidrocarburos se dispone de modelos numéricos y metodologías que permiten predecir la deriva de un vertido con un horizonte temporal de aproximadamente 3 días, y obtener información a más largo plazo con base en métodos estadísticos. Por lo tanto, hoy en día, no existen metodologías que permitan predecir la evolución de un vertido de hidrocarburos en el medio marino en el medio-largo plazo (15 días-1mes). En este estudio, se investiga la aplicación de técnicas de selección y clasificación combinadas con procesos estocásticos para realizar la predicción de las trayectorias de un derrame a medio-largo plazo.

Entre las técnicas de clasificación y selección existentes se pueden citar el método de análogos (Lorenz, 1969), *K-Medias* (MacQueen, 1967), redes auto-organizativas (Kohonen, 2000) y análisis de componentes principales

(Lorenz, 1969). Estas técnicas han sido aplicadas con éxito, en disciplinas como meteorología (Gutierrez *et al.*, 2004) y clima marítimo (Méndez *et al.*, 2009; Camus, 2009; Camus *et al.*, 2010), sin embargo no han sido utilizadas previamente para la predicción de la trayectoria de un derrame en el medio marino. Con base en la experiencia adquirida en esas disciplinas, dichas técnicas han sido investigadas y utilizadas para desarrollar la metodología objeto de este estudio.

Para la predicción de la trayectoria de derrames, se ha investigado la aplicación de la técnica esquemas de "Tipos de Tiempo" (*Weather Typing Schemes*) (Camus, 2009; Ancell *et al.*, 2009) y los procesos estocásticos, concretamente la cadenas de Markov (Sperandio y Coelho, 2006; Rincón, 2011). La técnica de *Weather Types* está basada en el concepto de climatología sinóptica tradicional y relaciona un estado atmosférico particular con un conjunto de variables locales climatológicas u oceanográficas. Esta técnica ha sido empleada principalmente en el campo de la meteorología, para estudios relacionados con climatología (Sheridan, 2001; Moron *et al.*, 2008; Guancho *et al.*, 2013) y cambio climático (Bardossy *et al.*, 1990; Semenov *et al.*, 1997). Aunque también se pueden encontrar trabajos donde se aplica para caracterizar el clima marítimo (Camus, 2009; Izaguirre, 2010) o el clima de transporte de sedimentos (Cánovas, 2012). Con respecto a la técnica estadística de las cadenas de Markov, esta se caracteriza por ser un proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende del evento inmediatamente anterior. Tiene múltiples aplicaciones en campos tan distintos como la física o la epidemiología. Asimismo, se ha aplicado con éxito en disciplinas como la meteorología, como lo demuestran los trabajos de Katz (1977), Pang (2001) y Gutierrez *et al.* (2004).

En este trabajo de investigación, las técnicas previamente descritas son investigadas y utilizadas para desarrollar una metodología que permita predecir estadísticamente la trayectoria de un vertido y calcular la probabilidad de contaminación. La metodología desarrollada se ha aplicado en el Golfo de Vizcaya y se ha utilizado para simular la trayectoria de manchas observadas durante el accidente del buque *Prestige* (2002).

2. OBJETIVOS

El objetivo fundamental de este trabajo es la investigación y el desarrollo de una metodología que permita predecir a medio-largo plazo la evolución de derrames en el mar, con el objetivo de mejorar los sistemas de prevención y respuesta frente a la contaminación por hidrocarburos en la costa y en el medio marino.

En caso de accidente, la metodología desarrollada permitirá predecir estadísticamente la evolución de las condiciones océano-meteorológicas a medio-largo plazo y con ello calcular la probabilidad de contaminación del vertido.

Para alcanzar este objetivo general se plantean los siguientes objetivos específicos:

- Revisar las bases de datos históricas (reanálisis) de forzamientos hidrodinámicos y meteorológicos disponibles, y estudiar su idoneidad para este estudio.
- Investigar diferentes técnicas estadísticas y seleccionar la más apropiada para relacionar situaciones océano-meteorológicas actuales con bases de datos de reanálisis.
- Desarrollo de una metodología basada en procesos estocásticos que proporcione la predicción océano-meteorológica y la deriva de un derrame a medio-largo plazo.
- Aplicación de la metodología desarrollada a una zona piloto con riesgo de contaminación.

3. MATERIALES Y METODOLOGÍA

3.1 Bases de datos de reanálisis

En este apartado se describen las bases de datos de reanálisis seleccionadas para el desarrollo y aplicación de la metodología.

- **Datos de viento:** se ha utilizado la base de datos de reanálisis SeaWind-EraInterim (Menéndez *et al.*, 2011), que proporciona velocidad y dirección del viento con una resolución temporal horaria en el período de tiempo comprendido entre 1989-2009. Esta base de datos ha sido forzada con los datos de reanálisis ERA-Interim mediante la utilización del modelo WRF-ARW 3.1.1 (Weather Research and Forecasting, WRF y Advanced Research dynamical solver, ARW) (Skamarock *et al.*, 2008). Los datos abarcan el dominio europeo y presentan una resolución espacial de 0.15° (~15 km).

- **Datos de presión atmosférica:** se ha utilizado el reanálisis global NCEP/NCAR (Kalnay *et al.*, 1996). Esta base de datos ha sido realizada por el National Center for Environmental Prediction (NCEP) y el National Center for Atmospheric Research (NCAR). Los datos han sido aportados por NOAA/OAR/ES-RLPSD, Boulder, Colorado, USA y se han obtenido de su página web <http://www.esrl.noaa.gov/psd/>. La malla empleada en el reanálisis global NCEP/NCAR, abarca prácticamente todo el globo terrestre (latitud: 90° N-90° S, longitud: 0° E-357.5° E). Los datos presentan una resolución espacial 2.5° y una resolución temporal 6-horaria de los campos de presión, cubriendo el período de tiempo desde 1948 hasta la actualidad.
- **Datos de corrientes de marea meteorológica:** se ha utilizado la base de datos de corrientes de marea meteorológica en el Sur de Europa procedentes del reanálisis GOS (Global Ocean Surges) (Abascal *et al.*, 2010; Abascal *et al.*, 2011). Esta base de datos ha sido generada por IH Cantabria para la costa suroeste de Europa mediante la utilización del modelo tridimensional de circulación ROMS (Regional Ocean Modelling System) (Shchepetkin *et al.*, 2005). El GOS 2.1 se caracteriza por haber sido forzado con campos de viento y presión del SeaWind-EraInterim. Presenta una resolución espacial de 1/8° (12 km) y una resolución temporal horaria, cubriendo el período de tiempo comprendido entre 1989 y 2009.

En la tabla siguiente se presenta de forma resumida la información de las bases de datos descrita en los párrafos anteriores. Para más información consultar el "Anejo I", donde se hace una descripción detallada de las bases de datos empleadas.

Tabla 1. Resumen de las bases de datos: NCEP/NCAR, SeaWind-EraInterim y GOS 2.1.

Bases de Datos	Forzamiento	Mallas	Período Tiempo	Resolución Espacial	Resolución temporal
NCEP/NCAR	Global	1948-2010	250 km	6 horario
SeaWind-EraInterim	ERA-Interim	Europa	1989-2009	15 km	1 hora
GOS 2.1	SeaWind-EraInterim	Europa	1989-2009	12 km	1 hora

3.2 Clasificación y selección de patrones atmosféricos

El primer paso en la metodología propuesta consiste en estudiar los patrones atmosféricos (viento y presión a nivel del mar) que caracterizan la zona de estudio y que afectan al transporte de un hidrocarburo derramado en el medio marino. Para ello, mediante técnicas de clasificación y selección, se obtendrán los estados de viento y los

patrones de circulación atmosférica estadísticamente representativos de la zona de estudio. Dada la escala temporal de la predicción (~ 1 mes) y la variabilidad temporal de las variables viento y presión, se trabaja con medias diarias del viento y medias mensuales de la presión.

Con el objetivo de reducir la dimensionalidad de los datos y aumentar la eficiencia de la técnica de clasificación (Barnett y Presendorfer, 1978; Gutierrez *et al.*, 2004;

Brands *et al.*, 2010), antes de realizar la clasificación, se aplica la técnica de análisis de componentes principales (PCA) a los datos de reanálisis (Lorenz, 1969; Camus *et al.*, 2011), descrita en detalle en el Anejo II.

El análisis de las componentes principales es una técnica para reducir la dimensionalidad de una muestra de datos que presenta redundancia de información reteniendo la máxima variabilidad existente. Ya que es habitual que en los espacios de alta dimensionalidad, las distintas componentes o dimensiones suelen estar correlacionadas y, por tanto, se emplea información redundante en su descripción. La compresión de la información se consigue mediante la proyección de los datos en un espacio de menor dimensión que el original conservando el máximo de la varianza de la muestra. La base del nuevo espacio se forma con aquellos vectores donde la muestra proyectada presenta mayor varianza (Camus, 2009). Para seleccionar el número óptimo de PCs debe existir un equilibrio entre la reducción de información deseada y la calidad de la aproximación resultante. Por lo que interesa quedarse con un número de PCs que expliquen un alto porcentaje de la varianza total de los datos, sin que se llegue a perder información. Y de este modo, cumplir el objetivo de reducir la dimensionalidad de los datos de partida sin pérdida de información.

Una vez aplicada la técnica de PCA, se realiza la clasificación y obtención de los estados de viento en la zona de estudio. Las técnicas de clasificación y selección extraen información básica de las bases de datos mediante la detección o selección de un grupo reducido de datos representativo de toda la base de datos de partida. Estas técnicas exploran el espacio de los datos pero con objetivos diferentes. Las técnicas de clasificación dividen el conjunto de datos en subconjuntos siguiendo un criterio de similitud, mientras que las técnicas de selección identifican un conjunto de elementos de la base original representativos de la diversidad de los datos.

El método estándar para la búsqueda de un conjunto de vectores análogos a un vector dado es el método de *k-vecinos*, que consiste en seleccionar el conjunto de los *k* vectores más próximos al vector dado según una métrica prefijada (normalmente la métrica Euclídea). Basándonos en el análisis previo de las diversas técnicas de clasificación, se ha utilizado una técnica más compleja que el tradicional método de *k-vecinos*, la técnica basada en *K-Medias* (MacQueen, 1967; Hastie *et al.*, 2001) (descrita en detalle en el Anejo III).

El algoritmo *K-Medias* divide el conjunto de datos de partida en un determinado número de subconjuntos, define cada uno de ellos por un prototipo, y forma cada grupo con los datos para los cuales ese prototipo es el más cercano. Para ello, en primer lugar se establece el número de grupos deseado para inicializar los prototipos (o centroides) y a continuación, el algoritmo procede de forma iterativa moviendo estos centroides hasta minimizar la varianza total intra-grupos. Esta varianza total intra-grupos significa que al final del proceso, la suma de

distancias entre el prototipo y los datos de entrada correspondientes a cada prototipo es mínima (Camus, 2009).

Por lo tanto, el objetivo de la aplicación de la técnica es obtener un número *M* de subconjuntos o grupos definidos cada uno de ellos por un prototipo o centroide $v_k = \{v_{1k}, \dots, v_{nk}\}$ de la misma dimensión que los datos de partida, siendo $k = 1, \dots, M$. El proceso de clasificación consiste en un ajuste iterativo de los centroides. Los centroides iniciales $\{v_{10}, v_{20}, \dots, v_{M0}\}$ con los que comienza el proceso se han seleccionado mediante la aplicación de la técnica *MaxDiss* (Kennard y Stone, 1969) (descrita en detalle en el Anejo III) a los datos de partida. Este algoritmo tiene como objetivo identificar aquel subconjunto *M*, que represente la mayor diversidad posible del total de la base *N* de datos. Este algoritmo comienza con la inicialización del subconjunto mediante la elección de un dato de la base de origen. Los siguientes elementos del subconjunto seleccionados, se corresponden con los datos aún pertenecientes a la base de datos de partida que presenta mayor disimilitud con respecto al subconjunto seleccionado.

Una vez inicializado el proceso de clasificación, en cada ciclo se identifican los datos de entrada más próximos a cada centroide y el nuevo centroide se redefine como la media de los vectores de entrada correspondientes a dicho centroide. Por ejemplo, si en la iteración (*r*+1), cada vector *x_i* es reasignado al grupo *j*, el centroide es actualizado según:

$$v_j^{r+1} = \sum_{x_i \in C_j} \frac{x_i}{n_j} \quad (\text{Ec.1})$$

donde n_j es el número de vectores de entrada pertenecientes al grupo *j*, y dicho grupo se puede definir como:

$$j = \min \{x_i - v_j, j = 1, \dots, M\} \quad (\text{Ec.2})$$

Después de varias iteraciones, el proceso converge cuando la pertenencia de los vectores a un grupo determinado se estabiliza.

Como resultado de la clasificación de las bases de datos de viento y presión se obtienen: (1) *K* grupos que corresponden a los estados de viento y (2) *M* grupos que corresponden a los patrones de circulación atmosférica (tipos de tiempo) más representativos de los datos de partida. Los patrones de circulación proporcionan información sobre el paso de borrascas o anticiclones por la zona de estudio. A partir de esta información se puede estudiar cómo estas condiciones atmosféricas afectan a la evolución de cada estado de viento. Es decir, dado un campo de viento, ¿evolucionará de forma diferente en función del patrón de circulación atmosférica inicial?, ¿afectará esto a la trayectoria del vertido? Para tener en cuenta estas cuestiones, en este trabajo se calculará la evolución de los estados de viento teniendo en cuenta el patrón de circulación atmosférica existente.

3.3. Predicción probabilística de la evolución del viento y corrientes

Una vez realizada la clasificación de la base de datos atmosféricos (viento y presión), el siguiente paso consiste en su utilización para obtener la predicción probabilística de la evolución del campo de viento y de corrientes a medio-largo plazo. Para ello se aplicará una técnica estadística basada en procesos aleatorios, conocida como cadenas de Markov (*en adelante CM*).

En el siguiente apartado se describe la técnica y los pasos a seguir.

3.3.1. Generación de la matriz de transición

Las *CM* son un tipo especial de procesos estocásticos, que se caracterizan por ser una sucesión de ensayos similares u observaciones en la cual cada ensayo tiene el mismo número finito de resultados posibles y en donde la probabilidad de cada resultado para un ensayo dado depende sólo del resultado del ensayo inmediatamente precedente y no de cualquier resultado previo. Estas poseen la siguiente propiedad (Rincón, 2011):

- **Propiedad de Markov:** suponiendo conocido el estado presente del sistema, los estados anteriores no tienen influencia en los estados futuros del sistema.

Una cadena de Markov es una sucesión de variables aleatorias $X_t, i \in \mathbf{N}$, que se caracteriza por tener S espacios de estados discretos y T conjuntos de instantes de tiempo también discretos, $T = \{t_0, t_1, t_2, \dots\}$, el proceso se puede expresar de la forma siguiente:

$$P \left[X_{t+1} = j / X_0, \dots, X_t \right] = P \left[X_{t+1} = j / X_t \right] \quad (\text{Ec.3})$$

que es la expresión algebraica de la propiedad de Markov para un tiempo T discreto.

Las *CM* están completamente caracterizadas por las probabilidades de transición en una etapa,

$$P \left[X_{t+1} = j / X_t = i \right], i, j \in S, t \in T \quad (\text{Ec.4})$$

En este caso se trabaja con *CM* homogéneas en el tiempo, que son aquellas en las que

$$\forall i, j \in S, \forall t \in T, P \left[X_{t+1} = j / X_t = i \right] = q_{ij} \quad (\text{Ec.5})$$

donde q_{ij} se llama probabilidad de transición en una etapa, desde el estado i hasta el estado j .

Dichas probabilidades satisfacen las siguientes condiciones:

$$\forall i, j \in S, q_{ij} \in [0, 1] \quad (\text{Ec.6})$$

$$\forall i \in S, \sum q_{ij} = 1 \quad (\text{Ec.7})$$

Los q_{ij} se agrupan en la denominada matriz de transición de la *CM*:

$$Q = \begin{pmatrix} q_{00} & q_{01} & \dots \\ q_{10} & q_{11} & \dots \\ \dots & \dots & \dots \end{pmatrix} = (q_{ij})_{i, j \in S} \quad (\text{Ec.8})$$

En caso de que la matriz de transición satisfaga las dos propiedades anteriores, se dice que esta matriz es estocástica.

Las transiciones entre estados se representan mediante grafos. A continuación, se muestra un ejemplo de una matriz de transición Q , con su grafo correspondiente. En este grafo se muestra la probabilidad de estar en la primera etapa pasar a la siguiente ($q_{01}=0.1$), la probabilidad de quedarse en la misma etapa ($q_{00}=0.9$ y $q_{11}=0.7$) y de estar en la segunda etapa pasar a la primera ($q_{10}=0.3$) (véase la figura 1).

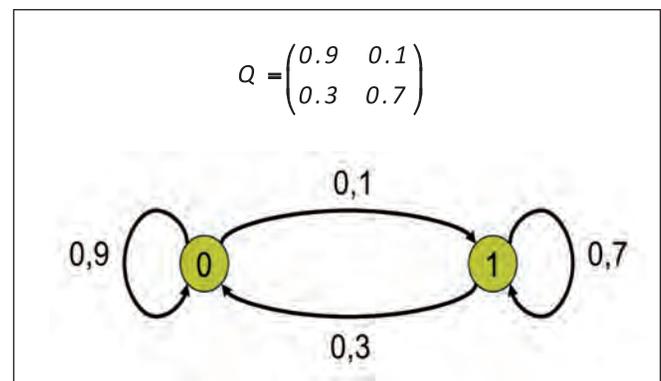


Figura 1. Ejemplo de grafo de una matriz de transición.

Por lo tanto, para la aplicación de las cadenas de Markov, tal y como se propone en este estudio, es necesario generar la matriz de transición. Dicha matriz se calcula a partir de la clasificación presentada en la sección anterior (3.2). En el caso de estudio, la matriz de transición se caracteriza por ser una matriz tridimensional, de $K \times K \times M$ dimensiones, donde K es el número de estados de viento y M el número de patrones de presión. Asumiendo que en un tiempo particular t , las condiciones atmosféricas corresponden al clúster ' i ', el elemento q_{ijm} de la matriz de transición proporcionará la probabilidad de pasar del estado de viento ' i ' al estado de viento ' j ' en el próximo instante de tiempo, teniendo en cuenta que el patrón de circulación atmosférica inicial ' M '.

3.3.2. Aplicación de cadenas de Markov

Una vez calcula la matriz de transición, el siguiente paso es la aplicación de cadenas de Markov para predecir el campo de viento y corrientes durante un horizonte de predicción t y con una resolución temporal diaria.

Es importante recalcar que las *CM* son un caso simple de un proceso estocástico, por lo que el resultado en cualquier etapa contiene algún elemento que depende del azar. En este estudio, para *acotar la aleatoriedad* en la cadena de Markov, la matriz de transición está condicionada al patrón de circulación atmosférica, es decir, el paso de un estado de viento a otro se hará teniendo en cuenta el patrón de circulación del estado inicial. No obstante, dada la aleatoriedad del proceso es importante repetirlo un número N de veces. Teniendo en cuenta que cada una de las N repeticiones presenta una longitud igual al período de tiempo (t) que se desea predecir. En este estudio, se desea predecir a medio-largo plazo, por lo que el período de tiempo será del orden de 15 días – 1 mes.

Por lo tanto, los pasos que presentaría el proceso serían los siguientes:

1. En primer lugar, a partir de los estados de viento se genera la matriz de transición (Ec.8), que está formada por M submatrices (una por cada patrón de circulación). Cada una de estas submatrices contiene la probabilidad condicionada (al patrón de circulación) de pasar de un estado de viento a otro.
2. Se parte de un estado inicial para comenzar el proceso, válido para las N repeticiones. La inicialización de la *CM* no es aleatoria, se selecciona el patrón de circulación atmosférica y el estado de viento análogo a la situación meteorológica actual (la cuál puede ser proporcionada por modelos numéricos de predicción, medidas en tiempo real, etc.).
3. Cada N repetición consta de un número determinado de pasos de tiempo (resolución temporal diaria) y tiene una longitud temporal igual al tiempo total de simulación. Dado un horizonte de predicción de t días y que se está trabajando con una resolución temporal diaria, el número de pasos de tiempo será t .
4. En cada paso de tiempo se calcula una probabilidad de forma aleatoria.
5. Dada la probabilidad y el estado i en un instante t del proceso, se pasa al estado j en el instante $t+1$ utilizando la probabilidad de paso de un estado a otro proporcionada por la matriz de transición.
6. En cada paso de tiempo se va guardando el estado de viento al que se llega.
7. El resultado es una matriz de $S \times N$, siendo S el número de pasos de tiempo y N el número de repeticiones. Esta matriz contiene la predicción probabilística de los campos de viento.
8. A partir de los campos de viento se obtienen las corrientes de marea meteorológica asociadas a dichos campos, obtenidas de la base de datos histórica.

Este proceso se repetirá N veces y cada repetición constará de S pasos de tiempo. El resultado final son los estados de viento y corrientes por los que se van pa-

sando en cada paso de tiempo para cada una de las N simulaciones. Se obtiene de este modo la predicción probabilística de la evolución temporal de los campos de viento y corrientes.

3.4. Predicción probabilística de la trayectoria de un derrame a medio-largo plazo

Para predecir probabilísticamente la trayectoria del derrame se utiliza la evolución de las condiciones oceanográficas calculadas previamente en la sección 3.3. De este modo, con las condiciones de viento y corrientes se fuerza un modelo de transporte de hidrocarburos y se obtiene la evolución del derrame a medio-largo plazo.

3.4.1. Modelo numérico de transporte de hidrocarburos

Para simular la trayectoria del vertido se utiliza el modelo numérico de transporte de hidrocarburos TESEO (Abascal *et al.*, 2007; Abascal *et al.*, 2008). El modelo TESEO es un modelo de transporte y degradación de hidrocarburos, desarrollado por IH Cantabria, para su utilización operacional en la lucha contra la contaminación por hidrocarburos en el medio marino. El modelo TESEO ha sido desarrollado en el marco de dos proyectos: el proyecto ESEOO, financiado por el Ministerio de Ciencia y Tecnología y el proyecto “Programa de predicción, protección y evaluación del vertido del buque *Prestige* en las costas de Cantabria”, financiado por el Gobierno de Cantabria.

Está constituido por un modelo de transporte y un modelo de envejecimiento del hidrocarburo. En este estudio, únicamente se utiliza el modelo de transporte, que se caracteriza por ser un modelo Lagrangiano bidimensional, que calcula el movimiento tanto de hidrocarburos como de objetos flotantes a la deriva.

El modelo de transporte está basado en el modelo PICHl, desarrollado por la Universidad de Cantabria durante el accidente del *Prestige* (Castanedo *et al.*, 2006). El transporte se simula mediante la suma de partículas que se mueven de forma independiente, por efecto del viento, oleaje, corrientes y la difusión turbulenta. El modelo Lagrangiano resuelve la siguiente ecuación:

$$\frac{dx_i}{dt} = U_a(x_i, t) + U_d(x_i, t) \quad (\text{Ec.9})$$

donde

- x_i es la coordenada (x, y) de la partícula i
- U_a es la velocidad advectiva de la partícula en esa coordenada, y
- U_d es la velocidad difusiva.

En este estudio, la velocidad advectiva, U_a , se calcula como la combinación lineal de la acción del viento y de las corrientes, siguiendo la siguiente expresión:

$$U_a(t) = U_c + C_D U_v \quad (\text{Ec.10})$$

donde

- U_c es la velocidad superficial de la corriente

- U_v es la velocidad del viento a 10 m de la superficie
- C_D es el coeficiente de arrastre del viento, con un valor del 3% (ASCE, 1996).

La velocidad turbulenta se obtiene mediante un proceso de Monte Carlo en el que la velocidad de difusión varía en el intervalo $[\vec{U}_d, \vec{U}_d]$ y se considera proporcional a los coeficientes de difusión (Hunter *et al.*, 1993). En cada paso de tiempo, el desplazamiento difusivo aleatorio de una partícula se define como:

$$|\vec{U}_d| = \sqrt{\frac{6D}{\Delta t}} \quad (\text{Ec.11})$$

donde D es el coeficiente de difusión comprendido en el intervalo 1-100 m²/s (ASCE, 1996).

Cabe destacar que el modelo de transporte ha sido validado y calibrado mediante la utilización de boyas de deriva lanzadas durante el accidente del *Prestige* y en diferentes ejercicios de lucha contra la contaminación realizados por Salvamento Marítimo (Abascal *et al.*, 2007; Sotillo *et al.*, 2008; Abascal *et al.*, 2009). Concretamente, se ha aplicado satisfactoriamente en ejercicios operacionales realizados tanto en el Mediterráneo (Ejercicios “Me-

diterráneo 2005”, “Mediterráneo 2008”), como en el Atlántico (Ejercicios “Gijón 2006”, “Finisterre 2006”, “Gascogne 2007”, “Vigo 2007” y “Santander 2010”).

3.4.2. Simulación de trayectorias y cálculo de la probabilidad de contaminación

La predicción probabilística de la evolución del vertido se obtiene aplicando el modelo numérico de transporte de hidrocarburos descrito previamente.

Para ello, se fuerza el modelo con el resultado de la predicción probabilística de los forzamientos océano-meteorológicos (viento y corrientes). Estos forzamientos se presentan como una matriz de $S \times N$, siendo S el número de pasos de tiempo y N el número de repeticiones. Por lo tanto, se tienen N condiciones de viento y corrientes que dan lugar, una vez se aplica el modelo de transporte, a N trayectorias equiprobables del derrame. Las N simulaciones son utilizadas para calcular la evolución del vertido en términos probabilísticos durante el período de predicción y determinar la probabilidad de contaminación.

En la figura 2 se muestra un esquema de la metodología para la obtención de la predicción probabilística de trayectorias:

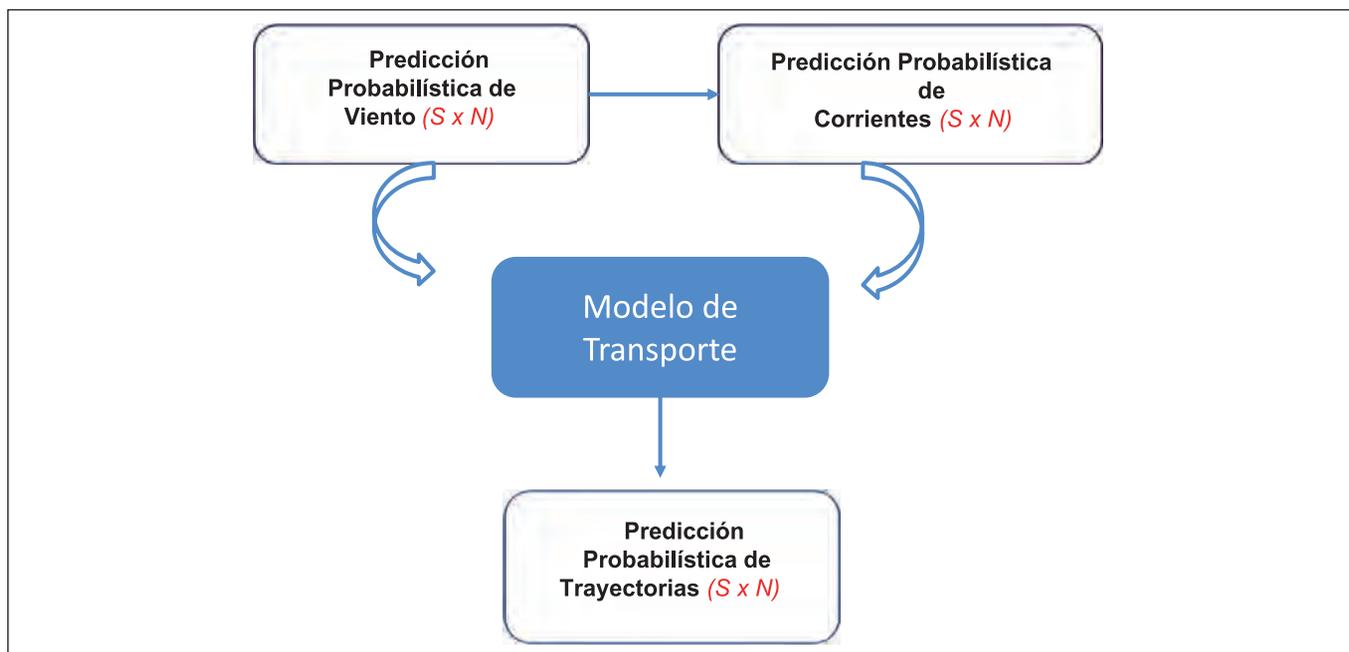


Figura 2. Esquema de la predicción probabilística de la evolución un vertido a medio-largo plazo.

3.5. Esquema general de la metodología

En caso de originarse un derrame de hidrocarburos al medio marino, la metodología desarrollada permitirá a partir de las condiciones océano-meteorológicas reales (en el momento del accidente), predecir probabilísticamente su evolución futura y con ello estimar la probabilidad de contaminación en el medio-largo plazo.

La metodología se estructura en las siguientes partes:

1. Clasificación y selección de patrones atmosféricos. La base de datos de reanálisis atmosférico es utilizada para clasificar y obtener los patrones de circulación atmosférica y los estados de viento estadísticamente representativos de la zona de estudio. Para ello, se sigue el siguiente procedimiento: (1) análisis de componentes

principales (PCA) para reducir la dimensionalidad de la base de datos atmosféricos y (2) aplicación de la técnica de clasificación *K-Medias* para seleccionar los patrones atmosféricos (estados de viento y patrones de circulación atmosférica o tipos de tiempo) representativos de la base de datos de partida.

2. Predicción probabilística de la evolución de viento y corrientes mediante la utilización de cadenas de Markov. Para obtener la predicción probabilística de la evolución de las condiciones de viento y corrientes se aplican técnicas estadísticas basadas en procesos estocásticos, concretamente en cadenas de Markov. Dicho procedimiento está basado en los siguientes pasos:

- **Generación de la matriz de transición.** A partir de las bases de datos de viento y presión previamente clasificadas se genera la matriz de transición, la cual proporciona para cada paso de tiempo la probabilidad de pasar de un estado de viento "i" a un estado de viento "j" teniendo en cuenta el patrón de circulación atmosférica inicial.
- **Selección del estado inicial de la cadena de Markov:** el estado inicial de la cadena de Markov se busca mediante el método de análogos, que relaciona las condiciones atmosféricas reales (en el mo-

mento del accidente) con el estado de viento y tipo de tiempo análogo resultante de la clasificación.

- **Predicción de la evolución del viento:** con base en la matriz de transición se realiza la predicción probabilística de los campos de viento para el período de predicción (S días).
- **Predicción de la evolución de las corrientes:** obtención de las corrientes superficiales asociadas a los campos de viento predichos.

La cadena de Markov se repite un número N de veces, por lo que se obtienen $S \times N$ condiciones de viento y corrientes.

3. Predicción probabilística de la evolución del vertido. A partir de los resultados obtenidos en el paso 2 ($S \times N$ condiciones de viento y corrientes), se aplica el modelo de transporte de hidrocarburos para simular a medio-largo plazo (S días) N trayectorias del vertido. De esta forma, se obtienen $S \times N$ simulaciones de trayectorias equiprobables que son utilizadas para calcular la probabilidad de contaminación originada por el vertido. Estos resultados se visualizan en forma de mapa de probabilidad.

En la figura 3 se muestra un esquema general de la metodología desarrollada para obtener la predicción probabilística de la evolución de la trayectoria de un derrame.

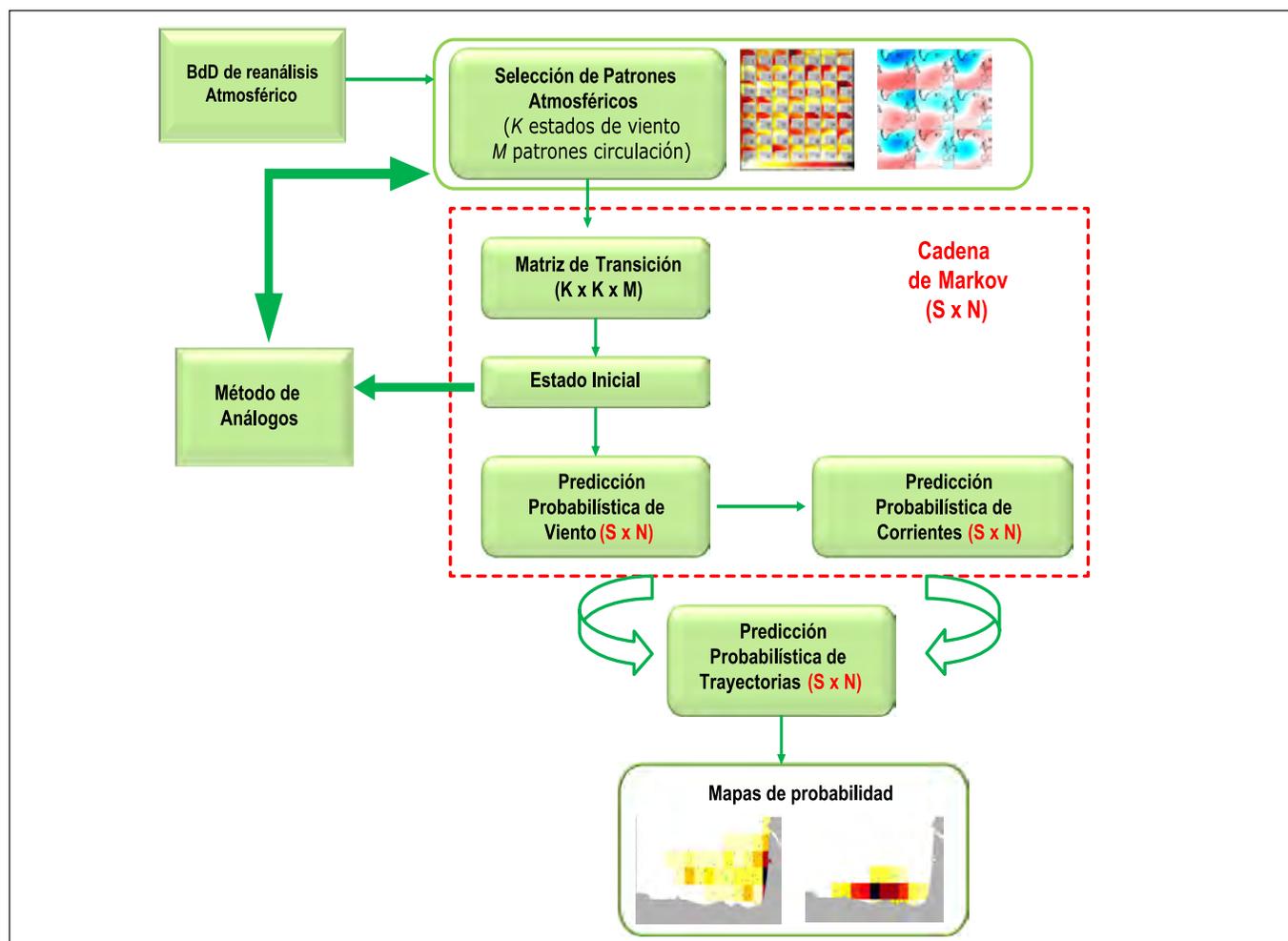


Figura 3. Esquema de la predicción probabilística de la evolución de las trayectorias de un derrame a medio-largo plazo.

4. RESULTADOS

En este apartado se muestran los resultados obtenidos al aplicar la metodología desarrollada al Golfo de Vizcaya. La metodología propuesta podría aplicarse para predecir la trayectoria de un hipotético derrame que se produjera hoy en día en el Golfo de Vizcaya. Sin embargo, con el objetivo de comparar los resultados obtenidos con datos reales se ha aplicado para simular la trayectoria seguida por manchas observadas en el Golfo de Vizcaya durante el accidente del buque *Prestige*.

El petrolero *Prestige* se hundió en noviembre de 2002 en frente de la costa de Galicia, derramando aproximadamente 63000 t de hidrocarburo (Castanedo *et al.*, 2006) y afectando más de 2000 km de costa (Ministerio de Medio Ambiente, 2005). El fuel derramado por el *Prestige* se vertió en tres mareas negras. La primera corresponde al fuel derramado por el buque en su trayectoria desde que comenzó el accidente hasta su hundimiento, entre el 13 y 19 de noviembre. La segunda marea negra (comprendida entre el 19/11/2002 y 10/12/2002) se originó por el hidrocarburo derramado

durante el hundimiento del barco. La tercera mancha (6/12/2002-8/01/2003) se formó por el fuel que continuaba derramando el buque después de su hundimiento (Ministerio de Educación y Ciencia, 2007). Las observaciones realizadas en este período muestran que la primera mancha superó la costa septentrional de Galicia el 25 de noviembre, afectando varias playas en la zona e introduciéndose en el Golfo de Vizcaya (Montero *et al.*, 2003).

En este estudio se va a simular la evolución de un vertido localizado en el límite entre Galicia y Asturias (7° O, 44.1° N) y coincidente con una de las manchas observadas en el Golfo de Vizcaya el día 27/11/2011 (Cedre, 2003). En la figura 4 se muestra el mapa de avistamientos de las manchas proporcionado por la institución francesa Cedre (*Centre de documentation, de recherche et d'expérimentations sur les pollutions accidentelles des eaux*) para esa fecha, indicándose con un rectángulo gris la posición de la mancha en las coordenadas mencionadas. Partiendo de esta posición y fecha inicial, se calculará la evolución de la mancha durante los 30 días siguientes y la correspondiente probabilidad de contaminación.

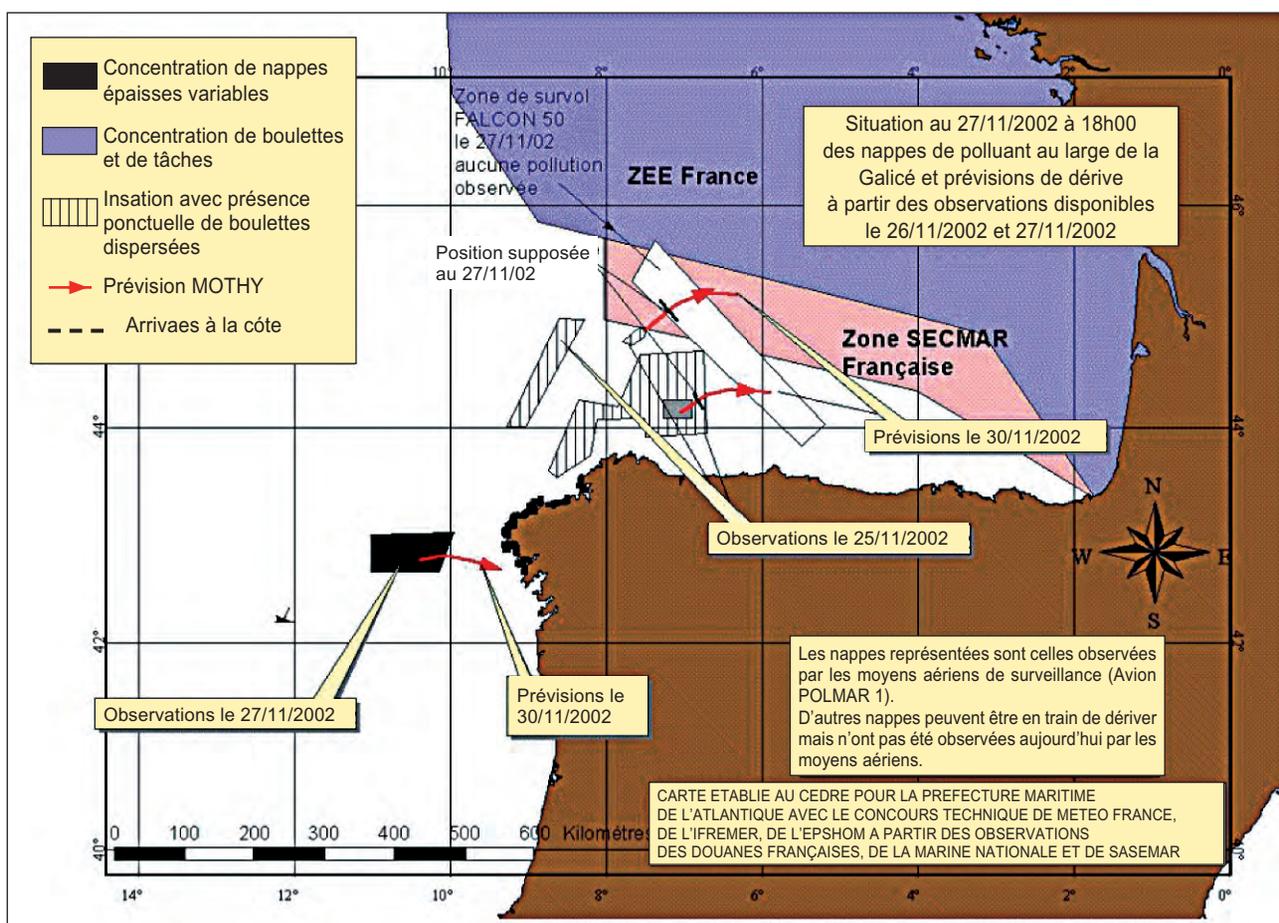


Figura 4. Mapa de avistamientos correspondiente al día 27/11/2002 (Cedre, 2003).

4.1. Clasificación y selección de patrones atmosféricos

El primer paso en la metodología es obtener los estados de viento y patrones atmosféricos que caracterizan la zona de estudio, en este caso el Golfo de Vizcaya. Para ello se van a utilizar las bases de datos descritas previamente (véase el apartado 3.1).

A continuación se presentan los resultados obtenidos.

4.1.1. Clasificación y selección de los datos de viento

Para la selección de los patrones de viento, se ha considerado una escala temporal de medias diarias de los datos horarios de viento del SeaWind-EraInterim disponibles entre 1989-2009. En cuanto al dominio espacial se ha seleccionado un área localizada en el Golfo de Vizcaya con una resolución espacial de aproximadamente 100 km. En la figura 5 se muestra el área y los nodos seleccionados para realizar la clasificación de los estados de viento.

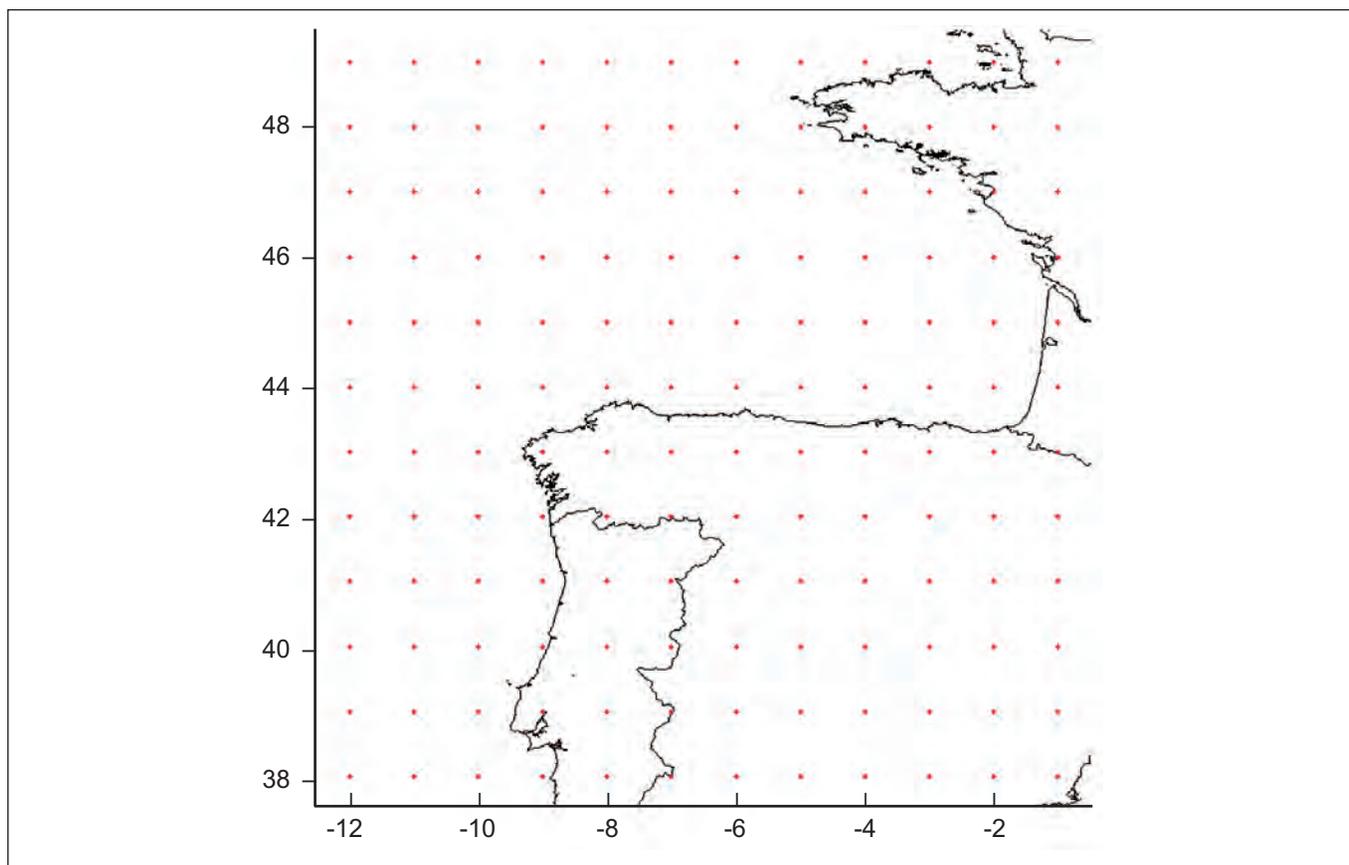


Figura 5. Dominio y nodos seleccionados de la base de datos de viento (SeaWind-EraInterim).

4.1.1.1. Análisis PCA

Con el objetivo de reducir la dimensionalidad de los datos y aumentar la eficiencia de la técnica de clasificación se aplica la técnica de análisis de componentes principales (PCA) a los datos diarios de viento del reanálisis (1989-2009) en el dominio indicado en la figura 5.

Para ello se parte, se parte de una matriz de PCs que presenta 7700 filas (longitud de los datos temporales) por 144 columnas (número de nodos de la malla). Las figuras 6 y 7 muestran, respectivamente, los resultados del análisis de PCA para la componente del viento en el eje x (V_x) y la componente del viento en el eje y (V_y). En estas figuras se observa en primer lugar el porcentaje de varianza explicada en función del número de componentes (gráfica a) y a continuación los errores cuadráticos medios (RMSE) que se cometen en función de la varianza explicada (gráfica b) y el número de componentes (gráfica c).

Para seleccionar el número óptimo de PCs debe existir un equilibrio entre la reducción de información deseada y la calidad de la aproximación resultante. Las figuras 5 y 6 muestran que con 21 modos se explica el 99% de la varianza. Estos valores de la varianza indican que existe una redundancia en los datos y que una proporción pequeña de las variables permite explicar un alto porcentaje de la varianza.

Por otro lado se observa que considerando 21 modos, el RMSE es menor de 0.5 m/s para ambas componentes. Con base en estos resultados, se han seleccionado 21 modos o componentes principales, los cuáles explican el 99% de la varianza total de los datos. De esta forma, se pasa de una matriz inicial de dimensiones 7700 x 144 a otra de 7700 x 21, cumpliendo el objetivo de reducir la dimensionalidad de los datos de partida sin pérdida de información.

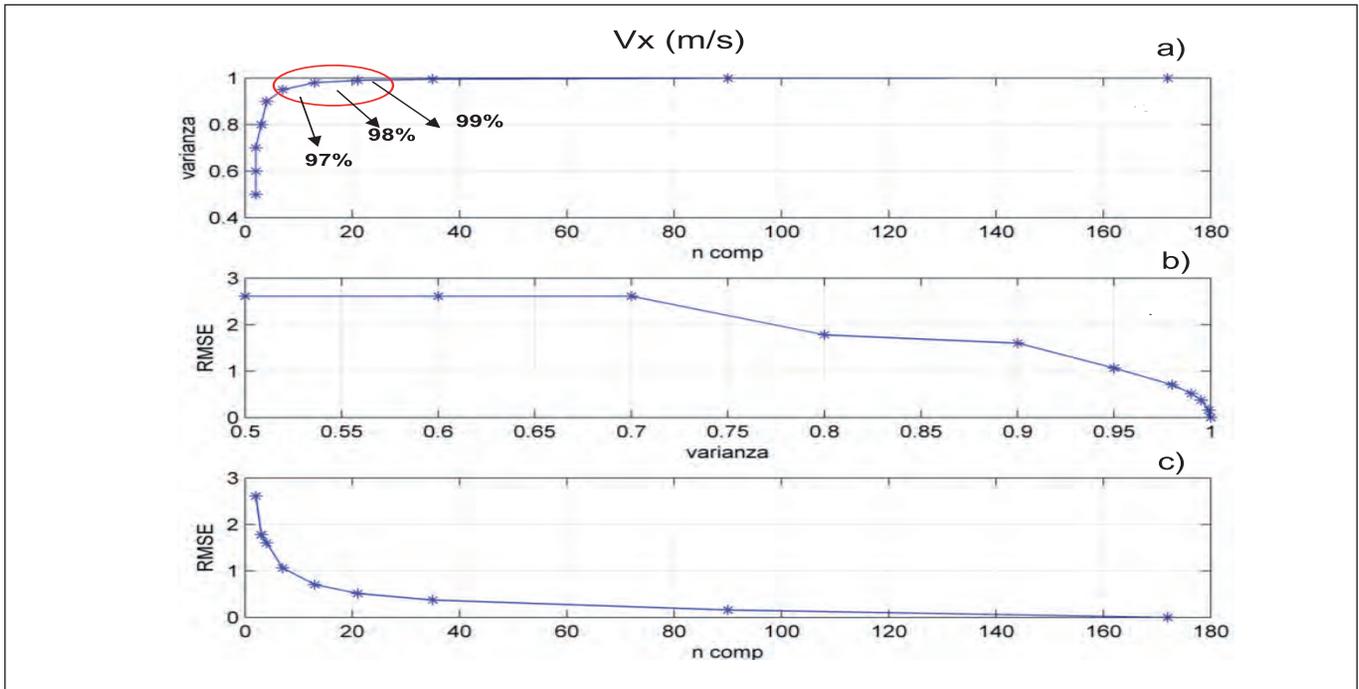


Figura 6. Resultados del PCA y errores de reconstrucción para la componente de la velocidad del viento en el eje x (variable V_x).

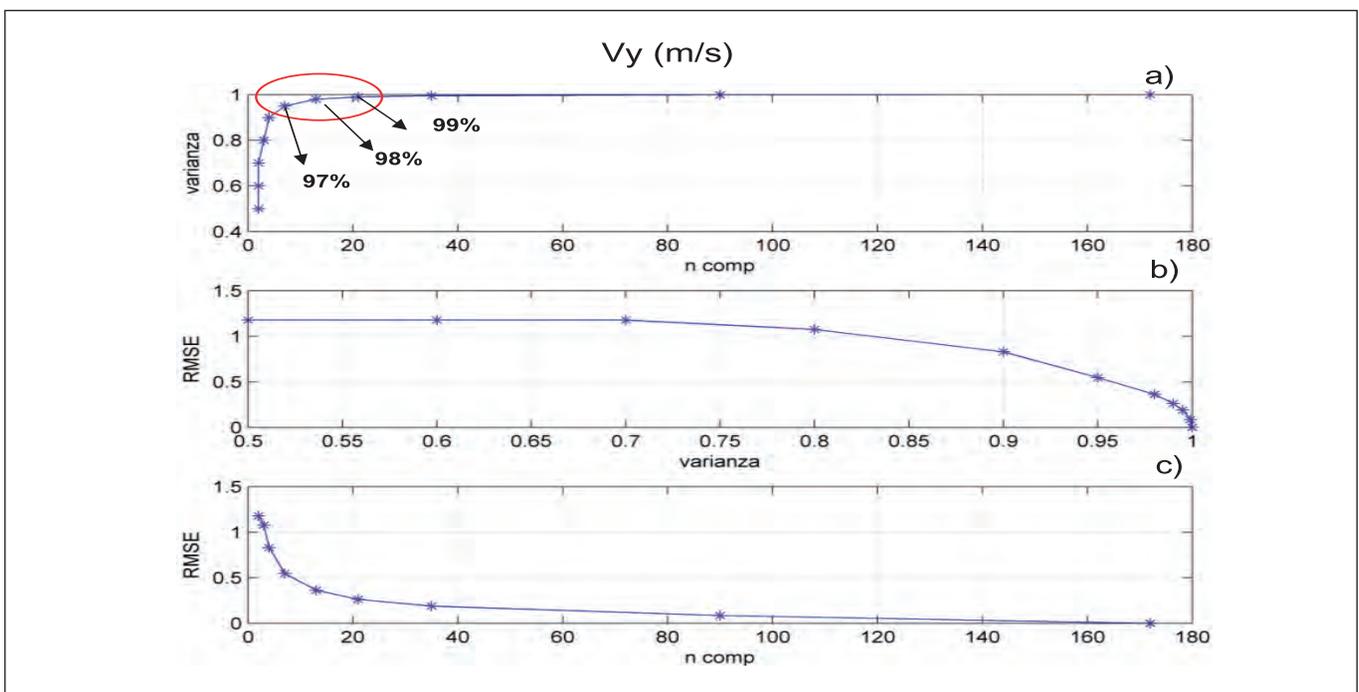


Figura 7. Resultados del PCA y errores de reconstrucción para la componente de la velocidad del viento en el eje y (variable V_y).

4.1.1.2. Clasificación K-Medias

Una vez aplicada la técnica de PCA, se realiza la clasificación y obtención de los estados de viento en la zona de estudio. Para ello se usa la técnica basada en *K-Medias* (MacQueen, 1967; Hastie *et al.*, 2001) (véase el Anejo III para más detalles).

El resultado de la clasificación de los datos medios diarios de la serie de reanálisis son los *K* patrones de

viento estadísticamente más representativos de los datos de partida y de la zona de estudio. Para seleccionar el número óptimo de grupos, se han realizado pruebas considerando 16, 25 y 49 grupos. Con base a los resultados obtenidos de las pruebas se han seleccionado 49 grupos, ya que se observa que a medida que se incrementa el número de grupos, estos grupos son más homogéneos en cuanto a los datos de la base de partida que represen-

tan, es decir, los datos que estén representados por un grupo i presentarán valores similares.

En la figura 8 se muestra el conjunto de los 49 grupos seleccionados. En esta figura se aprecia cómo la clasificación realizada recoge diferentes estados de viento que representan la variabilidad de la zona de estudio, tanto en dirección como en intensidad del viento. Por ejemplo, se

pueden observar estados de viento de componente SW (grupos 1, 2 y 3), vientos de componente W (grupo 14) o componente N (grupo 44). Asimismo se representan tanto casos extremos, donde el viento puede llegar a alcanzar los 20 m/s, por ejemplo en las posiciones 4 y 44, como períodos de calma como se observa en las posiciones 18 y 22.

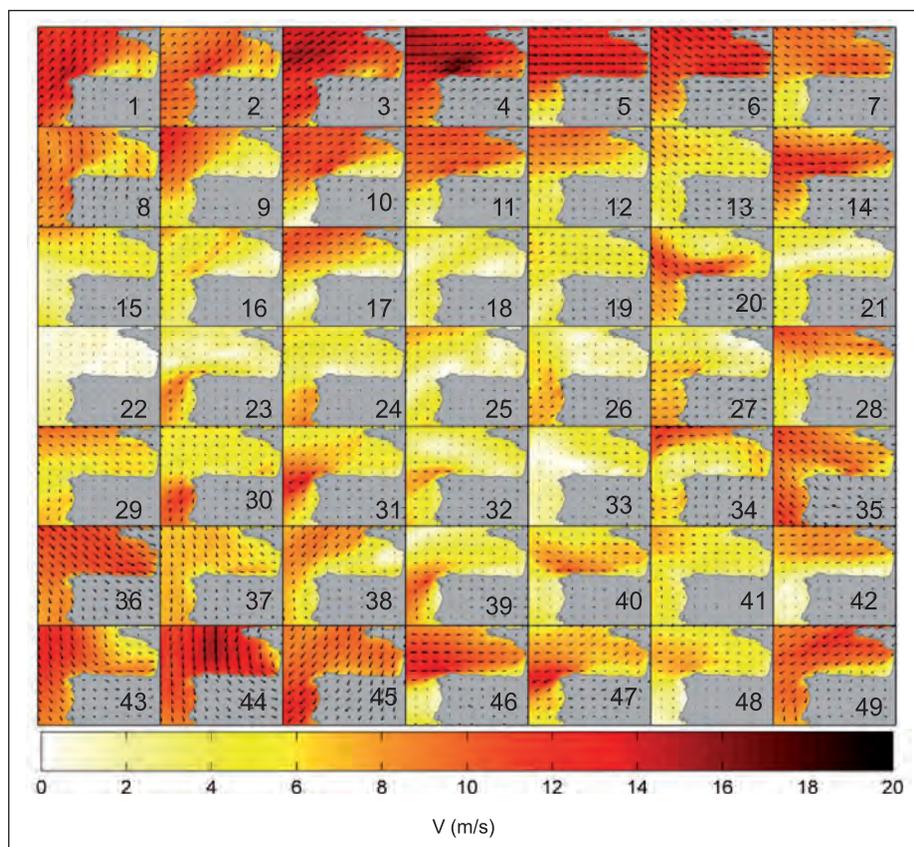


Figura 8. Clasificación de los datos medios diarios de la base de datos de SeaWind-EraInterim en 49 subconjuntos.

4.1.2. Clasificación y selección de los datos de presión

Para la selección de los patrones de circulación atmosférica, se han considerado las anomalías medias mensuales de la presión a nivel del mar (SLP, *sea level pressure*) obtenidas del reanálisis global NCEP-NCAR. El período de reanálisis utilizado en este trabajo comprende desde el año 1957 al año 2011, en total 55 años de datos. Cabe destacar, que se trabaja con este período de tiempo debido a la menor fiabilidad que presentan los datos anteriores al año 1957 (Kistler *et al.*, 2001; Guanche *et al.*, 2013). En este caso, las anomalías se estiman calculando las desviaciones mensuales de los 55 años de medias mensuales, lo que permite obtener las variaciones interanuales.

Con respecto al dominio, se ha seleccionado un área que abarca el noreste Atlántico cubriendo desde la latitud 25° a la 65°N y desde la longitud 52.5°E a la 15°W. El dominio ha sido seleccionado para abarcar la zona de

generación de borrascas al sur de Groenlandia, los anticiclones procedentes de las Azores y la oscilación del Atlántico Norte (*North Atlantic Oscillation*, NAO), que es el patrón dominante de la variabilidad atmosférica en el Hemisferio Norte, especialmente sobre el sector europeo Atlántico. En la figura 9 se muestra el área y los nodos seleccionados para realizar la clasificación de los patrones de circulación atmosférica.

4.1.2.1. Análisis PCA

El procedimiento seguido para la obtención de los patrones de circulación atmosférica es análogo al seguido para el caso de los estados de viento. Por lo tanto, se realiza un análisis de componentes principales para reducir la dimensionalidad a los datos de presión.

En primer lugar se han determinado los *EOF's* correspondientes a la zona seleccionada. Con el fin de garantizar que los *EOF's* seleccionados expliquen al menos el

97% de la varianza de las anomalías de presión medias mensuales, se han seleccionado las diez primeras componentes principales que explican el 97.4% de la va-

rianza. En la figura 10 se muestran los modos espaciales, mientras que en la figura 11 se muestran los modos temporales.

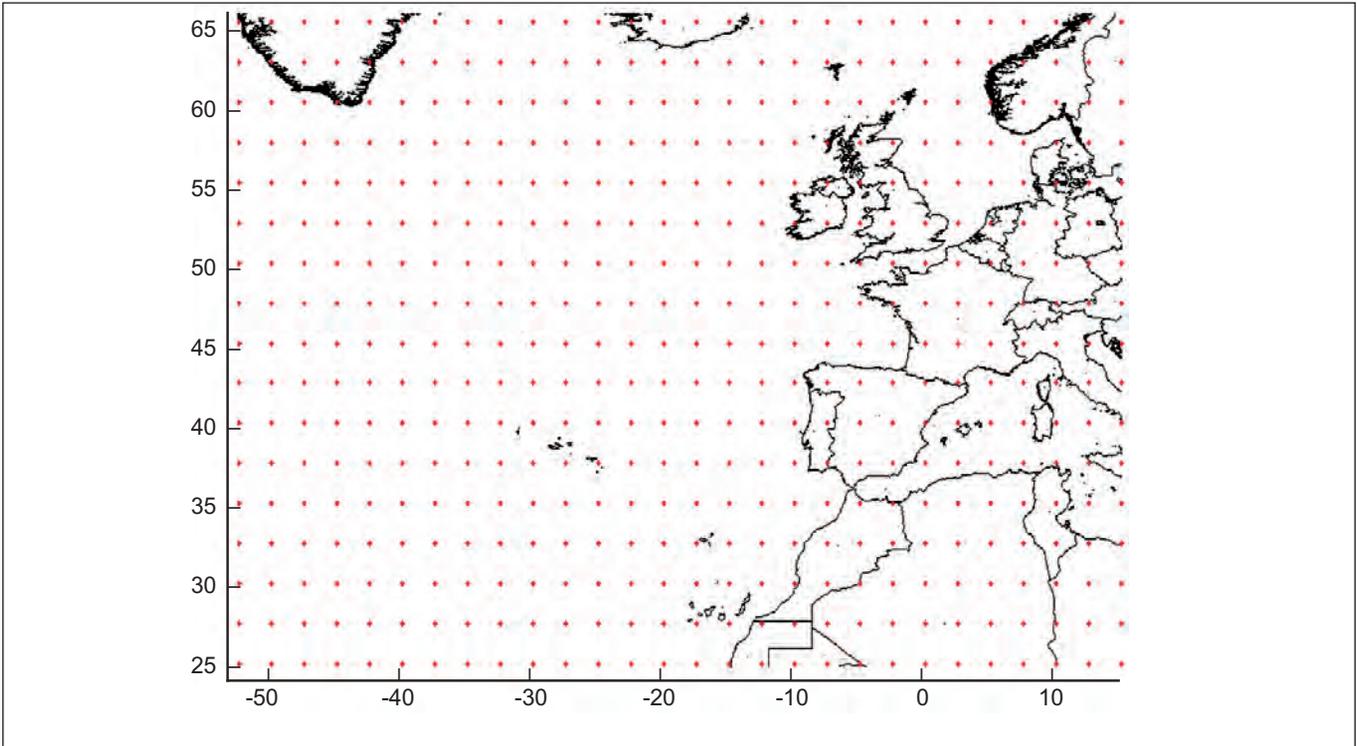


Figura 9. Dominio y nodos seleccionados de la base de datos global NCEP-NCAR.

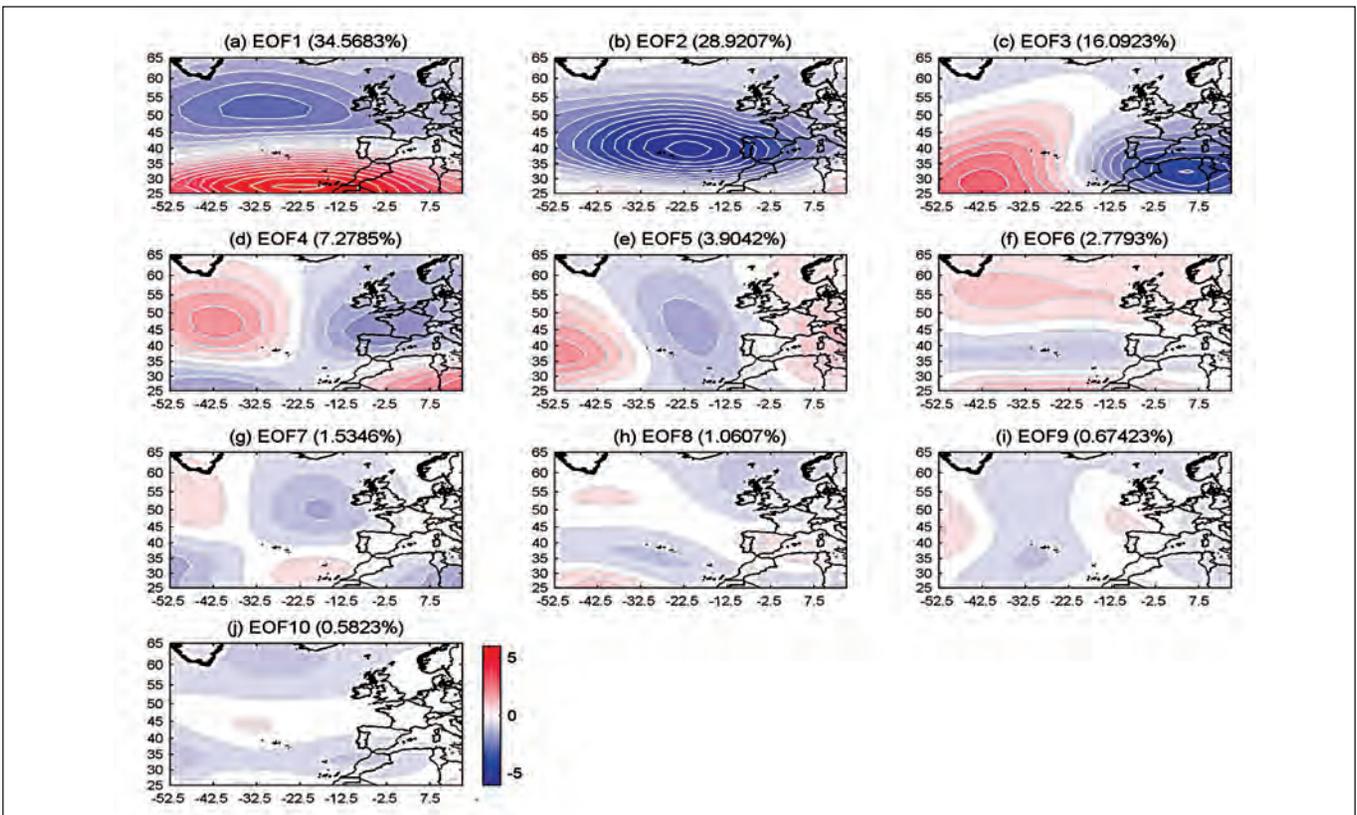


Figura 10. EOF's de las anomalías de presión medias mensuales en el Atlántico Norte.

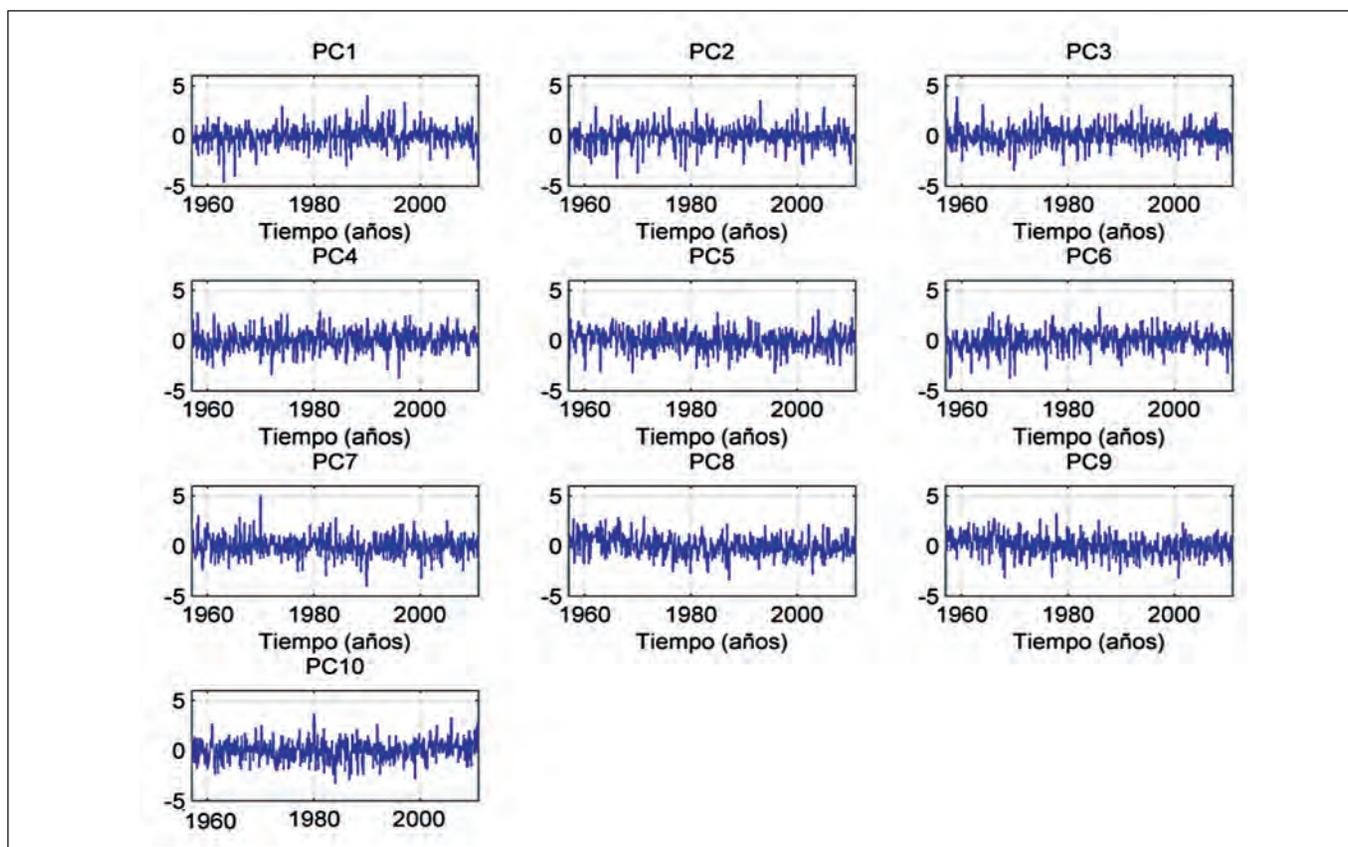


Figura 11. PC's de las anomalías de presión medias mensuales en el Atlántico Norte.

Como puede observarse en la figura 10, los dos primeros modos explican más del 63% de la varianza de los campos de anomalías, siendo el primer modo similar espacialmente al índice NAO (*North Atlantic Oscillation*) y el segundo al índice EA (*East Atlantic*), mientras que el resto de los modos espaciales muestran situaciones sinópticas independientes entre sí.

Como se ha comentado previamente, la NAO es el patrón dominante de la variabilidad atmosférica en el Atlántico Norte. La NAO consiste en un patrón bipolar de anomalías de presión, con un centro localizado en Islandia y el otro centro de signo opuesto localizado en las Azores. La fase positiva de la NAO se caracteriza por un acentuamiento de las bajas presiones sobre las latitudes altas del Atlántico Norte y de las altas presiones sobre el Atlántico Norte central, el Este de EEUU y el Oeste de Europa. La fase negativa presenta un patrón opuesto de las anomalías de presión.

El patrón del Atlántico Este presenta una estructura similar a la NAO, consiste en un dipolo norte-sur de centros de anomalías que se extiende por todo el Océano Atlántico Norte desde el este al oeste pero desplazado hacia el sureste respecto a la posición de la NAO.

4.1.2.2. Clasificación K-Medias

Como datos de entrada para realizar la clasificación y selección de los patrones de circulación atmosférica se

han empleado los diez modos temporales obtenidos mediante el análisis de componentes principales (97.4% de la varianza explicada).

Como resultado de la clasificación se obtienen M grupos, que corresponden a los patrones de circulación atmosférica o tipo de tiempo (WT, del inglés *weather type*) estadísticamente más representativos de los datos de partida. En la figura 12 se muestran las 9 situaciones sinópticas representativas del clima obtenidas de la clasificación. En ésta se han representado en azul las anomalías negativas, es decir, aquellas zonas en las que la presión es inferior a la presión media mensual, y en rojo las anomalías positivas.

A modo de ejemplo, se observa que el patrón de circulación atmosférica localizado en el centro de la primera columna (WT 2) representa una situación sinóptica con una anomalía de presión positiva centrada sobre las islas Azores y una anomalía de presión negativa centrada en el sur de Islandia, que correspondería espacialmente a una situación sinóptica de la fase positiva del índice NAO. Por otro lado, en la esquina superior izquierda (WT 1) se observa una situación muy distinta a la descrita anteriormente. Esta situación se encuentra caracterizada por una anomalía positiva centrada en el noroeste de Europa.

Se ha analizado la variabilidad de los estados de viento en función del tipo de tiempo. El objetivo de este análisis es comprobar si hay patrones redundantes, es

decir, si dos o más patrones de circulación representan las mismas condiciones de viento. La figura 13 muestra la probabilidad de ocurrencia de cada uno de los 49 estados de viento (véase la figura 8) obtenidos de la clasificación, en función de los 9 patrones de circulación (véase la figura 12). El número de estados de viento varía del 1 al 49, comenzando en la esquina superior izquierda (como se muestra en el WT 1). Se comprueba como la probabilidad de que se produzca un estado de viento u

otro varía en función del patrón de circulación. De este modo, para el primer patrón de circulación (WT 1) el estado de viento más probable es el que ocupa la posición 15 de la clasificación, que corresponde a un patrón de viento del SE con velocidades entre 5 y 8 m/s, mientras que para el noveno patrón de circulación (WT 9) el estado de viento más probable corresponde a la celda 49, que representa un estado de viento del NE con velocidades en torno a 12 m/s (ver figura 13).

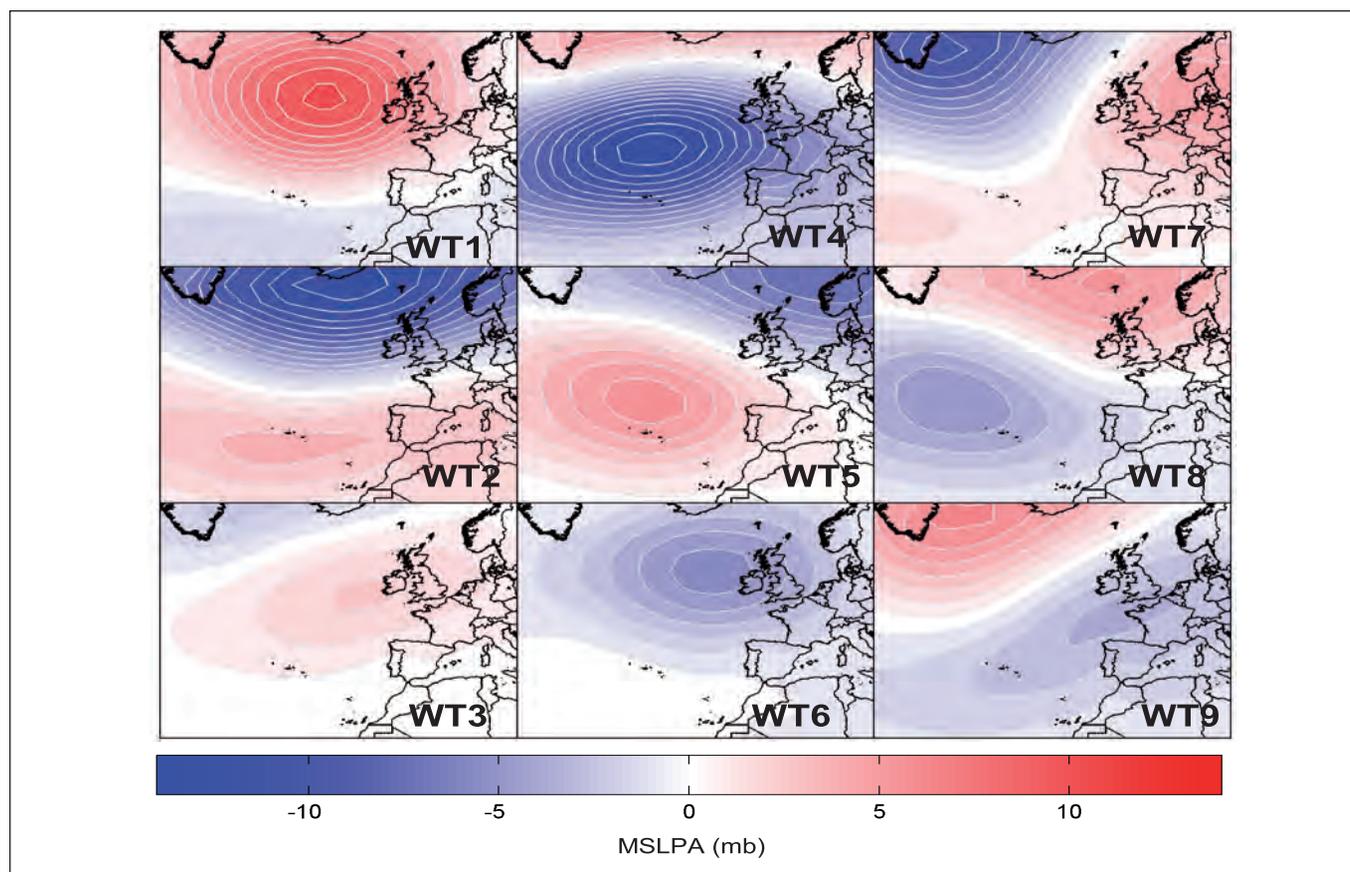


Figura 12. Clasificación de las anomalías medias mensuales de presión (MSLPA) de la base de datos NCEP-NCAR en 9 subconjuntos.

4.2. Predicción probabilística de la evolución de viento y corrientes

4.2.1. Generación de la matriz de transición

Una vez obtenidos los estados de viento y los patrones de circulación atmosférica (tipos de viento o *weather types*) el siguiente paso es utilizar esta información para predecir la evolución de las condiciones atmosféricas en el período de análisis (27/11/2002 – 25/12/2002) mediante la utilización de cadenas de Markov.

En primer lugar, a partir de los estados sinópticos de viento (véase la figura 8) se genera la matriz de transición, que tendrá unas dimensiones de $49 \times 49 \times 9$ (49 estados de viento y 9 patrones de circulación).

4.2.2. Selección del estado inicial de la cadena de Markov

La cadena de Markov se inicializa con el estado de viento y el patrón de circulación correspondiente al día del avistamiento de la mancha en el Golfo de Vizcaya, el 27/11/2002.

Para la fecha que se está considerando, se obtiene que el patrón de circulación es el que se encuentra en la posición 6 de la matriz de clasificación de presión (WT 6 en la figura 12). En la figura 14 se muestra la anomalía de presión del mes de noviembre del 2002, el patrón 6 en la clasificación y la matriz de transición asociada a este patrón de circulación ($49 \times 49 \times 1$). Comparando la anomalía de presión del mes de noviembre del 2002 con el patrón

de clasificación (WT 6) se observa que mantienen la misma estructura espacial, una anomalía negativa de presión centrada en el norte de Europa. Sin embargo, se observa que la intensidad de la anomalía es mayor en noviembre del 2002 que en el WT 6 (nótese que la escala de colores es diferente). Esto sugiere que sería necesario utilizar más grupos en la clasificación del campo de pre-

siones, para poder diferenciar claramente entre estructura espacial e intensidad. La matriz de transición expuesta en la figura 14 ($49 \times 49 \times 1$) contiene la probabilidad de pasar de un estado de estado “*i*” al estado “*j*” condicionado al patrón de circulación WT 6. Los colores más oscuros indican mayor probabilidad de tránsito entre estados.

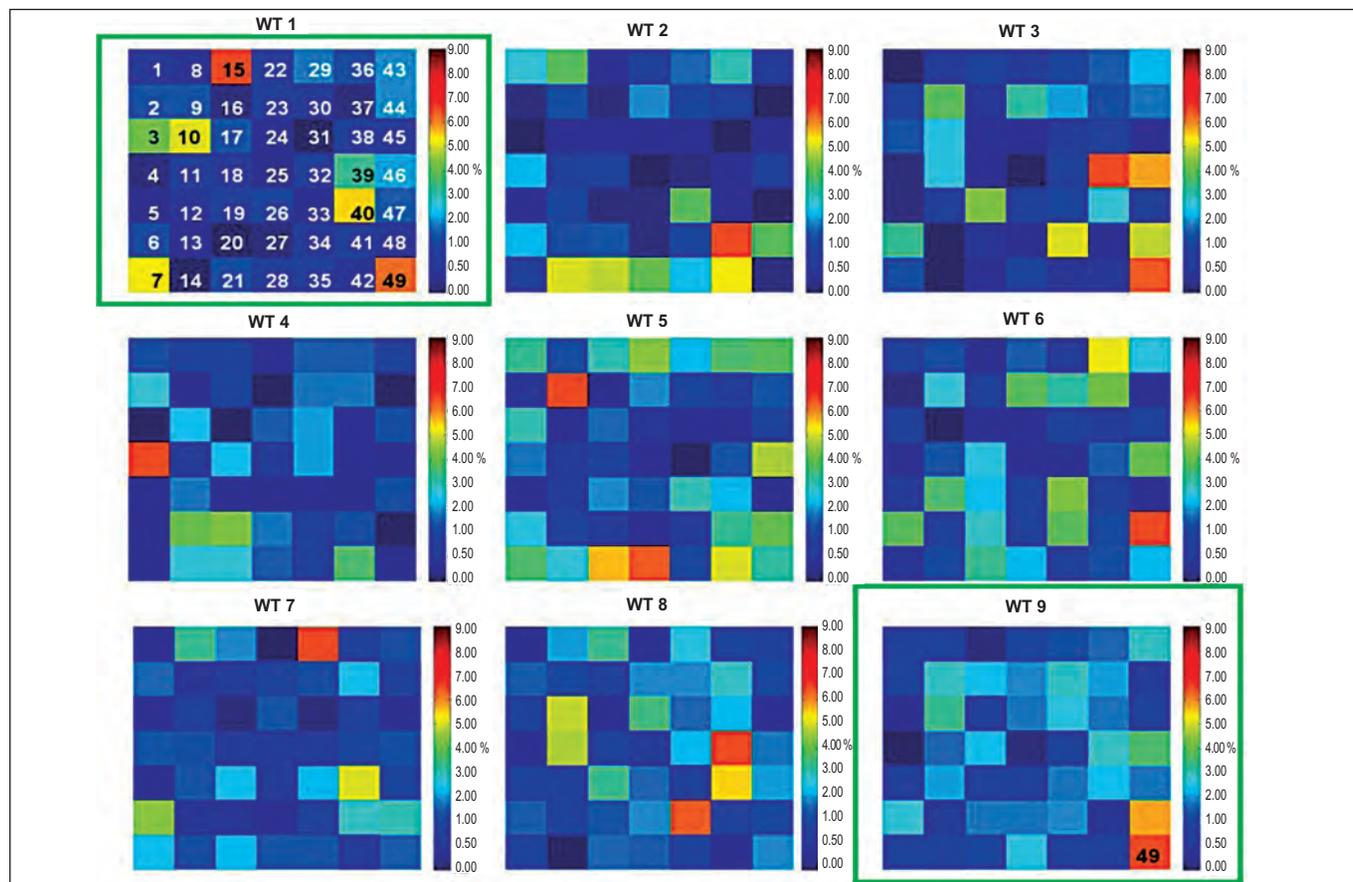


Figura 13. Probabilidad de ocurrencia de los estados de viento obtenidos de la clasificación en función del patrón de circulación.

A continuación, se selecciona el estado inicial de viento con el que comienza el proceso. Para el día en el que se inicia la predicción, se obtiene que el estado de viento análogo a la situación meteorológica de ese día es el que se encuentra en la celda 2 de la matriz de clasificación. En la figura 15 se puede observar el campo viento diario del día 27/11/2002 (izquierda) y el estado inicial de la cadena de Markov obtenido mediante técnicas de análogos (Grupo 2) (Figuras 14 y 15).

4.2.3. Aplicación de la cadena de Markov

A partir de este momento comienza el proceso de Markov, que consiste en lanzar *N* simulaciones para calcular la evolución del campo de viento inicial. Para este estudio se ha considerado *N*=1000. Cada una de las simulaciones proporciona la evolución del viento diario durante un período de 30 días (27/11/2002 – 25/12/2002), por lo que cada simulación consta de 30 iteraciones.

Cabe destacar que en cada paso de tiempo (o iteración) se calcula una probabilidad de forma aleatoria, y dada esta probabilidad y el estado de viento *i* en un instante *t* del proceso, se pasa al estado *j* en el instante *t*+1 utilizando la probabilidad de paso de un estado a otro proporcionada por la matriz de transición (véase la sección 3.3.2 para más detalle de la metodología).

El resultado del proceso es la predicción probabilística de los campos de viento para cada una de las *N* simulaciones. En la tabla 2 se recogen, a modo de ejemplo, los resultados obtenidos para algunas de las simulaciones del proceso.

Dado los resultados de la simulación 1, en la figura 16 se muestran los campos de viento medios diarios obtenidos para los seis primeros días de la predicción. En las gráficas se observa como los tres primeros días, el viento se caracteriza por ser de componente suroeste. Mientras que a partir del día 4 cambia de dirección, presentando una dirección de componente no-

roeste virando hacia el este el día 6. Además, el paso de un estado de viento a otro es coherente con las

condiciones climatológicas que se producen en el Golfo de Vizcaya.

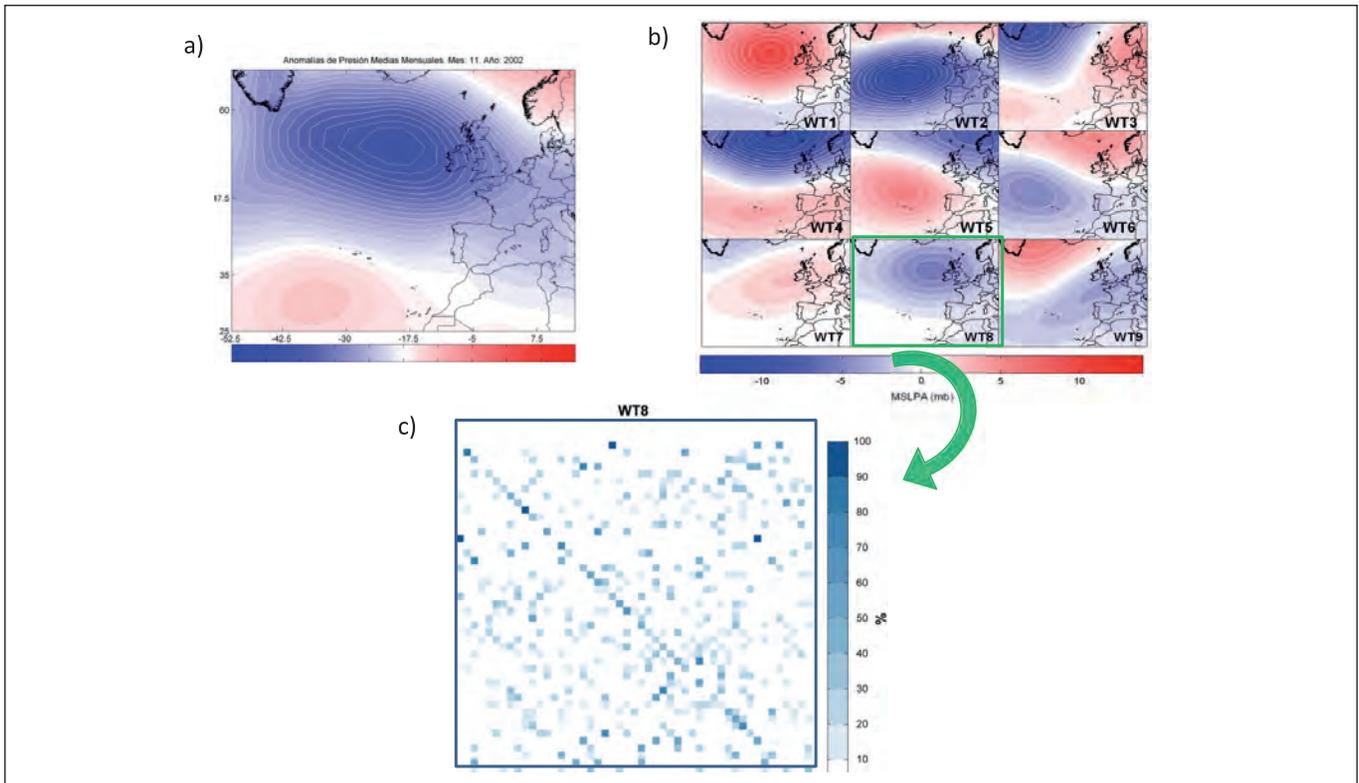


Figura 14. Matriz de transición con las probabilidades de paso de un estado de viento a otro (gráfica c), condicionada al patrón de circulación que ocupa la celda 8 de la clasificación (gráfica b). Anomalía del campo de presión (gráfica a) para la fecha de la predicción.

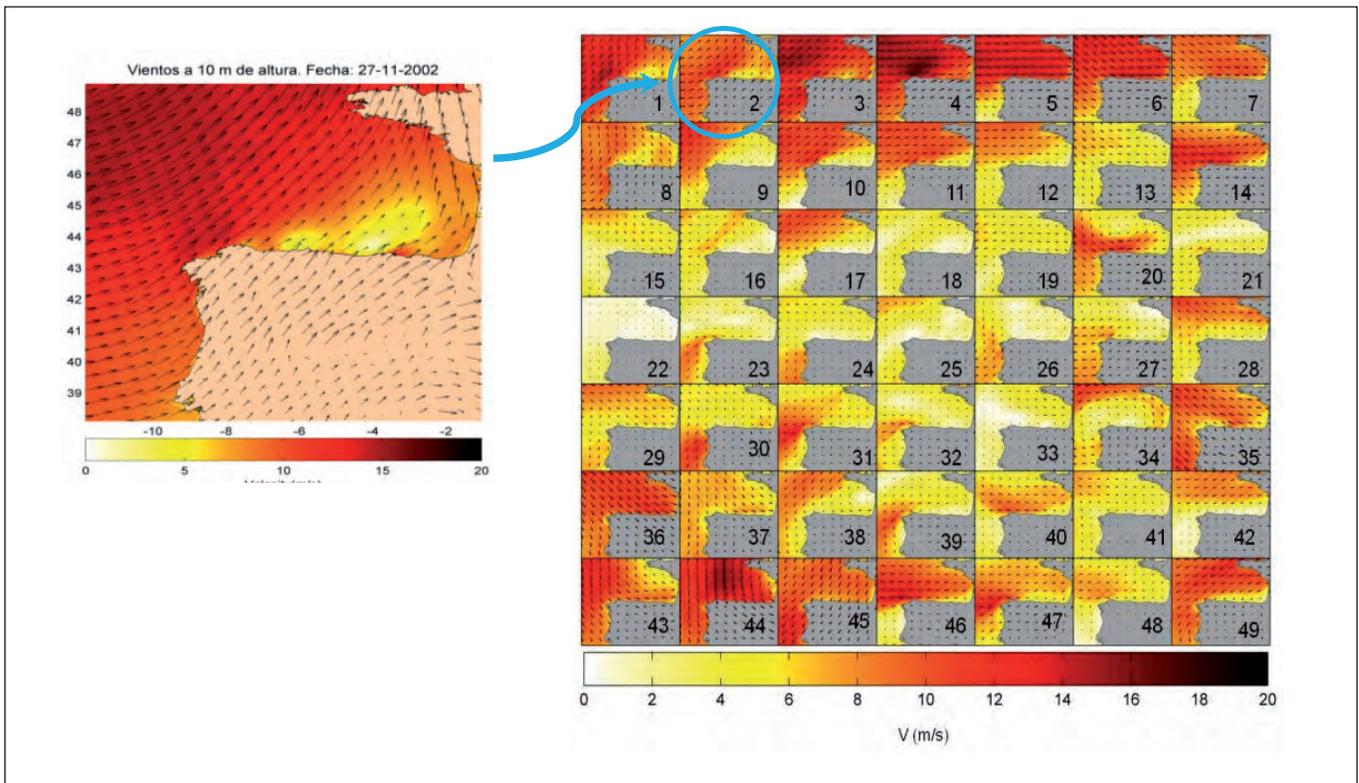


Figura 15. Estado inicial de viento con el que comienza el proceso de Markov.

Tabla 2. Ejemplo de resultados de la cadena de Markov.

AÑO	SIMU 1	SIMU 50	SIMU 100	SIMU 400	SIMU N
Día 1	20	43	21	18	21
Día 2	13	47	28	21	36
Día 3	6	8	9	21	28
Día 4	9	13	46	18	28
Día 5	30	20	3	5	48
Día 6	36	36	3	34	33
Día 7	41	41	49	46	36
Día 8	13	14	19	49	48
Día 9	37	42	19	46	23
Día 10	6	4	34	49	48
Día 11	34	14	34	19	48
Día 12	6	43	19	19	9
Día 13	33	40	34	34	48
Día 14	37	32	28	9	6

Día 15	37	16	18	34	37
Día 16	31	40	1	34	13
Día 17	38	11	22	31	37
Día 18	44	27	9	31	30
Día 19	19	37	12	30	36
Día 20	6	13	36	12	36
Día 21	6	13	8	21	23
Día 22	30	20	41	28	23
Día 23	34	21	29	21	19
Día 24	23	18	8	8	49
Día 25	21	12	20	49	19
Día 26	18	36	43	49	11
Día 27	36	14	36	15	34
Día 28	9	21	22	40	23
Día 29	34	21	12	23	19
Día 30	37	21	37	23	34

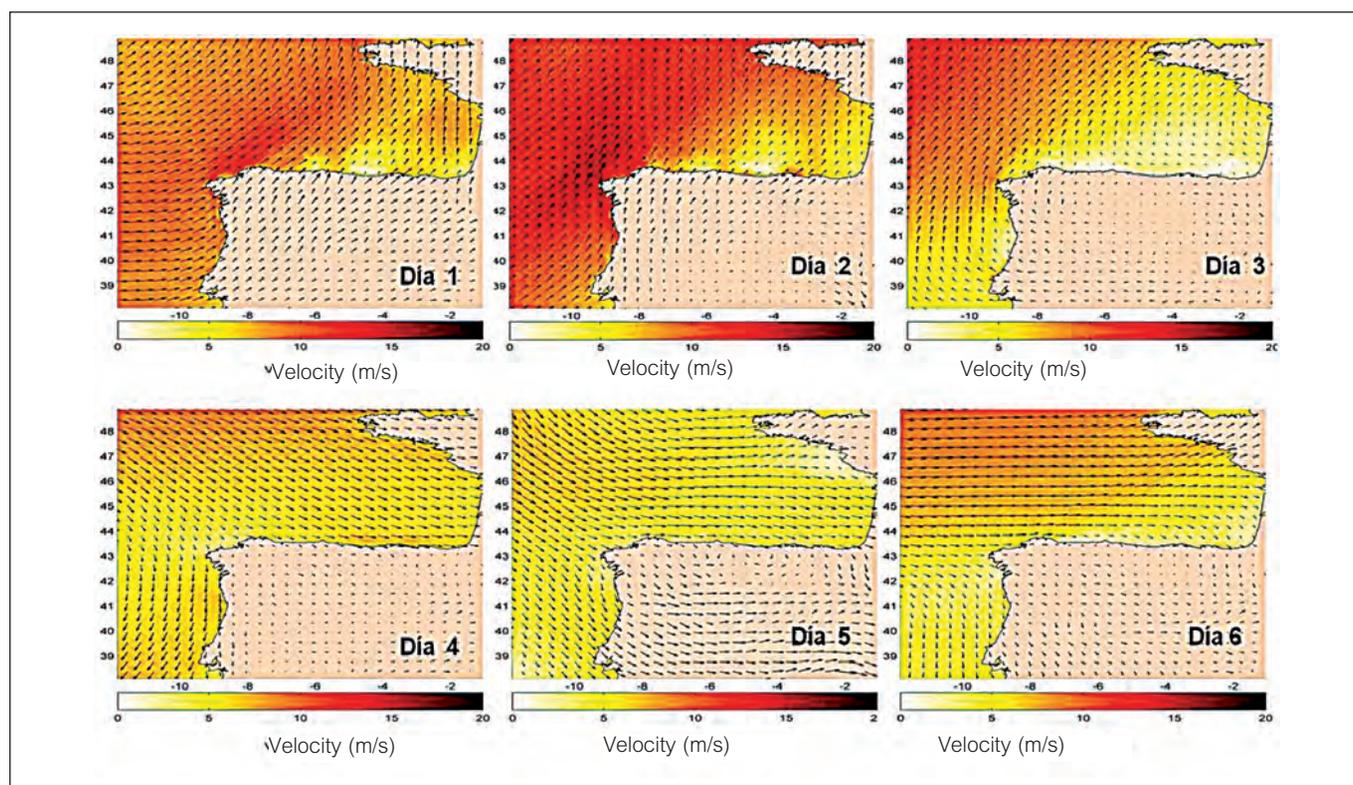


Figura 16. Resultado de la predicción probabilística de los campos de viento para los seis primeros días de predicción para la simulación 1.

Una vez conocida la evolución del viento, se asocian las corrientes medias diarias correspondiente a la misma fecha, y que en este caso, son obtenidas de la base de datos de corrientes de marea meteorológica GOS 2.1 (descrita previamente). En la figura 17 se muestra las corrientes de marea meteorológica correspondientes a los campos de viento mostrados en la figura 16.

Es importante destacar que las corrientes oceánicas son un proceso complejo que incluye las corrientes generadas por la marea astronómica, la marea meteorológica (gradientes de presión atmosférica y viento), los procesos baroclínicos y la circulación general oceánica. Debido a la gran complejidad que entraña la simulación de todos estos procesos, los reanálisis de corrientes son escasos y de

hecho, no se tiene conocimiento de que exista una base de datos de esas características para la zona de estudio. Sin embargo, existen reanálisis numéricos que consideran las dinámicas generadoras de corriente más importantes en

esta zona, como son la marea astronómica y la marea meteorológica. Dada la resolución temporal de trabajo (medias diarias) la marea astronómica no es relevante, por lo que en este trabajo se utilizan corrientes de marea meteorológica.

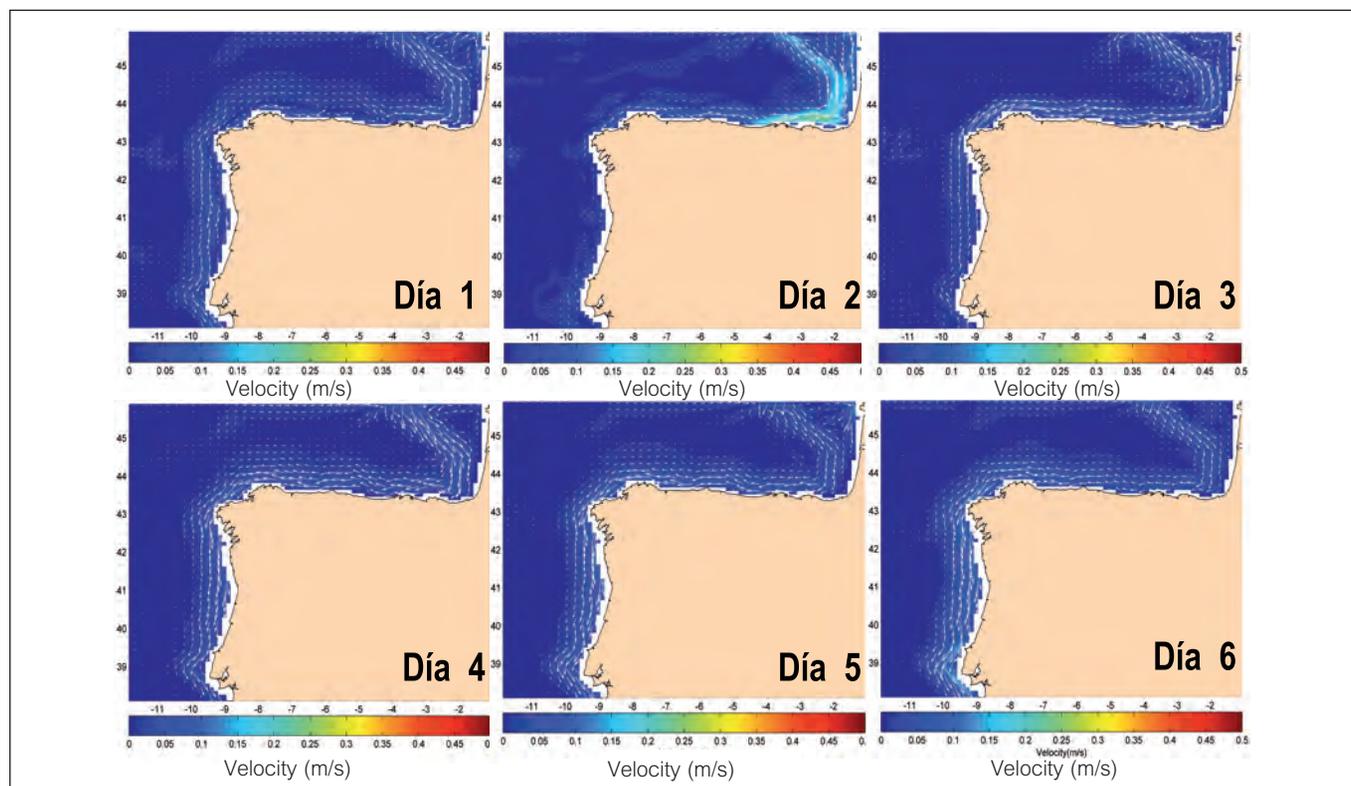


Figura 17. Resultado de la predicción probabilística de las corrientes medias de marea meteorológica para los seis primeros días de predicción para la simulación 1.

4.3. Predicción probabilística de las trayectorias del vertido

4.3.1. Evolución del derrame condicionado al tipo de tiempo del accidente

Una vez que se han obtenido los campos de viento y de corrientes, se aplica el modelo de transporte TESEO para predecir la evolución de la mancha observada (7°O , 44.1°N) durante 30 días. La simulación comienza el día del avistamiento, el 27/11/2002 y finaliza el 25/12/2002. Como resultado se obtiene la evolución de las N ($N=1000$) trayectorias equiprobables del derrame para los 30 días de simulación, las cuáles son utilizadas para calcular los mapas de probabilidad de contaminación originados por el derrame.

A modo de ejemplo, en la figura 18 se presentan las trayectorias calculadas para las simulaciones 1, 50, 100 y 500. En estas gráficas se observa en azul el punto donde se inicia la simulación (localización de la mancha) y en rojo la predicción a 30 días de la evolución del derrame (período de predicción: 27/11/2002 - 25/12/2002). Cabe destacar, que cada simulación presenta un recorrido dis-

tinto pero de igual probabilidad de ocurrencia. De este modo, en la figura se aprecia como para el caso de la simulación 1 el derrame se dirige hacia el noreste y permanece en el mar después de 30 días característica que comparte con las simulaciones 50 y 100. Mientras que en la simulación 500 éste se dirige hacia el este y al cabo de 30 días llega a la costa (Figura 18).

En la figura 19 se muestran conjuntamente N trayectorias equiprobables obtenidas para la predicción que comienza el día 27/11/2002 y finaliza el 25/12/2002. Por motivos de claridad en la figura, se han representado 500 de las 1000 trayectorias simuladas. Se aprecia como para el período de tiempo en que se realiza la predicción, la mayor parte de las trayectorias se dirigen hacia el este del Golfo de Vizcaya.

En la figura 20 se muestra la posición final de las mil trayectorias simuladas después de 7, 15 y 30 días de predicción. Se aprecia como el séptimo día la mayoría de las trayectorias se dirigen hacia el este y noreste, aunque también se observa que algunas de las trayectorias se dirigen hacia el suroeste viéndose afectada la costa gallega. Se puede observar que este comportamiento se

mantiene en el tiempo. Después de 30 días la mayor parte de las trayectorias se han dirigido hacia el este, afectando principalmente el litoral cantábrico y alcanzando la costa de Francia. Aunque la mayor parte de las simulaciones se concentra en el Golfo de Vizcaya, hay algunas que se

dirigen hacia el suroeste, alcanzando Galicia y Portugal. Cabe destacar, que cuando finaliza la predicción al cabo de 30 días, la mayor parte de las simulaciones continúan en el mar.

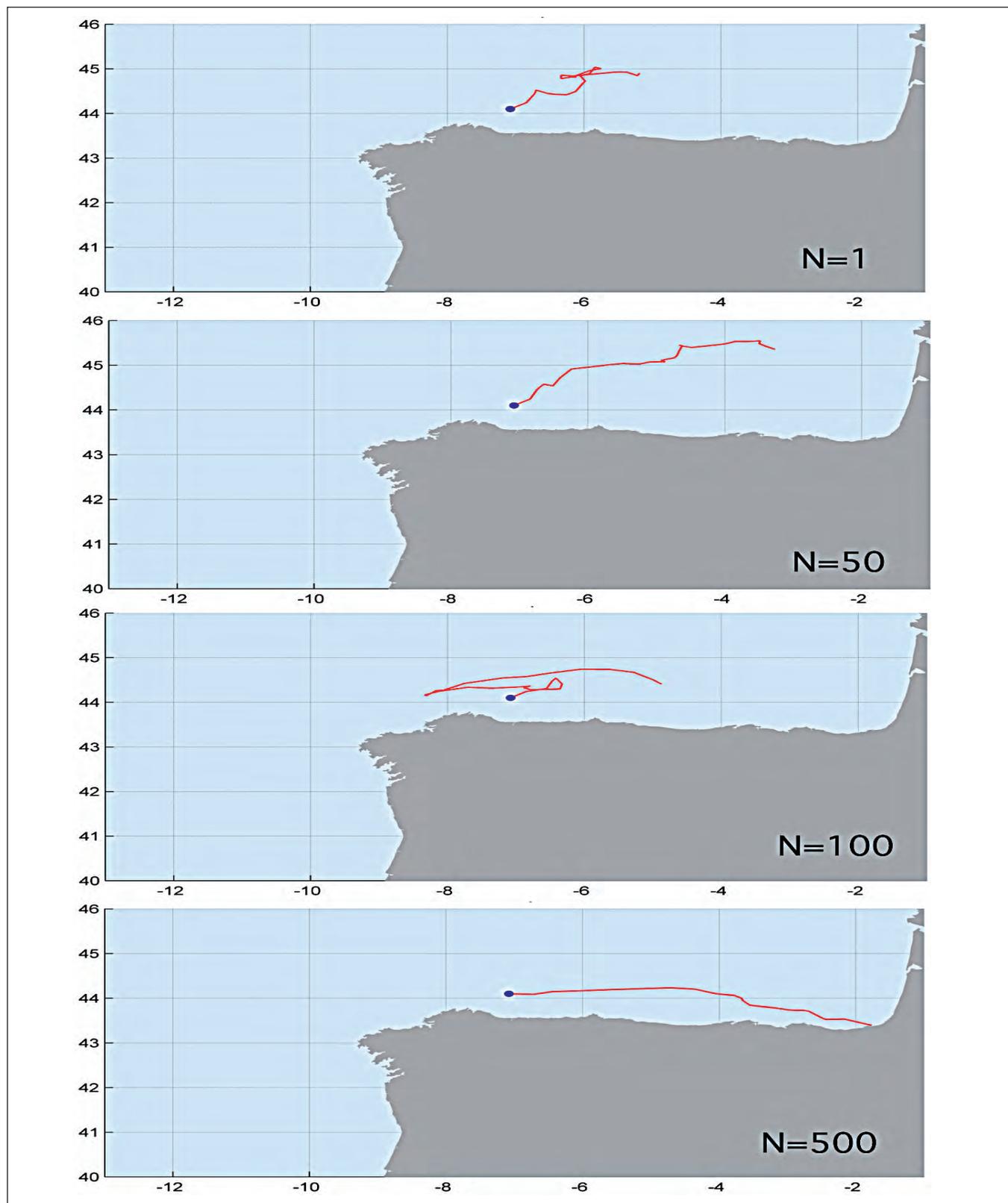


Figura 18. Resultado de la trayectoria del derrame para las simulaciones 1, 50, 100 y 500. En azul el punto de vertido y en rojo la trayectoria del derrame. Período de predicción: 27/11/2002 – 25/12/2002.

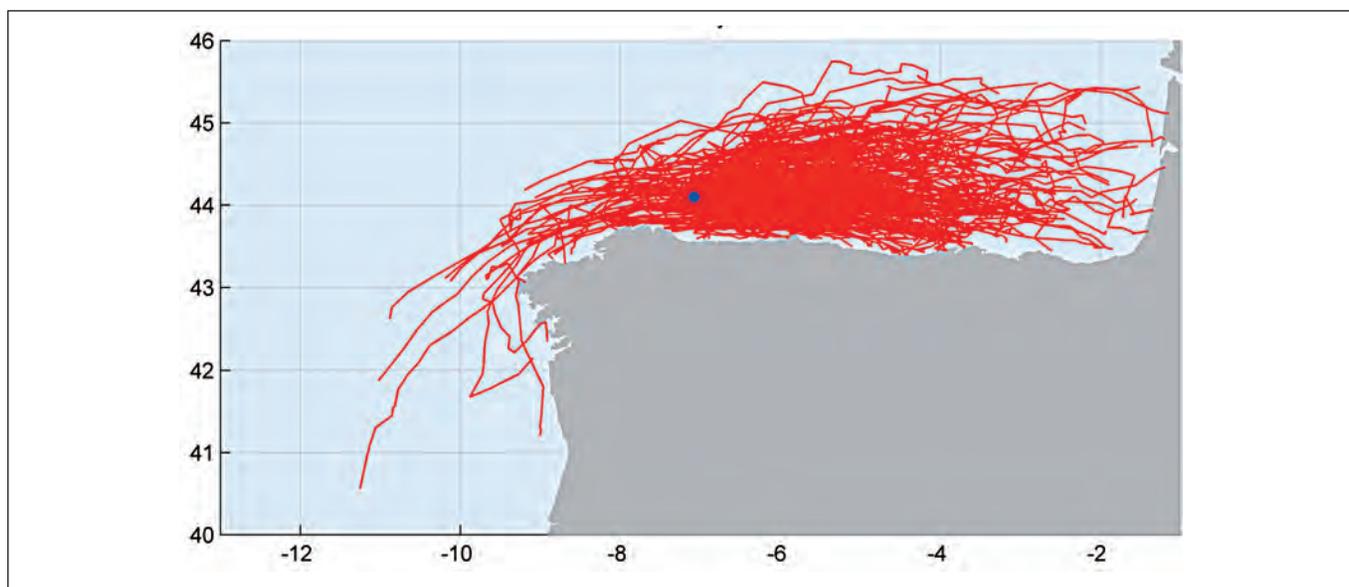


Figura 19. Resultado de las N trayectorias equiprobables. En azul el punto de vertido y en rojo las trayectorias del derrame para las N simulaciones.

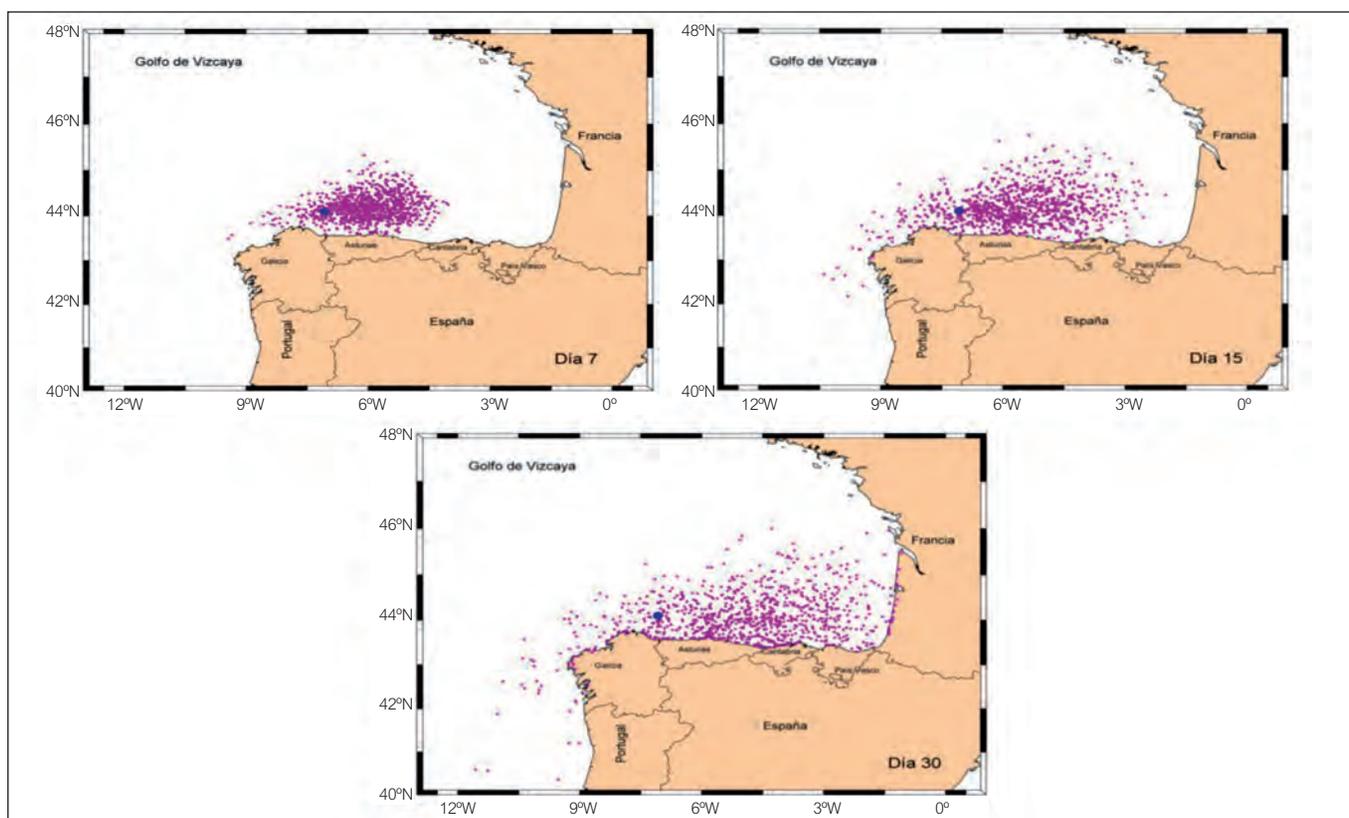


Figura 20. Posición final de las 1000 trayectorias después de 7 (arriba-izquierda), 15 (arriba-derecha) y 30 (abajo) días de simulación. El punto azul es el punto inicial del derrame y en magenta se representa la posición del derrame para cada una de las simulaciones

Las N ($N=1000$) trayectorias simuladas se utilizan para calcular la probabilidad de contaminación originada por el vertido en diferentes instantes temporales. A modo de ejemplo, en la figura 21 se muestra el mapa de probabilidad acumulada de contaminación para el día 30 de la simulación, junto con las observaciones de algunas de las manchas avistadas durante el accidente del *Prestige*. Es importante destacar que la validación de resultados pro-

probabilísticos es un proceso complejo, puesto que se necesitaría una gran cantidad de datos reales para poder estimar la probabilidad real de contaminación. Aunque no se dispone de tal cantidad de información, en este trabajo se han utilizado algunas de las observaciones del *Prestige* para comparar los resultados con datos reales, tal y como se muestra en la figura 21. Se observa que la mayor probabilidad de contaminación se localiza en el Golfo de

Vizcaya alcanzando después de 30 días la costa de Francia. También puede apreciarse que existe probabilidad de que el vertido se dirija hacia el suroeste afectando la costa de Galicia y derivando hacia el Sur, en dirección a

Portugal. Puede observarse que la mayor probabilidad de contaminación se produce en el litoral del mar cantábrico y en la costa de Asturias y Cantabria, lo que concuerda con las observaciones de las manchas (Figura 21).

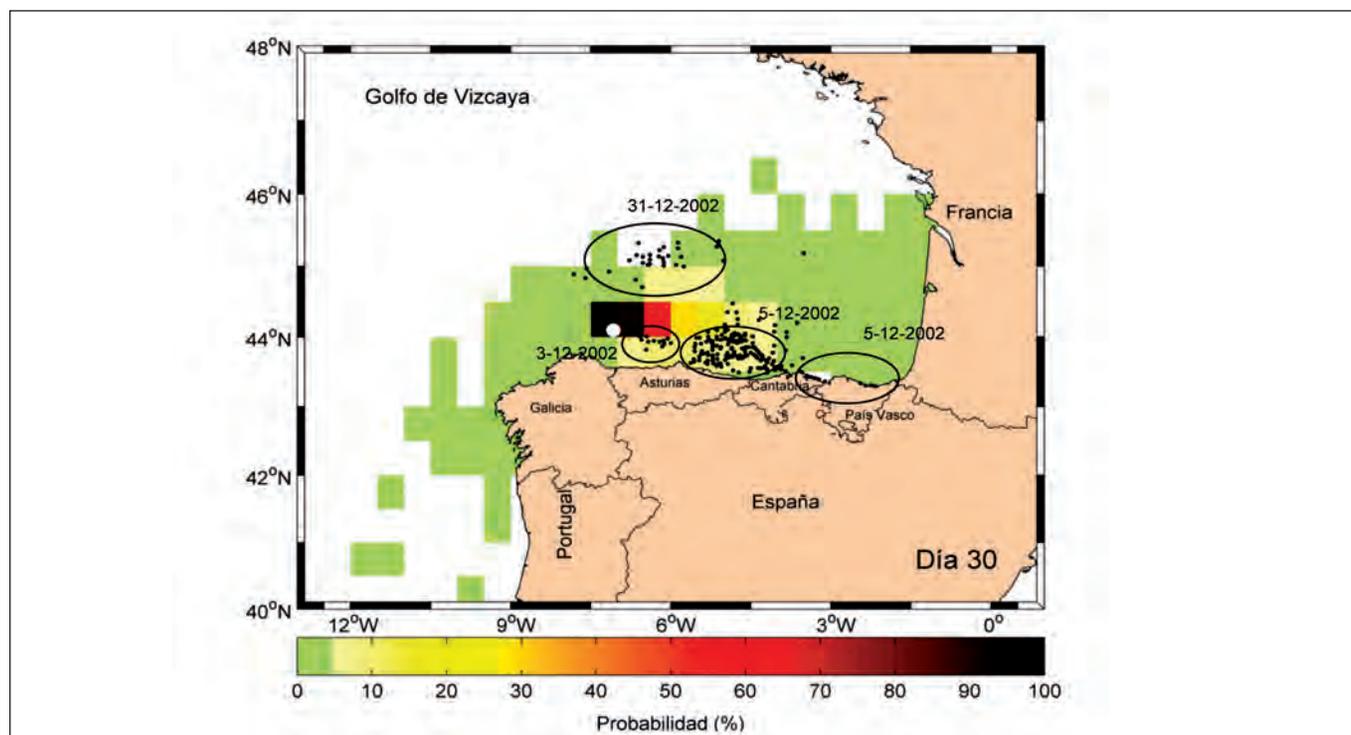


Figura 21. Mapas de probabilidad acumulada de contaminación para el día 30 de simulación. El punto blanco es el punto inicial del derrame numérico y los puntos negros son las observaciones de las manchas.

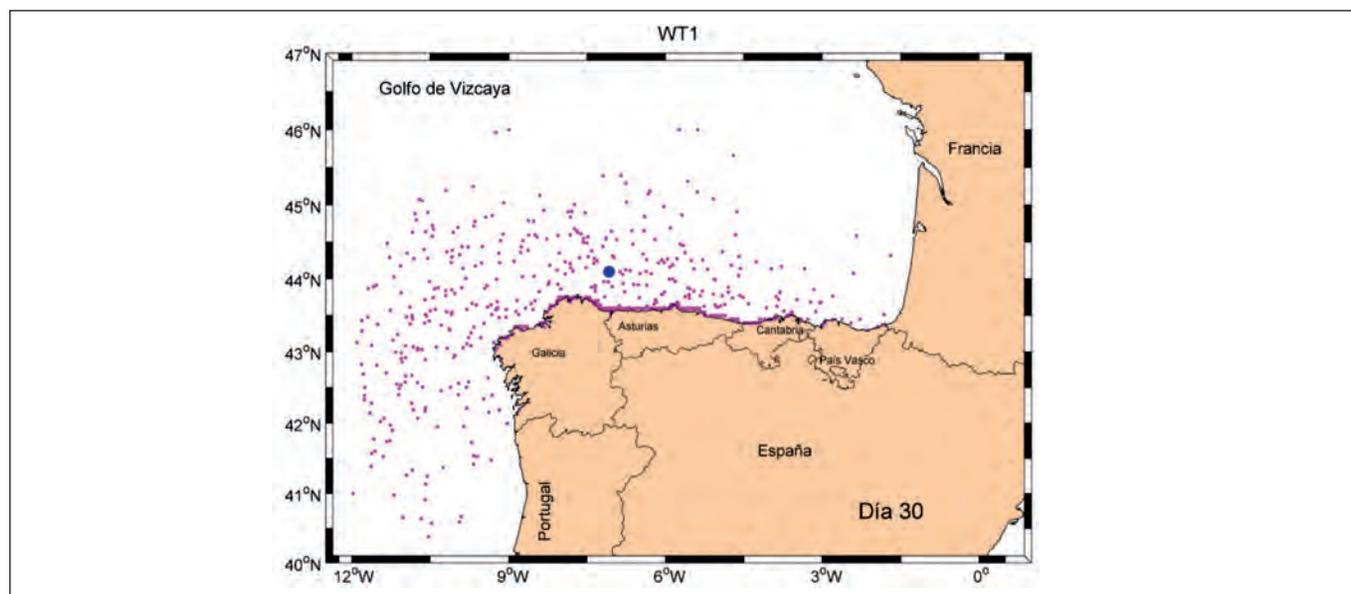


Figura 22. Mapas de probabilidad acumulada de contaminación para el día 30 de simulación. El punto blanco es el punto inicial del derrame numérico y los puntos negros son las observaciones de las manchas.

Como se ha comentado previamente, la simulación está forzada con viento y corrientes de marea meteorológica, procesos más relevantes en la trayectoria del vertido. No obstante, hay que tener en cuenta que hay pro-

cesos de la circulación general que no se están considerando, como por ejemplo la corriente dirigida hacia el polo (*Poleward Current*) que tiene influencia sobre el transporte en la zona (Sánchez y Gil, 2000) y

que contribuiría a introducir el vertido en el Golfo de Vizcaya.

4.3.2. Evolución del derrame en función de los diferentes tipos de tiempo

En este apartado se presenta la evolución de la mancha observada en el Golfo de Vizcaya (7°O , 44.1°N) durante 30 días variando el patrón de circulación atmosférica (o tipo de tiempo) inicial. Para ello se han realizado nueve simulaciones para el mismo período de tiempo (27/11/2002 - 25/12/2002), considerando el mismo estado inicial de viento (el utilizado en el apartado 4.3.1) y variando el pa-

trón de circulación inicial con los resultados de la clasificación (véase la figura 12). El objetivo de este análisis es estudiar cómo se vería afectada la trayectoria del vertido, si el patrón de circulación inicial (relacionado con el paso de borrascas y anticiclones), hubiera sido diferente.

Las figuras 22 a 29 muestran la posición final de las trayectorias simuladas ($N=1000$) después de 30 días para cada uno de los patrones de circulación atmosférica. Nótese que la simulación se realiza para el mismo punto de vertido y misma condición inicial de viento, la correspondiente al 27/11/2002. Se observa que la evolución de las trayectorias y, por lo tanto, las zonas afectadas por la contaminación varía en función del patrón inicial.

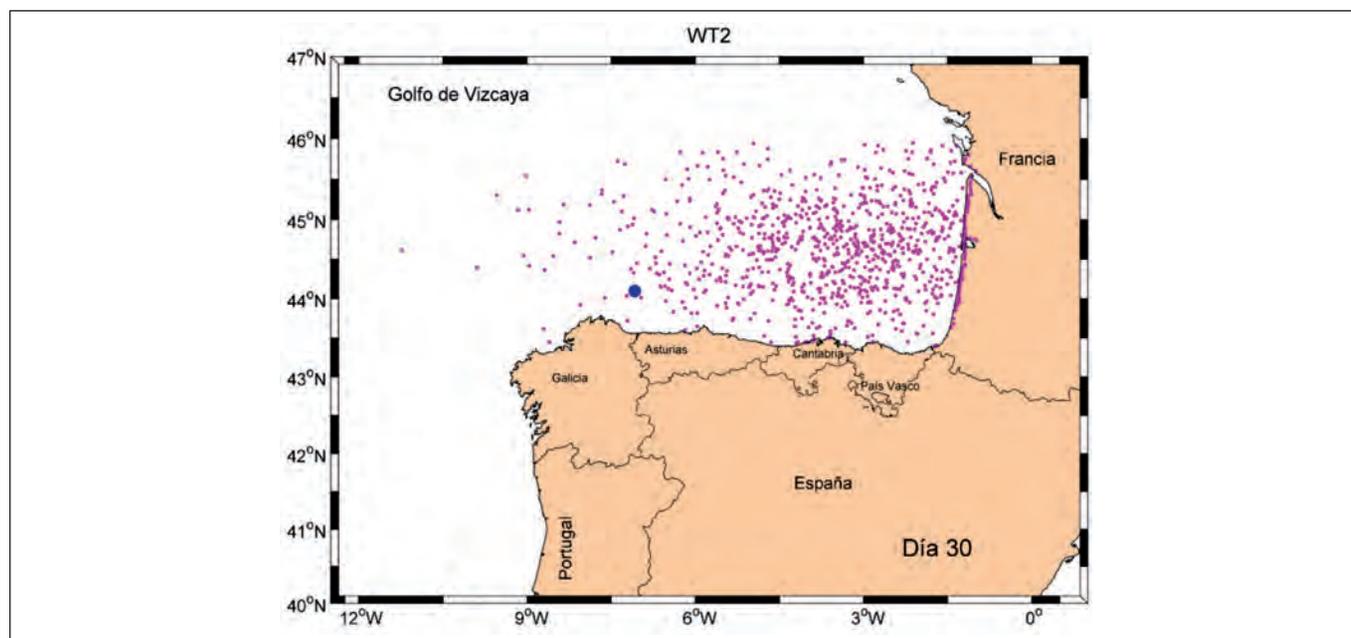


Figura 23. Evolución del derrame para el patrón de circulación 2 (WT 2).

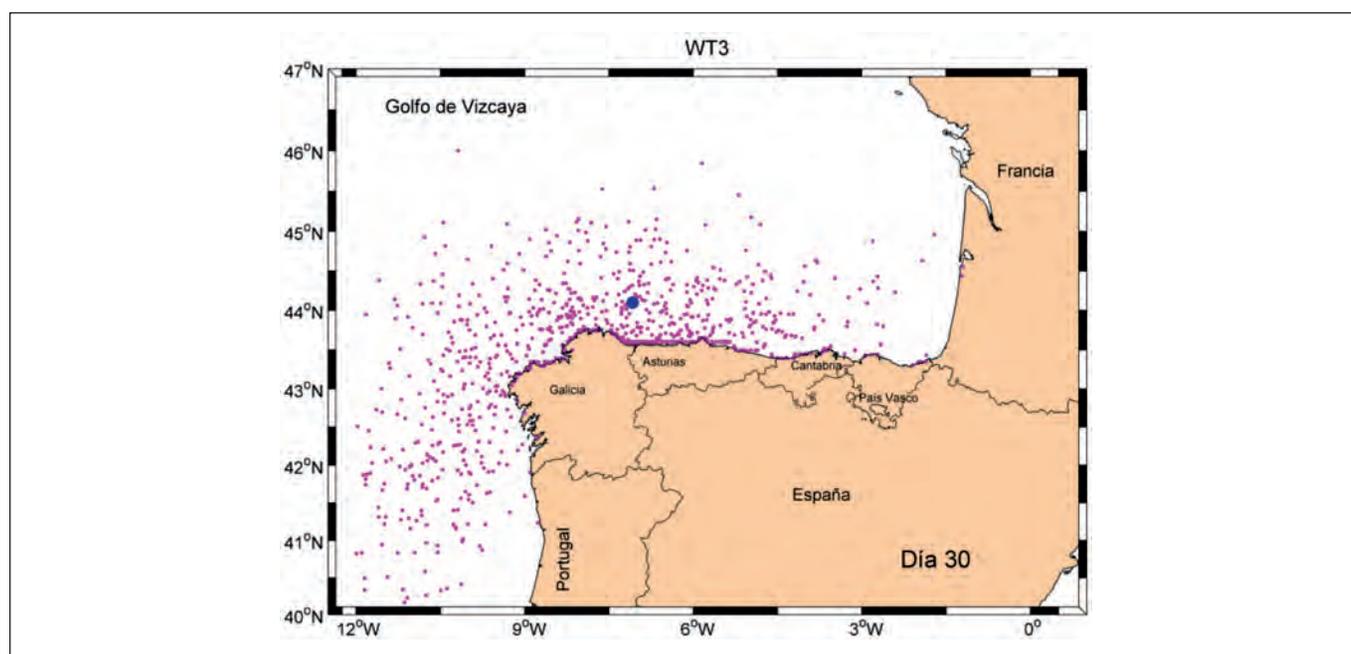


Figura 24. Evolución del derrame para el patrón de circulación 3 (WT 3).

Comparando, por ejemplo, las simulaciones presentadas en la figura 22 (WT 1) y figura 23 (WT 2) se observa como la evolución es totalmente diferente. Bajo un patrón de circulación atmosférica tipo WT 1, que representa una anomalía positiva de presión centrada en el Norte de Europa (véase la figura 12), la mayor parte de las trayectorias se dirigen hacia el oeste y suroeste, contaminando la costa de Galicia y dirigiéndose hacia Portugal. Por el contrario, bajo un patrón de circulación tipo WT 2, similar a la NAO (véase la figura 12), la

mayor parte de las trayectorias penetran en el Golfo de Vizcaya afectando principalmente la costa de Francia. Cabe destacar, que la mayor parte de las simulaciones permanecen en el mar después de 30 días. Nótese las diferencias con la evolución del vertido presentada en la sección anterior (correspondiente a WT 6), donde la mayor probabilidad de contaminación corresponde al litoral cantábrico, principalmente Asturias y Cantabria (véase la figura 21).

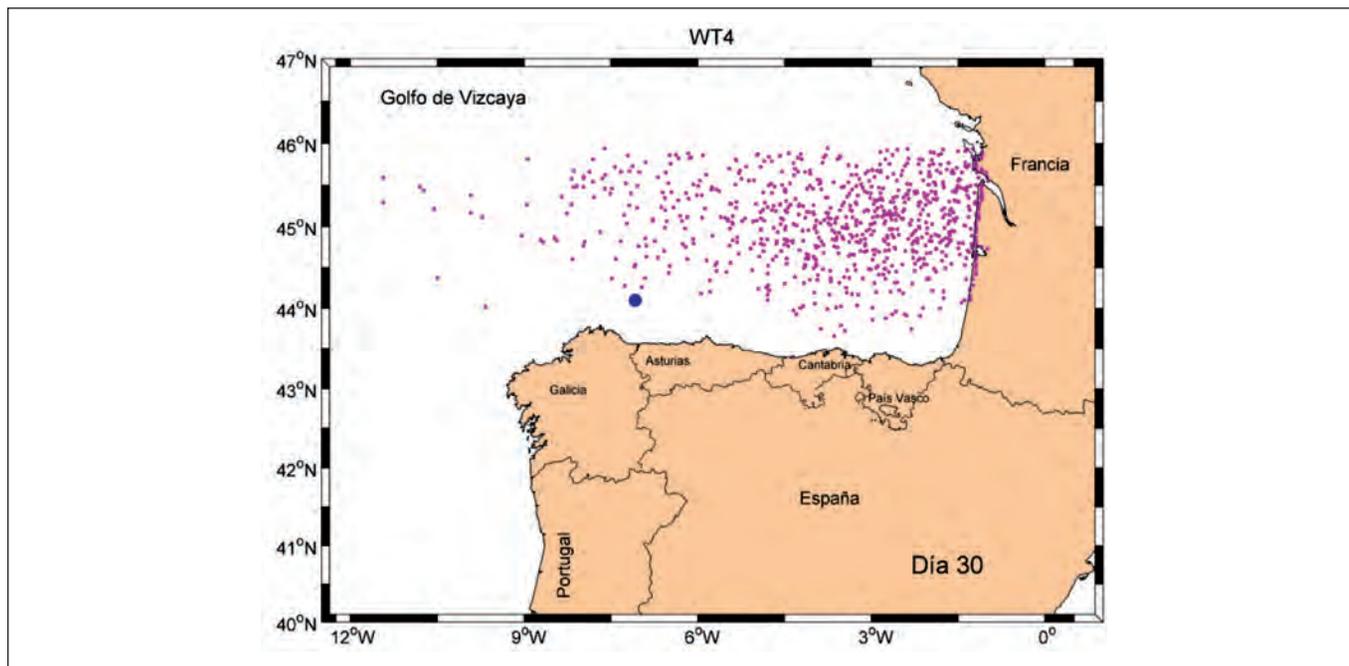


Figura 25. Evolución del derrame para el patrón de circulación 4 (WT 4).

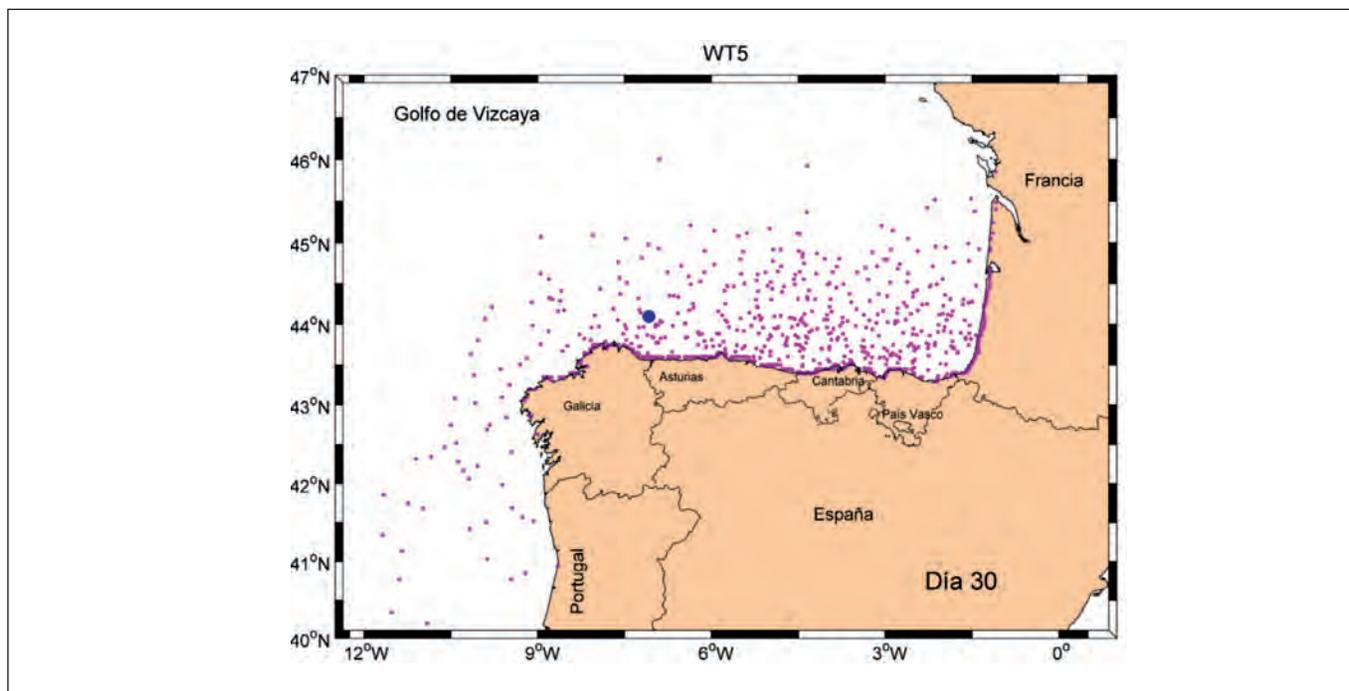


Figura 26. Evolución del derrame para el patrón de circulación 5 (WT 5).

En la figura 24 y 25 se observa la posición final de las trayectorias en función del patrón 3 y 4, que presentan un comportamiento similar al 1 (véase la figura 22) y 2 (véase la figura 23), descritos previamente. Nótese en la figura 25, que en este caso el derrame se ve restringido a la zona exterior del Golfo de Vizcaya, por lo que el litoral cantábrico no se ve contaminado. En la figura 26, se muestran los resultados en función del patrón 5 (WT 5). En este caso, el derrame afecta a todo el litoral del golfo,

desde la costa gallega hasta la costa francesa, e incluso parte de la costa portuguesa.

Para las tres condiciones sinópticas siguientes (WT 7, WT 8, WT 9), se observa que la mayor parte de las trayectorias simuladas permanece en el agua al cabo de 30 días, siendo el impacto en la zona costera menor para los patrones 7 y 8 (véase las figuras 27 y 28) y mayor para el patrón de circulación 9 (véase la figura 29).

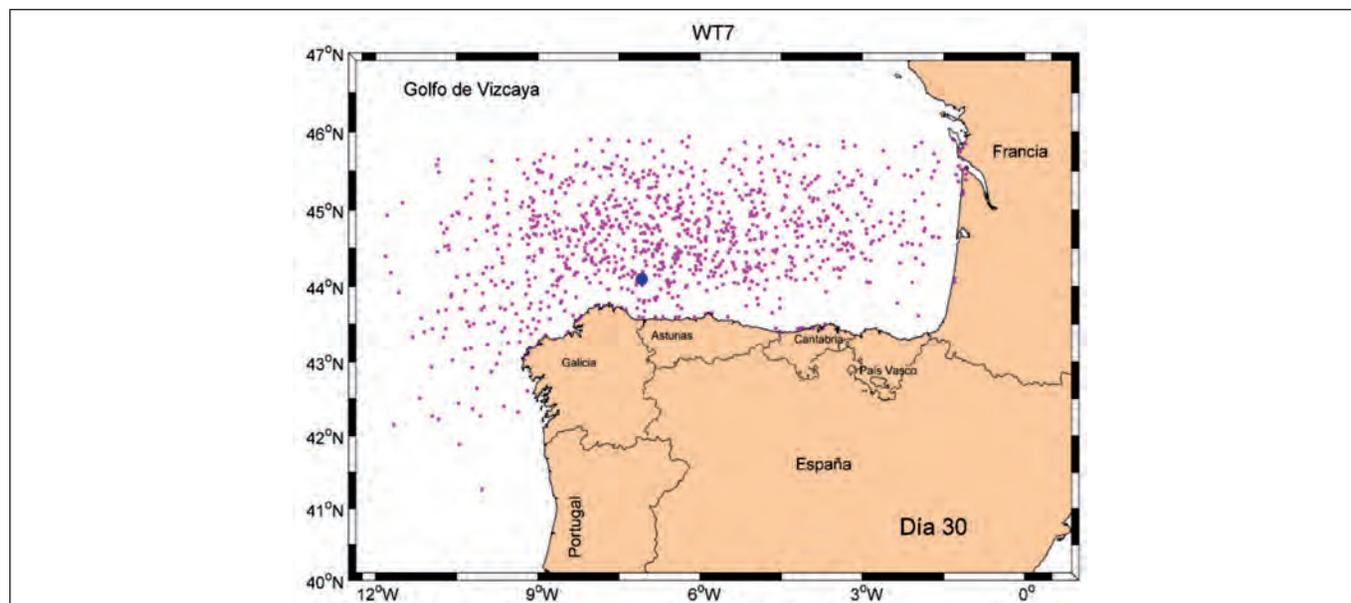


Figura 27. Evolución del derrame para el patrón de circulación 7 (WT 7).

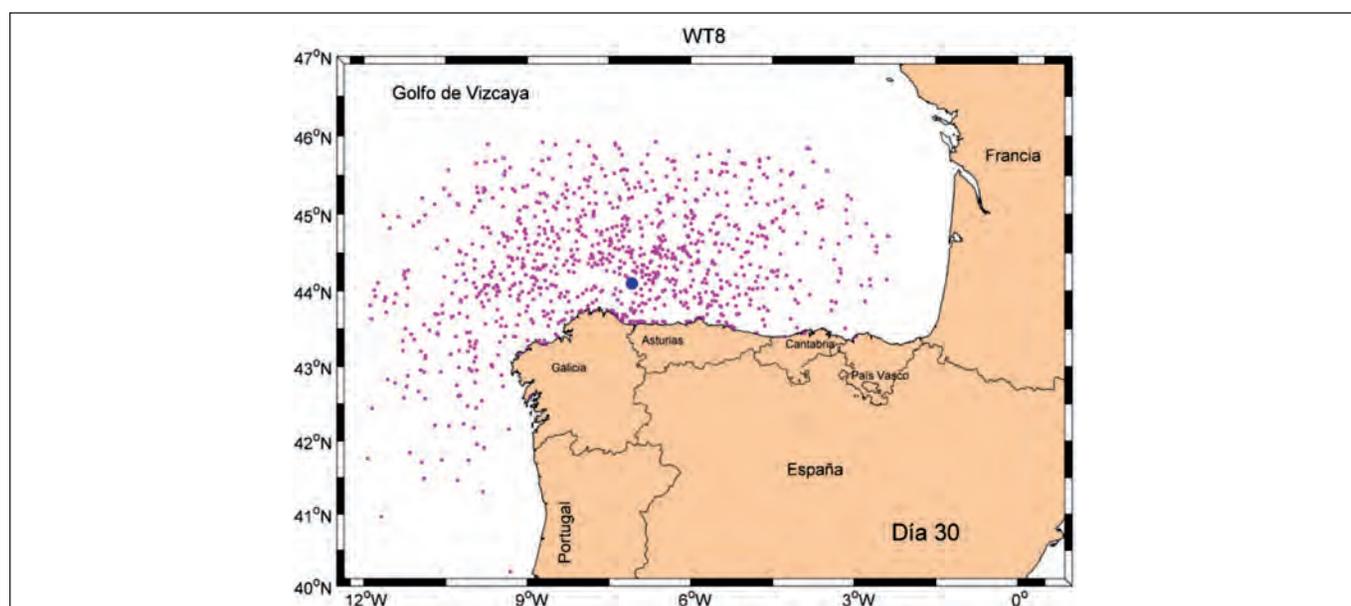


Figura 28. Evolución del derrame para el patrón de circulación 8 (WT 8).

Estos resultados muestran la importancia de tener en cuenta el patrón de circulación atmosférica del estado

inicial para simular estadísticamente la evolución del vertido y calcular la probabilidad de contaminación.

5. DISCUSIÓN

En este proyecto de investigación se ha desarrollado una metodología basada en métodos estadísticos para predecir a medio-largo plazo la evolución de un derrame de hidrocarburos en el medio marino. La metodología desarrollada combina técnicas de clasificación con procesos estocásticos, concretamente cadenas de Markov de primer orden, con el objetivo de relacionar las condiciones océano-meteorológicas actuales con bases de datos de reanálisis (históricas) y con ello calcular su evolución a medio plazo en términos probabilísticos. Una vez conocida la evolución del campo atmosférico y oceanográfico, esta información es utilizada para predecir la evolución del derrame.

La metodología se estructura en las siguientes partes:

1) Clasificación y selección de patrones atmosféricos. Las bases de datos de reanálisis de viento y presión a nivel del mar son clasificadas mediante la aplicación de análisis de componentes principales (*PCA*) y técnicas de clasificación (*K-Medias*). Como resultado se obtienen *K* estados de viento y *M* patrones de circulación atmosférica (o tipos de tiempo) representativos de la base de datos de partida.

2) Predicción probabilística de la evolución de viento y corrientes mediante cadenas de Markov.

- A partir de las bases de datos de viento y presión previamente clasificadas se genera la matriz de transición, la cuál proporciona para cada paso de tiempo la probabilidad de pasar de un estado de viento '*i*' a un estado de viento '*j*' teniendo en cuenta el patrón de circulación atmosférica del estado inicial.
- En caso de accidente, se simulará la evolución a medio-largo plazo de las condiciones océano-meteorológicas actuales mediante la aplicación de cadenas de Markov. El estado inicial de la cadena de Markov se busca mediante el método de análogos, que relaciona las condiciones atmosféricas reales (viento y presión) con el estado de viento y tipo de tiempo resultante de la clasificación. Con base en la matriz de transición se realizan *N* simulaciones de la evolución del campo de viento y corrientes en el período de predicción establecido (*S días*), obteniéndose un conjunto de *N x S* simulaciones de viento y corrientes.

3) Predicción probabilística de la evolución del vertido. A partir de los resultados obtenidos en el paso 2 se simulan *N* trayectorias de vertidos durante un período de *S* días. Las simulaciones se realizan con el modelo de transporte de hidrocarburos TESEO (Abascal *et al.*, 2007). De esta forma, se obtienen *S x N* simulaciones de trayectorias equiprobables que son utilizadas para calcular la probabilidad de contaminación originada por el vertido.

La metodología desarrollada se ha aplicado en el Golfo de Vizcaya, por lo que podría utilizarse para pre-

decir la trayectoria de un hipotético derrame que se produjera hoy en día en esta zona. Para ello se han analizado las bases de datos de forzamientos existentes y se han seleccionado las más adecuadas para el presente estudio: base de datos de viento en el período 1989-2009 procedentes del reanálisis SeaWind-EraInterim (Menéndez *et al.*, 2011), base de datos de presión a nivel del mar en el período 1948-actualidad procedentes del reanálisis NCEP/NCAR (Kalnay *et al.*, 1996) y base de datos de corrientes de marea meteorológica (gradientes de presión y viento), en el período 1989-2009 procedentes del reanálisis GOS (Abascal *et al.*, 2011). Dichas bases de datos se han clasificado, obteniéndose 49 estados de viento y 9 patrones de circulación atmosférica representativos de la zona de estudio. A partir de esta clasificación se ha generado la matriz de transición, de dimensiones 49 x 49 x 9, base de la cadena de Markov.

Con el objetivo de comparar los resultados obtenidos con datos reales se ha simulado una de las manchas observadas en el Golfo de Vizcaya durante el accidente del buque *Prestige*. Concretamente, se ha simulado una mancha observada el día 27/11/2002, en las coordenadas 7°O, 44.1° N (Cedre, 2003). El período de predicción considerado ha sido de 30 días (27/11/2002- 25/12/2002) con resolución temporal diaria. Los resultados obtenidos muestran que la mayor parte de las trayectorias se dirigen hacia el este, penetrando en el Golfo de Vizcaya y alcanzando después de 1 mes la costa de Francia. Las zonas con más probabilidad de contaminación son la costa de Asturias y Cantabria. Existe probabilidad, aunque menor de que el vertido se dirija hacia el SE alcanzando la costa de Galicia y derivando hacia Portugal. La validación de resultados probabilísticos es un proceso complejo, puesto que se necesita una gran cantidad de datos reales para poder estimar la probabilidad real de contaminación, información difícilmente disponible. No obstante, con el objetivo de validar la metodología, se han comparado los resultados obtenidos con observaciones del período *Prestige*, encontrándose un comportamiento acorde a la evolución del vertido en aquel período. Por otro lado, se ha estudiado como afecta el patrón de circulación atmosférica (relacionado con el paso de borrascas y anticiclones) a la evolución estadística de la trayectoria del vertido, encontrándose que la evolución del vertido y la probabilidad de contaminación depende del patrón de circulación atmosférica inicial. Nótese que las simulaciones han sido forzadas con datos de viento y corrientes de marea meteorológica, procesos más relevantes en la trayectoria del vertido, y que a la resolución temporal del trabajo (diaria), las corrientes de marea astronómica pueden despreciarse. No obstante, hay que tener en cuenta que hay procesos de la circulación general que no se están considerando, como por ejemplo la corriente dirigida hacia el polo (*Poleward Current*) que tiene influencia sobre el transporte en la zona (Sánchez y Gil, 2000) y que contribuiría a introducir el vertido en el Golfo de Vizcaya.

Los resultados obtenidos muestran la capacidad de la metodología desarrollada y su utilidad para proporcionar información necesaria en la planificación y respuesta ante un derrame de hidrocarburos al medio marino. Una aplicación inmediata, de la que ya hay demanda en el sector, es su integración en los sistemas operacionales de predicción de deriva para dar respuesta en tiempo real. En caso de accidente, el sistema operacional proporcionaría la predicción a corto plazo (2-3 días) basada en las técnicas ya existentes, y la predicción estadística a medio-largo plazo (para los próximos 15 días – 1 mes) con base en la metodología presentada. La incorporación de esta metodología en los sistemas operacionales de predicción supondrá una mejora en la planificación y respuesta frente a la contaminación por hidrocarburos en la costa y en el medio marino.

6. CONCLUSIONES

Del trabajo realizado se pueden extraer las siguientes conclusiones:

- En este trabajo se ha desarrollado una metodología que permite predecir estadísticamente la evolución de derrames en el mar a medio-largo plazo.
- La metodología desarrollada integra la utilización de bases de datos de reanálisis (viento, presión atmosférica y corrientes) de amplia cobertura espacial y temporal para estudiar la evolución en el pasado de las dinámicas. Las condiciones océano-meteorológicas actuales se relacionan con las bases de datos de reanálisis para calcular su evolución a medio plazo en términos probabilísticos.
- Se han utilizado técnicas de clasificación, técnicas de análogos y cadenas de Markov, ampliamente utilizadas en otras disciplinas como meteorología y clima marítimo, y se han aplicado de forma novedosa al campo de la oceanografía operacional y predicción del transporte de hidrocarburos.
- La metodología desarrollada integra las siguientes partes: (1) clasificación y selección de patrones atmosféricos, (2) predicción probabilística de viento y corrientes mediante Cadenas de Markov y (3) predicción probabilística de la evolución del

vertido. Como resultado se obtiene la predicción estadística de la trayectoria del vertido y su probabilidad de contaminación.

- La metodología se ha aplicado en el Golfo de Vizcaya y se ha utilizado para simular una de las manchas observadas en el Golfo de Vizcaya durante el período del *Prestige*. La comparación con la evolución de las manchas durante ese período, muestra que los resultados obtenidos concuerdan con las observaciones.
- Se ha analizado la influencia del patrón de circulación atmosférica (o tipo de tiempo) en la evolución estadística de la trayectoria del derrame, encontrándose que la evolución del vertido y la probabilidad de contaminación depende del patrón de circulación atmosférica inicial.
- La metodología desarrollada puede implementarse de forma operacional en los sistemas operacionales de predicción de deriva para ser utilizada en la respuesta ante un derrame de hidrocarburos al medio marino.
- La metodología desarrollada supone una innovación y una mejora los sistemas de prevención y respuesta frente a la contaminación por hidrocarburos en la costa y en el medio marino.

7. FUTURAS LÍNEAS DE INVESTIGACIÓN

A continuación, se plantean las futuras líneas de investigación que complementan el presente trabajo:

- Analizar la sensibilidad de los resultados al número de patrones de circulación atmosférica de la zona de estudio.
- Estudiar la evolución de la trayectoria y la probabilidad de contaminación del derrame, teniendo en cuenta la estacionalidad y la influencia de índices climáticos (como la NAO).
- Desarrollar un modelo logístico autoregresivo que permita la utilización de cadenas de Markov de mayor orden.
- Validación de la metodología con datos de boyas de deriva y más observaciones obtenidas durante el accidente del *Prestige*.

8. ANEJO I. BASE DE DATOS DE REANÁLISIS

AI.1. INTRODUCCIÓN

En este anejo se presenta la descripción de las bases de datos de reanálisis utilizados para la realización del trabajo: base de datos de viento en el período 1989-2009 procedentes del reanálisis SeaWind-EralInterim (Menéndez *et al.*, 2011), base de datos de presión a nivel del mar en el período 1948-actualidad procedentes del reanálisis NCEP/NCAR (Kalnay *et al.*, 1996) y base de datos de corrientes de marea meteorológica (gradientes de presión y viento), en el período 1989-2009 procedentes del reanálisis GOS (Abascal *et al.*, 2011).

AI.2. DATOS DE VIENTO

Con respecto a los datos de viento se ha utilizado la base de datos de reanálisis SeaWind (SW), desarrollada por IH

Cantabria. La base de datos SeaWind está constituida por los campos de viento horarios resultantes de la dinámica atmosférica durante un periodo histórico. En particular se han desarrollado dos reanálisis paralelos. El SeaWind-NCEP (SW-NCEP), que cubre el periodo 1948-2010, y el SeaWind-EralInterim (SW-Eral), durante el periodo 1989-2010. En este trabajo se ha utilizado la base de datos SeaWind-EralInterim debido a que presenta una mayor resolución espacial.

Para su elaboración se ha utilizado el modelo WRF (Weather Research and Forecasting, más detalles en: Skamarock *et al.*, 2008). En concreto la versión 3.1.1 del WRF-ARW (Advanced Research dynamical solver, ARW) que constituye uno de los modelos atmosféricos más modernos (versión Abril 2009), ampliamente utilizado por la comunidad científica (por ej. Lo *et al.*, 2008 muestra su aplicación en Estados Unidos). El dominio simulado cubre casi al completo la región europea con una resolución media de 30 Km para el SW-NCEP (figura AI.1) y de aproximadamente 15 Km para el SW-Eral (figura AI.2).

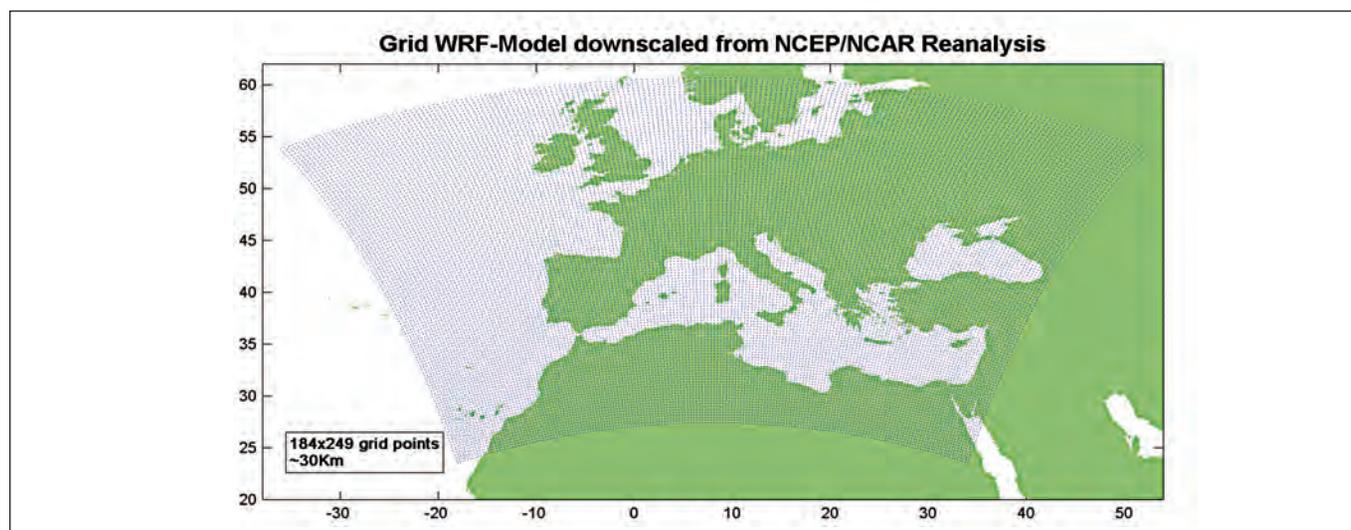


Figura AI.1. Dominio de la malla SW-NCEP.

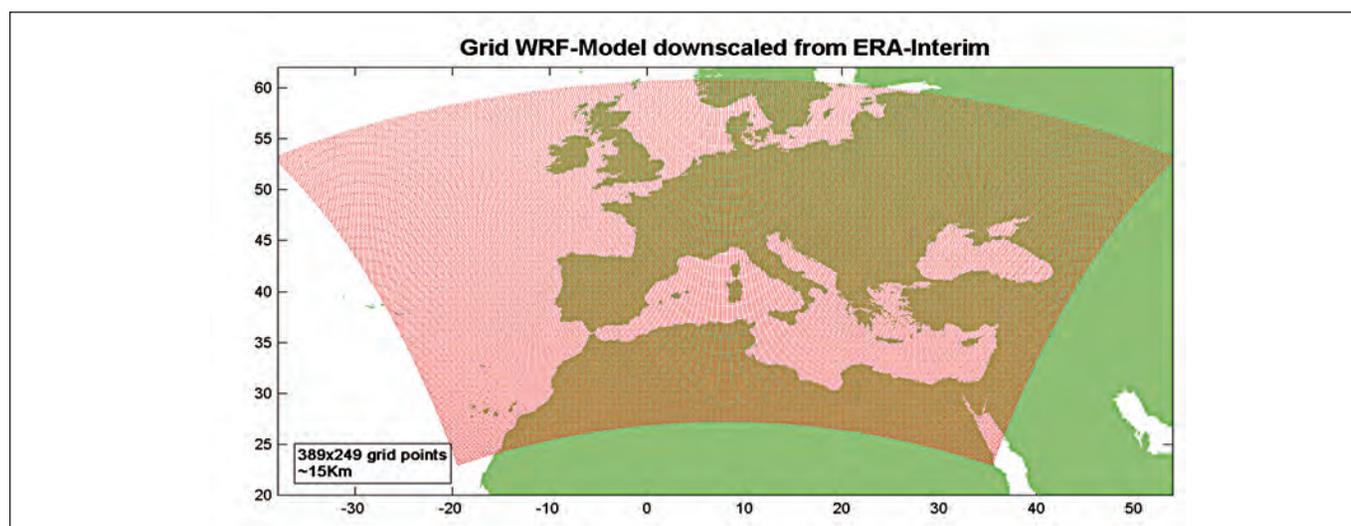


Figura AI.2. Dominio de la malla SW-Eral.

Los forzamientos de contorno e inicio utilizados proceden del reanálisis global NCEP-NCAR I (Kalnay *et al.*, 1996) para el SW-NCEP y del reanálisis global ERA-Interim (Dee y Uppala, 2009) para el SW-Eral. El desarrollo de estos dos reanálisis regionales en paralelo se justifica debido a que el primero, de resolución más grosera, permite el estudio de aspectos que requieren información de largo plazo, como análisis de cambios en las climatologías, mientras que el segundo permite obtener una información más detallada a nivel espacial a pesar de proporcionar información más corta en el tiempo.

Las bases de datos generadas han sido validadas con las observaciones procedentes de la red de boyas y de estaciones meteorológicas del Organismo Público Puertos del Estado (OPPE), REDEXT y REMPOR. Estas medidas disponen de información de la velocidad y dirección del viento a distintas alturas. También se ha utilizado

la información satelital procedente de las misiones: ERS-2(1995-2003), Envisat(2002-), GFO(2000-2008), Jason-1(2002-), Jason(2008-) y T/P (1992-2005). Previo control de calidad y homogeneidad, esta base de datos proporciona información de velocidad del viento a 10 metros de altura.

La validación mediante boyas se ha realizado utilizando el nodo de reanálisis más cercano, comparando los datos horarios. Cabe señalar que se ha considerado la aproximación del perfil de viento potencial ante una situación de atmósfera neutra con el parámetro potencial de 0,12 para los datos REDEXT y 0.143 para los datos REMPOR. En la figura AI.3 se muestra una comparación entre la boya de Villano-Sisargas (en negro) y la serie temporal de SW-HR (en verde). En esta figura se puede apreciar como el *downscaling* simula correctamente tanto la estructura temporal de la magnitud del viento como su dirección.

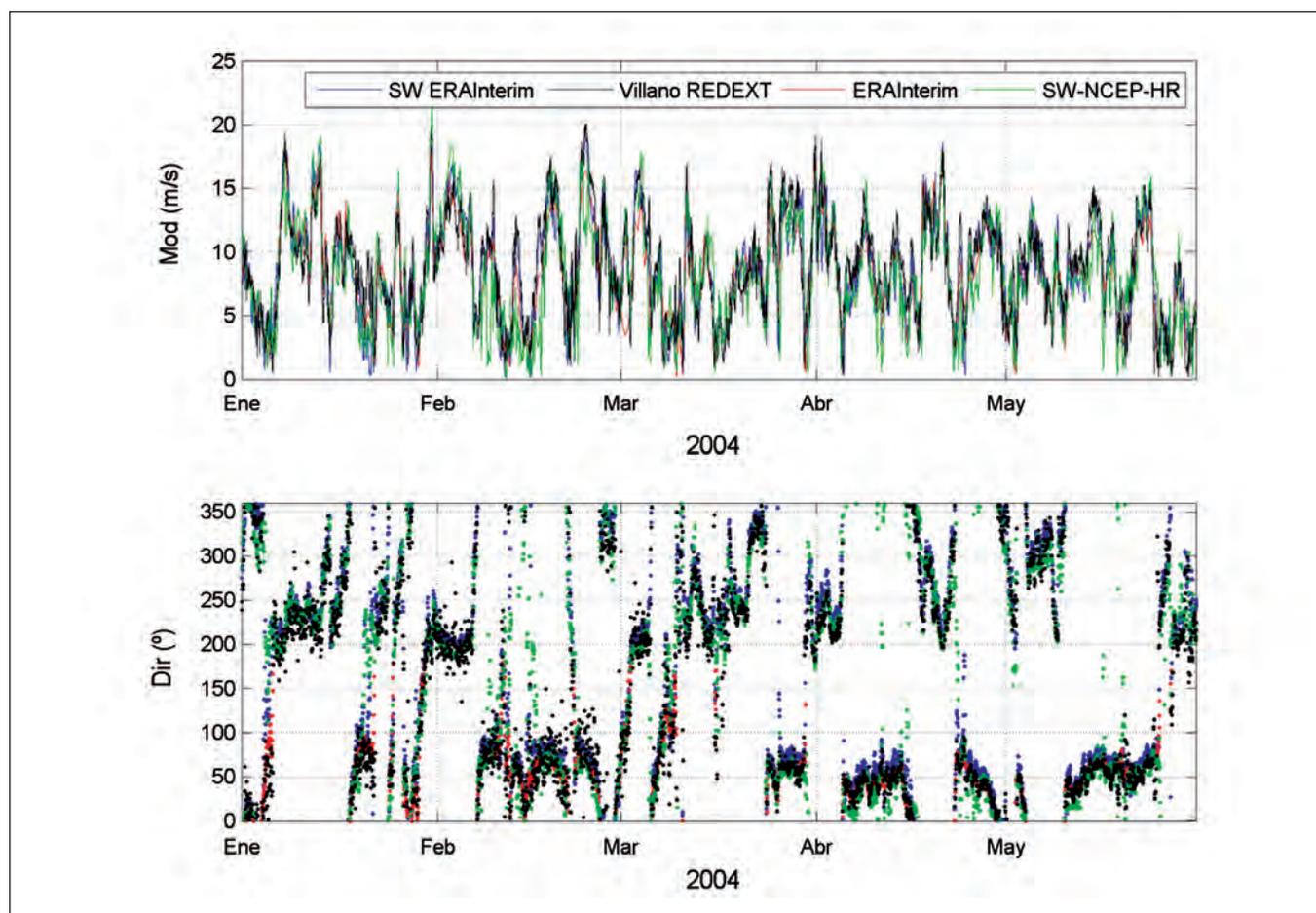


Figura AI.3. Series temporales de la boya Villano-Sisargas frente a diversas simulaciones procedentes de modelado atmosférico.

En la figura AI.4 se presentan para distintas boyas, los índices estadísticos y el diagrama de dispersión de los pares de datos. La escala de colores representa la densidad de datos.

AI.3. DATOS DE PRESIÓN

La base de datos de presión seleccionada para aplicar la metodología desarrollada a la zona piloto es el reanálisis global NCEP/NCAR (Kalnay *et al.*, 1996).

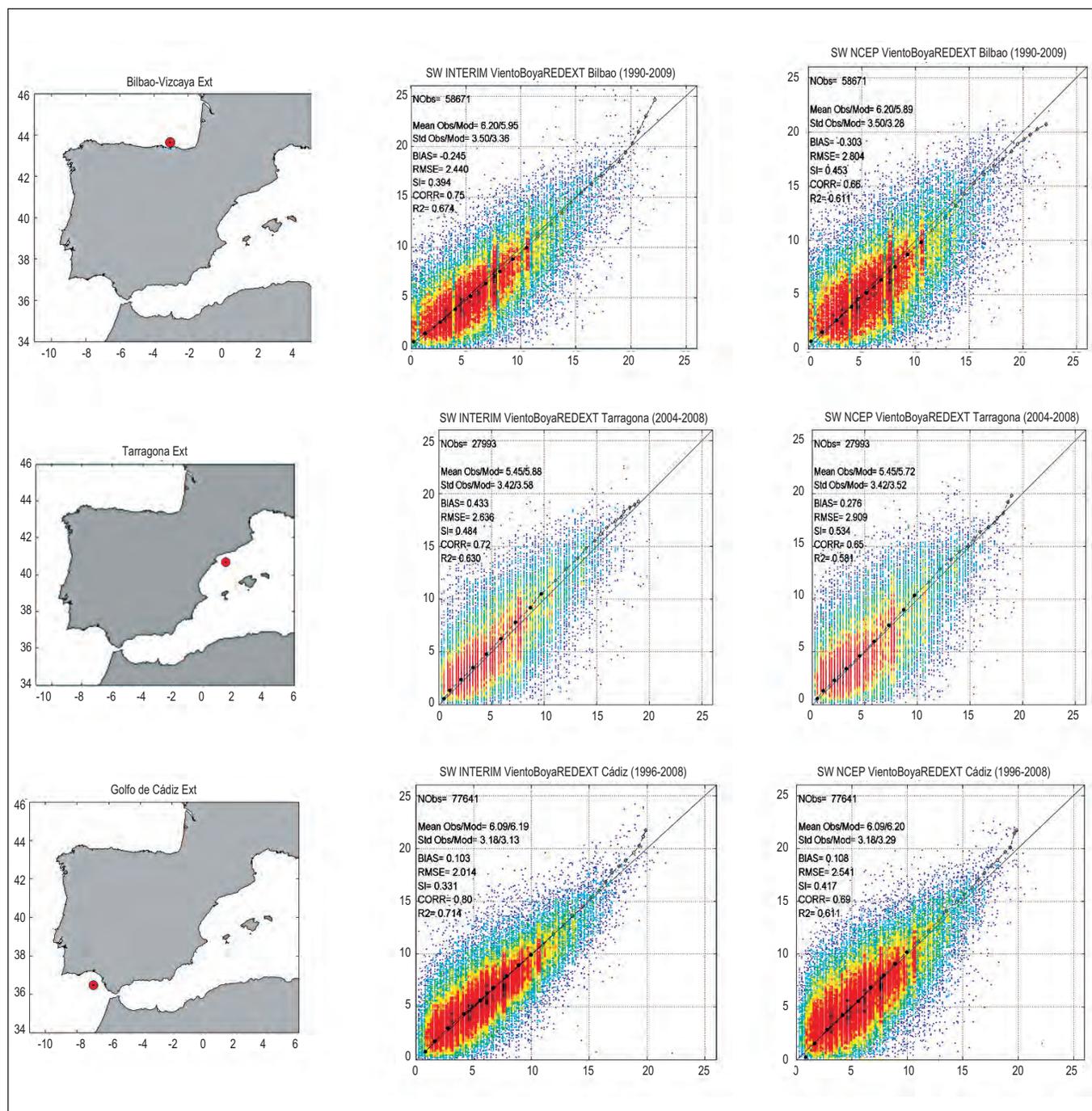


Figura AI.4. Validación de los datos de SW-NCEP y SW-Eral. Índices estadísticos y diagramas de dispersión.

Esta base de datos ha sido realizada por el National Center for Environmental Prediction (NCEP) y el National Center for Atmospheric Research (NCAR). Los datos han sido aportados por NOAA/OAR/ESRLPSD, Boulder, Colorado, USA y se han obtenido de su página web <http://www.esrl.noaa.gov/psd/>

La malla empleada en el reanálisis global NCEP/NCAR, abarca prácticamente todo el globo terráqueo (latitud: 90° N-90° S, longitud: 0° E-357.5° E). Los datos presentan una resolución espacial 2.5° y una resolución temporal 6-horaria de los campos de presión, cubriendo el periodo de tiempo desde 1948 hasta la actualidad.

AI.4. DATOS DE CORRIENTES DE MAREA METEOROLÓGICA

Para obtener las corrientes y el nivel del mar generados por la marea meteorológica existe la base de datos de reanálisis GOS (*Global Ocean Surges*) (Abascal et al., 2011) generada por el IH Cantabria para la costa suroeste de Europa (figura AI.5).

El modelo seleccionado para el cálculo de la marea meteorológica ha sido el modelo tridimensional de circulación ROMS (*Regional Ocean Modeling System*), descrito previamente. En este caso, dado el carácter barotrópico

del fenómeno a resolver, se ha utilizado la versión 2D del modelo.

En lo que se refiere a los forzamientos utilizados, éstos son: el reanálisis desarrollado por el organismo NCEP/NCAR y la base de datos ERA-Interim. Para obte-

ner la marea meteorológica con alta resolución espacial en la costa española se ha realizado un *downscaling* dinámico de los campos de viento y presión mediante el modelo WRF: proyecto SeaWind.

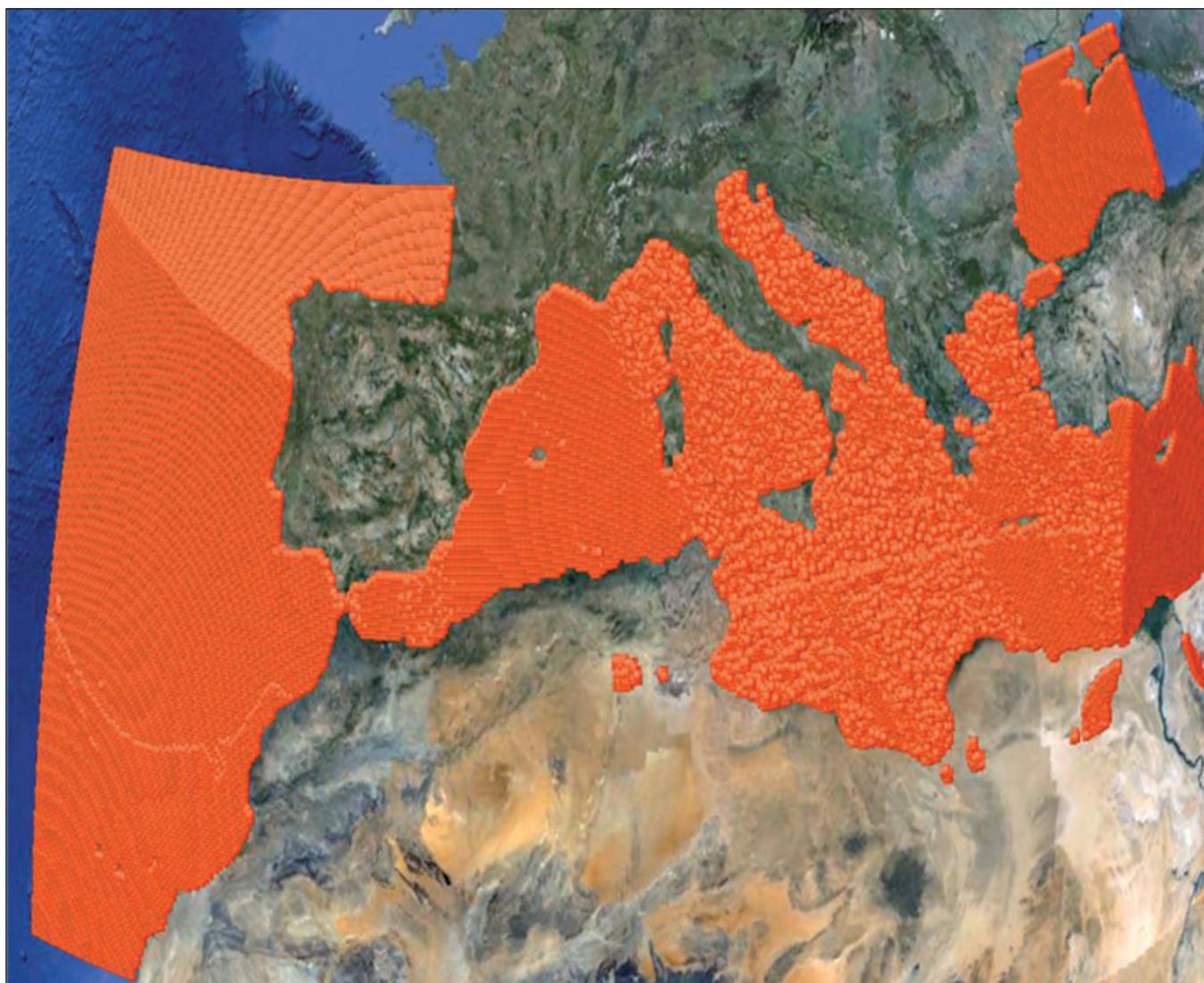


Figura A1.5. Dominio de la base de datos GOS.

Por lo tanto, para el Golfo de Vizcaya, existen dos bases de datos de marea meteorológica:

- GOS 1.1, forzado con campos de viento y presión de SeaWind-NCEP (30 km de resolución). Esta base de datos tiene resolución espacial de $1/8^\circ$ y temporal horaria. Cubre el periodo 1948-2009.

- GOS 2.1, forzado con campos de viento y presión de SeaWind-EraInterim (15 km de resolución). Esta base de datos tiene resolución espacial de $1/8^\circ$ y temporal horaria. Cubre el periodo 1989-2009.

En este trabajo se ha utilizado la base de datos GOS 2.1, por estar forzada con datos atmosféricos de mayor resolución espacial.

9. ANEJO II. ANÁLISIS DE COMPONENTES PRINCIPALES

AII.1. ANÁLISIS DE COMPONENTES PRINCIPALES

Generalmente, en los espacios de alta dimensionalidad, las distintas componentes o dimensiones suelen estar correlacionadas y, por tanto, se emplea información redundante en su descripción. El análisis de las componentes principales es una técnica para reducir la dimensionalidad de una muestra de datos que presenta redundancia de información reteniendo la máxima variabilidad existente. La compresión de la información se consigue mediante la proyección de los datos en un espacio de menor dimensión que el original conservando el máximo de la varianza de la muestra. La base del nuevo espacio se forma con aquellos vectores donde la muestra proyectada presenta mayor varianza (Camus, 2009).

Se parte de una muestra de N observaciones de un vector x_k :

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

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

$$X = \begin{pmatrix} x_{11} & \dots & x_{N1} \\ \vdots & & \vdots \\ x_{1n} & \dots & x_{Nn} \end{pmatrix} \quad (\text{Ec. 2})$$

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

$$X = \sum_{i=1}^n e_i x_i \approx \bar{X} = \sum_{i=1}^d f_i \bar{x}_i \quad (\text{Ec. 3})$$

El cálculo matemático consiste en estimar la matriz de varianzas y covarianzas de la muestra de datos. Los autovectores (o *Funciones Ortogonales Empíricas*, EOFs) de esta matriz son los nuevos vectores f_j , y los correspondientes autovalores (λ_j) indican la varianza explicada (la varianza de la muestra proyectada sobre el vector). Los coeficientes de cada elemento de la muestra en la nueva base se denominan *Componentes Principales* o *Principal Components* (PCs). Los patrones espaciales de los EOFs se pueden interpretar como los modos de variabilidad del campo de los datos, y como la nueva base es ortogonal, no están correlacionados.

En realidad se obtienen n vectores pero se eligen un número d , menor al original, que representan una determinada varianza de los datos. Los d vectores se eligen en orden decreciente de varianza, y por tanto, los vectores originales quedarán representados por PCs de dimensión d . Matricialmente, esto se puede representar como:

$$\bar{x}_k = \begin{pmatrix} \overline{x_{1k}} \\ \vdots \\ \overline{x_{dk}} \end{pmatrix} = F^T x_k = \begin{pmatrix} f_{11} & \dots & f_{1n} \\ \vdots & & \vdots \\ f_{d1} & \dots & f_{dn} \end{pmatrix} = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix} \quad (\text{Ec. 4})$$

Las columnas de la matriz F^T son los autovalores f_j , que se van a denominar *EOF_j(s)*, donde s puede ser el espacio geográfico. Por otro lado, los vectores proyectados en este espacio o componentes principales se van a denominar *PC_i(t)*. Por tanto, el campo original se puede expresar como:

$$X \approx \sum_{i=1}^d EOF(s) PC_i(t) \quad (\text{Ec. 5})$$

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

Al proyectar estos vectores al espacio original, se obtendrán una aproximación del vector original (mejor cuanto mayor sea la dimensión d del espacio proyectador).

$$x_k = F^T \bar{x}_k \quad (\text{Ec. 6})$$

10. ANEJO III. TÉCNICAS DE SELECCIÓN Y CLASIFICACIÓN

AIII.1. INTRODUCCIÓN

Las técnicas de clasificación y selección extraen información básica de las bases de datos mediante la detección o selección de un grupo reducido de datos representativo de toda la base de datos de partida. Estas técnicas exploran el espacio de los datos pero con objetivos diferentes. Las técnicas de clasificación dividen el conjunto de datos en subconjuntos siguiendo un criterio de similitud, mientras que las técnicas de selección identifican un conjunto de elementos de la base original representativos de la diversidad de los datos.

El método estándar para la búsqueda de un conjunto de vectores análogos a un vector dado es el método de *k-vecinos*, que consiste en seleccionar el conjunto de los *k* vectores más próximos al vector dado según una métrica prefijada (normalmente la métrica Euclídea). En esta tarea se ha analizado la utilización de técnicas más complejas que el tradicional método de *k-vecinos*. Concretamente, para la clasificación se ha analizado la técnica basada en *K-Medias* (descrita en el apartado AIII.1) y para la selección, el Algoritmo de Máxima Disimilitud (*MaxDiss*) (apartado AIII.2).

A continuación se describen ambas técnicas.

AIII.2 ALGORITMO DE MÁXIMA DISIMILITUD (MAXDISS)

Los algoritmos de máxima disimilitud consisten en la selección de un subconjunto de datos representativo de la muestra de datos de partida. Por tanto, si se dispone de una muestra de datos $X = \{x_1, x_2, \dots, x_N\}$ compuesta por N vectores n -dimensionales, el objetivo de este algoritmo consiste en obtener un número M de vectores $X = \{v_1, \dots, v_M\}$ de los datos de partida que representen la diversidad de estos datos. El algoritmo comienza con la inicialización del subconjunto $\{v_j\}$ mediante la transferencia de un dato de la base de partida. La selección del resto de elementos se realiza iterativamente, y en cada ciclo se transfiere al subconjunto aquel dato perteneciente a la base de datos de partida con la mayor disimilitud respecto al subconjunto seleccionado (Camus, 2009).

Este algoritmo fue descrito por Kennard y Stone (1969) y admite distintas versiones dependiendo del criterio considerado en la inicialización del subconjunto y en el criterio de selección del resto de datos del subconjunto.

El elemento inicial del subconjunto $\{v_j\}$ puede ser escogido:

- de forma aleatoria
- el elemento con mayor disimilitud de la base de datos de partida.
- el dato más próximo al centro de la base de datos.

Una vez seleccionado el primer elemento del subconjunto, la selección del resto de elementos se realiza en dos fases. En primer lugar, para cada dato aún perteneciente a la base de datos de partida, se calcula la distancia o disimilitud con todos los elementos del subconjunto y se define una única distancia entre este dato y el subconjunto, según distintos criterios. En segundo lugar, una vez que se dispone de la distancia entre cada dato de la muestra de partida y el subconjunto, se elige aquel dato con el máximo valor de esta distancia.

Por ejemplo, si en el subconjunto ya existen R ($R \leq M$) datos seleccionados, primero se calcula la disimilitud entre el dato i de la muestra de datos $N-R$ y los j elementos pertenecientes al subconjunto R :

$$d_{ij} = \|x_i - v_j\|; i = 1, \dots, N-R; j = 1, \dots, R \quad (\text{Ec.7})$$

Posteriormente se calcula la disimilitud i , subconjunto d entre el dato i y el subconjunto R . En segundo lugar, una vez calculadas las $N-R$ disimilitudes, el dato con la máxima disimilitud es seleccionado para ser incluido en el subconjunto.

AIII.3 TÉCNICA DE CLASIFICACIÓN DE LAS K-MEDIAS (K-MEANS)

Las técnicas de clasificación extraen información básica de las bases de datos mediante la detección o selección de un grupo reducido de datos representativo de la base de datos de partida. En este estudio se emplea la técnica de clasificación *K-Medias*. El algoritmo *K-Medias* divide el conjunto de datos de partida en un determinado número de subconjuntos, define cada uno de ellos por un prototipo, y forma cada grupo con los datos para los cuales ese prototipo es el más cercano. Para ello, en primer lugar se establece el número de grupos deseado para inicializar los prototipos (o centroides) y a continuación, el algoritmo procede de forma iterativa moviendo estos centroides hasta minimizar la varianza total intra-grupos. Esta varianza total intra-grupos significa que al final del proceso, la suma de distancias entre el prototipo y los datos de entrada correspondientes a cada prototipo es mínima (Camus 2009).

Por lo tanto, el objetivo de la aplicación de la técnica es obtener un número M de subconjuntos o grupos definidos cada uno de ellos por un prototipo o centroide $v_k = \{v_{1k}, \dots, v_{nk}\}$ de la misma dimensión que los datos de partida, siendo $k = 1, \dots, M$.

El proceso de clasificación consiste en un ajuste iterativo de los centroides. Los centroides iniciales $\{v_{10}, v_{20}, \dots, v_{M0}\}$ con los que comienza el proceso se han seleccionado mediante la aplicación de la técnica *MaxDiss*, descrita previamente, a los datos de partida. Este algoritmo tiene como objetivo identificar aquel subconjunto M , que represente la mayor diversidad posible del total de la base N de datos. Este algoritmo comienza con la inicialización del subconjunto mediante la elección de un dato de la base de origen. Los siguientes elementos del subconjunto seleccionados, se corresponden con los datos

aún pertenecientes a la base de datos de partida que presenta mayor disimilitud con respecto al subconjunto seleccionado.

Una vez inicializado el proceso de clasificación, en cada ciclo se identifican los datos de entrada más próximos a cada centroide y el nuevo centroide se redefine como la media de los vectores de entrada correspondientes a dicho centroide. Por ejemplo, si en la iteración $(r+1)$, cada vector x_i es reasignado al grupo j , el centroide es actualizado según:

$$v_j^{r+1} = \sum_{x_i \in C_j} x_i / n_j \quad (\text{Ec.8})$$

donde n_j es el número de vectores de entrada pertenecientes al grupo j , y dicho grupo se puede definir como:

$$j = \min\{x_i - v_j, j = 1, \dots, M\} \quad (\text{Ec.9})$$

Después de varias iteraciones, el proceso converge cuando la pertenencia de los vectores a un grupo determinado se estabiliza.

A modo de ejemplo y para ilustrar el objetivo de la técnica, en la figura AIII.1 se muestra la inicialización de centroides de un conjunto N de datos y la clasificación final. La figura a) muestra la inicialización de los centroides, marcados con círculos, y la definición de los mismos después de la primera iteración (puntos) mientras que los centroides actualizados después de esta primera iteración se han representado con puntos rellenos. En la figura b) se muestran los centroides y subconjuntos finales.

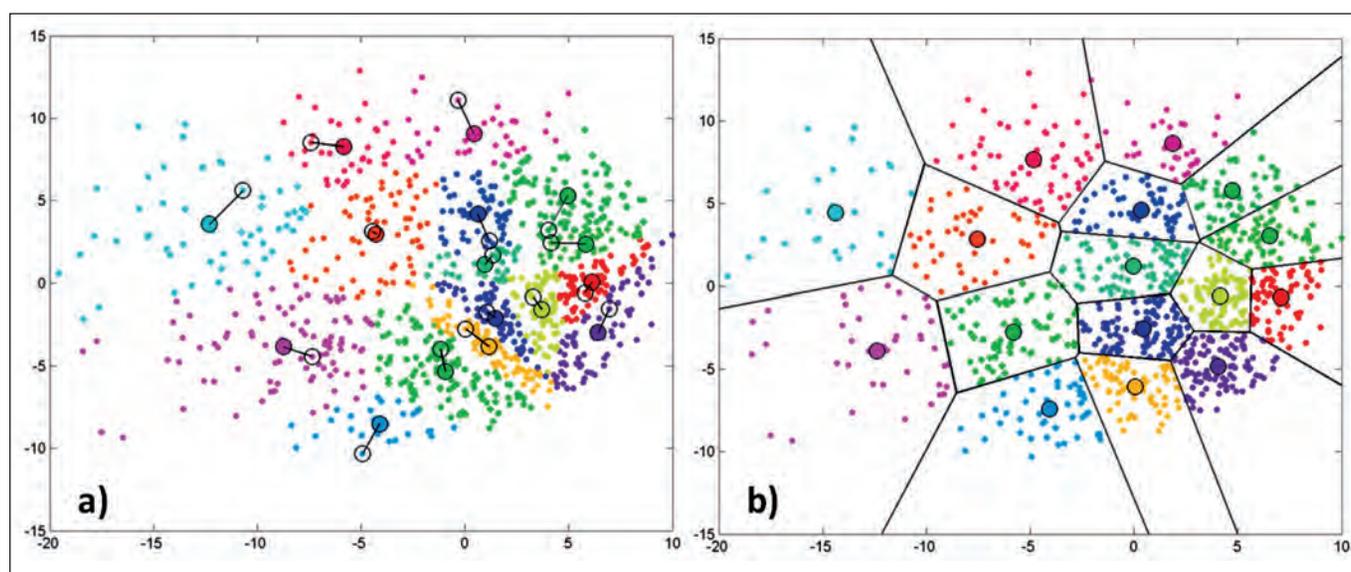


Figura AIII.1. Ejemplo de aplicación del algoritmo K-medias: a) Inicialización de los centroides (círculos) y definición de los mismos después de la primera iteración (puntos), b) Centroides y subconjuntos finales. (Fuente: Camus, 2009).

11. REFERENCIAS BIBLIOGRÁFICAS

- Abascal, A.J.; Castanedo, S.; Gutierrez, A.D.; Comerma, E.; Medina, R.; Losada, I.J. TESEO, an Operational System for Simulating Oil Spills Trajectories and Fate Processes. *ISOPE-2007 (International Offshore (Ocean) and Polar Engineering Conference)*, 2007 (3) 1751-1758.
- Abascal, A.J.; Castanedo, S.; Gutierrez, A.D.; Comerma, E.; Medina, R.; Losada, I.J. Description and application of the operational oil spill forecast system TESEO. *Proceedings of the 2008 International Oil Spill Conference*, Savannah, Georgia (EEUU), 2008.
- Abascal, A.J. Avances en el modelado de transporte de hidrocarburos. Aplicación en la respuesta a vertidos marinos accidentales. *Tesis Doctoral*. Universidad de Cantabria, 2009.
- Abascal, A.J.; Castanedo, S.; Mendez, F.J.; Medina, R.; Losada, I.J. Calibration of a Lagrangian transport model using drifting buoys deployed during the Prestige oil spill. *Journal of Coastal Research*, 2009 (25(1)) 80-90.
- Abascal, A.J.; Castanedo, S.; Medina, R.; Liste, M. Analysis of the reliability of a statistical oil spill response model. *Marine Pollution Bulletin*, 2010 (60) 2099-2110.
- Abascal, A.J.; Castanedo, S.; Cid, A.; Medina, R. High resolution dynamic downscaling of storm surge in Southern Europe (GOS, Global Ocean Surges). Twelfth Session of the GLOSS Group of Experts, *GLOSS GE XII and associated Workshops*, France-November 7-11, 2011.
- Alvarez Fanjul, E.; Losada, I.; Tintoré, J.; Menéndez, J.; Espino, M.; Parrilla, G.; Martínez, I.; Muñozuri, V.P. The ESEOO Project: developments and perspectives for Operational Oceanography at Spain. Proc. ISOPE-2007: The 17th International Offshore Ocean and Polar Engineering Conference. Lisbon, Portugal, *The International Society of Offshore Ocean and Polar Engineers (ISOPE)*, 2007 (3) 1708-1715.
- Al-Rabeh, A.H.; Lardner, R.W.; Gunay, N. Gulfspill Version 2.0: a software package for oil spills in the Arabian Gulf. *Environmental Modelling and Software*, 2000 (15) 425-442.

- Ancell, R.; Bermejo, M.; Oviedo, L. Cambios observados en los tipos de tiempo. *AME Boletín*, 2009 (25) 32-36.
- ASCE. State-of-the-art review of modeling transport and fate of oil spills, ASCE Committee on Modeling Oil Spills. Water Resources Engineering Division. *Journal of Hydraulic Engineering*, 1996 (122(11)) 594-609.
- Bardossy, A.; Capary, H.J. Detection of Climate Change in Europe by Analyzing European Atmospheric Circulation Patterns from 1881 to 1989, *Theoretical Applied Climatology*, 1990 (42) 155-167.
- Barker, C.H.; Galt, J.A. Analysis of Methods Used in Spill Response Planning: Trajectory Analysis Planner TAP II. *Spill Science and Technology Bulletin*, 2000 (6(2)) 145-152.
- Barnett, T.P.; Preisendorfer, R. Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmos. Sci.*, 1978 (35) 1771-1787.
- Beegle-Krausse, C.J. GNOME: NOAA's Next-generation spill trajectory model. OCEAN'S 99 MTA/IEEE, 1999 (3) 1262-1266.
- Brands, S.; Taboada, J.J.; Cofiño, A.S.; Sauter, T.; Schneider, C. Statistical downscaling of daily temperatures in the northwestern Iberian Peninsula from general circulation models: validation and future scenarios. *Climate Research*, 2011 (48) 163-176.
- Camus, P. Metodologías para la definición del clima marítimo en aguas profundas y someras: aplicaciones el corto, medio y largo plazo. *Tesis Doctoral*. Universidad de Cantabria, 2009.
- Camus, P.; Méndez, F.J.; Medina, R.; Cofiño, A.S. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, doi:10.1016/j.coastaleng.2011.02.003, 2010.
- Camus, P.; Cofiño, A.S.; Méndez, F.J.; Medina, R. Multivariate wave climate using self-organizing maps. *Journal of Atmospheric and Oceanic Technology*, 2011 (28(11)) 1554-1568.
- Cánovas, V. Análisis estadístico del transporte longitudinal de sedimentos en el corto medio y largo plazo. *Tesis Doctoral*. Universidad de Cantabria, 2012.
- Castanedo, S.; Medina R.; Losada, I.J.; Vidal, C.; Méndez, F.J.; Osorio, A.; Juanes, J.A.; Puente, A. The Prestige oil spill in Cantabria (Bay of Biscay). Part I: Operational forecasting system for quick response, risk assessment and protection of natural resources, *J. Coast. Res.*, 2006 (22 (6)) 1474-1489.
- Cedre, 2003. http://www.cedre.fr/fr/accident/prestige/carto_nov02.htm (último acceso, marzo 2008).
- Comerma, E. Modelado numérico de la deriva y envejecimiento de los hidrocarburos vertidos al mar. Aplicación operacional en la lucha contra las mareas negras. *Tesis Doctoral*. Universitat Politècnica de Catalunya, 2004.
- Condie, S.A.; Waring, J.; Mansbridge, J.V.; Cahill, M.L. Marine connectivity patterns around the Australian continent. *Environmental Modelling and Software*, 2005 (20) 1149-1157.
- Daniel, P. Operational forecasting of oil spill drift at METEO-FRANCE. *Spill Science and Technology Bulletin*, 1996 (3(1)) 53-64.
- Dee, D.P.; Uppala, S. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. R. Meteorol. Soc.*, 2009 (135) 1830-1841.
- Díaz, B.; Pavón, A.; Gómez-Gesteira, M. Use of a probabilistic particle tracking model to simulate the Prestige oil spill. *Journal of Marine Systems*, 2008 (72(1-4)) 159-166.
- González, M.; Uriarte, A.; Pozo, R.; Collins, M. The Prestige crisis: Operational oceanography applied to oil recovery, by the Basque fishing fleet. *Marine Pollution Bulletin*, 2006 (53) 369-374.
- Guanche, Y.; Mínguez, R.; Méndez, F.J. Autoregressive Logistic Regression Applied to Atmospheric Circulation Patterns. *Enviado a Climate Dynamics (en revisión)*, 2012.
- Guillen, G.; Rainey, G.; Morin, M. A simple rapid approach using coupled multivariate statistical methods, GIS and trajectory models to delineate areas of common oil spill risk. *Journal of Marine Systems*, 2004 (45 (3-4)) 221-235.
- Gutierrez, J.M.; Cofiño, A.S.; Cano, R.; Rodríguez, M.A. Clustering methods for statistical downscaling in short-range weather forecasts. *Monthly Weather Review*, 2004 (132) 2170-2183.
- Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. Springer, New York, 2001.
- Hunter, J.R.; Craig, P.D.; Phillips, H.E. On the use of random walk models with spatially variable diffusivity. *J. Comp. Phys.*, 1993 (106) 366-376.
- Izaguirre, C. Estudio de la variabilidad climática de valores extremos de oleaje. *Tesis Doctoral*. Universidad de Cantabria, 2010.
- James, I.D. Modelling pollution dispersion, the ecosystem and water quality in coastal waters: a review. *Environmental Modelling and Software*, 2002 (17) 363-385.
- Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J.; Zhu, Y.; Chelliah, M.; Ebisuzaki, W.; Higgins, W.; Janowiak, J.; Mo, K. C.; Ropelewski, C.; Wang, J.; Leetmaa, A.; Reynolds, R.; Jenne R.; Joseph, D. The NCEP/NCAR 40-years reanalysis project, *Bulletin of the American Meteorological Society*, 1996 (77 (3)) 437-471.
- Katz, R.W. Precipitation as a Chain-Development Process, *Journal of Applied Meteorology*, 1997 (16(7)) 671-677.
- Kennard, R.W.; Stone, L.A. Computer aided design experiments. *Technometrics*, 1969 (11) 137-148.
- 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.; Fiorino, M. The NCEP-NCAR 50-year reanalysis: Monthly means. CD-Rom and documentation. *American Meteorological Society*, 2001 (82) 247-268.
- Kohonen, T. *Self-Organizing maps*, Springer-Verlag, Berlin, 3rd ed, 2000.
- Korotenko, K.A.; Mamedov, R.M.; Mooers, C.N.K. Prediction of the Dispersal of Oil Transport in the Caspian Sea Resulting from a Continuous Release. *Spill Science and Technology Bulletin*, 2000 (6(5)) 323-339.
- Lardner, R.W.; Zodiatis, G.; Loizides, L.; Demetropoulos, A. An operational Oil Spill model in the Levantine Basin (Eastern Mediterranean Sea). *International symposium on Marine Pollution*, Monaco, 1998.
- Lo, J.C.-F.; Yang, Z.-L.; Pielke, R. A. Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model. *J. Geophys. Res.*, 2008 (113), D09112, doi:10.1029/2007JD009216.
- Lorenz, E. N. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, 1969 (26) 636-646.
- MacQueen, J.B. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1967 (1) 281-297.
- Maier-Reimer, E. On tracer methods in computational hydrodynamics. Engineering applications of computational hydraulics. *Pitman Advanced*, 1982.
- Méndez, F.J.; Camus, P.; Cofiño, A.; Medina, R. Analyzing the multi-dimensional wave climate with self organizing maps. OCEANS 2009, IEEE, Bremen, 2009.
- Menéndez, M.; Tomás, A.; Camus, P.; García-Díez, M.; Fita, F.; Fernández, J.; Méndez, F.J.; Losada, I.J. A methodology to evaluate regional-scale offshore wind energy resources. OCEANS '11, IEEE, Santander, 2011.

- Ministerio de Medio Ambiente, Dirección General de Costas. La catástrofe del Prestige. Limpieza y restauración del litoral norte peninsular, 2005 288 páginas.
- Ministerio de Educación y Ciencia. Ciencia and Prestige. La investigación desarrollada a raíz del accidente del Prestige. Ed. Oficina Técnica de Vertidos Marinos (Univesidad de Vigo), Comisión de Coordinación Científica, Dirección General de Investigación Ministerio de Educación y Ciencia, 2007 67 páginas.
- Miranda, F.; Braunschweig, F.; Leitao, P.; Neves, R.; Martins, F.; Santos, A. MOHID 2000, a coastal integrated object oriented model. *Hydraulic engineering software VIII*, WIT press, 2000.
- Montero, P.; Blanco, J.; Cabanas, J.M.; Maneiro, J.; Pazos, Y.; Moroño, A.; Balseiro, C.F.; Carracedo, P.; Gomez, B.; Penabad, E.; Pérez-Muñuzuri, V.; Braunschweig, F.; Fernandes, R.; Leitao, P.C.; Neves, R. Oil spill monitoring and forecasting on the Prestige-Nassau accident. Proc. Environment Canada's 26th Artic and Marine Oil spill (AMOP) *Technical Seminar*, Ottawa, Canada, 2003 1013-1029.
- Moron, V.; Robertson, A.W.; Ward, M.N.; Ndiaye, O. Weather types and rainfall over Senegal. Part I: Observational Analysis, *Journal of Climate*, 2008 (21) 266-287.
- Nittis, K.; Perivoliotis, L.; Korres, G.; Tziavos, C.; Thanos, I. Operational monitoring and forecasting for marine environmental applications in the Aegean Sea. *Environmental Modelling and Software*, 2006 (21(2)) 243-257.
- Pang, W.; Forster, J.J.; Troutt, M.D. Estimation of wind speed distribution using markov chain Monte Carlo techniques. *Journal of Applied Meteorology*, 2001 (40) 1476-1484.
- Ratsimandresy, A.W.; Sotillo, M.G.; Carretero, J.C.; Albiach, J.C.; Álvarez Fanjul, E.; Hajji, H. A 44-year high-resolution ocean and atmospheric hindcast for the Mediterranean Basin developed within the HIPOCAS Project. *Coastal Engineering*, 2008 (55) 827-842.
- Reed, M.; Ekrol, N.; Rye, H.; Turner, L. Oil spill contingency and response (OSCAR) analysis in support of environmental impact assessment offshore Namibia. *Spill Science and Technolgy Bulletin*, 1999 (5) 29-38.
- Rincón, L. Introducción a los procesos estocásticos. Departamento de Matemáticas, Facultad de Ciencias UNAM, México, 2011.
- Sánchez, F.; Gil, J.. Hydrographic mesoscale structures and Poleward Current as a determinant of hake (*Merluccius merluccius*) recruitment in southern Bay of Biscay. *ICES Journal of Marine Science*, 2000 (57) 152-170.
- Semenov, M.A.; Barrow, E.M. Use of a stochastic weather generator in the development of climate change scenarios, *Climatic Change*, 1997 (35) 397-414.
- Shchepetkin, A.F.; McWilliams, J.C. Regional Ocean Model System: a split-explicit ocean model with a freesurface and topography-following vertical coordinate. *Ocean Modelling*, 2005 (9) 347-404.
- Sheridan, S.C. The redevelopment of a weather-type classification scheme for North America, *Int. J. Climatology*, 2001 (22) 51-68.
- Skamarock, W. C.; Klemp, J. B.; Dudhia, J.; Gill, D. O.; Baker, D. M.; Duda, M. G.; Huang, X.-Y.; Wang, W.; Powers, J.G. A description of the advanced research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 2008 125p.
- Skognes, K.; Johansen, O. Statmap - a 3-dimensional model for oil spill risk, Assessment, Environmental Modelling and Software, 2004(19) 727-737.
- Sotillo, M.G.; Alvarez Fanjul, E.; Castanedo, S.; Abascal, A.J.; Menendez, J.; Emelianov, M.; Olivella, R.; Garcia-Ladona, E.; Ruiz-Villareal, M.; Conde, J.; Gómez, M.; Conde, P.; Gutierrez, A.D.; Medina, R.. Towards an operational system for oil-spill forecast over Spanish waters: Initial developments and implementation test. *Marine Pollution Bulletin*, 2008 (56(4)) 686-703.
- Spaulding, M.L.; Howlett, E.; Anderson, E.; Jayko, K. OILMAP: A global approach to spill modeling. 15th Annual Arctic and Marine Oilspill Program, Technical Seminar, Edmonton, Alberta, Canada, 1992.
- Sperandio, M.; Coelho, J. Constructing Markov models for reliability assessment with self-organizing maps. 9th International Conference on Probabilistic Methods Applied to Power Systems, Sweden, 2006.

Conflicto de intereses

Los autores hemos recibido ayuda económica de FUNDACIÓN MAPFRE para la realización de este proyecto. No hemos firmado ningún acuerdo por el que vayamos a recibir beneficios u honorarios por parte de alguna entidad comercial o de FUNDACIÓN MAPFRE.