UNIVERSIDAD DE CANTABRIA

ESCUELA DE DOCTORADO DE LA UNIVERSIDAD DE CANTABRIA
DOCTORADO EN INGENIERÍA DE COSTAS, HIDROBIOLOGÍA Y GESTIÓN DE SISTEMAS ACUÁTICOS (IH2O)

TESIS DOCTORAL

AVANCES EN METODOLOGÍAS ESTADÍSTICAS PARA LA SIMULACIÓN EN EL MEDIO-LARGO PLAZO DE DERRAMES DE HIDROCARBUROS EN EL MAR

PhD THESIS

ADVANCES IN STATISTICAL METHODOLOGIES FOR MID-LONG TERM SIMULATION OF OIL SPILLS IN THE SEA

Presentada por: **HELIOS CHIRI**

Dirigida por: **Dra. ANA JULIA ABASCAL SANTILLANA**

Dra. SONIA CASTANEDO BÁRCENA

Santander, Noviembre de 2019

Alla mia famiglia

"*The journey of a thousand miles begins beneath one's feet*"

Lao-Tzu

Agradecimientos

La realización de una tesis doctoral no es solo el resultado de los esfuerzos del doctorando, sino también de las aportaciones de otras personas que, directa o indirectamente, voluntaria o involuntariamente, colaboran a su conseguimiento.

En mi caso, siento de deber agradecer, en primer lugar, mis directoras de tesis Ana y Sonia. Muchas gracias por vuestra disponibilidad, vuestro apoyo, vuestras ideas, vuestra guía, vuestras correcciones y vuestros consejos. Sonia, gracias también por haber sido la primera persona en creer en mis posibilidades al finalizar la etapa de master y en haberme abierto las puertas del instituto y de tu grupo de investigación. Ana, te merecerías una sección de agradecimientos aparte solo para ti. Te agradezco de corazón haber sido mi guía a lo largo de estos últimos casi seis años. Me has abierto la puerta de tu despacho, todas y cada una de las veces que he necesitado algún apoyo, hasta en los momentos de mayor apuro, siempre con una sonrisa y siempre haciéndome volver a mi puesto con un problema menos. De verdad, no podía haber pedido mejor superior.

Gracias también a mis compañeros de grupo en el instituto. A lo largo de estos años todos habéis demostrado una gran disponibilidad a ayudar y a aconsejarme. Sois todos unos fenómenos y espero que os vaya realmente bien!!

Quiero dar gracias a Paula Camus y Fer por haberme dado, en más de una ocasión, preciosos consejos que me han ayudado mucho en el desarrollo de la tesis.

También quiero agradecer a Robert Weisberg y, en particular, a Yonggang Liu por haberme recibido tan amablemente en su departamento de la University of South Florida por

mi estancia pre-doctoral. Ha sido realmente una experiencia fantástica, desde varios puntos de vistas.

Me gustaría agradecer, además, el instituto IHCantabria por haberme dado la posibilidad y las herramientas para realizar mi doctorado y por haber financiado este camino.

Muchas gracias a todos mis amigos y amigas por ayudar a desconectar y recargar pilas de vez en cuando, en particular a Camilo, Carlos, Oscar, con los cuales he compartidos prácticamente las 24 horas durante varios años.

Un grazie molto speciale alla mia famiglia che, anche se fisicamente un po' distante, sento in realtà tanto vicino. Saper di poter contare ciecamente su quattro persone trasmette un senso di serenità e sicurezza che aiuta molto durante le difficoltà di un percorso lungo e complicato come quello di un dottorato.

Por último, quiero dar las gracias a Nuria por haberme acompañado muy de cerca en este camino y, sobre todo, por ayudarme a ver las cosas desde otro punto de vista.

Contents

CONTENTS I
LIST OF TABLES V
LIST OF FIGURES VI
RESUMEN XI
ABSTRACT 1

CHAPTER I. INTRODUCTION. 5
1.1 Background 5
1.2 State of the art 10
1.2.1 Oil spill numerical modeling 10
1.2.1.1 Far field 10
1.2.1.2 Near field 15
1.2.2 Application of statistical techniques in met-ocean fields 17
1.2.2.1 Statistical modeling 18
1.2.2.2 Data-mining for pattern extraction 19
1.2.3 Stochastic oil spill modeling for oil spill risk assessment 20
1.2.4 Conclusions 22
1.3 Objectives 23
1.4 Layout of the thesis 26

CHAPTER II. STATISTICAL SIMULATION OF CURRENT PATTERNS USING A LOGISTIC REGRESSION MODEL. APPLICATION CASE IN THE GULF OF MEXICO. 27
2.1 Introduction 27
2.2 Study site 28
2.3 Data 31
2.4 Methods 33
2.4.1 Logistic regression general framework 33
2.4.2 Response variable characterization 36
2.4.2.1 Principal component analysis 37

2.4.2.2 Wavelet spectrum analysis 37
2.4.2.3 K-means 40
2.4.3 Model set-up 42
2.5 Results 46
2.6 Conclusions 52

CHAPTER III. MID-LONG TERM PREDICTION OF OIL SPILLS BASED ON LOGISTIC REGRESSION MODELING OF MET-OCEAN FORCINGS. APPLICATION CASE IN THE BAY OF BISCAY. 53
3.1 Introduction 53
3.2 Methodology 55
3.2.1 Statistical wind and current forecast 56
3.2.1.1 Wind and current pattern selection 56
3.2.1.2 Logistic Regression model 58
3.2.2 Probabilistic trajectory forecasts 61
3.3 Study site 64
3.4 Data 66
3.4.1 Currents 66
3.4.2 Wind 67
3.4.3 Drifters 68
3.5 Statistical wind and current forecasts in the Bay of Biscay 69
3.5.1 Wind and current pattern selection 70
3.5.2 Logistic Regression model 73
3.6 Probabilistic trajectory forecasts 78
3.7 Discussion 85
3.8 Conclusions 88

CHAPTER IV. DEEP OIL SPILL HAZARD ASSESSMENT USING STOCHASTIC MODELING BASED ON MET-OCEAN SPATIO-TEMPORAL PATTERNS. APPLICATION TO THE NORTH SEA 89
4.1 Introduction 89
4.2 General methodological framework 91
4.2.1 Reanalysis database selection 93
4.2.2 Met-ocean data setup 94
4.2.3 Data dimensionality reduction (PCA) 95

4.2.4 PCs setup for spatio-temporal clustering 96
4.2.5 Spatio-temporal pattern extraction (k-means) 97
4.2.6 Stochastic 3D oil spill modeling 98
4.2.7 Statistical analysis 98
4.3 Application to a hypothetical deep spill in the North Sea 99
4.3.1 Study area 100
4.3.2 Reanalysis database selection 102
4.3.2.1 Currents 102
4.3.2.2 Wind 103
4.3.3 Met-ocean data setup 104
4.3.4 Data dimensionality reduction (PCA) 104
4.3.5 PCs setup for spatio-temporal clustering 105
4.3.6 Spatio-temporal pattern extraction (k-means) 105
4.3.7 Spill numerical simulation 108
4.3.8 Results 109
4.3.9 Comparison with traditional stochastic modeling technique 111
4.4 Conclusions 114

CHAPTER V. CONCLUSIONS AND FUTURE RESEARCH. 115
5.1 Conclusions 115
5.2 Future research 120

REFERENCES 123

APPENDIX. IMPROVEMENTS AND DEVELOPMENTS OF NUMERICAL TOOLS FOR OIL SPILL SIMULATION. 147
A1. Introduction 147
A2. Advances in the TESEO oil spill model (far field) 148
A2.1 3D oil transport 149
A2.2 Sedimentation 151
A2.3 Natural dispersion 151
A3. Development of a blowout model (near field) 153
A3.1 Jet model 154

A3.2 Oil droplet size distribution module 155
A.3.2.1 Droplet size distribution 155
A.3.2.2 Use of chemical dispersant 157
A4. Near field – far field coupling 158

List of Tables

Table 1. Statistical inference for different nested model (C: constant value, AR#: autoregressive term of order #, WSC GoM: wind stress curl in the GoM, WSC CAR: wind stress curl in the Caribbean Sea; #wl: temporal lag of # weeks). For each examined model, covariates, number of parameters, degrees of freedom (df), deviance increment (ΔDev) chi-square distribution value associated and significance are shown. The final selected model highlighted in bold. ..46
Table 2. Sensitivity (TPR), Precision (PPV) and Accuracy (ACC) for each category and average values. 51
Table 3. Current and wind data setup for the extraction of simultaneous characteristic patterns. 58
Table 4. Summary of buoys deployed during the Prestige accident. 69
Table 5. Evaluation of the performance of different predictors and predictor combinations in the LR model of wind and current patterns. 74
Table 6. Liu and Weisberg (2011) ss score considering the distance between the drifter's trajectories and the simulated areas with probability 0.1% and 2%. 82
Table 7. 3D current and wind data setup for PCA application (M0). 96
Table 8. Matrix M1 obtained as a result of applying PCA to M0. 96
Table 9. Matrix M2 obtained by reshaping M0 to consider the temporal evolution of the PCs. 97

List of Figures

Figura R1. Esquema de la metodología para la simulación integrada de derrames de hidrocarburos en el medio-largo plazo. xx
Figura R2. Patrones espaciales identificados de las corrientes geostróficas en el Golfo de México. Velocidad de la corriente en m/s. xxvi
Figura R3. Evolución de las probabilidades de ocurrencia de los 8 patrones de corrientes en el GoM a escala anual con base en (panel a) la serie histórica y (panel b) los resultados del modelo estadístico. xxviii
Figura R4. Probabilidad media anual de ocurrencia de los 8 patrones de corrientes en el GoM en base a (panel de la izquierda) la serie histórica y (panel de la derecha) los resultados del modelo estadístico. xxviii
Figura R5. Diagrama de flujo del caso de aplicación presentado en este capítulo, en el marco de la metodología general objetivo de la tesis..... xxxii
Figura R6. Patrones de viento y corrientes superficiales seleccionados. Los vectores rojos representan el viento, los vectores negros y los mapas de colores se refieren a las corrientes. xxxv
Figura R7. Mapas de probabilidad obtenidos para las 8 simulaciones efectuadas. En cada panel, se muestra , además, las trayectorias reales de las boyas de deriva (líneas negras)..... xxxvii
Figura R8. Diagrama de flujo del caso de aplicación presentado en este capítulo, en el marco de la metodología general objetivo de la tesis..... xxxix
Figura R9. Ejemplo de un patrón de evolución a los 30 días de viento (vectores rojos), corrientes superficiales (vectores negros y mapas de colores) y corriente media en la columna (indicadas sus componentes en la esquina inferior izquierda de cada panel). El rectángulo naranja indica la localización del derrame profundo. xlii
Figura R10. Mapa de probabilidad de contaminación obtenido con la metodología propuesta, basada en 125 patrones meteo-oceanográficos..... xliv
Figura R11. Mapa de probabilidad de contaminación obtenido con la metodología tradicional, basada en 1000 patrones meteo-oceanográficos aleatorios xliv

Figure 1. Left panel: Painting of a deep sea roughneck working on an offshore rig during the 1950s (Swann, 2007). Right panel: Nowadays robotic well intervention (source: https://www.oceaneering.com/) 7

Figure 2 Upper panel: evolution of the number of tanker spills per year (blue bars) vs evolution of total oil and gas loaded per year (in million metric tons, purple line) (source: ITOPF, 2019). Lower panel: evolution of the number of platform spills per year (yellow bars) vs evolution of oil production per year (in billion barrels, grey line) (source: Bureau of Ocean Energy Management, 2016) 7

Figure 3. Map showing location and date of the 10 major accidental oil spill in history. 10

Figure 4. Graphical scheme of the main weathering processes affecting oil at sea (source: ITOPF, 2011b). 11

Figure 5. Graphical representation of the main processes occurring during a blowout spill (source: Zheng and Yapa, 2002). 16

Figure 6. Flowchart of the general methodology for statistical mid-long term oil spill simulation, which development and definition represents the goal of this thesis. 25

Figure 7. Diagram of the proposed methodology to obtain an autoregressive logistic regression (ALR) model for the Loop Current Patterns (LCPs) simulation 36

Figure 8. Top: First 2 and last 2 EOFs of the geostrophic current in the Gulf of Mexico. The variance accounted for by each mode is shown in each graph. Current units in m/s. Bottom: First 2 and last 2 PCs plots and the associated wavelet power spectra. The PC variance is normalized to be 0.5. Rectified wavelet power spectra are shown as filled contours in base 2 logarithm. The regions of greater than 90% confidence are shown with black contours. Cross-hatched regions on either end indicate the “cone of influence”, where edge effects become important. 39

Figure 9. Selected Loop Current patterns representing the categories of the response variable in the ALR model. Current speed units in m/s. 41

Figure 10. Panel a: Evolution of historical annual frequency of occurrence of the selected LCPs. Panel b: Evolution of annual probability of occurrence of the selected LCPs based on Model V simulation. Panel c: Evolution of annual probability of occurrence of the selected LCPs based on Model VII simulation. Panel d: Evolution of annual probability of occurrence of the selected LCPs based on Model VIII simulation. 48

List of Figures

Figure 11. Left: Annual average probability of occurrence of the 8 extracted Loop Current patterns, based on historical data (left) and on the model simulation (right). 49
Figure 12. Original (left) and normalized (right) confusion matrix from the ALR model simulation. 52
Figure 13. Flowchart of the methodology case study presented in this chapter in the framework of the general methodology pursued in the thesis. 55
Figure 14. Proposed steps for a mid-long term prediction of oil spills. 56
Figure 15. Study area and trajectories of the drifters used in the study. Grey lines represent Isobaths 300 m and 2500 m. 66
Figure 16. K-means extracted daily patterns of wind and surface currents. Red vectors represent wind, while black vectors and background color maps refer to surface currents. Distances among vectors do not represent the actual resolution of the data: for clarity's sake only the data of one of every two wind grid points and of one for every six current grid points are represented. 72
Figure 17. Confusion matrix of the defined Logistic Regression model for wind and current patter simulations. 76
Figure 18. Histograms of pattern prediction based on 100 LR model simulations for the first 9 days of the 5-month forecast period. The true pattern is indicated in brackets in each plot title and in bold in the X-axes. 78
Figure 19. Cumulative probability maps compared to actual drifter trajectories. 83
Figure 20. Instantaneous probability maps compared with drifter 16291's track. The contour plots shown represent the density (as percentages) of the simulated particles at each time interval. 84
Figure 21. Cumulative probability maps obtained using drifter's16291 deploying position (white triangle) as initial location and two different simulation time frames: January 15th to February 8th, 2002 (panel a) and January 15th to February 8th, 2005 (panel b). 88
Figure 22. Flowchart of the methodology case study presented in this chapter in the framework of the general methodology pursued in the thesis. 91
Figure 23. Flow chart of the proposed methodology. 93
Figure 24. Sketch showing that different degrees of horizontal dispersion between the water column and the sea surface. 95
Figure 25. Study area map. Also shown the locations of offshore platforms in the area (orange circles) (OSPAR, 2015). The red circle represents the considered spill location. 101

Figure 26. Evolution of Willmott's index of agreement d with the numbers of patterns considered. 106
Figure 27. Representation of one of the 125 spatio-temporal patterns of 30-day evolution of wind and 3D currents at daily frequency in the North Sea. Each snapshot corresponds to one day. Black arrows with background color maps represent surface currents, red arrows refer to wind. Subsurface current information is shown as a text in the lower left corner in each snapshot. The location of the current grid cell to which subsurface currents information refers (corresponding to the spill location) is indicated as a filled orange rectangle. Distances among vectors do not represent the actual resolution of the data: for clarity's sake only the data of one of every five grid points, for both wind and surface currents, are represented. 107
Figure 28. A snapshot from one on the spill simulations with two levels of zoom. 109
Figure 29. Probability, expressed as a percentage, of oil contamination in 0.25º x 0.25º gridded cells over the 30 day period following the beginning of the release, based on the proposed methodology. 110
Figure 30. Probability, expressed as a percentage, of oil contamination in 0.25º gridded cells over the 30 day period following the beginning of the release, based on 1000 random met-ocean scenarios. 113
Figure 31. Map of the probability difference between the probabilities of contamination obtained with the proposed methodology and the ones obtained with the Monte Carlo approach considering 1000 simulations. Positive values refer to overestimation and negative values refer to underestimation of the proposed methodology results respect to the ones obtained using the Monte Carlo approach. 113
Figure 32. Overall methodological framework proposed in this thesis for the mid-long term statistical simulation, considering both mid-long term predictions and stochastic modeling for oil spill risk assessments. 119
Figure 33. DOR-IFT relationship as obtained by (Venkataraman et al., 2013). 158

Resumen

1. Introducción

A lo largo de la cadena de valor del petróleo existen diferentes actividades que conllevan una cierta probabilidad de que se produzca un derrame de hidrocarburo al mar. En particular, la fase de exploración y producción y la fase de transporte, en ambos casos entendidas en mar, son aquellas que, en este sentido, tienen un mayor riesgo. Durante la fase de exploración y producción, se excavan pozos en el fondo marino en busca de reservas profundas de hidrocarburos. Los accidentes que ocurren en estas situaciones son, generalmente, altamente catastróficos, produciéndose violentas explosiones (o *blowouts*, en inglés) acompañadas por grande flujo de agua, petróleo, gas o una mezcla de estos. Vista su particular situación, extinguir un derrame de este tipo es altamente complejo, por lo que se trata de derrames que suelen durar varias semanas o meses. Durante la fase de transporte, el producto es trasladado desde el lugar de extracción, mediante buques petroleros u oleoductos, a su destino, normalmente una refinería costera. También en este caso, debido a los grandes volúmenes de producto involucrados, un accidente y el consecuente derrame representaría una enorme amenaza. De hecho, los derrames de hidrocarburo al mar, especialmente cuando se trata de vertidos grandes, tienen grandes consecuencias tanto ambientales, como sociales y económicas. Una característica común, observada con los grandes accidentes, tanto de tipo

blowout como debido a accidente de buques petroleros, es que el producto derramado persiste en el medio marino durante largos periodos de tiempo (del orden de meses) antes de ser controlado o, como en la mayoría de los casos, de alcanzar la costa.

Debido al progreso tecnológico que, como en muchos otros ámbitos, ha avanzado muy rápidamente en la industria petrolera, a pesar de un aumento continuo de los volúmenes de hidrocarburo manejado a nivel global, el número de grandes accidentes ha disminuido tendencialmente. No obstante, como hizo recordar el dramático accidente del 2010 de la plataforma *Deepwater Horizon* en el Golfo de México, clasificado como el mayor accidente de la historia, un derrame catastrófico puede ocurrir en cualquier momento. Por este motivo, las administraciones públicas y las grandes empresas de la industria del petróleo necesitan estar preparadas para, en caso de accidente, proporcionar una respuesta rápida y eficaz, enfocada a minimizar las consecuencias del derrame. Para ello, se desarrollan y se ponen a punto planes de contingencia que, en una situación de emergencia, facilitan el proceso de toma de decisiones. Estos planes de contingencia se apoyan fuertemente en dos herramientas: los modelos numéricos de trayectorias y degradación de vertidos de hidrocarburos y el análisis del riesgo derivado de un derrame (u OSRA, de su traducción al inglés *oil spill risk assessment*). Los modelos numéricos de vertidos de hidrocarburos se utilizan para predecir la evolución de un derrame en el mar una vez que se conozca la posición, distribución y propiedades iniciales del mismo y son forzados normalmente por predicciones numéricas de campos de viento y corrientes (también de oleaje si es relevante, sobre todo en zonas costeras). Los OSRAs consisten en el estudio integral de la probabilidad de ocurrencia de un determinado tipo de derrame y de sus esperables consecuencias.

Hoy en día, en caso de aplicación a grandes derrames, se presentan tres problemas que limitan los beneficios de estas dos herramientas:

I. Las predicciones de evolución de derrames proporcionadas por los modelos numéricos no van, comúnmente, más allá de 5-7 días, siendo limitadas en este sentido por el horizonte temporal de las predicciones meteo-oceanográficas que alimentan el modelado. Esta escala temporal está en fuerte desacuerdo con los periodos de deriva observados en muchos grandes accidentes que pueden durar hasta varios meses.

II. En caso de analizar el riesgo asociado a un derrame subsuperficial, incluyendo la tipología *blowout*, no existe una metodología que permite tener en cuenta tanto los procesos 3D que afectan en la columna de agua como los procesos 2D superficiales. Los pocos trabajos disponibles en la literatura de OSRA aplicado a vertidos profundos, simplemente trasladan en la vertical la localización horizontal del derrame hasta la superficie y aplican una metodología tradicional 2D, obviando todas las dinámicas 3D de la columna de agua que pueden alterar el resultado final del análisis.

III. Relacionado con el punto anterior, el desarrollo de una metodología de análisis probabilística de riesgo para derrames en profundidad conlleva un fuerte aumento de recursos computacionales para tener en cuenta, aparte de las dinámicas superficiales, todos aquellos procesos 3D específicos de la evolución subsuperficial del derrame en la columna, además de los forzamientos oceanográficos en la columna de agua.

Esta tesis está dirigida a solventar estas limitaciones, buscando el apoyo de diferentes técnicas estadísticas. En relación a la primera problemática, se propone un modelado estadístico para predecir condiciones meteo-oceanográficas en el medio-largo plazo (meses), para poder de este modo aprovechar las potencialidades del modelado numérico de hidrocarburos en un horizonte temporal más acorde al tiempo real de deriva de los grandes derrames. En cuanto a los puntos II y III, se desarrolla una metodología integral para el análisis de riesgo de contaminación, aplicable tanto a derrames superficiales como subsuperficiales (incluyendo *blowouts*), basada en la selección estadística de aquellos patrones meteo-oceanográficos más significativos de cada zona de estudio.

Estado del conocimiento

Modelado numérico de derrames de hidrocarburos

La generación actual de modelos numéricos para la simulación de hidrocarburos ha alcanzado un elevado nivel de madurez, existiendo hoy en día formulaciones y algoritmos para modelar la mayoría de procesos conocidos que afectan la evolución del hidrocarburo en el mar. Entre los modelos de vertidos de hidrocarburos existentes en el estado del conocimiento podemos mencionar Oil Spill Contingency and Response Model u OSCAR (Reed et al., 2000), Spill Impact Model Application Package/Oil Modeling Application Package o SIMAP/OILMAP (French-McCay et al., 2015; Spaulding et al., 1992), GNOME/ADIOS (Lehr et al., 2002, 2000, 1992; Zelenke et al., 2012), OILTRANS (Berry et al., 2012), MEDSLIK-II (De Dominicis et al., 2013), MOHID (Janeiro et al., 2008; Martins

et al., 2001), MOTHY (Daniel et al., 2003), TESEO (Abascal et al., 2007) y VOILS (Azevedo, 2010; Azevedo et al., 2014). La mayoría de ellos permite simular tanto la evolución superficial como subsuperficial de un derrame. El método más extendido para simular el transporte de un derrame es mediante el movimiento lagrangiano de una serie de partículas independientes equivalentes a la mancha simulada. Entre los principales procesos de degradación simulado por estos modelos están: el esparcimiento o *spreading* mediante los algoritmos de (Mackay et al., 1980) o (Lehr et al., 1984); la evaporación, considerando alternativamente la formulación tradicional (Stiver and Mackay, 1984) o el método de las pseudo-componentes (Fingas, 1996, 1999); la emulsión, comúnmente simulada mediante la formulación de (Mackay and Zagorski, 1982); la dispersión natural por efecto del oleaje rompiente, mediante Delvigne (1993), Delvigne and Hulsen (1994) and Delvigne and Sweeney (1988) o el reciente algoritmo propuesto por Li et al. (2017a, 2017b).

Debido a las condiciones y dinámicas que caracterizan un evento de *blowout*, para la simulación numérica de este tipo de accidentes es necesario el uso de modelos de campo cercano, cuyo resultados vendrán posteriormente acoplados a un modelo 3D de derrames de hidrocarburo para la simulación del campo lejano. Entre los modelos que en la actualidad son los referentes para la simulación de *blowouts*, se encuentran CDOG (Yapa et al., 2001), DEEPBLOW (Johansen, 2000) and OILMAP-DEEP (Spaulding et al., 2000; Zheng et al., 2003; Zheng and Yapa, 2002). Los principales procesos incluidos en todos estos modelos son: las dinámicas de chorro, la expansión de la fase gaseosa de la pluma, el efecto de un ambiente estratificado y la formación de hidratos de gas. Dos aspectos relacionados a este tipo de accidente que han sido estudiados recientemente son la distribución del tamaño de las partículas de petróleo que se generan durante la explosión y el efecto de emplear dispersantes

químicos. La importancia de simular correctamente el tamaño de las partículas de hidrocarburo reside en la consiguiente velocidad de ascenso hacia la superficie por diferencia de densidad con el agua del mar: a mayor tamaño, mayor velocidad de ascenso y menor tiempo transcurrido por la partícula en la columna de agua. Partículas con un tamaño suficientemente pequeño pueden incluso quedar atrapadas en la columna sin alcanzar la superficie. El efecto de usar dispersantes químicos influye justamente en la distribución del tamaño de las partículas, disminuyendo la tensión superficial agua-hidrocarburo y produciendo entonces partículas mucho más pequeñas.

Modelado estocástico para el análisis de riesgo por derrames de hidrocarburos

El riesgo asociado a un derrame de hidrocarburo es una función de las probabilidades de ocurrencia de diferentes escenarios de derrame y de sus potenciales consecuencias en el área de estudio considerada. La variabilidad del derrame se considera considerando diferentes duraciones, posiciones, volúmenes de productos y tipos de hidrocarburos, comúnmente con base en registros históricos. Para determinar las consecuencias de cada tipo de derrame considerado, se simulan los mismos con un modelo numérico de vertidos de hidrocarburo considerando diferentes condiciones meteo-oceanográficas, características de la zona de estudio, como forzamientos ambientales de la evolución del derrame. Este tipo de modelado de derrames, enfocado a considerar las variabilidad ambiental general de un área de estudio, es estocástico, ya que hay que realizar diferentes simulaciones usando en cada una diferente combinación de forzamientos meteo-oceanográficos (escenarios meteo-oceanográficos). Para considerar la variabilidad ambiental real en el dominio de interés, la

metodología normalmente empleada consiste en elegir aleatoriamente la fecha inicial de la simulación numérica y considerar los forzamientos asociados a estas fechas desde bases de datos de reanálisis (método de Monte Carlo). Para considerar correctamente toda la variabilidad meteo-oceanográfica en cada caso de aplicación, se necesitan considerar muchas condiciones ambientales, lo que conlleva un elevado número de simulaciones numéricas del derrame, normalmente cientos o miles. Esta técnica, aunque muy costosa computacionalmente, ha demostrado ser efectiva y se ha aplicado ampliamente a derrames superficiales (2D) (Abascal et al., 2010; Canu et al., 2015; Fingas, 2017; Liubartseva et al., 2015). Para casos de derrames subsuperficiales, las referencias literarias son prácticamente inexistentes. Los pocos trabajos existentes suelen adoptar simplificaciones para reducir el problema de 3D a 2D (Barker, 2011; Ji et al., 2011). A pesar de esto, recientemente las asociaciones IOGP (*International Association of Oil and Gas Producers*) e IPIECA (*International Petroleum Industry Environmental Conservation Association*), referentes a nivel mundial de diferentes temáticas relacionadas con el mundo de la industria del petróleo, han publicado una guía para OSRAs específica para derrames profundos. En esta guía se recalca profundamente la importancia de incluir en el modelado los procesos 3D que afectan al derrame en la columna y la necesidad de considerar no solo la variabilidad ambiental en la superficie, sino también la de toda la columna de agua.

Aplicación de técnicas estadísticas en los campos meteo-oceanográficos

El modelado estadístico en estudios atmosféricos y oceanográficos ha ganado popularidad sobre todo en la última década, evidenciando como ventajas más relevante

respecto al modelado numérico el ahorro de recurso computacional y de tiempos de cálculo (Cid et al., 2017; Diez-Sierra and del Jesus, 2019; Prasad et al., 2009). Entre las técnicas existentes, los modelos de regresión han sido empleados por diferentes autores para predecir el valor de una variable en función de valores anterior de la misma variable y/o en base a otras variables (Casas-Prat et al., 2014; Mishra et al., 2017; Wang et al., 2012). En particular, la regresión logística (o LR, del inglés *logistic regression*) ha demostrado su eficacia cuando se trabaja con campos espaciales. En un modelo de LR aplicado a campos mete-oceanográficos espaciales, la variable independiente (o variable respuesta) es representada por un determinado número de patrones espaciales que identifican las posibles alternativas o valores que puede asumir la variable respuesta del modelado, mientras las variables dependientes (o covariables) son todas aquellas que pueden tener una influencia en la variable respuesta. Entre los trabajos presentes en la literaturas que emplean modelos de LR para simular estadísticamente campos meteo-oceanográficos hay que mencionar Prasad et al. (2009), Guanche et al. (2014), Antolínez et al. (2016), Cardenas et al. (2017).

Otras herramientas estadísticas que, en concordancia con el creciente desarrollo de extensas bases de datos de reanálisis numéricos en los campos atmosférico y oceanográfico, han sido altamente empleadas por la comunidad científica para manejar de forma eficaz la gran cantidad de datos, en muchos casos disponibles, son las técnicas de minería de datos y de extracción de patrones (Camus et al., 2011). K-medias (Hastie et al., 2001), en particular, ha sido empleada en muchos trabajos en este ámbito para extraer patrones espaciales de variables atmosféricas (Diez-Sierra and del Jesus, 2017; Omrani et al., 2016; Rueda et al., 2017; Sönmez and Kömüşcü, 2011) u oceanográficas (Hsieh et al., 2004; Solabarrieta et al., 2015; Solidoro et al., 2007; Yi et al., 2013). Recientemente Bárcena et al. (2015) y Núñez et

al. (2019) han explorado las capacidades de k-medias para extraer patrones temporales de variables meteo-oceanográficas en un punto. Esta técnica no ha sido empleada hasta ahora para extraer patrones temporales en un dominio, es decir, patrones espacio-temporales.

Conclusiones

El estado del conocimiento actual en el modelado de derrames de hidrocarburos evidencia que existen varios modelos numéricos muy avanzados que permiten simular tanto vertidos superficiales como profundos, incluyendo el efecto de los principales fenómenos de degradación que puede experimentar un hidrocarburo derramado en el mar. Hasta la fecha, las predicciones de estos modelos para la respuesta en tiempo real en caso de un derrame, alcanzan un tiempo horizonte máximo de 5-7 días, de acuerdo con el alcance de las previsiones numéricas meteo-oceanográficas. Sin embargo, la deriva de un derrame en el mar puede durar hasta varios meses, por lo que se pueden investigar metodología estadísticas para ampliar la posibilidad de simular vertidos en medio-largo plazo.

Por otro lado, se ha evidenciado como, hoy en día, el modelado estocástico de derrames en el marco de OSRAs, cuyo objetivo es calcular las consecuencias de un derrame considerando la principal variabilidad ambiental de una determinada zona de estudio, ha estado únicamente centrado en simulaciones de derrames superficiales en 2D. No obstante, la crecente tendencia de la industria petrolera en la búsqueda de reservas de hidrocarburos *offshore*, hace que sea necesario desarrollar metodologías que permitan un eficiente modelado estocástico 3D, a pesar del fuerte aumento de recursos computacionales que puede

significar el modelado integral de los procesos en la columna de agua y de la evolución superficial.

Objetivos

El objetivo general de esta tesis es la definición de un marco integral para la simulación de derrames, superficiales o profundos, en el medio-largo plazo (aproximadamente 1-6 meses), considerando tanto la predicción de un derrame específico como el modelado estocástico para análisis de riesgo. El marco general para dar respuesta a este objetivo se presenta gráficamente en la Figura R1.

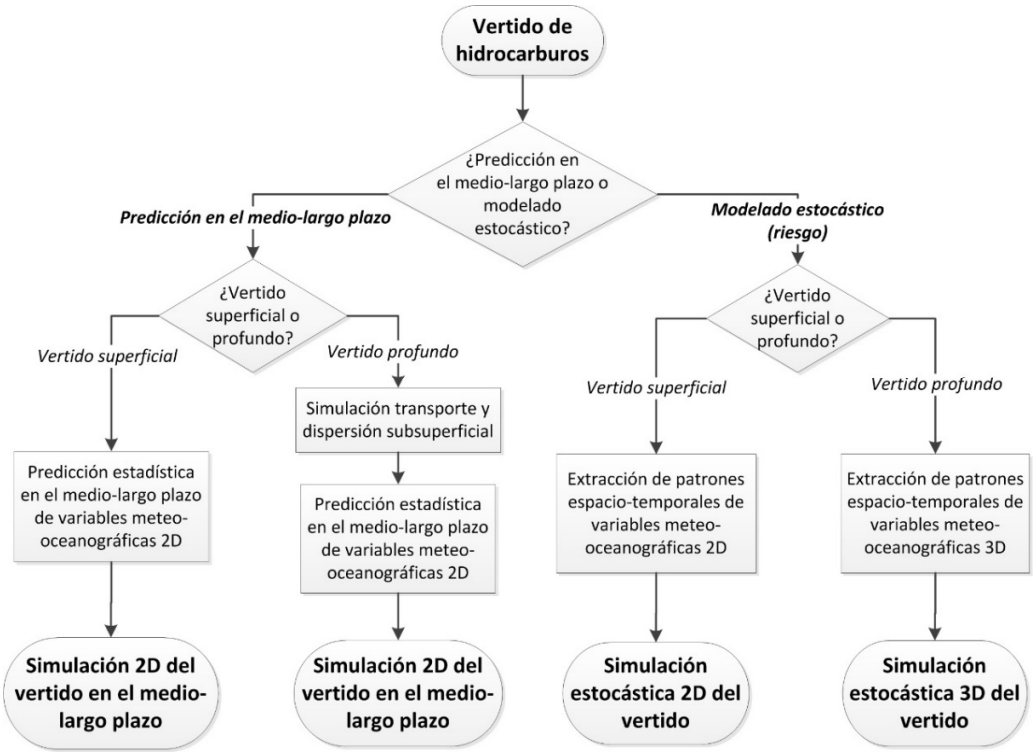

Figura R1. Esquema de la metodología para la simulación integrada de derrames de hidrocarburos en el medio-largo plazo.

Para alcanzar dicho objetivo general, se definen los siguientes objetivos específicos:

1. Investigar la aplicación del modelado estadístico, en particular enfocado en regresión logística, para la simulación estadística de secuencias de patrones meto-oceanográficos.
2. Definir una metodología para la predicción de la evolución de derrames de hidrocarburo en el medio-largo plazo, basada en la predicción estadística de patrones meto-oceanográficos.
3. Definir una metodología para el modelado estocástico de peligrosidad de vertidos de hidrocarburos, aplicable tanto para derrames en superficie como para derrames profundos, basada en la extracción de patrones de evoluciones temporales de campos espaciales meteo-oceanográficos, y que permita reducir el coste computacional requerido por las metodologías tradicionales.

Las metodologías desarrolladas irán a mejorar y completar las herramientas disponibles para la lucha a la contaminación marina por derrames de hidrocarburos, tanto en cuanto a la respuesta en tiempo real como en el ámbito de la planificación basada en análisis de riesgo.

Estructura del documento

En cuanto a la metodología para la predicción de derrames en el medio-largo plazo, es importante tener en cuenta las diferentes tipologías de dinámicas meteo-oceanográfica que

determinan la evolución de un derrame, en función de la zona en la que se manifiesta el accidente. Se han considerado las diferentes situaciones de áreas dominadas por la circulación oceánica general (como el Golfo de México) y áreas en las que el viento local juega un papel importante en la circulación y la deriva de un derrame (como el Golfo de Vizcaya).

Para ello, en el Capítulo 2 se presenta la metodología para la simulación estadística de patrones ambientales a través de un modelo de regresión logística, aplicada al Golfo de México, zona que en 2010 sufrió uno de los mayores derrames de hidrocarburos de siempre, el accidente de la plataforma *Deepwater Horizon*. En el Capítulo 3, la metodología es aplicada para simular patrones de viento y corrientes y trayectorias de derrames en el medio-largo plazo en el Golfo de Vizcaya donde, en 2002, se verificó el derrame del buque petrolero *Prestige*. El objetivo específico 1 se aborda en los Capítulos 2 y 3, mientras el objetivo específico 2 se trata en el Capítulo 3.

En el Capítulo 4, se presenta la metodología para la simulación estocástica de derrames en el medio-largo plazo y es aplicada para evaluar la peligrosidad de un derrames profundo en el Mar del Norte, región que registra una elevada actividad de la industria del petróleo en mar abierto. El objetivo específico 3 es abordado en este capítulo.

Por último, en el Capítulo 5, se proponen las principales conclusiones de la tesis y las posibles futuras líneas de investigación individuadas.

Además, en el Anexo, se presentan las mejoras que se han implementado en herramientas numéricas para la simulación de derrames de hidrocarburos.

2. Simulación estadística de patrones de corrientes mediante un modelo de regresión logística. Caso de aplicación a las corrientes del Golfo de México*

Introducción

En este capítulo se analiza la aplicabilidad de los modelos de regresión logística para simular estadísticamente patrones de corrientes marinas y su evolución. Como caso de aplicación y para una descripción más clara de los pasos metodológicos seguidos, el estudio se centra en las corrientes del Golfo de México (GoM), ya objeto de múltiples estudios tanto por su peculiar dinámica como por su relevancia para las industria petrolera del Golfo de México. En esta región, en 2010, la plataforma *Deepwater Horizon* sufrió un accidente que ocasionó el mayor derrame de hidrocarburos accidental de la historia. La simulación estadística de corrientes representa un elemento clave para predecir en el medio-largo plazo la evolución de derrames offshore condicionados por los procesos de circulación general. Con el presente capítulo se aborda el primer objetivo específico de la tesis.

Datos

Los datos de corrientes empleados derivan de *Archiving, Validation and Interpretation of Satellite Oceanographic Data* (AVISO+), una base de datos de anomalías

* Los resultados de este trabajo han sido publicados en: Chiri, H., Abascal, A.J., Castanedo, S., Antolínez, J.A.A., Liu, Y., Weisberg, R.H., Medina, R., 2019. Statistical simulation of ocean current patterns using autoregressive logistic regression models: A case study in the Gulf of Mexico. Ocean Model. 136, 1–12. https://doi.org/10.1016/j.ocemod.2019.02.010

del nivel del mar obtenida de satélites producida por *Ssalto/Duacs* con el soporte de *Cnes* (http://www.aviso.altimetry.fr/duacs/). Es una base de datos a escala global, en una malla regular con resolución de 1/4º y frecuencia diaria. A partir de esta base de datos, siguiendo el proceso presentado en Liu et al. (2008), es posible obtener datos de corrientes geostróficas. Los datos usados en este estudio corresponden al dominio del GoM, [22.8, 31.8] Norte y [98.8, 81.8] Oeste, sin considerar los nodos en profundidades < 100m, por la calidad de los mismos (Vignudelli et al., 2011).

Los datos de índices climáticos han sido obtenidos del *Royal Netherlands Meteorological Institute* (http://climexp.knmi.nl/).

El rotacional de los campos de viento ha sido calculado a partir de los datos de viento del *North American Regional Reanalysis* (NARR) del *National Center for Environmental Prediction* (NCEP) (Mesinger et al., 2006). Los datos de viento NARR cubren la región de Norte América, desde 1979 hasta hoy, con una resolución de, aproximadamente, 32 km, 45 niveles verticales y frecuencia tri-horaria. En el presente trabajo se han considerados los datos en dos subdominios: en el GoM, [22.8, 31.8] Norte y [98.8, 81.8] Oeste, y la Cuenca de las Caimán en el Mar del Caribe, [17.0, 21.4] Norte y [89.0, 78.3] Oeste.

Los datos de presión al nivel del mar proceden de la base de datos *Climate Forecast System Reanalysis* (CFSR) de NCEP (Saha et al., 2010). Para este trabajo se han utilizado los valores medios mensuales con resolución de 1º en el dominio [0, 65] Norte y [-65, 15] Este. A partir de estos datos se han calculado las anomalías mensuales respecto a los valores promedios para cada mes del año.

Metodología

El primer paso de la metodología consiste en caracterizar la variable respuesta, es decir, en este caso, los diferentes patrones espaciales que resumen la variabilidad principal de las corrientes en el GoM. Para ello, se ha propuesto una metodología de tres pasos que consiste en aplicar análisis de componentes principales (PCA), análisis del espectro Wavelet de las componentes principales (Liu et al., 2016) y de allí la resolución temporal de los patrones a extraer y, por último, k-medias. Como resultado de este primer paso se han definido ocho patrones espaciales de las corrientes geostróficas en el GoM, a escala temporal bi-semanal (Figura R2). Estos ocho patrones representan los posibles valores que puede asumir la variable respuesta como resultado del modelo LR. Una vez obtenidos los patrones, la serie original de corrientes en el GoM es reconstruida a partir de éstos patrones, con base en un criterio de mínimas distancias.

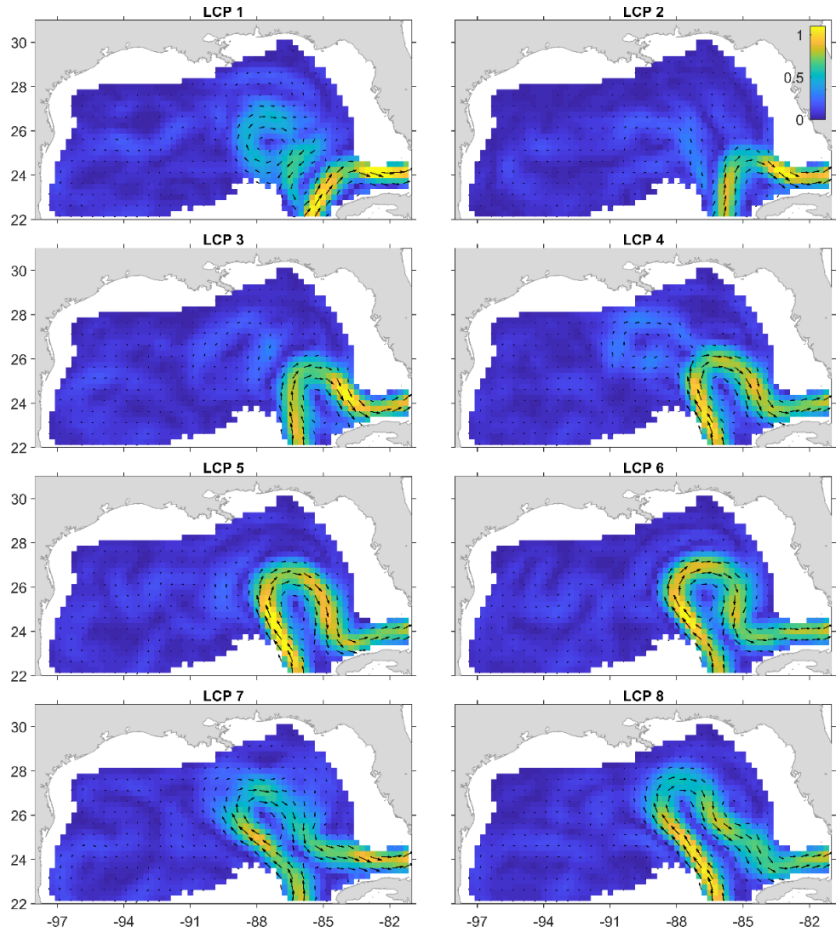

Figura R2. Patrones espaciales identificados de las corrientes geostróficas en el Golfo de México. Velocidad de la corriente en m/s.

El segundo paso de la metodología es definir las covariables del modelo LR. Con base en estudios previos sobre las dinámica de las corrientes en el GoM (Alvera-Azcárate et al., 2009; Lugo-Fernández, 2007; Muller-Karger et al., 2015; Oey et al., 2003; Zeng et al., 2015), los elementos que se han considerado como potenciales predictores del modelo LR han sido: la memoria del sistema representada mediante términos autorregresivos, varios índices climáticos, el rotacional del viento sobre el GoM, el rotacional del viento en la Cuenca de las Caimán y las anomalías de la presión al nivel del mar sobre el Atlántico Norte,

considerando la eventualidad de posibles desfases temporales entre estas variables y su potencial efecto sobre la variabilidad de las corrientes. Para valorar la significancia de estos elementos como predictores del modelo LR, se ha empleado el test estadístico de razón de verosimilitud. Sobre la base de un modelo nulo, este test valora la mejora del ajuste del modelo a la serie de patrones al añadir, una por una, las diferentes covariables consideradas. Como resultado de este análisis, las variables elegidas finalmente como covariables del modelo LR han sido: el rotacional del viento sobre el GoM, el rotacional del viento sobre la Cuenca de las Caimán con 6 semanas de desfase temporal, las anomalías de presión sobre el Atlántico Norte y 3 términos autorregresivos. El modelo LR así definido contiene 141 coeficientes que han sido ajustados considerando la serie original de patrones mediante la librería Python *scikit-learn* (Pedragosa et al., 2011).

Resultados

Una vez definido y calibrado el modelo estadístico, el mismo se ha usado para simular el periodo completo de 24 años. Para tener en cuenta la naturaleza probabilística del modelo, se han ejecutado 100 simulaciones independientes y se ha considerado el promedio de las probabilidades obtenidas. Los resultados obtenidos se comparan con la serie real en Figura R3. El panel superior representa los datos reales históricos y el panel inferior los resultados del modelo. En cada columna las probabilidades de cada patrón están acumuladas por años, de modo que la suma de las probabilidades en cada año sea 1. Como se puede apreciar el modelo ha sido capaz de captar la gran variabilidad interanual de los patrones de corrientes en el GoM. Para poder valorar las capacidades del modelo a escala intra-anual, en Figura R4

se muestra la comparación entre las probabilidades medias anuales, a escala mensual, obtenidas de los datos históricos (a la izquierda) y de los resultados del modelo (a la derecha). Se pueden apreciar nuevamente las capacidades del modelo.

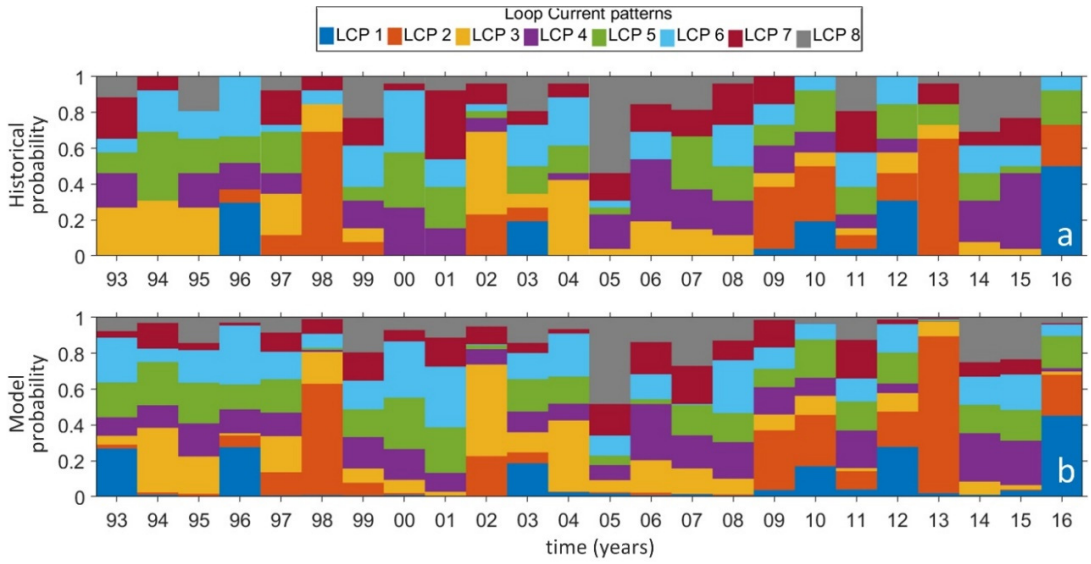

Figura R3. Evolución de las probabilidades de ocurrencia de los 8 patrones de corrientes en el GoM a escala anual con base en (panel a) la serie histórica y (panel b) los resultados del modelo estadístico.

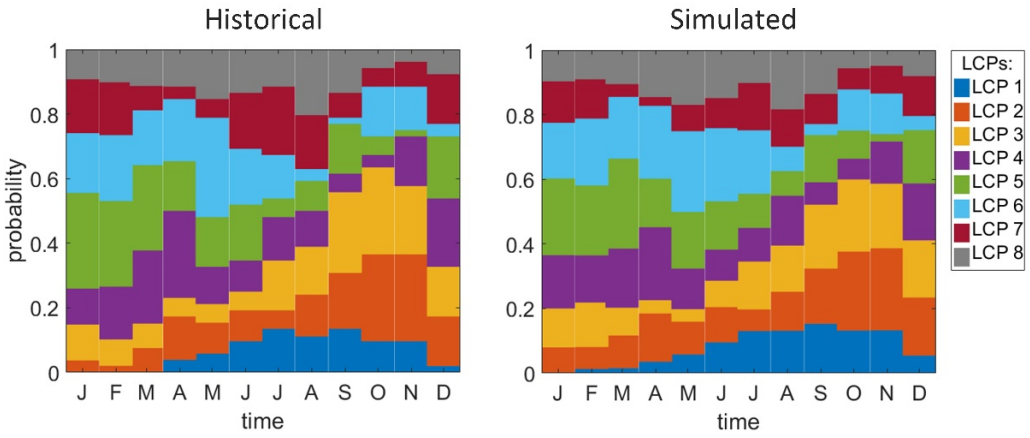

Figura R4. Probabilidad media anual de ocurrencia de los 8 patrones de corrientes en el GoM en base a (panel de la izquierda) la serie histórica y (panel de la derecha) los resultados del modelo estadístico.

Para realizar un análisis de tipo más cuantitativo de los resultados del modelo LR, se han comparado, para cada paso de tiempo, el patrón real con el patrón simulado para el total del periodo de 24 años considerado. Como indicadores de la comparación entre serie de patrones real y simulada se han considerado *recall*, precisión, *F-score* y exactitud, coeficientes estadísticos comúnmente empleados para comparar valores reales con valores recreados. Los valores obtenidos para estos cuatro coeficientes han sido, respectivamente, 0.79, 0.78, 0.78 y 0.94.

Conclusiones

En este capítulo se ha presentado una metodología para simular de forma estadística, mediante un modelo de regresión logística, patrones de corrientes. El método se ha aplicado a en el Golfo de México, donde se han extraído 8 patrones representativos de corrientes geostróficas. Se ha definido un modelo de regresión logística considerando como covariables el rotacional del viento sobre el Golfo de México, el rotacional del viento sobre la Cuenca de las Caimán con 6 semanas de desfase temporal, las anomalías de presión sobre el Atlántico Norte y 3 términos autorregresivos. El modelo ha demostrado su capacidad para simular la evolución de los patrones extraídos, tanto a escala inter-anual como intra-anual. Como indicadores de la comparación entre serie de patrones real y simulada se han considerado *recall*, precisión, *F-score* y exactitud obteniendo, respectivamente, 0.79, 0.78, 0.78 y 0.94. Establecida la capacidad de un modelo de regresión logística de reconstruir adecuadamente

la evolución de patrones de corrientes, se puede desarrollar en una metodología, basada en la predicción estadística de los forzamientos meteo-oceanográfico mediante un modelo de regresión logística, para la predicción en el medio-largo plazo de derrames de hidrocarburos en el medio marino.

3. Predicción de derrames de hidrocarburos en el medio-largo plazo basada en modelado estadístico de los forzamientos meteo-oceanográficos mediante un modelo de regresión logística*

Introducción

Una vez analizada la simulación estadística de corrientes en áreas dominadas por la circulación general, el siguiente paso ha sido avanzar en el modelado estadístico en zonas donde la dinámica del derrame se ve afectada tanto por la corriente como por el efecto local del viento. Para ello, se ha desarrollado una metodología que permite predecir estadísticamente patrones conjuntos de viento y corrientes, y con ello poder predecir la evolución del derrame en el medio-largo plazo.

Para poder valorar la eficacia de la metodología propuesta, ésta ha sido aplicada para el caso de un derrame superficial (Figura R5) en el Golfo de Vizcaya donde, en 2002, como consecuencia de un accidente del buque petrolero *Prestige*, se produjo uno de los más severos derrames del Norte Atlántico. Los resultados obtenidos han sido comparados con trayectorias de boyas de deriva empleadas durante el accidente para seguir las manchas de fuel. En el presente capítulo se tratan el primero y el segundo objetivo específico de la tesis.

* Los resultados de este trabajo han sido publicados en: Chiri, H., Abascal, A.J., Castanedo, S., Medina, R., 2019b. Mid-long term oil spill forecast based on logistic regression modelling of met-ocean forcings. Mar. Pollut. Bull. 146, 962–976. https://doi.org/10.1016/j.marpolbul.2019.07.053.

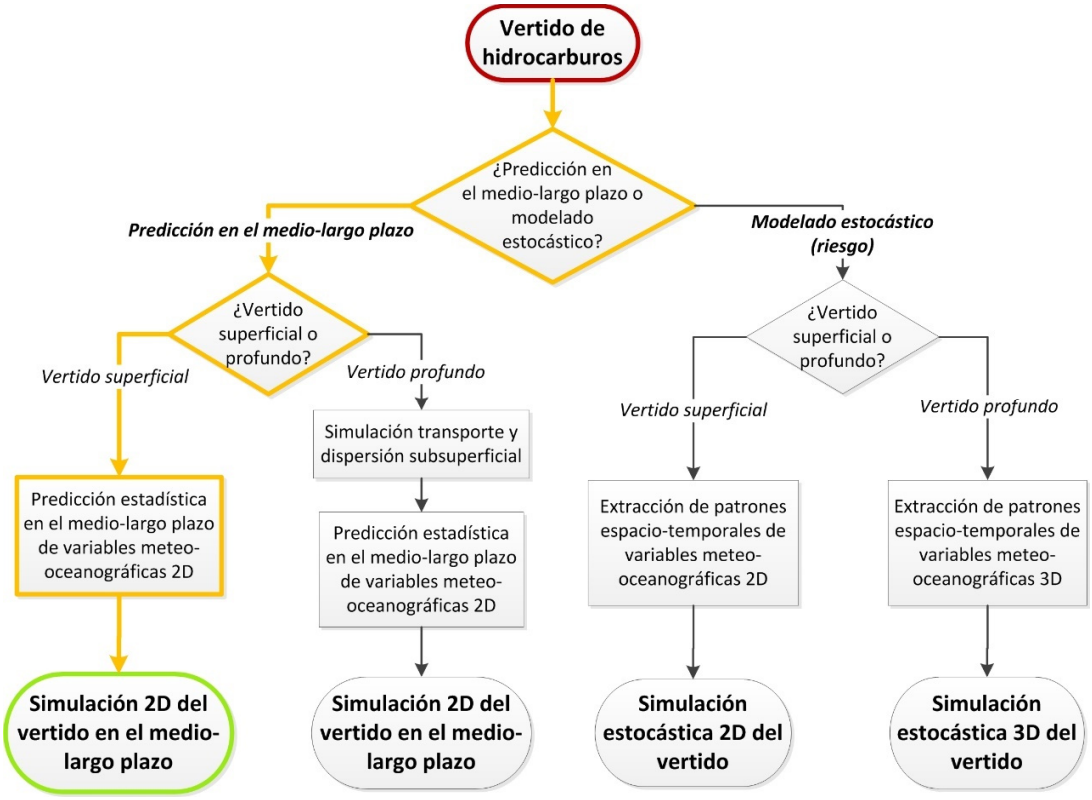

Figura R5. Diagrama de flujo del caso de aplicación presentado en este capítulo, en el marco de la metodología general objetivo de la tesis.

Datos

Los datos de corrientes utilizados provienen de la bases de datos GLORYS12v1 de CMEMS (Lellouche et al., 2018), un reanálisis oceanográfico numérico a escala global, del 1993 al 2016 (24 años) con una resolución espacial de aproximadamente 8 km, 50 niveles verticales y frecuencia diaria. Para este trabajo se han considerado los datos de la capa más superficial (profundidad de 0.5 m) en el dominio [-14º, -1º] Norte y [40º, 48.5º] Este.

Los datos de viento se han obtenido del reanálisis ERA-Interim de ECMWF (Dee et al., 2011), base de datos atmosférico global que cubre dese el 1979 hasta la actualidad,

resolución espacial de 80 km y frecuencia diaria, considerando el mismo dominio que las corrientes.

Se han utilizados datos de trayectorias reales de 12 boyas de derivas para validar los resultados finales del modelo. Estas boyas habían sido lanzadas en diferentes fechas y localizaciones en el Golfo de Vizcaya, entre finales de 2002 y principio del 2003, para trazar diferentes manchas de hidrocarburo derramado durante la emergencia del Prestige. Es importante destacar que el periodo comprendido entre el primer lanzamiento y la última recogida de estas boyas fue de 5 meses.

Metodología

La metodología propuesta se compone de dos apartados principales. El primer apartado tiene como objetivo predecir estadísticamente, mediante un modelo de regresión logística, las condiciones de viento y corrientes en el medio-largo plazo. Dichas condiciones serán expresadas en términos de una serie de patrones espaciales característicos previamente seleccionados a partir de bases de datos de reanálisis ambientales. El segundo apartado consiste en utilizar un modelo numérico de simulación de vertidos de hidrocarburo para predecir la evolución del derrame en el medio-largo plazo, considerando las condiciones ambientales anteriormente predichas.

Como se ha indicado en el capítulo anterior, el primer paso para la aplicación de un modelo es la selección de patrones representativos de las variables en análisis. En este caso se pretende simular condiciones simultáneas de viento y corrientes superficiales. Después del necesario pre-proceso de los datos y de un análisis de sensibilidad para establecer el número

de patrones necesarios, a través de PCA y k-medias se han obtenidos 18 patrones característicos de vientos y corrientes superficiales a escala diaria, representados en la Figura R2. En este caso, para la definición de las covariables del modelo estadístico se ha empleado el método de la validación cruzada, especialmente indicado para minimizar la posibilidad de sobre ajustar el modelo a los datos históricos. El modelo obtenido incluye como covariables los gradientes espaciales de la presión al nivel del mar sobre el Golfo de Vizcaya y la estacionalidad. Posteriormente, se ha utilizado el modelo de regresión logística así definido para predecir las condiciones de viento y corrientes superficiales en el Golfo de Vizcaya durante el periodo de 5 meses correspondiente al registro de trayectorias de las boyas de deriva. Debido a que los resultados del modelo de regresión logística son probabilísticos, se han realizado 100 simulaciones independientes. Posteriormente, siguiendo el segundo apartado de la metodología, se ha empleado el modelo numérico de simulación de derrames de hidrocarburo TESEO (Abascal et al., 2007) con el cual se han realizado 8 simulaciones, considerando las fechas iniciales, la localización inicial y el periodo de deriva correspondientes a los datos de boyas disponibles. A pesar de disponer de datos de trayectorias de 12 boyas, han sido suficientes 8 simulaciones ya que, en algunos casos, varias boyas fueron lanzadas conjuntamente. Disponiendo de 100 evoluciones de viento y corrientes independientes y equiprobables, para cada simulación de TESEO se han realizado 100 ejecuciones independientes considerando, en cada caso, 1000 partículas numéricas. Por último, a partir de las 100 ejecuciones realizadas para cada simulación, se han calculado mapas de probabilidad apoyándose en una malla auxiliar de resolución 0.25º.

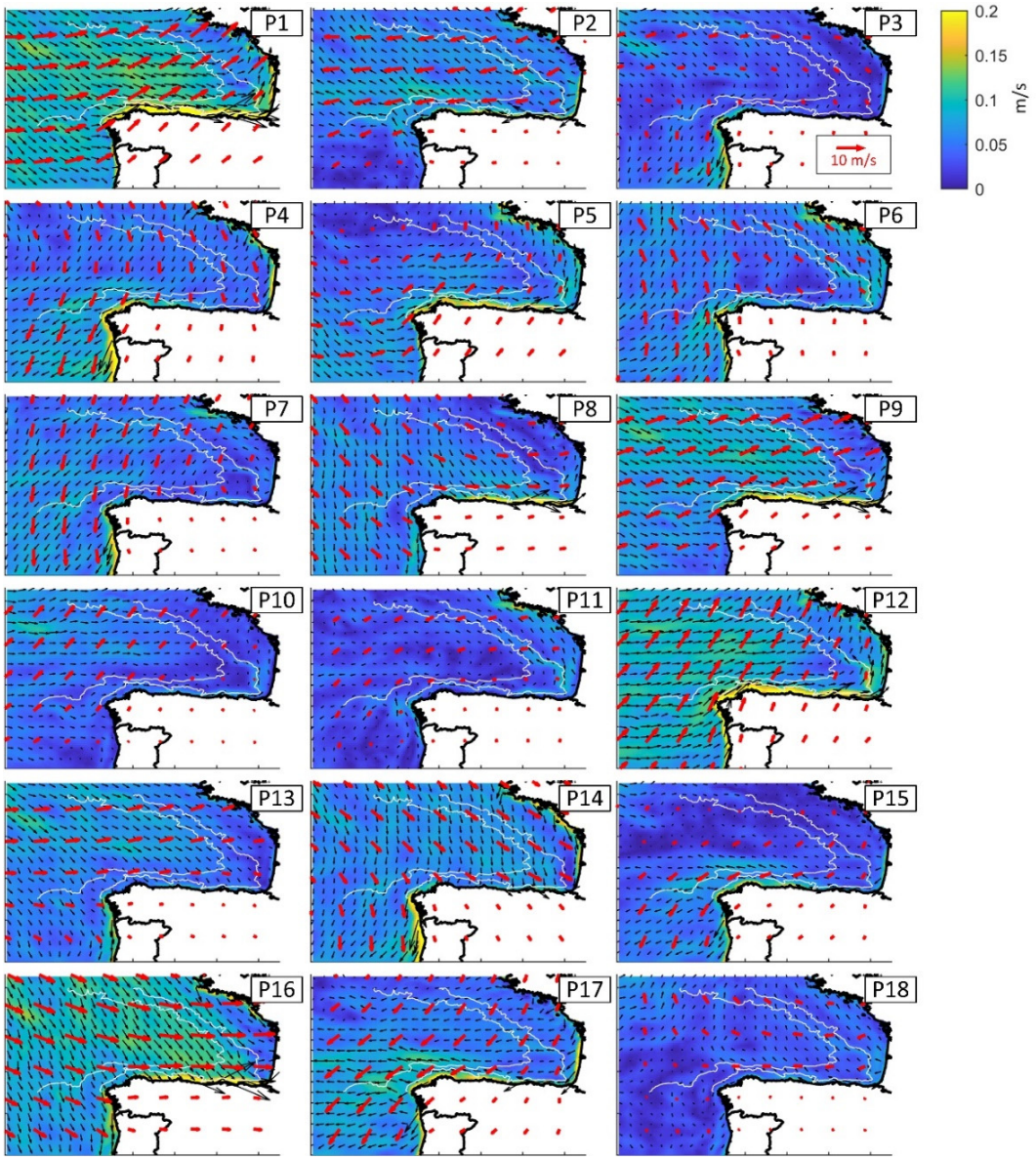

Figura R6. Patrones de viento y corrientes superficiales seleccionados. Los vectores rojos representan el viento, los vectores negros y los mapas de colores se refieren a las corrientes.

Resultados

En la Figura R7 se representan los mapas de probabilidades obtenidos. En cada panel, se representan también las trayectorias reales de las boyas de deriva para valorar los mapas

obtenidos. Como es posible observar, en todos los casos las trayectorias de las boyas están contenidas, en su gran mayoría, en los mapas proporcionados. Además, en la mayoría de los casos, las trayectorias observadas ocupan las zonas más centrales de los mapas obtenidos, donde las probabilidades estimadas son mayores.

Conclusiones

En este capítulo se ha presentado una metodología para la predicción probabilística de derrames en el medio-largo plazo. La metodología, basada en la predicción estadística de patrones medio ambientales en el medio-largo plazo mediante un modelo de regresión logística, ha sido aplicada en el Golfo de Vizcaya. Los resultados obtenidos en el caso de aplicación han sido comparados con trayectorias reales de boyas de deriva lanzadas durante el accidente del *Prestige*, en 2002. Los resultados muestran las capacidades predictivas de la metodología propuesta y el beneficio que la misma podría aportar a la lucha contra la contaminación marina. En este capítulo se da respuesta a los objetivos específicos 1 y 2 de la presente tesis.

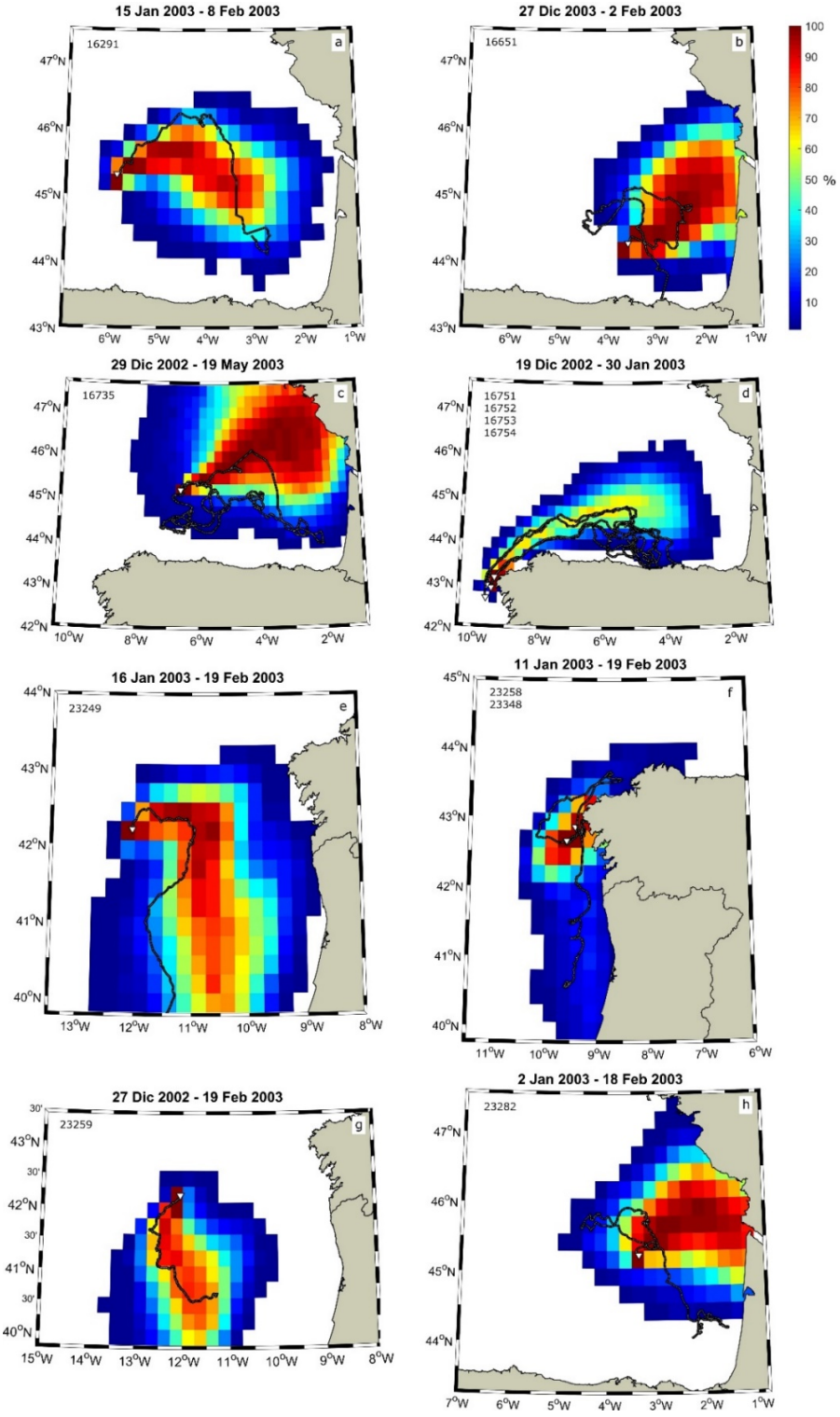

Figura R7. Mapas de probabilidad obtenidos para las 8 simulaciones efectuadas. En cada panel, se muestra, además, las trayectorias reales de las boyas de deriva (líneas negras).

4. Análisis de peligrosidad de derrames de hidrocarburos profundos mediante modelado estocástico basado en patrones espacio-temporales de variables meteo-oceanográficas

Introducción

Los capítulos 2 y 3 se han enfocado al desarrollo de nuevas metodologías que contribuyen a la mejora de los sistemas de respuesta en tiempo real, a través de la predicción de la evolución de un derrame en el medio-largo plazo. En el presente capítulo nos vamos a centrar en el análisis de riesgo de derrames.

El análisis de riesgo es una herramienta fundamental para minimizar las consecuencias de un derrame de hidrocarburo y representa una pieza clave en muchos planes de contingencia. Una parte del análisis de riesgo de contaminación de un derrame es el modelado estocástico del mismo, para evaluar probabilísticamente sus potenciales consecuencia considerando las posibles condiciones medioambientales de la zona de estudio considerada. En este capítulo se presenta una metodología novedosa para el modelado estocástico de un derrame de hidrocarburo basada en la selección estadística de patrones espacio-temporales de variables meteo-oceanográficas. Las aportaciones en el estado del conocimiento de la metodología propuesta son: 1) es aplicable a vertidos profundos, teniendo en cuenta la evolución del derrame en la columna de agua y en superficie; 2) se basa en técnicas de clasificación que permiten una selección óptima de patrones meteo-

oceanográficos y 3) optimiza el coste computacional, algo crítico en este tipo de estudios. La metodología es aplicable a cualquier parte del mundo, basándose en datos ambientales procedentes de reanálisis numéricos.

La metodología propuesta se aplica para analizar la peligrosidad de un hipotético derrame profundo en el Mar del Norte (Figura R8), una de las regiones del mundo con mayor presencia de instalaciones de industria petrolera *offshore*. Con el presente capítulo se busca alcanzar el tercer objetivo específico de la tesis.

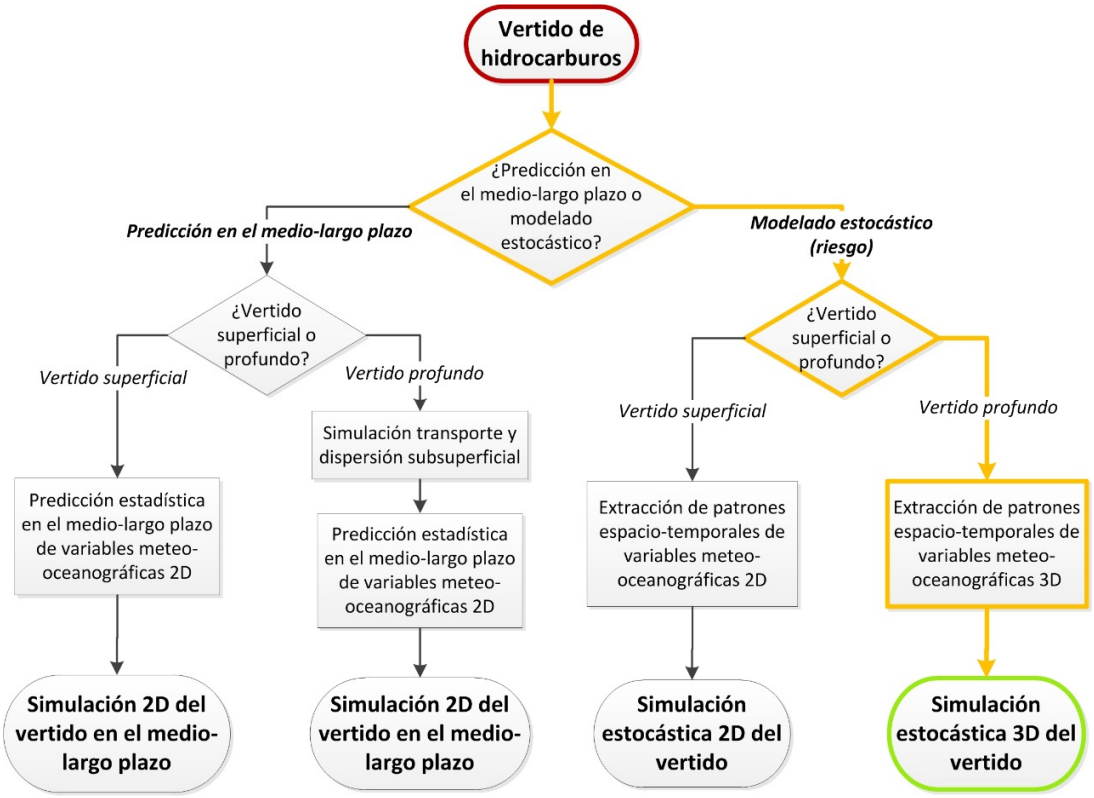

Figura R8. Diagrama de flujo del caso de aplicación presentado en este capítulo, en el marco de la metodología general objetivo de la tesis.

Data

En el caso de estudio de la metodología presentada, se han considerado datos numéricos de viento y corrientes 3D en el Mar del Norte. Los datos de corrientes proceden del reanálisis Global Ocean Forecasting System (GOFS) 3.1 del *Naval Research Laboratory* (NRL) de Estados Unidos. Esta base de datos global cubre 22 años (1994-2015) con resolución espacial de aproximadamente 9 km, frecuencia 3-horaria y 39 niveles verticales. Los datos de viento considerados son los del Climate Forecast System Reanalysis (CFSR) del *National Centers for Environmental Prediction* (NCEP) de Estados Unidos. CFSR cubre 41 años (1979-actualidad). Los datos de viento son proporcionados con resolución de 0.3º hasta 2010 y 0.22º desde 2011, cuando se implementó una nueva versión del modelo atmosférico, con frecuencia horaria. Para asegurar la homogeneidad de los datos empleados, se han considerado solo los datos hasta 2010. Así, el periodo de solape de las dos bases de datos finalmente considerado ha sido desde 1994 hasta 2010 (17 años).

Metodología

En caso de analizar la peligrosidad de un derrame profundo, la metodología propuesta se basa en tres hipótesis:

1. La velocidad vertical de ascenso de las partículas de hidrocarburo es significativamente mayor que la velocidad horizontal de las mismas partículas en la columna de agua.
2. El gradiente vertical de velocidad de la corriente subsuperficial es pequeño, por lo tanto se puede considerar una velocidad promediada en vertical.

3. En superficie, la deriva de Stokes es despreciable frente al efecto del viento y las corrientes, pudiendo excluirse del proceso el efecto del oleaje.

Una vez reducida la dimensionalidad de los datos mediante PCA y organizado los mismos para considerar las evoluciones temporales de los mismos, mediante k-medias se han extraído 125 patrones espacio-temporales de viento, corrientes superficiales y corriente media en la columna de agua, con una duración de un mes y frecuencia diaria. El número de patrones seleccionado ha sido el resultado de un análisis de sensibilidad considerando el índice de Willmott et al. (1985). A modo de ejemplo, en Figura R9 se representa uno de los 125 patrones espacio-temporales obtenidos. Una vez reconstruida la serie original de 17 años de los datos en términos de estos 125 patrones, ha sido posible determinar, para cada uno de ellos, su probabilidad de ocurrencia.

Posteriormente, se ha empleado el modelo numérico de vertidos de hidrocarburo TESEO para simular, por cada una de los 125 patrones meteo-oceanográficos seleccionados, la evolución durante 30 días del accidente considerado. Se ha supuesto un derrame continuo durante 48 horas, desde una posición central en el dominio de estudio a una profundidad de 71 m, y una distribución de tamaño de partículas similar a Chen y Yapa (2003). Por último, con base es una malla auxiliar de 0.25º de resolución, se ha estimado la probabilidad de contaminación superficial en todo el dominio de estudio, teniendo en cuenta los resultados de cada simulación y las probabilidades de ocurrencia de los diferentes patrones ambientales considerados.

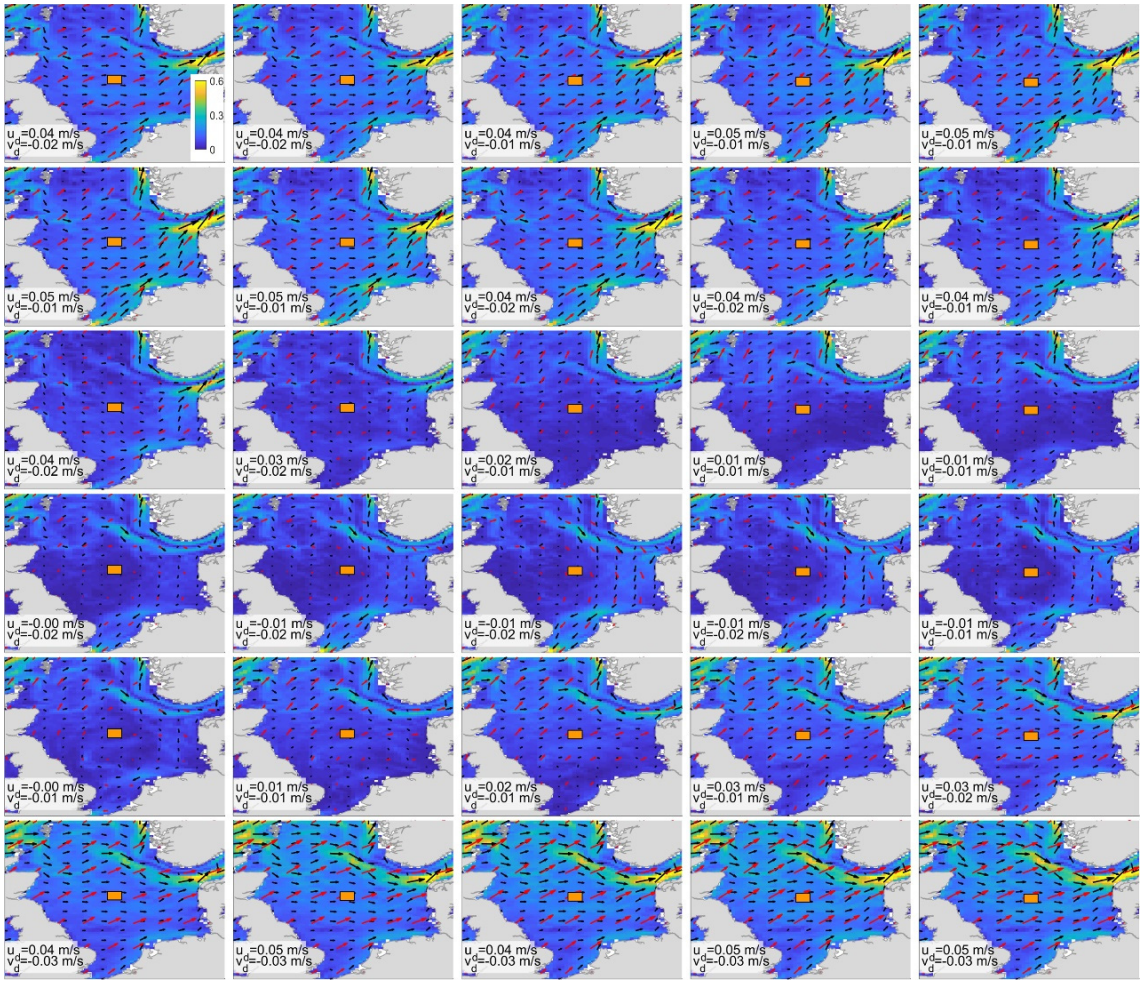

Figura R9. Ejemplo de un patrón de evolución a los 30 días de viento (vectores rojos), corrientes superficiales (vectores negros y mapas de colores) y corriente media en la columna (indicadas sus componentes en la esquina inferior izquierda de cada panel). El rectángulo naranja indica la localización del derrame profundo.

Resultados

El mapa de peligrosidad obtenido se representa en Figura R10. Los colores más fríos identifican zonas con menor probabilidad de ser afectadas por el derrame considerado y los colores más cálidos zonas con mayor peligro. Como se puede observar en el mapa, el área

con mayor probabilidad de contaminación se sitúa en la zona más oriental del dominio, particularmente las zonas costeras de Dinamarca.

Para validar la metodología desarrollada, se han comparado los resultados con la metodología tradicional de Monte Carlo, ampliamente utilizada y contrastada. Para ello, basándose en varios estudios previos (Canu et al., 2015; Elliott, 2004; Liubartseva et al., 2015) se han elegido aleatoriamente 1000 diferentes condiciones ambientales, considerando las mismas bases de datos de viento y corrientes empleadas anteriormente. Los resultados obtenidos en este caso, se representan en Figura R11. Comparando estos resultados con los obtenidos con la metodología propuesta, se puede ver como en ambos casos se pueden sacar conclusiones muy similares en cuanto a la probabilidad de contaminación esperable en la zona de estudio. Esta comparación evidencia las capacidades de la metodología propuesta y las ventajas en términos de coste computacional requerido respecto a una metodología de tipo tradicional.

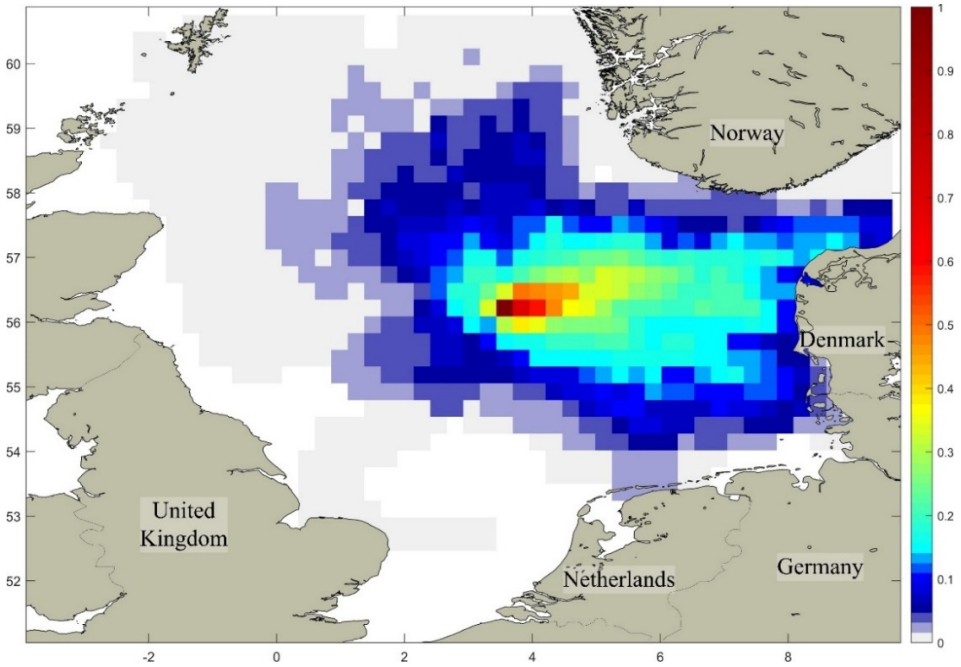

Figura R10. Mapa de probabilidad de contaminación obtenido con la metodología propuesta, basada en 125 patrones meteo-oceanográficos.

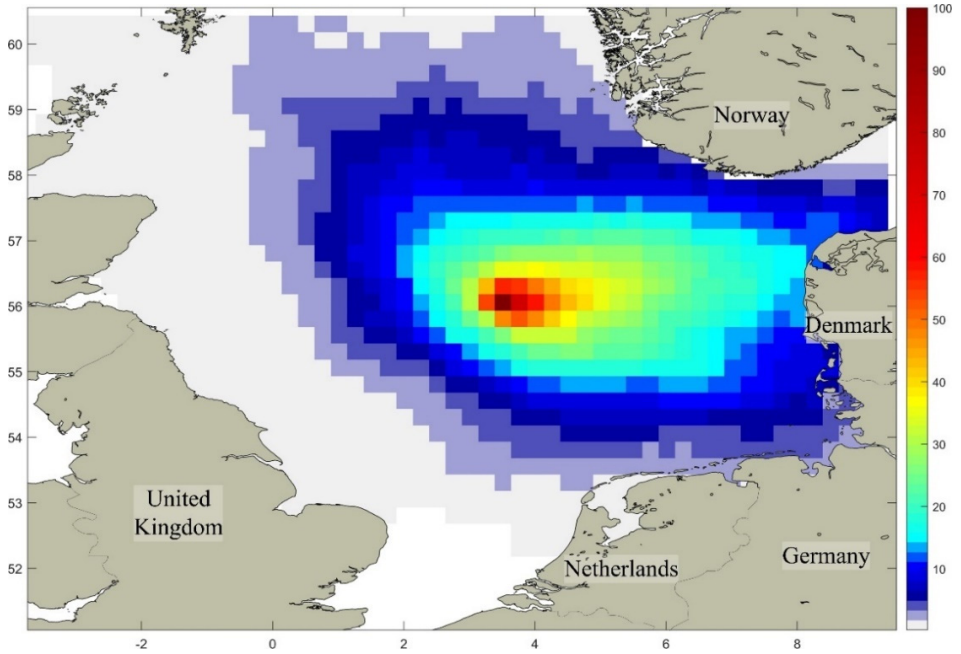

Figura R11. Mapa de probabilidad de contaminación obtenido con la metodología tradicional, basada en 1000 patrones meteo-oceanográficos aleatorios

Conclusiones

En el presente capítulo se ha presentado una metodología para el modelado estocástico de derrames de hidrocarburos, basada en la selecciones de patrones espacio-temporales meteo-oceanográficos característicos. Dicha metodología puede ser empleada en el análisis de riesgo de derrames. Respecto a las metodologías tradicionales, el método presentado 1) puede ser aplicado a derrames profundos, 2) se basa en técnicas de extracción de patrones que permite optimizar la selección de condiciones ambiental consideradas y 3) permite mejorar el coste computacional, elemento crítico especialmente para aplicaciones a derrames en profundidad. Los resultados obtenidos en su aplicación a un derrame hipotético en el Mar del Norte evidencian sus capacidades y el menor coste computacional requerido respecto a las metodologías de modelado estocástico tradicionales. Con este capítulo se cumple el objetivo específico 3 de la presente tesis.

Abstract

Oil spill modeling represents one of the most effective tools to minimize the consequences of oil spills in the sea. Numerical oil spill models are used: a) to predict the evolution of specific spills for real time response activities and b) to probabilistically assess the risk in a preparedness framework. Although this field has, in general, experimented relevant developments during the last decades, there are still some important gaps that need to be worked through. Oil spill predictions for real time response are typically limited to 5-7 days, due to the time coverage of met-ocean forecasts used as environmental forcings of oil spill models. However, major historical accidents have shown that the spilled oil may drift across the ocean during several months. In this sense, new methodologies need to be explored in order to achieve oil spill predictions in the mid-long term. On the other side, stochastic modeling techniques for oil spill risk assessment are generally thought for vessel-related oil spill and are mainly focused on 2D processes and sea surface dynamics variability. Once a spill scenario is defined, several simulations are required in order to estimate its consequences taking into account the overall environmental variability of the considered study area. To do so, a technique usually adopted in several literature studies focused on vessel related spills is to randomly change the initial date of the spill simulation hundreds or thousands times so to statistically include most of the expectable surface environmental conditions of the study area. However, there is still lack of methodologies when considering deep spills, in which case water column processes and forcings add an extra complexity in

the stochastic simulation process that may, eventually, lead to critical level of computational costs.

In the aim fulfilling these gaps, the main objective of this thesis is to develop a general framework for oil spill modeling at the mid-long term, based on the support of different statistical techniques. To reach this main goal, three specific objectives are defined: 1) to investigate the capabilities of logistic regression models in simulating met-ocean patterns, 2) to define a methodology for the mid-long term forecast of oil spills, based on the statistical prediction of met-ocean patterns and 3) to develop a methodology for stochastic modeling in oil spill risk assessment, suitable both for surface and subsurface spills, based on the selection of temporal evolution patterns of met-ocean fields, allowing for a reduction of computational costs and calculation time with respect to traditional stochastic modeling techniques.

Regarding the logistic regression modeling of environmental patterns, in the framework of the mid-long term oil spill prediction methodology, it is important to consider the variety of dynamics that may affect the evolution of a spill in the sea, depending on the considered area, i.e. areas mainly dominated by general circulation, such as the Gulf of Mexico, and areas in which local winds also play an important role in spill evolution, such as the Bay of Biscay. A first logistic regression model is built to simulate sequences of surface current patterns in the Gulf of Mexico, region that suffered, in 2010, the *Deepwater Horizon* platform spill. Different covariates are analyzed using a statistical test. The obtained results show that the statistical model is capable of reproducing both the inter-annual and the intra-annual variability of surface current patterns in the Gulf of Mexico. Another logistic regression model is used to statistically simulate patterns of wind and currents in the Bay of Biscay where, in 2002, the *Prestige* accident occurred. Then, TESEO oil spill model is used

to produce mid-long term spill simulation considering the met-ocean predictions obtained with the statistical model. The obtain results are compared with drifter trajectories data collected during the *Prestige* oil spill accident, highlighting the potential of the proposed methodology.

After analyzing the capabilities of logistic regression models to simulate met-ocean patterns and defining a method for mid-long term prediction of oil spills, a statistical methodology is proposed to provide stochastic modeling of deep spills, based on the extraction of spatio-temporal patterns from met-ocean reanalysis databases. The methodology is applied in the North Sea, one of the world region with higher presence of offshore oil and gas activities, to assess the hazard of a hypothetical deep spill. 125 patterns of 1-month evolutions of wind, surface and water column currents are extracted as representatives of the local met-ocean variability. TESEO is then used to run 3D simulations under each of the 125 scenarios identified. Considering the probability of occurrence of each of the considered scenarios, a probability oil contamination map is then generated. Similar probabilistic pollution results are obtained by randomly selecting 1000 scenarios, following the traditional oil spill hazard assessment approach, highlighting the benefits of the proposed methodology of reducing the computational costs and the time required to assess the likely consequences of an oil spill.

The obtained results and the developed methodologies allow to establish a general framework for the simulation of oil spills in the mid-long term, both for prediction of future evolutions and for hazard assessment, based on proved statistical techniques. This framework fulfills the main objective of this thesis and serves as a further step in the fight against marine pollution.

CHAPTER I.

INTRODUCTION.

1.1 Background

Across the oil value chain, different activities exist that involve a possibility of an oil spill in the sea. The most dangerous operations in this sense occur within the stage of offshore exploration and production and during the transportation phase. Offshore exploration and production includes drilling exploratory wells in search of deep reservoirs and developing the required infrastructure around those wells showing a commercial interest. Any unexpected complication during these operations may eventually lead to a blowout, the most destructive and feared oil spill type. A blowout is an uncontrolled flow of water, oil and gas (or a mixture of these) from a wellbore that cannot be controlled with previously installed barriers. Difficulties related to capping an offshore blowout make that this kind of spills may persist during months before the introduction of oil in the sea is controlled. Following, in the

oil value chain, is the transportation phase, in which the extracted crude may be transported through tankers or underwater pipelines to its destination, in general coastal refineries. During transportation, oil releases to the sea may occur as a result of pipeline cracks (Sun et al., 2015) or tanker accidents (Castanedo et al., 2006; Goerlandt and Kujala, 2011).

Oil spills, in particular large spills, produce critical environmental, sociological and economic impacts. Environmental consequences have been deeply studied and has been shown that beside the direct effect on individual organisms, the worst damage is at the long term on an ecosystem scale (Peterson, 2003). An oil spill may also have major social impacts, like uncertainty about short-term and long-term effects of the spill on ecosystems and human communities, loss of trust for parties responsible for protecting from the threat, public distrust of oil corporations and increased levels of collective stress (Picou et al., 2009). Oil spill affections from the economic perspective span from implication on the fishing industry (Surís-Regueiro et al., 2007), damages to the tourism industry, loss in coastal real estate values and cleanup costs (Smith et al., 2010).

However, during the last decades a great effort has been put by the oil and gas industry in technology development focused on oil spill prevention and mitigation (Figure 1). As a result of this effort, in spite of the steadily increasing worldwide oil activities, the frequency of oil spills is decreasing, considering both the offshore exploration/production activities (blowouts) and the sea surface accident from oil tankers (Bureau of Ocean Energy Management, 2016) (Figure 2).

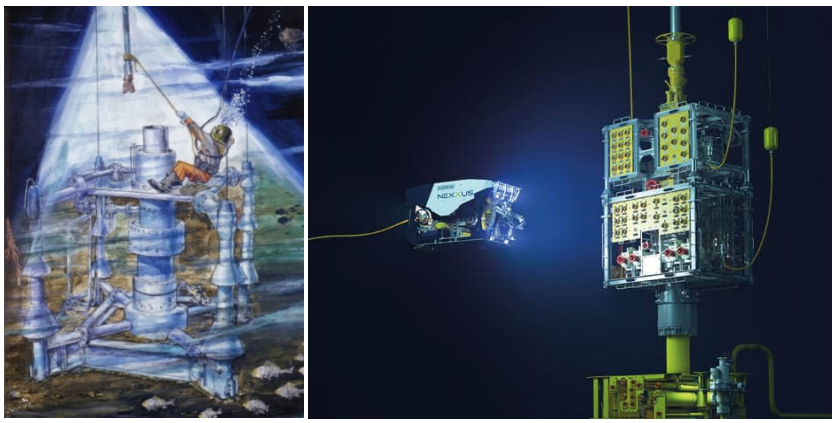

Figure 1. Left panel: Painting of a deep sea roughneck working on an offshore rig during the 1950s (Swann, 2007). Right panel: Nowadays robotic well intervention (source: [https://www.oceaneering.com/)](https://www.oceaneering.com/)

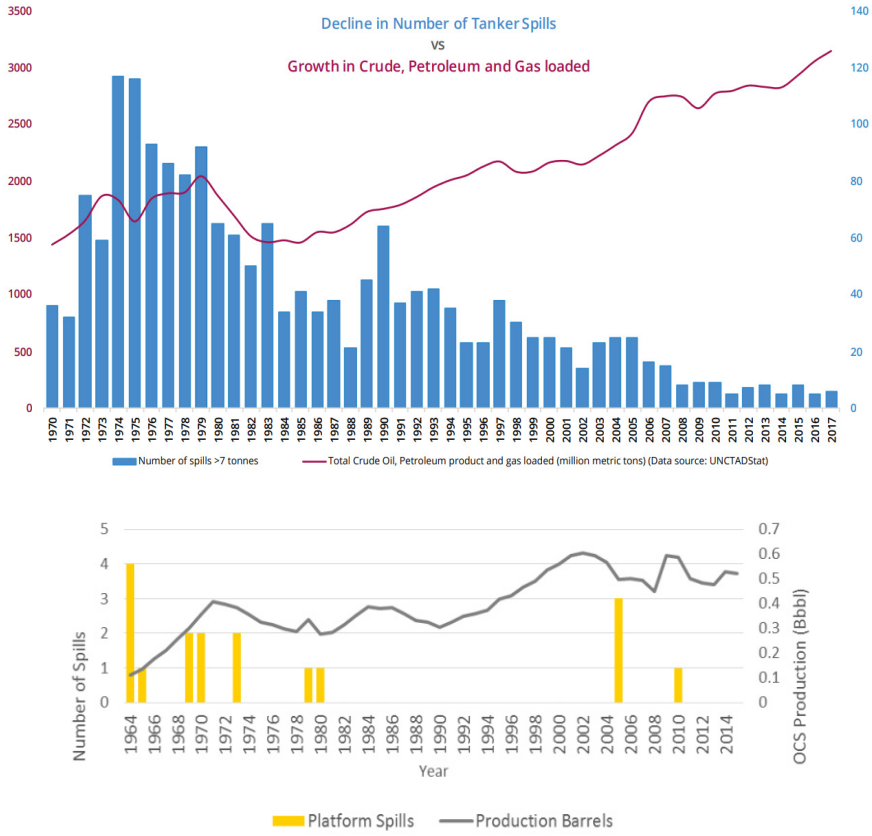

Figure 2 Upper panel: evolution of the number of tanker spills per year (blue bars) vs evolution of total oil and gas loaded per year (in million metric tons, purple line) (source: ITOPF, 2019). Lower panel: evolution of the number of platform spills per year (yellow bars) vs evolution of oil production per year (in billion barrels, grey line) (source: Bureau of Ocean Energy Management, 2016)

Offshore exploration/production spills and oil tanker spills are less frequent than small ship, pipeline and storage/refinery releases, but they are also more dangerous because of the much larger reservoirs from which the oil spills (Eckle et al., 2012). As such, this kind of oil spills requires a deeper consideration and analysis. An interesting feature observed in the aftermath of major tanker spills and blowout accidents is that the spilled substance may drift across the ocean during very large periods of time (in the order of months) before being totally controlled or, as, unfortunately, in most cases, reaching the coast (González et al., 2006; Kim et al., 2015; Liu et al., 2011). A map with the locations of the ten major accidental oil spills in history is presented in Figure 3, with the year of each accident. As shown, most important historical spills were originated from blowouts or tanker accidents. It is also worth noting that the biggest historical spill, the Deepwater Horizon blowout in the Gulf of Mexico, is relatively very recent (2010), respect to the oil and gas industry time scale. The occurrence of such a tragic spill in such recent times, needs to remember that, despite the positive and encouraging spill frequency statistics, the risk of oil pollution is always present and the worse spill in history could be the next one.

This is why govern administrations and oil and gas companies have always to be ready to respond quickly and efficiently in case of an oil spill. This preparedness is generally reached by means of the development and definition of contingency plans. A well-developed contingency plan allows to make crucial decisions and optimize the mobilization of the available resources without hesitation and in short time during the stressful situation of an unfolding accident (ITOPF, 2011a). Two important components of these plans that have already thoroughly demonstrated their efficacy, are represented by numerical models of oil

spill fate and trajectory and by methodologies for the probabilistic analysis of oil spill risk (e.g., Abascal et al. (2010), Berry et al. (2012), Coppini et al. (2011), French and Rines (1997) and Zhao et al. (2014)). Numerical oil spill models are mathematical models that allow to predict the trajectory of the spilled slicks and (if the used model permits it) the evolution of the oil weathering. Probabilistic analyses of oil spill risk, often referred to as oil spill risk assessments (OSRAs), consist of the comprehensive study of the likelihood of an oil spill to occur and its probable consequences. As will be justified in the next section, nowadays, in case of apply these two tools to major oil spills, the following key limitations occur:

1) When an oil spill model is used to predict the future evolution of a major oil spill, the temporal horizon of weather and ocean forecasts used as the model environmental forcing inputs makes impossible the proper simulation of entire spill evolution, often during several months.
2) Existing methodologies for stochastic oil spill modeling are mainly centered on surface spills and 2D processes. Considering deep spills involves a significant increment of complexity, due to all those processes affecting the subsurface spill evolution and the higher quantity of environmental forcings required, eventually leading to critical computational costs.

With this work, we try to make some advances to solve these problems by means of different statistical methods that, regarding the first point, allow to predict weather and ocean conditions at the mid-long term (order of months) so to permit oil spill simulations at this temporal horizon, as requested in case of major spill emergencies, and, regarding the second point, allow to consider a reduced number of 3D met-ocean scenarios by specifically extract those most representative from the used environmental reanalysis databases.

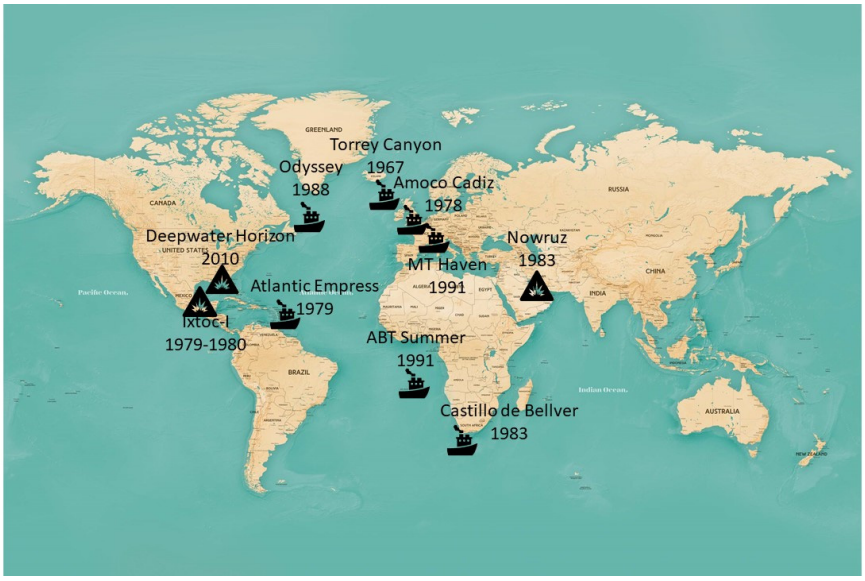

Figure 3. Map showing location and date of the 10 major accidental oil spill in history.

1.2 State of the art

This section of the document presents a review of the two main components of the thesis: oil spill modeling and the stochastic spill simulation for OSRA. A state of the art specific of those statistical techniques employed to develop the thesis objectives, are presented in the corresponding sections in the document.

1.2.1 Oil spill numerical modeling

1.2.1.1 Far field

The review of oil fate and transport modeling presented next covers, in a general way, the historical evolution of the employed techniques and algorithms. Particular attention,

however, is given on the current generation of oil spill models and the latest investigations in the field.

Once spilled to the sea, oil suffers a series of physical, chemical and biological processes that influence its evolution in the water (Figure 4). In the last decades, several researchers have dedicated their efforts to study these processes in the aim of contributing to the field of the oil spill modeling. As a result of these investigations, several oil spill models have been developed with varying degrees of sophistication and different approaches for the considered processes.

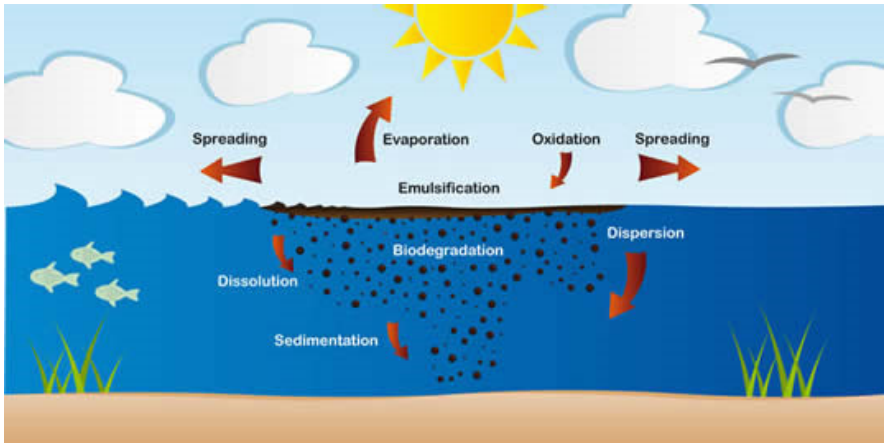

Figure 4. Graphical scheme of the main weathering processes affecting oil at sea (source: ITOPF, 2011b).

Three generations of oil spill models can be identified in the literature, in accordance with historical periods and the advances achieved in the state of the art. The first generation date back to the 1970s and most models were limited to simulate the 2D movement of oil slicks at the sea surface due to the effect of wind, surface currents and waves (Zafirakou, 2019). Since from the beginning, the most widely used method to simulate the oil transport

has been the lagrangian method (Lynch et al., 2015), according to which the slick is considered as a cloud of independent particles that moved for the combined effect of wind, current and waves (advective transport) and a turbulent dispersion component. Very few from the first models also considered the oil spreading but, in general, only through the balance of force, although the importance of chemical processes was already under investigation (Fay, 1969). A review of the first generation of oil spill models can be found in Stolzenbach et al. (1977). Models from the second generation showed a significant progress in oil weathering modeling, particularly in the areas of evaporation of the oil and emulsification. The importance of spreading was more widely recognize, also due to the relevance of the slick area for other processes like evaporation (Huang, 2005). The three-regime spreading theory by Fay (1971), (1969) was the most widely used algorithm to simulate the spreading phenomenon. Approaches of different degree of complexity were considered for evaporation. Some models considered evaporation as percentage of oil mass loss from tabulated values (Blaikley et al., 1977). Other models used more advanced approaches based on first evaporation formulations (e.g., Aravamudan et al. (1982), Fay (1971)). Emulsification was considered by means of simple formulations derived from experimental studies (Mackay et al., 1980) or with more complex mathematical models as a function of asphaltene and resin contents (Kolpack et al., 1977). Other weathering processes like dissolution and natural dispersion were taken in consideration in very few models. However simple and approximated method were used, like Payne et al. (1983) that modeled dissolution in the same way as evaporation with a component approach, or Spaulding et al. (1983) that combined dissolution and dispersion in the same oil mass loss estimation. Dispersion was

considered through tabulated values as a function of sea state and time after spill (Spaulding, 1988).

The most recent generation of oil spill models deeply improve the precursor models both in the advective transport and in the fate simulation algorithm. During the 1990s and 2000s, following the ongoing growth of offshore deep exploration and production, the focus of many oil spill modelers started to move towards the modeling of subsurface oil evolution. At the same time, more mature investigations in many oil weathering processes allowed the implementation of more robust and accurate oil fate algorithms. Some of the state-of-the art oil spill models of the current generation are Oil Spill Contingency and Response Model or OSCAR (Reed et al., 2000), Spill Impact Model Application Package /Oil Modeling Application Package or SIMAP/OILMAP (French-McCay et al., 2015; Spaulding et al., 1992), GNOME/ADIOS (Lehr et al., 2002, 2000, 1992; Zelenke et al., 2012), OILTRANS (Berry et al., 2012), MEDSLIK-II (De Dominicis et al., 2013), MOHID (Janeiro et al., 2008; Martins et al., 2001), MOTHY (Daniel et al., 2003), TESEO (Abascal et al., 2007) and VOILS (Azevedo, 2010; Azevedo et al., 2014). Most of these are 3D models allowing to simulate an oil spill through the whole water column and can be applied to both surface and subsurface releases.

Spreading is generally modelled using the empirical based algorithm by Mackay et al. (1980), which represents a variant of the gravity-viscous equation by Fay (1971), (1969). Lehr et al. (1984) proposed a revised model to account for the observed non-symmetrical spreading of oil slicks. In this model the shape of the slick during spreading is considered an elongated ellipse in the direction of the wind, rather than circular as predicted by Fay formulation. This approach is considered in ADIOS model (NOAA, 1994).

Evaporation is a key factor for oil mass loss at the surface, especially for the lighter oils, occurring soon after the spill. One of the most common methods used to estimate this loss is the analytic algorithm proposed by Stiver and Mackay (1984). A different approach, the so called 'pseudo-components method' for the evaporative loss was proposed more recently by Fingas (1996), (1999). Jones (1997) found that this method was more realistic but, due to the computational costs and the high data requirements, it is less popular than the Stiver and Mackay's method. The pseudo-component approach is implemented in SINTEF weathering model (Daling et al., 1997), while the Stiver and Mackay's algorithm is used in NOAA's ADIOS weathering model.

The commonly implemented scheme for emulsification is the one provided by Mackay and Zagorski (1982). This is based on a mousse viscosity coefficient and an emulsification coefficients which determine the rate at which emulsification is predicted to occur. The maximum fraction of water incorporated in the spilled oil is either provided by the user or specified in the model oil database.

Natural dispersion, sometimes referred to as entrainment, is the physical phenomenon in which a surface oil slick is broken, for effect of breaking waves, in droplets of different sizes that, depending on these sizes may resurface in a very short time or persist under the sea surface. An algorithm widely used in oil spill models to simulate the entrainment rate of oil from the surface was presented by Delvigne (1993), Delvigne and Hulsen (1994) and Delvigne and Sweeney (1988). Based on these studies, the dispersed mass due to breaking waves is a function of the wave energy dissipation rate, the fraction of sea surface covered by oil and an empirical coefficient that depends on oil type and weathering status. The size distribution of the generated droplets is defined by an empirical power law fit to laboratory

experimental observations. This approach is used in ADIOS model (NOAA, 1994), the SINTEF oil weathering (Aamo et al., 1993; Daling et al., 1997), OSCAR (Aamo et al., 1997; Reed et al., 1995; Reed and Rye, 1995) and OILMAP (Spaulding et al., 1992). Recently, (Li et al., 2017a, 2017b) have proposed a new approach to the natural dispersion problem. The authors propose an empirical model based on three dimensionless groups; the dimensionless oil flux, the Weber number (*We*) and the Ohnesorge number (*Oh*). The Weber number is the ratio of disruptive momentum (hydrodynamic) forces to restorative interfacial tension forces while the Ohnesorge number is the ratio of viscous to interfacial tension forces (Spaulding, 2017). The dimensionless flux rate at the surface is expressed in terms of *We* and *Oh*. This method has not been implemented yet in any oil spill model.

1.2.1.2 Near field

Following the growth of offshore exploration and extraction of crude oil from deep reservoirs of the 1970s and 1980s, the importance of understanding the nearfield dynamics of a blowout for an efficient oil spill response quickly emerged. A schematic scenario of this kind of spill is shown in Figure 5. Among the first works focused on understanding the processes occurring in an underwater blowout. Mcdougall (1978) proposed a first scheme of the structure of a blowout plume as a combination of gas bubbles and oil droplets. Fannelop and Sjoen (1980) proposed a first numerical solution for the hydrodynamic of a blowout plume, considering underwater spreading and the expansion of the gas portion of the plume. Rye (1994) proposed the eulerian model BLOW based on Fannelop and Sjoen (1980) and including oil dissolution and the effect of a stratified ambient. Yapa and Li (1997) developed a first version of the Comprehensive Deepwater Oil and Gas model (CDOG), a lagrangian

blowout model that, after several improvements, represents one of the reference model up to date.

Following the progressively movement of the offshore oil and gas industry towards deeper waters, new processes emerged that were to be taken in consideration in the blowout models, such as the formation of gas hydrates due to high pressures and low temperatures and the gas dissolution, no depreciable in shallow waters (Fogg and Gerrar, 1990; Sloan and Dekker, 1998). These deep water processes were implemented in CDOG (Yapa et al., 2001) and in BLOW (Johansen, 2000), with the latter that was adapted to lagrangian approach in order to more easily include the new dynamics while also changed its name to DEEPBLOW. In the same period, a new module was developed for OILMAP model to simulate underwater blowout, called OILMAP-DEEP (Spaulding et al., 2000; Zheng et al., 2003; Zheng and Yapa, 2002) including all the mentioned processes.

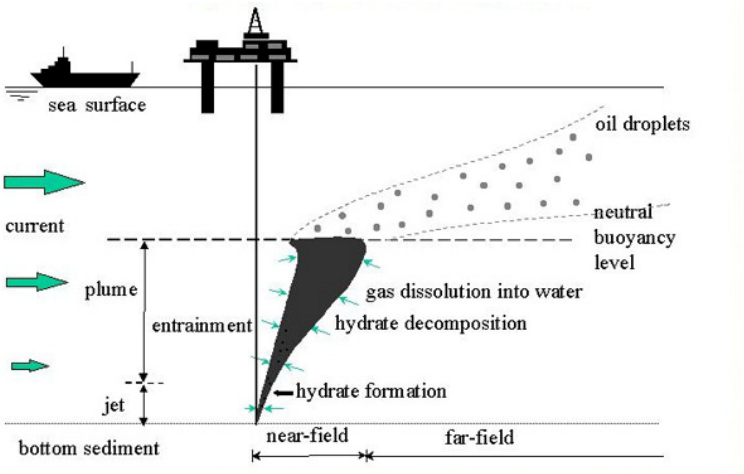

Figure 5. Graphical representation of the main processes occurring during a blowout spill (source: Zheng and Yapa, 2002).

Two aspects related to blowouts investigated recently have been the size distribution of oil droplets formed as a consequence of the high degrees of turbulence of a blowout, and the effect of using chemical dispersant in the release zone. The importance of properly simulate the oil droplet size distribution lays in the fact that a smaller droplets rise more slowly towards the surface and hence may be transported away from the release point by currents before reaching the surface. Moreover, droplet small enough may be trap underwater for long time enhancing dissolution and biodegradation. In this regard, the study of Johansen et al. (2013) propose the definition of the size distribution by means of a characteristic diameter, function of the modified Weber number, and a Rossin-Rammler distribution function. This approach has been widely adopted by different blowout models. Chemical dispersants were massively used for the first time in 2010 during the *Deepwater Horizon* emergency to reduce the size of oil droplets and enhance dissolution and degradation. Up to date, advanced algorithms to consider the effect of chemical dispersant have not been published yet and their effect has been introduced in blowout model in a simplified form, in general reducing the oil-water interfacial tension in the equation for the characteristic droplet diameter.

1.2.2 Application of statistical techniques in met-ocean fields

The presented numerical tools, during an oil spill accident, are used to predict the future evolution of the oil, in order to organize response activities. Such predictions, however, have a temporal limit imposed by the met-ocean forecasts used as model forcings and, in

general, can predict only 5-7 days ahead. In this sense, statistical met-ocean modeling may offer an alternative tools to try expand the time coverage of oil spill simulations.

1.2.2.1 Statistical modeling

Statistical modeling has been explored in atmospheric and marine fields, especially during the last decade, as an alternative to numerical modeling, for the well-known relative benefits in terms of computational efforts and time consumption (e.g., Camus et al., 2019; Cid et al., 2017; Diez-Sierra and del Jesus, 2019; Prasad et al., 2009). Regression model have been used to predict future events based on previous states and/or other variables by several authors (Casas-Prat et al., 2014; Mishra et al., 2017; Wang et al., 2012). In particular, Logistic Regression (LR) has shown to be very effective when dealing with met-ocean spatial fields. In a LR model, the independent (or response) variable is represented by a set of spatial patterns of the considered variable and the dependent (or predictor) variables are those that have an influence on the response variable. Prasad et al. (2009) applied LR to predict precipitation anomalies in India considering 36 environmental predictors. Guanche et al. (2014) used LR to simulate atmospheric circulation patterns on the Northeast Atlantic using monthly sea level pressure anomalies and long-term trend effects as predictors, including a Markov Chain in the model to consider the autoregressive component of the analyzed system. A similar model was applied by Antolínez et al. (2016) to European Atlantic wave climate, with a complementary model to better estimate the persistence probability of the patterns. More recently, Cardenas et al. (2017) were able to simulate daily wind states on the Northeast Atlantic by applying a LR model that considered seasonality, pressure anomalies over the study domain and a second-order Markov Chain as predictors.

1.2.2.2 Data-mining for pattern extraction

In the last decades, corresponding with the development of extensive meteorological and oceanographic numerical reanalyses, different data mining techniques have begun to be applied in these environmental fields, in order to make the most of such amount of data. Clustering and selection algorithms were particularly investigated (Camus et al., 2011). Among them, k-means has been one of the most popular clustering technique, used to extract relevant features from the original data allowing for a more manageable identification of important properties of the considered data. The k-means algorithm computes a set of k prototypes or centroids, each of them characterizing a group of data, formed by the vectors in the database for which the corresponding centroid is the nearest one (Hastie et al., 2001). K-means clustering has demonstrated to be able to extract the most relevant spatial pattern from atmospheric (e.g. Diez-Sierra and del Jesus (2017), Omrani et al. (2016), Rueda et al. (2017), Sönmez and Kömüşcü (2011)) and oceanographic hindcasts (e.g. Hsieh et al. (2004), Solabarrieta et al. (2015), Solidoro et al. (2007), Yi et al. (2013)). Bárcena et al. (2015) and Núñez et al. (2019) favorably applied k-means to met-ocean data series from one-point location to extract patterns of temporal evolutions of the considered data at this point. This clustering technique has never been used to identify patterns of temporal evolutions of met-ocean variables over a spatial domain, i.e. spatio-temporal patterns.

1.2.3 Stochastic oil spill modeling for oil spill risk assessment

One of the most important tools to help minimizing the consequences of an accidental spill is represented by oil spill risk assessments. The oil spill risk can be defined as a function of the likelihoods of a series of oil release scenarios and their potential environmental and socio-economic consequences in a considered study area. The oil release scenarios can be selected considering the historical variability in terms of duration, location and potential volumes and type of hydrocarbons spilled, and a likelihood is associated to each of them. Numerical modelling of the evolution of the selected release scenarios under specific met-ocean conditions characteristic of the study area (met-ocean scenarios) is required to determine their consequences (IPIECA and IOGP, 2013). The oil spill modelling carried on for each considered release scenario is stochastic since a large number of simulations are required to consider the local atmospheric and oceanographic variability (Fingas, 2017). An oil spill simulation is run for each met-ocean scenario considered, so the number of spill simulations needs to be large enough to reflect the met-ocean variations of the considered domain.

The general approach used in the literature to consider the met-ocean variability is a Monte Carlo method to iteratively randomly change the initial date of the spill simulations and consider the corresponding met-ocean conditions from historical reanalysis databases. This method brings, generally, to the selection of hundreds or thousands of met-ocean conditions to properly consider the environmental variability and, hence, leads to the execution of hundreds or thousands oil spill numerical simulations with the consequent high levels of computational effort. Abascal et al. (2010) used TESEO oil spill model (Abascal et

al., 2017, 2016, 2007) to simulate oil slick trajectories from different surface spill locations in the Bay of Biscay, applying a Monte Carlo technique to select 400 hypothetical spill dates from a 44-year hindcast dataset of wind and waves. Liubartseva et al. (2015) forced MEDSLIK-II model (De Dominicis et al., 2013) to evaluate the oil spill hazard in the Southern Adriatic and Northern Ionian Seas by randomly selecting the spill's date, in the period 2009-2012, and its location, within the most busy maritime routes for a total of more than 5000 simulations for each month of the considered period. Canu et al. (2015) calculate hazard maps for the island of Sicily (Italy) by performing an ensemble of 730 simulations, each driven by a slightly different (1 day shifted) circulation field extracted from the available 2-year period met-ocean database. Similarly to the mentioned ones, most of the oil spill risk studies from the literature are focused on vessel related spill analysis and hence limited to 2D surface spill modeling.

The oil spill modeling state-of-the-art review showed that, during the last decades, many advances have been accomplished in the numerical modeling field (including both near field and field modeling). However, there is still the lack of a methodology to assess the oil pollution hazard from a deep spill. Among the few available works of oil spill risk assessment related to offshore deep releases, in general, a physical simplification is applied consisting in the assumption that the transport of the spill begins at the sea surface just above the actual deep spill site, so the problem is redirect to the 2D case (e.g. Barker (2011), Ji et al. (2011)). In other occasions, only a limited number of 3D met-ocean scenarios are considered for the risk assessment (French-McCay et al., 2018). Such simplifications may be related to the complexity and high computational costs required when dealing with extensive 3D current data fields.

However, the International Association of Oil and Gas Producers (IOGP) and the International Petroleum Industry Environmental Conservation Association (IPIECA) have recently published an internationally accepted guideline for oil spill risk assessment planning for offshore installations (IPIECA and IOGP, 2013). In this guidance document, the relevance of including the subsurface spill modeling when assessing the oil spill related risk for an offshore installation is highlighted. To do so, together with wind data, 3D current information have to be considered. The document also emphasizes that the number of met-ocean scenarios should be large enough to represent the meteorological and 3D oceanographic variability of the considered study site. However, accomplishing with these criteria means dealing with an extremely high amount of met-ocean data and with very computational demanding numerical simulations, which would eventually bring to a very time and resource consuming process for assessing the risk related to an offshore deep oil spill.

1.2.4 Conclusions

As the state-of-the-art review of oil spill modeling shows, an ongoing trend exists focused on improving numerical oil spill models for the simulation of oil spills. The latest advances are especially centered on allowing realistic 3D simulations of oil spills, including deep blowout accidents, in which case a near field model and a near field – far field connection is required. In spite of these advances in the numerical simulations, there is still a lack of methodologies focused on expanding the benefits of oil spill models beyond the

typical short-term imposed by met-ocean forecasts coverage (5-7 days). A methodology for mid-long term oil spill predictions would match the actual drift period observed during past major oil spills (several months). In this regard, statistical modeling, in particular logistic regression modeling, have proved its capacities in different fields, and may represent a promising tool to be explored for oil spill modeling applications.

On the other side, existing methodologies for oil pollution hazard analyses are mainly focused on vessel-related oil spill and, hence, generally only involve surface simulations and 2D processes. In these studies, a stochastic modeling approach is used in order to consider the overall environmental variability that may affect a spill evolution in a given study area. In general, hundreds or thousands of met-ocean scenarios are selected to simulate the spill evolutions under each of these conditions. Methodologies for stochastic modeling of deep oil spills, integrating surface and subsurface processes, are still lacking. The adding complexity of considering the whole 3D evolution of a spill in a stochastic modeling leads to critical computational requirements and very large calculation time.

1.3 Objectives

The present thesis is focused on trying to fill the state-of-the-art gaps identified in the previous sections. The main objective of this thesis is to define and to develop a generic and comprehensive framework for the simulation, at the mid-long term, of oil spills in the sea, considering both the prediction of the future evolution of an actual spill and the stochastic modelling for assessing the risk of a hypothetical spill (Figure 6). Surface and subsurface

spills, including blowouts, are considered in the design of the general methodology. For the development of both branches of the methodology (oil spill prediction and stochastic modeling for risk assessment), statistical methods are used to solve the problems previously mentioned.

To achieve this main goal, the following specific objectives need to be fulfilled:

1) To investigate the capacities and the viability of statistical modeling, in particular of logistic regression, to simulate met-ocean patterns, with the final aim of allowing mid-long term oil spill simulations.
2) To define a methodology for the mid-long term forecast of oil spill, based on the statistical prediction of met-ocean patterns
3) To define a methodology for stochastic modeling in oil spill risk assessment, suitable both for surface and subsurface spills, based on the selection of temporal evolution patterns of met-ocean fields, allowing for a reduction of computational costs and calculation time with respect to traditional stochastic modeling techniques.

Moreover, the implementation in TESEO model of numerical processes and algorithms for the simulation of the evolution of a spill through the water column, as well as a linking module between near field and far field modeling, are presented in this thesis.

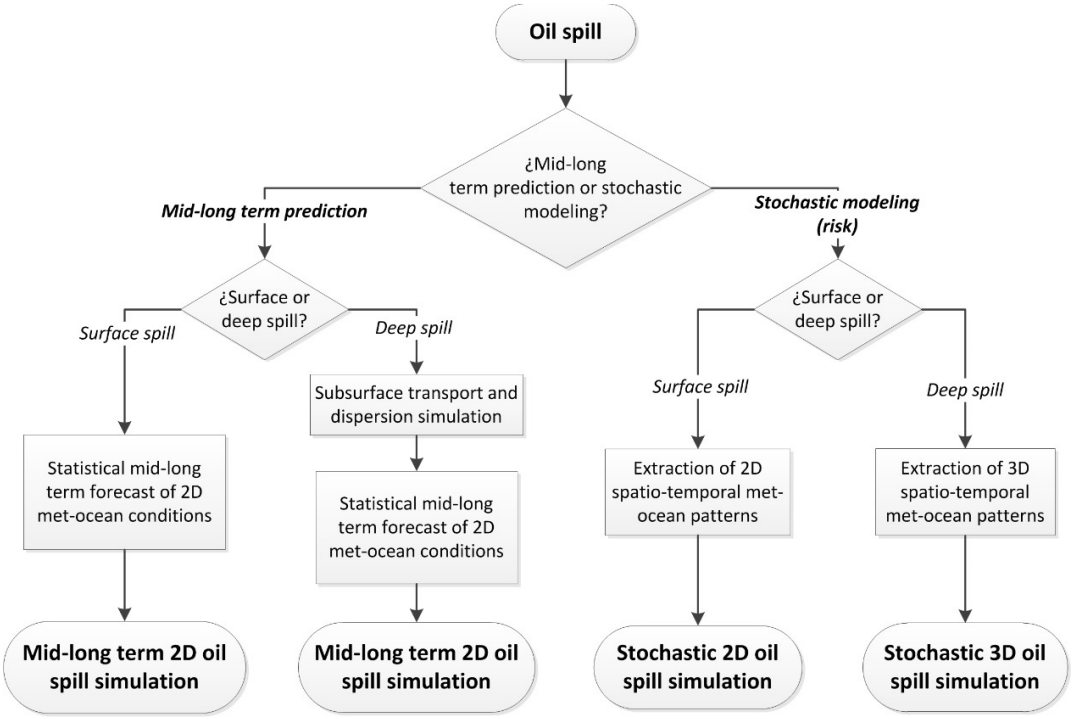

Figure 6. Flowchart of the general methodology for statistical mid-long term oil spill simulation, which development and definition represents the goal of this thesis

1.4 Layout of the thesis

Regarding the mid-long term oil spill prediction methodology, it is important to consider the different type of met-ocean dynamics that affect a spill evolution, depending on the area. The two cases of areas dominated by general ocean circulation (like in the Gulf of Mexico) and areas in which the local wind plays also an important role in the oil drift (like in the Bay of Biscay) have been considered. For this, in Chapter 2 a methodology for the statistical simulation of ocean currents in the Gulf of Mexico using a logistic regression model is presented. In this region, in 2010, the Deepwater Horizon offshore oil platform originated the wort accidental spill in history. In Chapter 3, the methodology is applied to simulate wind and current patterns in the Bay of Biscay where, in 2002, occurred the *Prestige* tanker oil spill. Taking advantage of the obtained wind and current pattern forecast, a method is developed for the mid-long term oil spill prediction and applied to simulated the trajectory of several drifter used during the *Prestige* accident.

Once presented the advances in the mid-long term prediction of oil spills, in Chapter 4 a methodology is provide for the stochastic oil spill modeling in the framework of risk assessment, based on met-ocean pattern extraction. The methodology is applied to assess the hazard of a deep spill in the North Sea, one of the most active region in the offshore oil and gas field.

Finally, in Chapter 5, the main conclusions of the thesis are summarized and future researches are proposed.

CHAPTER II.

STATISTICAL SIMULATION OF CURRENT PATTERNS USING A LOGISTIC REGRESSION MODEL. APPLICATION CASE IN THE GULF OF MEXICO[1].

2.1 Introduction

The main environmental variables to be considered in oil spill modeling vary depending on the considered area. In some area, dynamics affecting oil spill evolutions are dominated general ocean circulation, like in the Gulf of Mexico. In other cases, as in the Bay of Biscay, the local wind effect plays also a key role in current variability and in the oil drift.

[1] The findings of this chapter have been published in: Chiri, H., Abascal, A.J., Castanedo, S., Antolínez, J.A.A., Liu, Y., Weisberg, R.H., Medina, R., 2019a. Statistical simulation of ocean current patterns using autoregressive logistic regression models: A case study in the Gulf of Mexico. Ocean Model. 136, 1–12. https://doi.org/10.1016/j.ocemod.2019.02.010

Then, a methodology for the statistical simulation of environmental conditions have to consider this variability. In this chapter, the methodology for statistical modeling of ocean patterns by means of a logistic regression model for general circulation dominated areas is presented and applied to current patterns in the Gulf of Mexico (GoM). In this region, in 2010, the offshore oil platform *Deepwater Horizon* suffered an accident that eventually led to the worst accidental oil spill in history. The GoM currents have been object of several studies due to their particular spatial variability and their implication in the very active offshore oil and gas industry in the area. The statistical simulation of ocean currents here presented represents a key element for the mid-long term prediction of offshore oil spills affected by general circulation processes. This chapter aim to reach the first specific objective of the present thesis.

2.2 Study site

Ocean circulation in the GoM is dominated by the Loop Current (LC) system. The LC is a part of the North Atlantic Ocean's western boundary current, between the Yucatan Current and the Florida Current, leading to the Gulf Stream. After entering in the GoM through the Yucatan Strait, the LC flows following an anticyclonic looping path before exiting the GoM through the Strait of Florida. The main feature of the LC is the position of this looping path, which can alternatively penetrate northwestward toward the Mississippi River Delta or retract toward Cuba (e.g. Sturges and Lugo-Fernández, 2005). Additionally, it has been observed that large intrusions generally produce the shedding of anticyclonic

eddies that start travelling westward in the GoM, at the same time that the LC began to retract back to the south toward a more direct entrance-exit path (e.g. Lugo-Fernández, 2016).

Multiple studies, based on different kinds of analysis, share a similar conclusion that certain LC events tend to be affected by analogous previous LC states. Alvera-Azcárate et al. (2009) analyze altimetry-derived LC geostrophic component over a period of approximately 13 years. They observe that large eddy shedding tend to produce large LC retraction, followed by a long period before the LC penetrates enough for a new shedding and vice versa, the shedding of a small eddy results in a limited LC retreat and in a shorter time until the next intrusion and eddy shedding. Lugo-Fernández (2007) applies different statistical analyses on observational databases from published sources in his aim to understand if the LC dynamics represents a chaotic system or not. He obtains that the system have a memory of about 10–18 months, the time it takes for the LC to complete an entire cycle consisting of inrushing, shedding, and returning to its initial position. Liu et al. (2016b), analyzing the temporal evolution of 16 LC patterns extracted from a 23 years altimetry-derived database, identify some characteristic pattern sequences recurrent over the analyzed historical period.

In order to analyze inter-annual variability of the LC, some recent studies tried to link the LC variability with large-scale climate indices. Lugo-Fernández (2007) argues that LC presents oscillations in time scales of 3-5 years which agrees with the time scale of North Atlantic Oscillation (NAO) and of El Niño-Southern Oscillation (ENSO). Muller-Karger et al. (2015) analyzed monthly Atlantic Multidecadal Oscillation (AMO) index (Enfield et al., 2001) and the monthly Multivariate ENSO Index (MEI) (Wolter and Timlin, 2011) with several oceanographic variables in the GoM but do not find significant correlation between

the indices and any of the considered variables. Zeng et al. (2015b) after extracting three LC pattern through a SOM application, compared the frequency of occurrence of a specific pattern considered as an eddy shedding situation with various climate indices, including Oceanic Niño Index (ONI), NAO, Southern Oscillation Index (SOI) and Pacific Decadal Oscillation (PDO). These authors also analyze the influence on the frequency of occurrence of the selected pattern of the wind stress curl in the GoM, Caribbean Sea and Bahamas area. They find that the correlation of the frequency of occurrence of the considered LC pattern is 0.6 with the six-month moving averaged ONI considering a 90-day lag and 0.83 with the wind stress curl in the Caribbean Sea. Oey et al. (2003) also investigate the affection on the LC of the upstream conditions in the Caribbean Sea and even further in the North Atlantic. Through numerical experiments, they find that the wind stress curl in the Caribbean Sea spin up eddies that, drifting later towards the GoM, highly affect the LC eddy shedding capacity. Their results also reveal a significant influence on the LC eddy shedding frequency of the interactions between the North Atlantic and the Caribbean Sea through the Greater and Lesser Antilles Passages.

In the present study, taking advantage of this previous knowledge of the LC and its variability, we will consider as possible covariates of the ALR model for the LC: the memory of the system, the impact of the mentioned climate indices (NAO, AMO, ENSO, MEI, ONI, SOI and PDO) considering a temporal lag up to 1 year, the local effect of the wind stress curl in the GoM and the possible upstream affection of the wind stress curl in the Caribbean Sea. In order to consider a possible connection to the North Atlantic variability, we will also include in the analysis the sea level pressure anomalies over the North Atlantic. This variable

is often used when dealing with the variability of the North Atlantic (e.g. Guanche et al., 2014; Klotzbach and Gray, 2008).

2.3 Data

Current data in the GoM are a derived product of altimetry data from the Archiving, Validation and Interpretation of Satellite Oceanographic Data (AVISO+). Several recent works have shown that this represents a good product to study the geostrophic component of the LC (Liu et al., 2011; 2016a; 2016b; Weisberg and Liu, 2017). AVISO+ is a multimission gridded sea level anomaly data set produced by the *Ssalto/Duacs* with support from the *Cnes* (http://www.aviso.altimetry.fr/duacs/) and distributed by the Copernicus Marine and Environment Monitoring Service (CMEMS, http://www.marine.copernicus.eu). It is a global product with horizontal resolution of 1/4º and daily temporal resolution. The delayed-time data are used for the period between 1 January 1993 through 6 May 2016, and the near real-time version are used for the remaining time (7 May 2016 to 31 December 2016). The anomaly fields are combined with the mean dynamic topography (MDT_CNES-CLS13, Rio et al., 2014) produced by CLS Space Oceanography Division and distributed by Aviso, with support from Cnes. Surface geostrophic currents were then calculated as the gradient of the absolute sea level, similar to that in the work of Liu et al. (2008). The geostrophic current data used in this work are limited to the GoM region, within latitude and longitude ranges of [22.8, 31.8] North and [98.8, 81.8] West, respectively, and with the omission of coastal ocean

regions with water depth < 100 m where conventional altimetry data are not as reliable as in the open oceans due to a number of factors explained by Vignudelli et al. (2011) (Figure 8).

The climate indices data were obtained from Climate Explorer website of the Royal Netherlands Meteorological Institute (http://climexp.knmi.nl/). As mentioned in Section 2.2, based on previous studies of the LC variability, in this work we consider the following climate indices: NAO, AMO, ENSO, MEI, ONI, SOI and PDO.

Wind stress curl was calculated data from the National Center for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR, 2005; Mesinger et al., 2006) wind data. NARR database includes several meteorological variables from 1979 to present with a horizontal resolution of approximatively 32 km, 45 vertical layers, and 3-hourly frequency, covering the whole North American Region. NARR is a result of the NCEP Eta model with the Regional Data Assimilation System (RDAS) and significantly improves the accuracy of temperature, winds and precipitation compared to the NCEP-DOE Global Reanalysis 2. In this work we analyze the local effect of the wind in the GoM and the upstream effect in the Caribbean Sea. As shown by other authors, from the point of view of the hydrodynamic flow, two different sub-areas of the Caribbean Sea can be identified: the Venezuela-Colombia Basin and the Cayman Basin, separated by a shallow ridge extending from Honduras and Nicaragua to Hispaniola, called the Nicaraguan Rise (e.g. Alvera-Azcárate et al, 2009). Following Zeng et al. (2015b), with the aim of considering only the most direct effect of the upstream atmospheric situation, the wind over the Cayman Basin is considered in this study. Hence, for the GoM region wind stress curl was calculated within latitude and longitude ranges of [22.8, 31.8] North and [98.8, 81.8] West, respectively, and for the Caribbean Sea within latitude and longitude ranges of [17.0, 21.4] North and [89.0,

78.3] West, respectively. In order to reduce the dimensionality of the data, we apply Principal Component Analysis (PCA) to the data and keep the principal components (PCs) retaining the 95 % of the total variance, resulting in 68 and 41 PCs for the GoM and the Caribbean Sea domains, respectively.

Sea level pressure data are provided by the NCEP Climate Forecast System Reanalysis (CFSR, Saha et al., 2010). In this study we use CFSR monthly mean data with 1º spatial resolution, within latitude and longitude ranges of [0, 65] North and [-65, 15] East, respectively. In this domain we calculated the monthly anomalies with respect to the average months. Similarly to that for the wind stress curl data, we reduce the dimensionality of the data by keeping the first 10 PCs of the obtained anomaly series, which account for 95 % of the total variance.

2.4 Methods

2.4.1 Logistic regression general framework

Logistic regression is a special case of generalized linear model (GLM). GLMs are statistical models widely used to investigate effects of explanatory variables on a response or dependent variable. In practice, these models link a linear combination of a group of predictor variables to a function of the response variable. Logistic regression models are widely used when the response variable to be modelled is categorical, i.e., it can take exclusively one from a set of possible values. In the case of logistic regression, the function of the response

variable linked to the covariates is the *logit* function. This function represents the natural logarithm of the odds of each possible outcome with respect to an arbitrary baseline category. Based on this definition, if $Y \epsilon \{1, \dots, n\}$ is the set of n ocean current patterns, and the last category J is the baseline, a generic multinomial logistic regression model for the j-th pattern can be expressed as follows:

$$log\left(\frac{P(Y_t = j|X_t)}{P(Y_t = J|X_t)}\right) = \alpha_j + \beta_j X_t, \qquad j = 1, \dots, n-1 \tag{1}$$

where X_t is the vector of covariates at time t, β_j refers to the vector of coefficients for each of the values of the covariates at time t, and α_j represents the intercept of the model. The left-hand-side of equation 1 is the logit of a generic category of the response variable, the right-hand-side is the linear combination of the predictors.

As mentioned in Section 2.2, the influences of the LC variability have been studied in previous studies, including the memory of the LC system, possible relationship with some climate index phases and the response to atmospheric variations. Here we set up an ALR model taking all these features into account (Figure 7).

The influence of the memory of the system is explored by adding to the model autoregressive terms as a Markov Chain of an order equal to the number of previous states to be considered. The implementation in an ALR model is achieved by using a contrast matrix, such as the Helmert matrix (de Vries et al., 1998). A Markov Chain of order d can be implemented as follow:

$$X_t^{AR_d} = \sum_{i=1}^{d} \sum_{k=1}^{n-1} Z_k^{t-i} \gamma_{ik} \tag{2}$$

where Z_k^{t-i} is the dummy variable obtained through the Helmert matrix transformation, and represents the previous Loop Current pattern (LCP) at time *t-i*. γ_{ik} is the coefficient associated with LCP *k* at time *t-i*. Note that the autoregressive component adds a set of $(n-1) * d$ new parameters for each category of the model.

The influence of climate indices or atmospheric conditions on the *logit* of the LCPs can be included in the model as follows:

$$X_t^C = \sum_{i=1}^{n_c} C_i \beta_i^C \tag{3}$$

where C_i is the vector of the values of the n_C climate indices considered or of the n_C principal components of the atmospheric field considered, at time *t*, and β_i^C is the vector of associated coefficients.

Eqs. (2) and (3) show how to introduce the effect of the considered variables in a logistic regression model. However, the actual significance of these variables must be carefully assessed. After defining the response variable of our model, in Section 2.4.3 a statistical based method is described to assess the significance of each of these components. An outline of the methodology followed to obtain the LCPs and to define the components of the ALR model is presented in Figure 7.

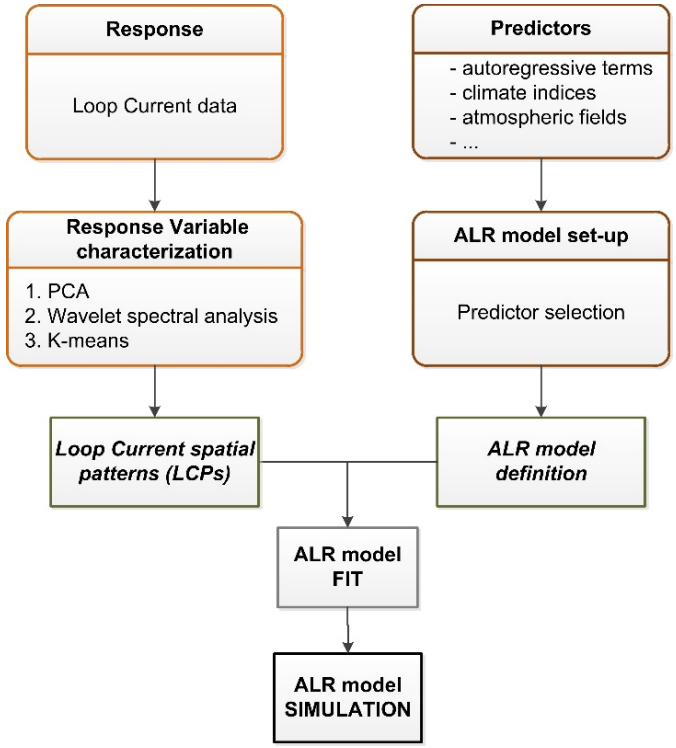

Figure 7. Diagram of the proposed methodology to obtain an autoregressive logistic regression (ALR) model for the Loop Current Patterns (LCPs) simulation

2.4.2 Response variable characterization

The response variable in an ALR model is categorical, i.e., it is made up of different categories that the dependent variable can exclusively assumes. In our case, the LC is the response variable and the categories are represented by a set of its typical spatial patterns (LCPs).

In order to select these representing LCPs we follow a three-step clustering methodology which aims to extract patterns that homogenously span the space of the original

data considering the most significant temporal scale for the pattern selection. The three phases of this clustering methodology are presented next (Figure 7).

2.4.2.1 Principal component analysis

In order to avoid spatially correlated variables that may disturb the LC data clustering, the PCA is applied to the altimetry-derived daily data of geostrophic current components. Through this analysis, original data are redefined with respect to new orthogonal dimensions, or empirical orthogonal functions (EOF; the eigenvectors of the covariance matrix of that dataset). PCs are the time evolution of the weights of the EOFs. The obtained EOFs and the associated PCs are ordered based on the variance of the data explained by each component. The first 84 PCs, accounting for 95% of the total variance are then considered. As examples, the first two and last two EOFs and the associated PCs obtained are shown in Figure 8. The percentage of the total variance accounted for by each component is shown in the individual EOF plot. As expected, the first two EOFs show patterns coherent with the LC system, while the last modes represent smaller scale spatial variations and higher frequencies affecting quite homogeneously the whole basin.

2.4.2.2 Wavelet spectrum analysis

As a second step in the clustering process, a joint PC-wavelet analysis similar to that in Liu et al. (2016b) is performed over the selected leading PCs. The aim of this step of the clustering methodology, consisting in a rectified wavelet spectral analysis (Liu et al., 2007a) on the obtained PCs, is to identify the temporal scale under which no major variability of the GoM geostrophic currents occurs. In Figure 8 we show, as examples, the results of this

analysis for the first two and last two considered PCs. The rectified wavelet spectra for the first two PCs (accounting for 15.9% and 12.3% of the total variance, respectively) show that variations of these GoM current components are mainly in the seasonal to annual time scales (2 months to 1 year). The variations of the last two PCs are generally faster, being the most energetic between 14 days (red horizontal line in the wavelet spectral analysis panels of Figure 8) and 1 year. We have performed this analysis for the total of the 84 considered PCs and found that, in all cases, the major variability for each PC occurs at a time scale greater than or equal to 14 days.

Therefore, in order to reduce the source of unexpected noise that could negatively affect the statistical model, the 84 daily PCs of the LC are rescaled to bi-weekly frequency by calculating the average values from the daily series.

Please note that in this work the application of PCA has a dual purpose: first, to allow the joint PC-wavelet analysis with the aim of exploring the temporal scales of the GoM geostrophic current variability and eventually choosing a consistent time scale for the pattern extraction and, second, to reduce the current data dimensionality and simplify the clustering process.

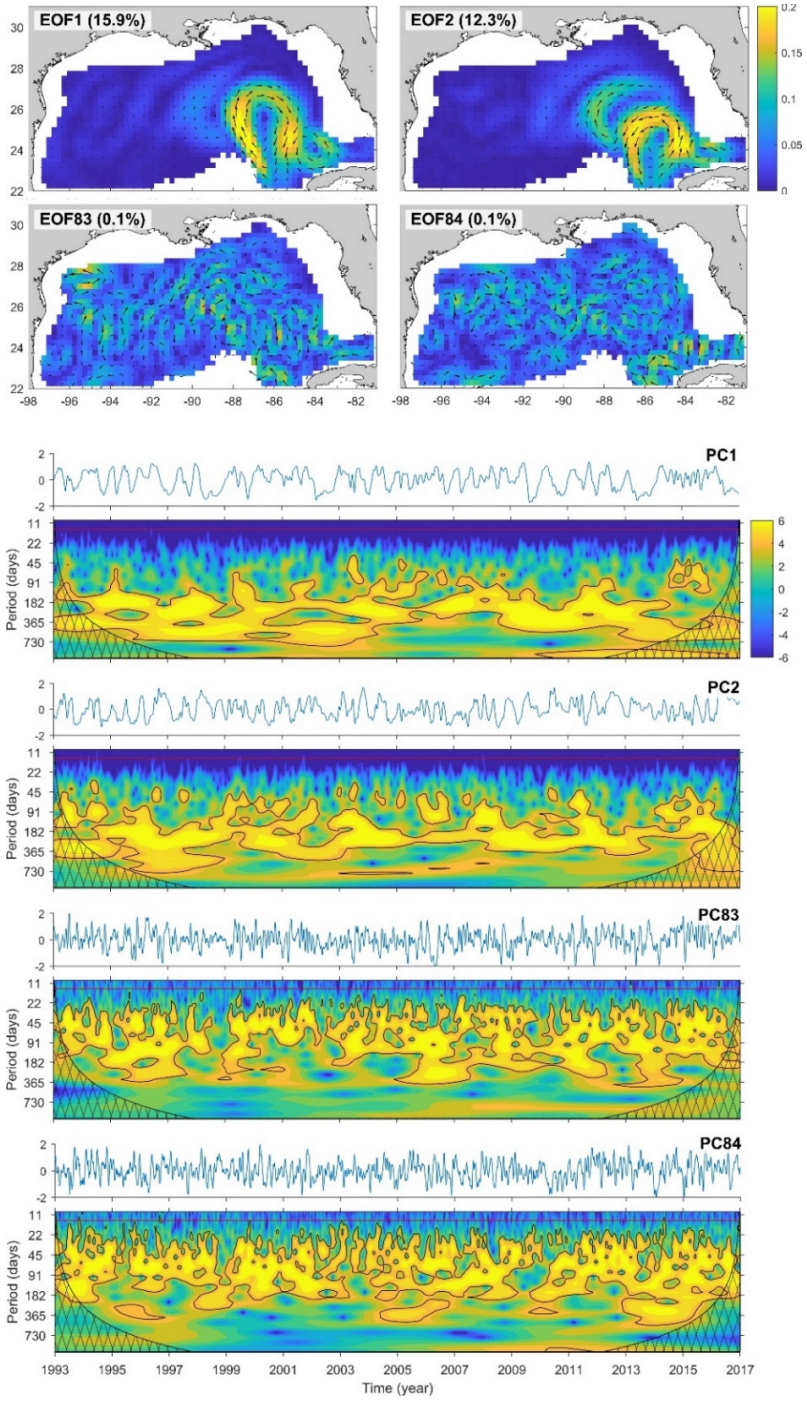

Figure 8. Top: First 2 and last 2 EOFs of the geostrophic current in the Gulf of Mexico. The variance accounted for by each mode is shown in each graph. Current units in m/s. Bottom: First 2 and last 2 PCs plots and the associated wavelet power spectra. The PC variance is normalized to be 0.5. Rectified wavelet power spectra are shown as filled contours in base 2 logarithm. The regions of greater than 90% confidence are shown with black contours. Cross-hatched regions on either end indicate the "cone of influence", where edge effects become important.

2.4.2.3 K-means

As a last step of the pattern extraction, the K-means algorithm technique is applied to the bi-weekly LC PCs. The Maximum Dissimilarity algorithm was used to pre-select initial centroids, ensuring K-means technique to correctly describe the diversity of the data (Camus et al., 2011; Antolínez et al., 2018). Several tests were performed in order to find an equilibrium between the spatial variability of the current and the number of groups to be considered. As a result of the clustering, 8 centroids are selected representing 8 LCPs (Figure 9). These typical LC conditions extracted from the 24-year geostrophic current dataset, will represent the 8 categories in which the ALR model will classify the LC. The order of the current patterns have been arranged following a similarity criteria, but only for an easier interpretation of the analysis of the results. From the perspective of the ALR model, the order of the categories of the response variable does not affect the results. The patterns show a broad variability of the LC and cover well different situations between the maximum retraction (LCP 1) and the deepest intrusion (LCP 10). It is interesting to see how for LCP 1, LCP 2 and LCP 3, representing retraction or mid intrusion stages, the LC path stays close to the southwestern part of the west Florida shelf, while the loop tends to curve westward as its path advances deeper in the GoM (LCP 4, LCP 5 and LCP 6), in accordance with the findings of Weisberg and Liu (2017). The obtained spatial LCPs also agree well with the results of Liu et al. (2016b) that extract 16 spatial pattern with a 4x4 SOM application for sea level and current data over the eastern part of the GoM. Alvera-Azcárate et al. (2009) describe extreme LC intrusion as an intrusion that extends up to 92ºW or 27ºN. LCP 7 and LCP 8 show these situations, highlighting that the used clustering algorithm spans the space of the original data

enough to represent these extreme conditions. Once the characteristic patterns are obtained, the original 24-year series of GoM geostrophic current is expressed in a sequence of BMUs, in which each time step of the GoM geostrophic current is expressed by the most similar LCP extracted. The frequencies of occurrence of the LCPs in the historical series is, in order from LCP 1 to LCP 8, the following: 6.5 %, 12.6 %, 13.6 %, 14.1 %, 15.7 %, 14.5 %, 11.8 % and 11.2 %.

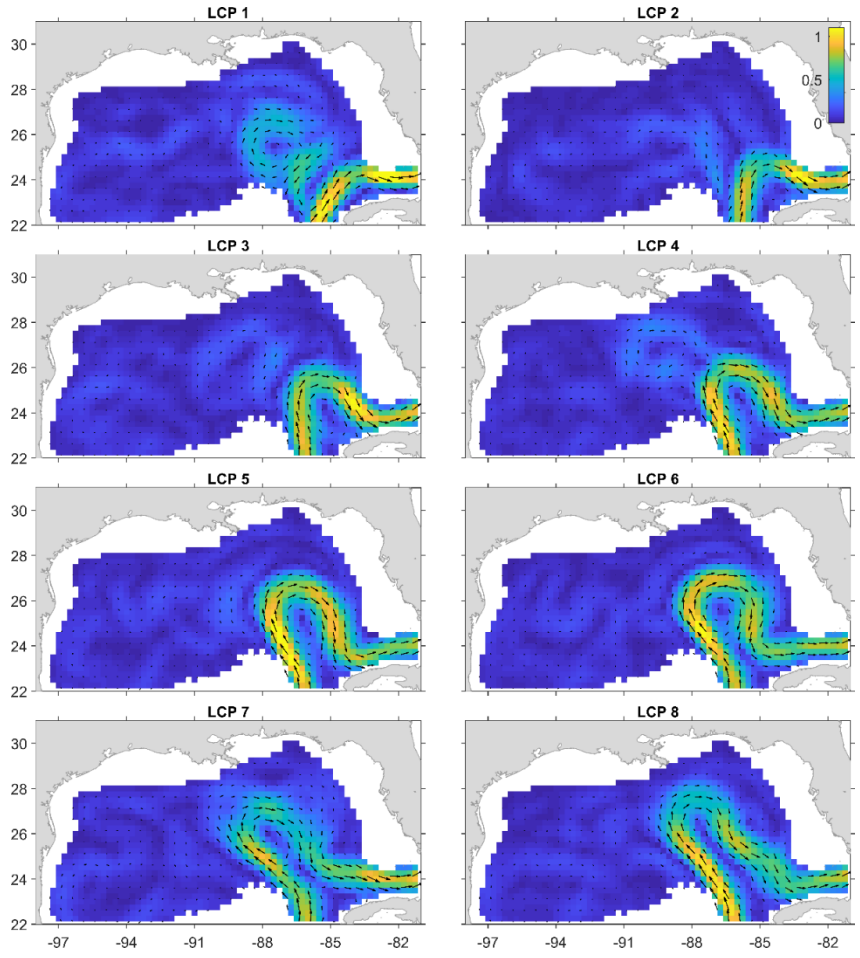

Figure 9. Selected Loop Current patterns representing the categories of the response variable in the ALR model. Current speed units in m/s.

2.4.3 Model set-up

Building an ALR model is the hardest part of applying logistic regression. Although in some applications the inclusion of certain predictors can be easily guessed, i.e., the wind effect in shallow water areas or tidal harmonic components in narrow channel flows, in general it may be challenging to select the set of explanatory variables that actually influence the current of interest. Moreover, it must be considered that even if a model with several well-chosen features is likely to fit the data better, as the number of feature increases, possible effects and interactions among some covariates increase as well (Agresti, 2007). Generally, it is preferable to keep the model as simple as possible for a better interpretation of the results.

Statistical inference-based algorithms can be used to build the model. One particularly useful when studying ocean currents is the forward stepwise method. It consists of starting from the simplest model possible, i.e., considering just the constant value, and then adding sequentially new explanatory variables by testing if they produce a significant improvement in the model fitting quality or not. Several significance tests can be used like Wald's test or Aikake's information criteria. Here, following Guanche et al. (2014), we propose the use of the likelihood ratio statistic. This method compares the deviance ratio *ΔDev* (the change of fitting quality for two nested models) and the chi-square distribution with *Δdf=Δnp* * (n-1) degree of freedom. *Δnp* is the number of parameters added in the more complex model. Assuming a confidence level α=95%, when ΔDev>$X^2_{0.95,\Delta df}$ the increasing of fitting quality of the model including the new covariates is statistically significant.

As mentioned earlier in Section 2.4.1, the aim of this analysis is to assess the influence of autoregressive terms and inter-annual variability on the LCP evolution. The results of the

forward stepwise method based on likelihood ratio statistic are shown in Table 1. Model *0*, the *Null* model, only considers the intercept value. Nesting to Model *0*, Model *I* includes the effect of an autoregressive term, i.e., it checks if the manifestation of a current pattern is influenced by the previous current spatial situation or not. The increment in deviance obtained by adding this effect $\Delta Dev_I=1781.7>X^2_{95\%,8}$ allows to accept the hypothesis that the first autoregressive term is significant in the 24-year series of LCPs. The autocorrelation of the series is further explored by progressively adding a second, third and fourth autoregressive term in Model *II*, Model *III* and Model *IV*, respectively. As shown in Table 1, the second and third term are significant as well, even if the improvement achieved in the model fit decreases as the order of the autoregressive terms increases. Based on the statistical test, the effect of including a fourth autoregressive term is not significant, being the improvement in the deviance between Model *IV* and Model *III* $\Delta Dev_{IV}=25.7<X^2_{95\%,29}$. This finding, that three previous steps at a bi-weekly interval are significant to determine the next LC status, is in concordance with the results of the study by Zeng et al. (2015a) in which the authors develop a model based on an autoregressive artificial neural network to predict the GoM sea surface height and find that six steps, at weekly scale, is the optimal delay for their model.

Once the autoregressive component of the model has been assessed, we test if any of the atmospheric variables and climate indices investigated in the previous studies on the LC variability can improve the goodness-of-fit of the ALR model. Since the LC patterns are calculated based on bi-weekly averaged current data, all the predictors are first resampled at the same time scale for a concordance among LCP series and predictor time series. In Model

V, we add to Model *III* the wind stress curl in the GoM, and obtain $\Delta Dev_V=652.6>X^2_{95\%,90}$; hence the contribution of the new predictor is significant. In the next model we include the effect of the wind stress curl in the Caribbean Sea, which results not significant ($\Delta Dev_{VI}=330.0<X^2_{95\%,131}$). However, considering a delay between the LCP series and the series of wind stress curl in the Caribbean, we find that the effect of the wind in the Caribbean Sea with a 6-week delay significantly improves the goodness-of-fit of the ALR model, as shown by Model *VII* in Table 1. In the same way, spanning different time lag of the North Atlantic MSLPAs with respect to the LCP series, we find that Model *VIII*, including North Atlantic MSLPAs with a 4-week time lag significantly improves the model ($\Delta Dev_{VIII}=128.3>X^2_{95\%,141}$). To further improve the capability of the model, we consider the previously introduced climate indices, each one with a possible time lag up to 1 year. However, none of these indices with none of the considered lags is found to be statistically significant at all (not shown here). As an example, in Table 1 we show Model *IX* which includes ONI with 6-week lag as a covariate. Following the test, the increment in the deviance is not significant ($\Delta Dev_{IX}=9.8<X^2_{95\%,142}$). Based on these results, we choose Model *VIII* as the optimal ALR model for the statistical simulation of the LCP series.

Hence, considering as covariates three autoregressive terms, the PCs of the wind stress curl in the GoM, the PCs of the wind stress curl in the Caribbean Sea with a 6-week lag and the PCs of the North Atlantic MSLPAs, based on (1), (2) and (3) the final equation for our ALR model can be written as follows:

$$P(Y_t = j|Y_{t-1}, Y_{t-2}, Y_{t-3}, X_t^{WG}, X_t^{WC}, X_t^{SlpaNA})$$

$$= \frac{exp\left(\alpha_j + \sum_{i=1}^{3}\sum_{k=1}^{7} z_k^{t-i}\gamma_{ik} + \sum_{i=1}^{68} C_i^{WG}\beta_i^{C^{WG}} + \sum_{i=1}^{41} C_i^{WC}\beta_i^{C^{WC}} + \sum_{i=1}^{10} C_i^{SlpaNA}\beta_i^{C^{SlpaNA}}\right)}{\sum_{k=1}^{7} exp\left(\alpha_j + \sum_{i=1}^{3}\sum_{k=1}^{7} z_k^{t-i}\gamma_{ik} + \sum_{i=1}^{68} C_i^{WG}\beta_i^{C^{WG}} + \sum_{i=1}^{41} C_i^{WC}\beta_i^{C^{WC}} + \sum_{i=1}^{10} C_i^{SlpaNA}\beta_i^{C^{SlpaNA}}\right)} \quad (4)$$

where Y_t represent the LCP representing the LC state at the time t, X_t represents the conditions of a covariate at the time t, C_i specifically refers to the *i-th* PC of the considered covariate, superscript acronyms specify the covariate: wind stress curl in the GoM (*WG*), wind stress curl in the Caribbean Sea (*WC*) and MSLPAs over the North Atlantic (*SlpaNA*).

The proposed model has a total of 141 coefficients: one for the intercept, 7 for each autoregressive component, 68 corresponding to the number of PCs of the wind stress curl in the GoM, 41 for the PCs of the wind stress curl in the Caribbean Sea and 10 corresponding to the PCs of the North Atlantic MSLPAs. These coefficients are adjusted to maximize the fit of the model to the historical data. Several available statistical software allow for an easy implementation and fit of a logistic regression model. In this work *scikit-learn* library (Pedregosa et al., 2011) of Python is used. We apply the *one-vs-rest* scheme, *l2* penalty option, a tolerance default value of 1e-4 and the *liblinear* solver algorithm.

Once the predictors have been selected and the corresponding coefficients have been adjusted, we use the fitted model to simulate the sequence of LCPs for the whole 24-year period. Considering the probabilistic nature of the statistical model, we run the model 100 times and obtain the average values of the probabilities.

Table 1. Statistical inference for different nested model (C: constant value, AR#: autoregressive term of order #, WSC GoM: wind stress curl in the GoM, WSC CAR: wind stress curl in the Caribbean Sea; #wl: temporal lag of # weeks). For each examined model, covariates, number of parameters, degrees of freedom (df), deviance increment (ΔDev) chi-square distribution value associated and significance are shown. The final selected model highlighted in bold.

Model	Covariates	N. param	df	ΔDev	$X^2_{95\%}$	Significant
0	C	1	5000			
I	C+AR1	8	4944	181.7	74.5	Yes
II	C +AR1+AR2	15	4888	99.7	74.5	Yes
III	C +AR1+AR2+AR3	22	4832	88.6	74.5	Yes
IV	C +AR1+AR2+AR3+AR4	29	4776	25.7	74.5	No
III	C +AR1+AR2+AR3	22	4832			
V	C+AR1+AR2+AR3+WSC GoM	90	4288	652.6	599.4	Yes
VI	C+AR1+AR2+AR3+WSC GoM+ WSC CAR	131	3960	333.0	371.2	No
V	C+AR1+AR2+AR3+WSC GoM	90	4288			
VII	C+AR1+AR2+AR3+WSC GoM+ WSC CAR 6wl	131	3960	375.1	371.2	Yes
VIII	**C+AR1+AR2+AR3+WSC GoM+ WSC CAR 6wl+MSLPAs 4wl**	**141**	3880	128.3	101.9	Yes
IX	C+AR1+AR2+AR3+WSC GoM+ WSC CAR 6wl+MSLPAs 4wl+ ONI 6wl	142	3872	9.8	15.5	No

2.5 Results

In order to check the capabilities of the selected model and also to show the improvement brought by the different predictors, we present in Figure 10 the comparison among the historical series of LCP annual probability of occurrence (panel *a*) and the ones obtained by the simulations provided by Model V (in panel *b*), Model VII (in panel *c*) and by the finally selected Model VIII (panel *d*). In each bar of the figure, the probabilities of occurrence of the 8 LCPs are stacked for every year so the yearly sum of probabilities is one. As can be seen, the results of the selected model (panel d) compare favourably with the historical series of LCPs (panel *a*). To check how the model capabilities improve by progressively including these covariates, in panel b we show the results of Model V,

considering the three autoregressive terms and only the wind stress curl in the GoM. The model is capable of reproducing some of the variability found in the historical series but, in general, a lot of inconsistencies exist with respect to the historical series. Model VII, considering also the effect of the wind stress curl in the Caribbean Sea with a 6-week lag, has greatly improved skills to reproduce the LCP series variability. This is clear, for example, during the period 1998 – 2001 that the Model V is unable to replicate and that is instead simulated quite favourably by the Model VII. Clear improvements can be seen also for the periods 2004 – 2005, 2009 – 2010 and for the year 2014. By also including the variability provided by the MSLPAs over the North Atlantic with a 4-week lag, Model VIII is able to further expand the capability of the statistical model to reconstruct the LCPs original series, as shown in panel *d*.

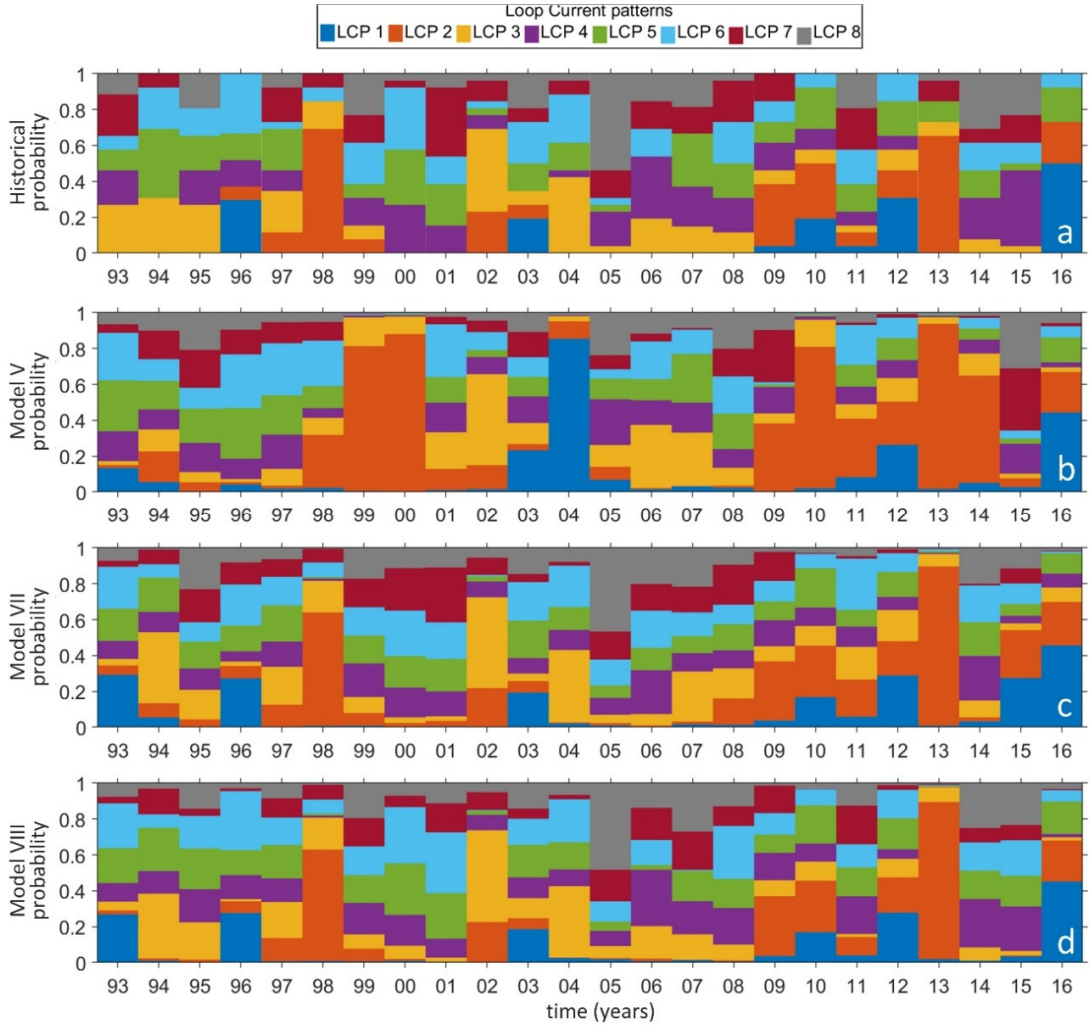

Figure 10. Panel a: Evolution of historical annual frequency of occurrence of the selected LCPs. Panel b: Evolution of annual probability of occurrence of the selected LCPs based on Model V *simulation. Panel c: Evolution of annual probability of occurrence of the selected LCPs based on Model* VII *simulation. Panel d: Evolution of annual probability of occurrence of the selected LCPs based on Model* VIII *simulation.*

To explore the intra-annual variability of the LCP evolution and to check for the skills of the selected ALR model, in Figure 11 we show the comparison between historical (left panel) and modelled (right panel) annual average probability of occurrence of LCPs at monthly scale. The main features in the intra-annual scale are replicated quite well by the model. Historical frequency of retraction stage patterns, like LCP 1 and LCP 2, are higher in

summer and autumn, with the maximum retraction (LCP 1) more likely between June and November. This feature is well reflected by the model results. Both historical and modelled LCP probabilities show that weak intrusion stages (LCP 3 and LCP 4) are common situations from autumn to early winter and in early spring, with LCP 3 more likely in autumn and LCP 4 in early spring. The historical frequency of occurrence of the intermediate stages LCP 5 and LCP 6 is higher in autumn and winter, with a minimum in late summer. Again, the model is shown to be capable of reproducing such variations. As represented by both panels, the deeper intrusions (LCP 7 and LCP 8) have a bimodal behaviour. They are more likely to occur in winter and in summer and less so in spring and autumn. The periodicity of these LCP agrees with the frequencies found by other authors for eddy shedding events, often associated with deep intrusion states (Chang and Oey, 2012; Hall and Leben, 2016). The comparison with the seasonality of the LCPs extracted by the historical data highlights the capability of the model in properly simulating the intra-annual variability as well.

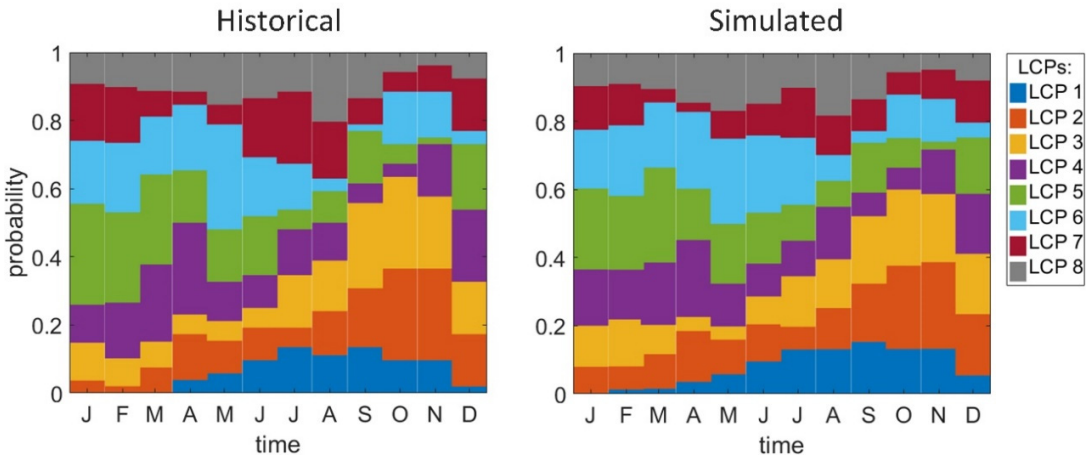

Figure 11. Left: Annual average probability of occurrence of the 8 extracted Loop Current patterns, based on historical data (left) and on the model simulation (right).

The previous analysis of the model performance suggests that the proposed ARL model is capable of reproducing the probability of occurrence of the extracted LCPs. Next, in order to quantify this capability, a series of indices generally used when assessing a statistical classifier are considered: Sensitivity, Precision, F-score and Accuracy. Sensitivity, or True Positive Rate (TPR), is the proportion of real positive cases that are correctly predicted as positive (true positives). Precision, or Positive Predictive Value (PPV) is defined as the proportion of predicted positive cases, correctly (true positives) or incorrectly (false positives), that are actually true positives. F-score can be interpreted as a weighted average of the Precision and Sensitivity. Accuracy (ACC) is the percentage of real positive and real negative correctly predicted. The values obtained for these parameters are shown in Table 2 and confirm the general good capabilities of the model (average values greater than 0.75). At the same time these parameters allow for a detailed assessment of the skills of the model in reproducing each singular LCP. The model performs better in simulating LCP 1, LCP 2 and LCP 3, corresponding with retraction and weak intrusion LC stages (see F-score, for example). In particular, TPR values equal to 1 for LCP 1 and LCP 2 remark that the model is able to find all the samples from the historical series corresponding to these two patterns. On the other side, the model weakest performance occurs with LCP 7, however with the values of the parameters greater than 0.5 in any case. ACC values are higher than 0.90 for all categories. These results confirm that the proposed ALR model is capable of simulating the LCP sequence, particularly for the patterns of LC retraction phase.

Table 2. Sensitivity (TPR), Precision (PPV) and Accuracy (ACC) for each category and average values.

index	LCP1	LCP2	LCP3	LCP4	LCP5	LCP6	LCP7	LCP8	avg
TPR	1.00	1.0	0.80	0.70	0.73	0.79	0.58	0.73	0.79
PPV	0.82	0.90	0.81	0.74	0.76	0.69	0.80	0.76	0.78
F-score	0.90	0.95	0.80	0.72	0.75	0.74	0.67	0.74	0.78
ACC	0.99	0.99	0.95	0.92	0.92	0.92	0.93	0.94	0.94

These four examined parameters are obtained from the confusion matrix. This matrix breaks down the overall model results compared to actual samples for each category. So it is useful to analyse it in detail. The confusion matrix for our case is shown in Figure 12, left panel as absolute counts and right panel as percentage. The confusion matrix relates the categories predicted by the models to the actual ones. Take LP3 (third row) for an example, the matrix indicates that the model correctly predicts the pattern (true positives) 68 times, but it misses the correct prediction 17 times (false negatives), assigning 7 times to LCP 1, once to LCP 2 and so on. An interesting feature extractable from Figure 12 is that when the model get "confused", i.e., it does not predict the true pattern, it tends to pick up some other spatial pattern not very different to the actual one. An exception to this is given by LCP 3 that, when missed by the model, is in most cases replaced by LCP 1 or LCP 8, actually quite different to LCP 3 (Figure 9).

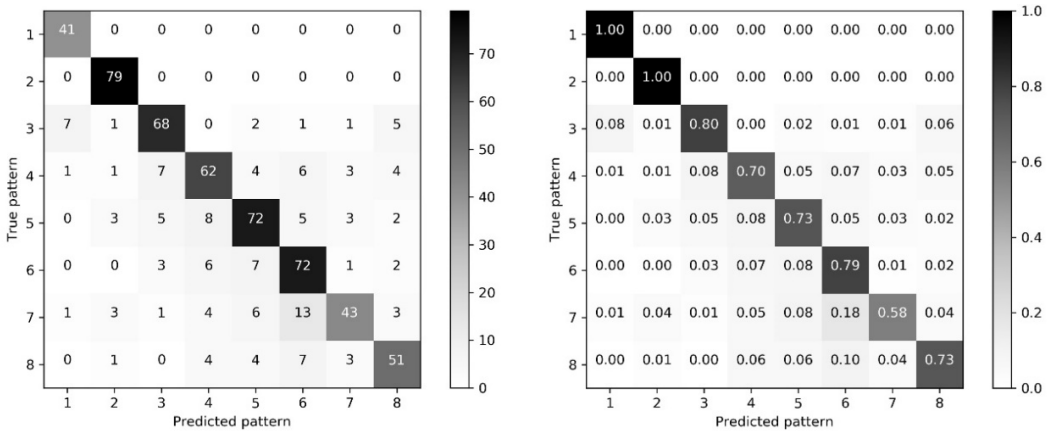

Figure 12. Original (left) and normalized (right) confusion matrix from the ALR model simulation.

2.6 Conclusions

In this chapter, a methodology for the statistical simulation of ocean patterns using a logistic regression model has been presented. The methodology has been applied to the Gulf of Mexico, where 8 representative geostrophic current patterns have been extracted. A logistic regression model has been built considering, as covariates, the wind stress curl over the Gulf of Mexico, the wind stress curl over the western Caribbean Sea with a 6-week lag, the monthly sea level pressure anomalies over the North Atlantic and three autoregressive terms. The model has demonstrated its ability to reproduce the inter-annual and intra-annual variability of the pattern sequence. Recall, Precision, F-score and Accuracy indexes have been evaluated to assess the capacity of the model obtaining, respectively, 0.79, 0.78, 0.78 y 0.94. Once the capacity of the proposed statistical model has been proved, it can be implemented in a methodology for the mid-long term prediction of oil spills in general circulation affected areas, based on the logistic regression modeling of current patterns.

CHAPTER III.

MID-LONG TERM PREDICTION OF OIL SPILLS BASED ON LOGISTIC REGRESSION MODELING OF MET-OCEAN FORCINGS. APPLICATION CASE IN THE BAY OF BISCAY[2].

3.1 Introduction

Once analyzed the statistical simulation of ocean currents for areas dominated by general circulation, the next step consists in developing statistical modeling for those regions in which ocean circulation and oil spill dynamics are also strongly affected by local wind

[2] The findings of this chapter have been published in: Chiri, H., Abascal, A.J., Castanedo, S., Medina, R., 2019b. Mid-long term oil spill forecast based on logistic regression modelling of met-ocean forcings. Mar. Pollut. Bull. 146, 962–976. https://doi.org/10.1016/j.marpolbul.2019.07.053

effect. For that, in this chapter a methodology is developed for the statistical prediction of simultaneous patterns of wind and surface currents and, taking advantage of this prediction, for the mid-long term simulations of oil spills. In order to test the capabilities of the proposed techniques, the methodology is applied in the Bay of Biscay where, in 2002, the *Prestige* tanker oil spill occurred. The obtained results are compared with the trajectories of several drifters deployed in the area after the accident in the aim of tracking the oil slick evolution.

Within the pursued general framework, main goal of the thesis, in this chapter the case of a mid-long term prediction of a surface spill is presented (Figure 6). However, a similar methodology could be used in case of a deep spill, considering a previous phase of subsurface transport modeling and applying the proposed methodology once the surface is reached. With this chapter, the first and second specific objectives of the thesis are approached.

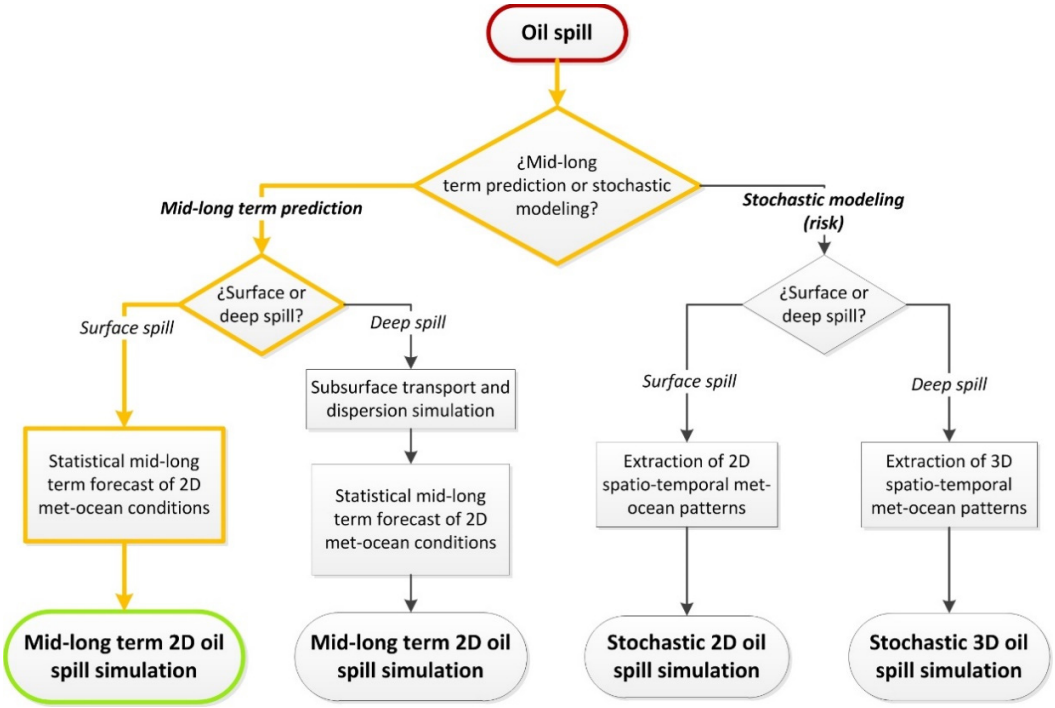

Figure 13. Flowchart of the methodology case study presented in this chapter in the framework of the general methodology pursued in the thesis.

3.2 Methodology

The proposed methodology for the mid-long term forecast of oil spill trajectories consists of two main parts (Figure 14). In the first part, a LR statistical model is used to generate *N* equally likely mid-long term forecasts of wind and current patterns. In the second part of the methodology, given a spill location, a state-of-the-art oil spill model simulates *N* equally likely oil trajectories under the wind and current conditions obtained in the previous step. The results of the methodology are presented as predicted probabilistic trajectory maps. These two main pieces of the methodology are described in detail in the following sections.

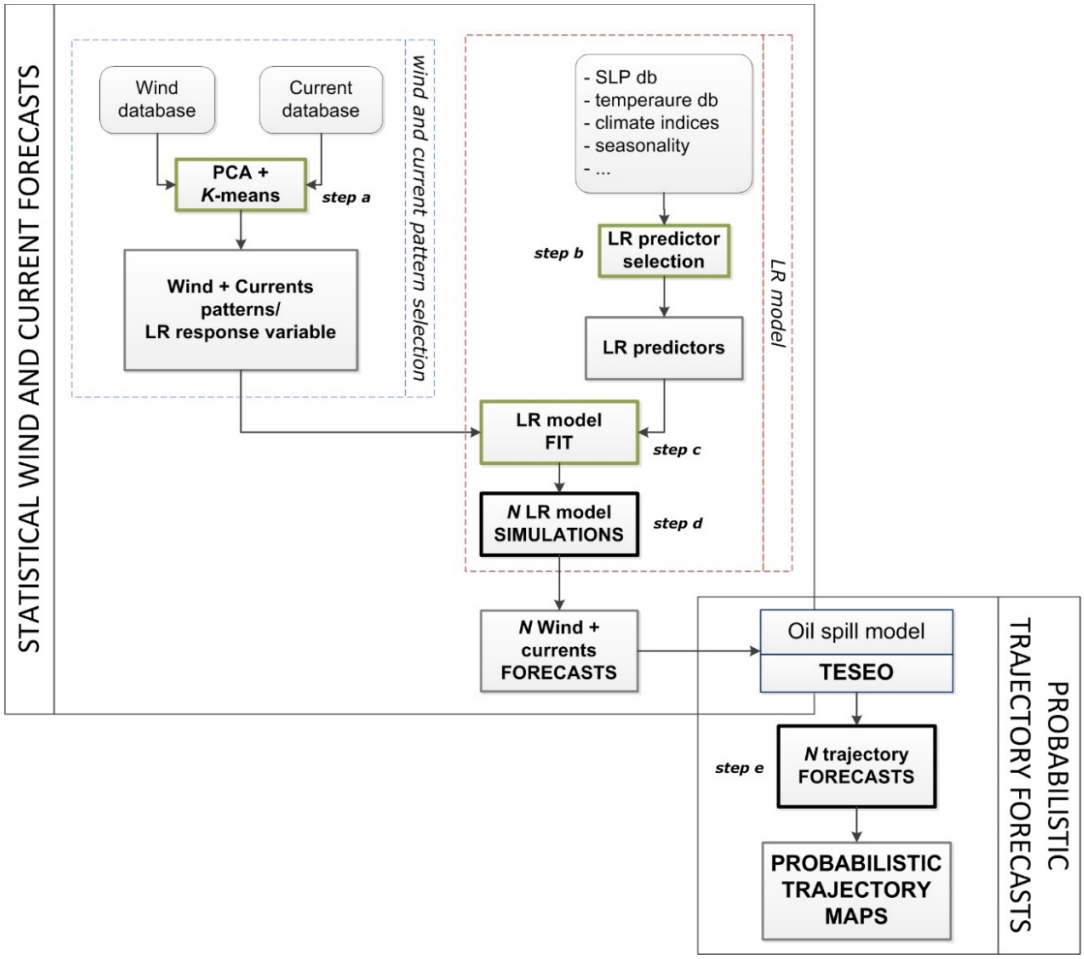

Figure 14. Proposed steps for a mid-long term prediction of oil spills.

3.2.1 Statistical wind and current forecast

3.2.1.1 Wind and current pattern selection

The initial step of the first part of the methodology consists in the selection of a set of spatial patterns of concurrent wind and current states in the study domain obtained from extended reanalysis databases (step *a* in Figure 14). These patterns describe the main variability of simultaneous wind and current conditions in the study area and will define the

categories of the LR model's response variable. Principal Component Analysis (PCA) and *k*-means clustering have demonstrated to be a convenient combination of data mining techniques for classification of atmospheric and oceanographic fields (e.g., Bárcena et al., 2015; Camus et al., 2014; Conil and Hall, 2006; Perez et al., 2015). Since the objective of this step of the methodology is to establish characteristic patterns of simultaneous wind and current conditions, PCA must be applied to a unique matrix merging both variables. If n_c and n_w are the number of numerical nodes of the considered current and wind grids, respectively, and T is the common number of time steps of both series, then the value of the u and v components of currents and wind can be expressed as Uc_i^t, Vc_i^t, Uw_j^t and Vw_j^t, respectively, where $i = 1,\ldots, n_c$, $j = 1,\ldots, n_w$ and $t = 1,\ldots, T$. The matrix to be built for the PCA application will have dimensions $\left(T \times (2n_c + 2n_w)\right)$ (Table 3). In order to give the same relative weight to the variables, they must be standardized for each numerical grid point, i.e. for each matrix column, before using PCA. Based on previous works (Antolínez et al., 2016; Camus et al., 2016), the leading principal components (PCs) obtained that explain 95% of the original data variance are kept. *K*-means is then applied to these PCs using the maximum dissimilarity algorithm for the initialization of the centroids for a deterministic classification and a better representation of the initial subsets (Camus et al., 2014). As a result of the classification process, several representative groups of concurrent wind and current conditions are identified. The number of groups to be extracted depends on any application case, mainly due to local met-ocean variability. In general, the number of patterns to be chosen should be large enough to consider the main environmental features of the study area but taking also

into account that a lower number of groups will enhance the performance of the LR model used in the next step.

Once the representative patterns are selected, the original reanalysis series of wind and current data used for pattern extraction, are projected into a single series of patterns of concurrent wind and current states, by matching each time stamp of the original series with the closest among the found centroids (Best Matching Units, BMUs).

Table 3. Current and wind data setup for the extraction of simultaneous characteristic patterns.

U components of current				V components of current				U components of wind				V components of wind			
Uc_1^1	Uc_2^1	…	$Uc_{n_c}^1$	Vc_1^1	Vc_2^1	…	$Vc_{n_c}^1$	Uw_1^1	Uw_2^1	…	$Uw_{n_w}^1$	Vw_1^1	Vw_2^1	…	$Vw_{n_w}^1$
Uc_1^2	Uc_2^2	…	$Uc_{n_c}^2$	Vc_1^2	Vc_2^2	…	$Vc_{n_c}^2$	Uw_1^2	Uw_2^2	…	$Uw_{n_w}^2$	Vw_1^2	Vw_2^2	…	$Vw_{n_w}^2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Uc_1^T	Uc_2^T	…	$Uc_{n_c}^T$	Vc_1^T	Vc_2^T	…	$Vc_{n_c}^T$	Uw_1^T	Uw_2^T	…	$Uw_{n_w}^T$	Vw_1^T	Vw_2^T	…	$Vw_{n_w}^T$

3.2.1.2 Logistic Regression model

LR is a statistical technique that has been proved to properly describe the relationship between a categorical response variable and a linear combination of predictor (categorical or continuous) variables. This statistical model links the predictor variables to the response variable through a function called *logit*, representing the natural logarithm of the odds of each possible outcome with respect to an arbitrary baseline category (Guanche et al., 2014). Based on this definition, if $Y \epsilon \{1, \dots, n\}$ is the set of n wind and current patterns, and the last category J is the baseline, a generic multinomial LR model for the j-th pattern can be expressed as follows:

$$log\left(\frac{P(Y_t = j|X_t)}{P(Y_t = J|X_t)}\right) = \alpha_j + \beta_j X_t, \qquad j = 1, \dots, n-1 \tag{5}$$

where X_t is the vector of considered predictors at time t, β_j refers to the vector of coefficients for each of the predictors at time t, and α_j represents the intercept of the model. The left-hand-side of Equation 1 represents the *logit* of a generic category of the response variable, and the right-hand-side is the linear combination of the considered predictors.

If the predictor variables are other environmental fields (e.g. sea level pressure or water temperature fields) or climate indices, they can be introduced in Equation 1 as follows:

$$X_t^C = \sum_{i=1}^{n_c} C_i \beta_i^C \tag{6}$$

where C_i is the vector of the values of the i-th considered climate index or of the i-th PC of the considered environmental field, at time t, n_c is the total number of considered PCs, climate indices, or the sum of both, and β_i^C is the vector of the associated coefficients. Besides climate indices or other environmental variables, a feature that is always worth checking as a feasible predictor when dealing with met-ocean fields is seasonality. Guanche et al. (2014) used seasonality as a predictor of a LR model to simulate sea level pressure patterns. Cardenas et al. (2017) also found seasonality beneficial for the LR modelling of wind states. Seasonality can be introduced as predictor in Equation 1 as follows:

$$X^S = \beta_0^S + \beta_1^S \cos(wt) + \beta_2^S \sin(wt) \tag{7}$$

where X_s represents the effect of seasonality in the model, $w=2\pi/T$ is the angular frequency with T set to one year and β_0^S, β_1^S and β_2^S are three coefficients associated to this predictor.

The correct selection of the model's predictors (step *b* in Figure 14) requires some knowledge of the study area, especially of the main drivers of local atmospheric and ocean circulations. Different methods such as the Likelihood Ratio test, the Wald test and the Hosmer-Lemeshow test, among others (Hosmer and Lemeshow, 2000), exist that support this selection by assessing the benefit brought by each predictor. However, a more practical way to assess predictor relevance is the cross-validation of different models, each of them including a different combination of potential predictor variables. By calculating a performance statistic for each of the analysed models, it is possible to select the one with the most powerful predictor combination. A robust metric useful to measure LR performance and commonly used to assess the skill of a classifier model is the *F*-score (Diez-Sierra and del Jesus, 2017; Powers, 2011). This index includes the recall and precision information of a model and can range between 0 and 1. The benefit of using cross-validation to select the predictors of a LR model is that it eliminates the overfitting problem, since the data used for validation are not included in the fitting stage. When the predictor variables are selected, the model is fitted, or trained, with the historical pattern series (BMU series) (step *c* in Figure 14). This is achieved by solving an optimization problem consisting in finding the values of the predictor coefficients that minimize the loss function (Fan et al., 2008).

Once the LR model is built (i.e. the relevant predictors are selected) and trained (i.e. the predictor coefficients are calibrated), future conditions of wind and currents can be predicted, in terms of the preselected patterns, conditioned to any possible value of the

predictors. LR model results are probabilistic, i.e. a probability between 0 and 100 % is assigned to each of the patterns given any predictor values. Due to its probabilistic nature, when the LR model is used to forecast time series of wind and currents, a Monte Carlo approach is implemented to simulate several (*N*) equally likely wind and current pattern sequences (step *d* in Figure 14).

3.2.2 Probabilistic trajectory forecasts

In the second part of the methodology, a state-of-the-art lagrangian oil spill model is used to predict the trajectory of the spill, forced by environmental conditions predicted in the previous step. In this case, the oil spill model TESEO (Abascal et al., 2007) was used, but any lagrangian trajectory model could be employed (step *e* in Figure 14). TESEO simulates the drift of an oil slick by tracking independent numerical particles. The surface movement of each particle is defined by the combined effect of advective and diffusive velocities as:

$$\frac{d\vec{x}_i}{dt} = \vec{u}_a(\vec{x}_i, t) + \vec{u}_d(\vec{x}_i, t) \quad (8)$$

where $\vec{u}_a(\vec{x}_i, t)$ represents the advective velocity at the *i*-th particle position at instant *t* and $\vec{u}_d(\vec{x}_i, t)$ is the diffusive velocity at the same location and at the same time. The advective velocity is generally considered as the combined effect of currents, wind stress and wave-induced drift. As indicated previously, given the relative relevance of wind and currents over waves for drift dynamics in the study area, Stokes drift was not integrated in

trajectory calculation. Thus, the advective component of the particles' velocity can be written as:

$$\vec{u}_a = \vec{u}_c + C_D \vec{u}_w \quad (9)$$

$\vec{u}_c$ being the current velocity, $\vec{u}_w$ the wind velocity 10 m above sea level and C_D the wind drag coefficient, generally varying between 2.5 and 4.4 % (ASCE, 1996). Depending on study site, the value of this coefficient should be defined based on previous calibration studies focused on the same area. Alternatively, a mean C_D value can be adopted. The diffusive component of the velocity is simulated using a Monte Carlo sampling in the range of velocities $[-\vec{u}_d, \vec{u}_d]$, assumed to be proportional to the diffusion coefficient D (Hunter et al., 1993):

$$|\vec{u}_d| = \sqrt{\frac{6D}{\Delta t}} \quad (10)$$

The diffusion coefficient D typically ranges between 1 (for closed basins) and 100 m^2/s (for open seas). The model has been previously calibrated and validated with drifter buoys in the study area (Abascal et al., 2009b, 2009a; Sotillo et al., 2008).

To consider the N equally likely met-ocean predictions obtained in the first part of the methodology, starting from a spill location, the oil spill model is run N times, one for each of the environmental forcing series previously obtained. Throughout each of the N trajectory simulations, a numerical grid (hereafter termed as probability grid) is used to calculate the number of particles laying in each of its cells and eventually to calculate probabilistic trajectory maps. Two types of probabilistic maps can be provided: instantaneous

probability maps and cumulative probability maps. In the first case, a map can be generated for every specific time during the trajectory simulation. In each instantaneous map, the shown probability is given by the percentage of particles counted in each cell of the probability grid with respect to the total particles used, for that specific time, and averaged among the *N* simulations. This kind of maps represent the search area where an oil slick is likely to be at a certain time from the beginning of the simulation. In the second case, a unique probability map provides information about the area where it is more likely for the whole oil trajectory to pass. To generate cumulative probability maps, for each of the *N* simulations, a probability of 100 is assigned to the cells where at least one particle has passed through during the whole simulation and a probability of zero to the remaining cells. The final probability assigned to each cell of the probability grid is calculated as the average probability obtained in the *N* simulations.

In order to test the performance of the proposed methodology, it was applied to predict the Prestige oil spill accident. On November 13th, 2002, during a strong storm the tanker Prestige suffered a severe structure failure when it was 28 miles offshore the North-western Spanish region of Galicia. After the accident, due to the rough weather conditions, the engine was damaged and the tanker was towed away from the coast, while the 77000 tons of heavy oil that it was carrying began to leak. After 6 days, on November 19th, the Prestige broke in half and sank at 145 nautical miles offshore. The spill then continued until February 13th, 2003, three months after the accident, when the last leaking holes were blocked. The overall affected coastline spanned from Northern Portugal, to the south, to Northern France, to the north, where the oil kept beaching several months after the accident (Daniel et al., 2004; Montero et al., 2003). During the emergency a set of drifters were deployed in the

vicinity of the oil slicks providing accurate trajectory records over a period of five months. By applying the methodology to this case study, we first predicted met-ocean patterns during the whole 5-month period of the drifter data, and next we simulated the trajectories of each of the buoys.

3.3 Study site

The Bay of Biscay is located at mid-latitudes in the Northeast Atlantic, between the west French coast and the Spanish North coast (Figure 15). The continental shelf is narrow in the South (~20 km), increasing northward on the French side, where it reaches 170 km in front of Brittany. The Bay of Biscay is in the North Atlantic Eastern Boundary system and is bounded by the North Atlantic Current to the North and the Azores Current to the South. As usual in the eastern boundary regions, the Bay of Biscay presents a poleward flow on the shelf and slope (Kersalé et al., 2016).

The primary source of large scale atmospheric variability in the area is given by the interaction of the high-pressure area called *Azores High* and the low-pressure area called *Iceland Low*, which confers a seasonal cycle to the area, with south-westerly winds during autumn and winter and north-easterly winds in spring and summer (Borja et al., 2019). This atmospheric seasonality also affects the ocean circulation, especially on the shelf and along the slope (Pingree and Le Cann, 1990). Over the abyssal plain the ocean circulation is in general very weak (1-2 cm/s) and mainly anticyclonic (Charria et al., 2013; Pingree, 1993).

Along the slope, the main oceanographic feature is represented by a poleward current, a generally weak current (5-10 cm/s) with seasonal variations along the slope (Pingree and Le Cann, 1990). This current reaches its maximum intensity during winter, passing in front of north Spain and is called Iberian Poleward Current or Navidad Current (Pingree and Le Cann, 1992a). Pingree and Le Cann, (1992b) describe how eddies of approximately 100 km in diameter (called *Swoddies* by the authors) can be induced by the instability caused by the poleward current interacting with the bottom topography.

On the Bay of Biscay's shelf, the currents are mainly the result of a combined effect of tides and wind (Pingree and Le Cann, 1989). Density driven flows are significant on the French shelf, in the vicinity of the Loire and Gironde estuaries (Lazure and Jegou, 1998). Due to the mentioned seasonal atmospheric cycles, the ocean circulation on the shelf presents a strong seasonality in its flow: equatorward in spring and summer, and poleward in winter and autumn (Charria et al., 2013; Pingree and Le Cann, 1990).

As shown by these studies, the atmospheric circulation over the Bay of Biscay has a strong influence on the ocean's circulation with more or less established seasonal variations.

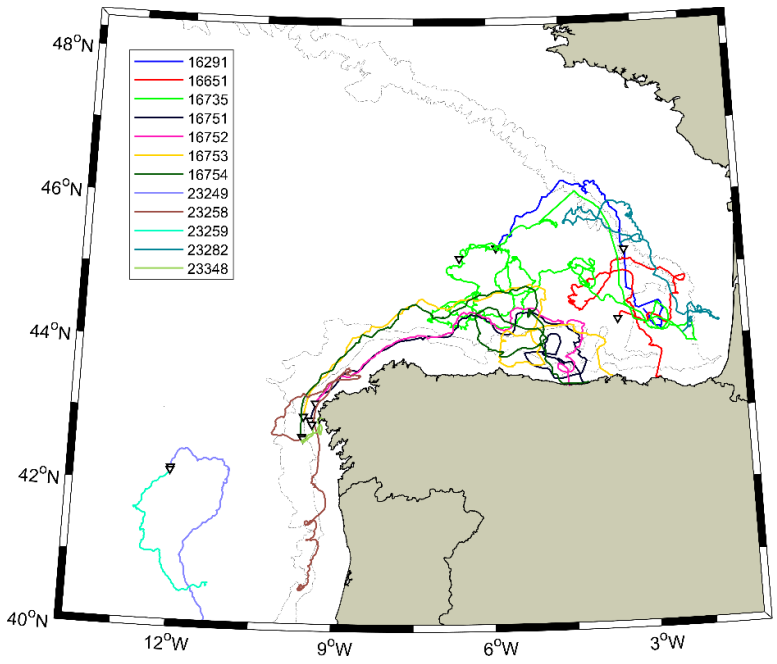

Figure 15. Study area and trajectories of the drifters used in the study. Grey lines represent Isobaths 300 m and 2500 m.

3.4 Data

3.4.1 Currents

Ocean current data for this application were obtained by CMEMS' GLORYS12v1 (Lellouche et al., 2018). This is a numerical ocean global reanalysis covering 24 years (1993 – 2016), with a 1/12º (approximatively 8 km) resolution, 50 vertical levels and a daily frequency. It was generated with the NEMO platform, driven at the surface by ECMWF ERA-Interim reanalysis. Along track altimeter data (Sea Level Anomaly), satellite Sea Surface Temperature, Sea Ice Concentration and in situ temperature and salinity vertical

profiles are assimilated jointly. Current zonal and meridional velocities are provided with a daily mean frequency. In this study, surface currents were considered, coming from the first vertical depth of the dataset, fixed at a 0.5 m depth. Current data were cut to the study's domain [-14º, -1º] longitude and [40º, 48.5º] latitude, for a total of 16171 (103 x 157) GLORYS12v1 grid points.

3.4.2 Wind

In order to reduce inaccuracies in final trajectory results caused by divergences among the wind and current databases used, we chose to take wind data from the same atmospheric dataset used by CMEMS to force the GLORYS12v1 reanalysis, i.e. the ECMWF ERA-Interim reanalysis (Dee et al., 2011), which is the latest global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecast (ECMWF). This reanalysis spans from 1979 to present with a daily mean frequency and is updated on a real-time basis. The ERA-Interim improves several aspects with respect to the previous ERA-40 reanalysis (Uppala et al., 2005), such as the incorporation of a four-dimension variational data assimilation approach (4D-Var), an increased spatial resolution up to 80 km, and improved model physics. The spatial resolution of the data set is approximately 80 km on 60 vertical levels, from the surface up to 0.1 hPa. The same domain limit as that of the current data has been considered.

3.4.3 Drifters

During the Prestige accident, a set of 12 satellite-tracked drifters was deployed to collect data of oil slick paths. Most of them were deployed in front of Galicia, where the accident occurred, in order to track the Prestige's main spill (Caballero et al., 2008; Daniel et al., 2004; Montero et al., 2003). Two different models of drifter were used, the SC40 and the PTR models. The SC40 buoys used consisted of a spherical main body of 40 cm in diameter, an ARGO transmitter and four pieces of lead added in order to reduce their buoyancy, thereby increasing their efficiency in tracking the spills (García-Ladona et al., 2005). The PTR buoy model's main body had a height of 30 cm and a square 10 cm x 10 cm section, as well as an ARGO satellite-based tracking system. Once deployed in water, the submerged part of this drifter is approximately 20 cm (Girin et al., 2006). Detailed information on each drifter, including their deploying position and date, drifting time and buoy model is shown in Table 4. A first group of four drifters was released near the Galician coast, one month after the accident, to track the so-called "main spill". These drifters moved toward the North-East, entering the Bay of Biscay and reaching the central Spanish coast after 1 – 1.5 months. Another group of five buoys was deployed between January 2nd and January 27th, 2003. One of them was deployed in the middle of the Bay and drifted inside it during 1.5 months without reaching the coast. Two of them were deployed next to the Galician coast and another two where the tanker sank. The last group of buoys was deployed in the framework of an AZTI-CEDRE (French Centre of Documentation, Research and Experimentation on Accidental Water Pollution) cooperation (Girin et al., 2004; Gouriou et al., 2004). This group was composed of four surface drifters of the PTR model type. As

shown in Table 4, the overall drifter data time covers from December 19th, 2002 to May 19th, 2003, i.e. exactly 5 months.

Table 4. Summary of buoys deployed during the Prestige accident.

Drifter ID	Drifter model	Institution	Initial position (ºlon, ºlat)	Initial date	Final date	Total time (days)
16291	PTR	AZTI/CEDRE	-5.868, 45.311	15-Jan-2003	08-Feb-2003	24
16651	PTR	AZTI/CEDRE	-3.528, 44.291	27-Dec-2002	02-Feb-2003	38
16735	PTR	AZTI/CEDRE	-6.593, 45.175	29-Dec-2002	19-May-2003	141
16751	SC40	LIM(UPC)[1]/ICM(CSIC)[2]	-9.447, 42.915	19-Dec-2002	30-Jan-2003	42
16752	SC40	LIM(UPC)/ICM(CSIC)	-9.356, 43.155	19-Dec-2002	19-Jan-2003	31
16753	SC40	LIM(UPC)/ICM(CSIC)	-9.581, 42.969	19-Dec-2002	30-Jan-2003	42
16754	SC40	LIM(UPC)/ICM(CSIC)	-9.604, 42.688	19-Dec-2002	01-Feb-2003	44
23249	SC40	ICM(CSIC)	-12.046, 42.207	16-Jan-2003	19-Feb-2003	34
23258	SC40	ICM(CSIC)	-9.580, 42.662	11-Jan-2003	19-Feb-2003	39
23259	SC40	ICM(CSIC)	-12.054, 42.174	27-Jan-2003	19-Feb-2003	23
23282	SC40	ICM(CSIC)	-3.350, 45.249	02-Jan-2003	18-Feb-2003	47
23348	SC40	ICM(CSIC)	-9.416, 42.861	11-Jan-2003	25-Jan-2003	14

[1] Maritime Engineering Laboratory of the Technical University of Cataluña.

[2] Marine Research Institute (National Spanish Research Council).

3.5 Statistical wind and current forecasts in the Bay of Biscay

Following, the first block of the proposed methodology is applied to obtain statistical forecasts of wind and current states in the study area for the period December 2002 - May 2003.

3.5.1 Wind and current pattern selection

The initial step of the methodology is the selection of a set of simultaneous wind and current states in the study area. First, the wind database, covering a larger time period, is cut to the time frame covered by the current database (1993-2016). PCA is then applied to the 24-year daily wind and currents datasets. To do so, meridional and zonal components of both datasets are previously merged in a unique matrix and standardized. As a result of applying PCA, the dimensionality of the original data is reduced to 376 PCs, explaining 95% of the total variance of the original datasets.

K-mean clustering is then applied to these leading PCs. A sensibility analysis is carried out to explore the variability represented by the obtained pattern when varying the number of groups between 8 and 80. If the local variability of wind and currents is better described with a higher number of groups, it must be considered that a lower number of categories of the response variable will enhance the effectivity of the LR model to be applied next. In this case, as a result of the analysis, it is found that with 18 groups were enough to represent all the main features of the atmospheric and ocean circulation in the study area (Figure 16). In Figure 16 red arrows represent the wind, while black arrows and the background color maps refer to surface currents. Patterns P1, P5, P9, P12 and P16 reflect wind and current conditions characteristic of autumn and winter seasons (Charria et al., 2013), with westerly and south-westerly winds and an intensified eastward current over the Spanish shelf that may be related to the Iberian Poleward Current. Each pattern shows some differences in direction or speed of both wind and currents. The current over the shelf flows in the opposite direction in P4, P15 and P17 in which north to north-easterly winds

predominate, consistent with met-ocean conditions typical of summer and spring in the Bay of Biscay (Charria et al., 2013; Solabarrieta et al., 2015). Four patterns show calm wind conditions over the whole considered domain, matched by weak current velocities (P3, P10, P11 and P18). The remaining patterns represent intermediate and varied wind and current conditions that don't have a clear seasonal preference and have an influence on the interannual variability of the Bay's dynamics.

Finally, the wind and current pattern closest to the actual conditions of that day is assigned to each time stamp of the 24-year daily series (from 1993 to 2016), and the BMU series is obtained.

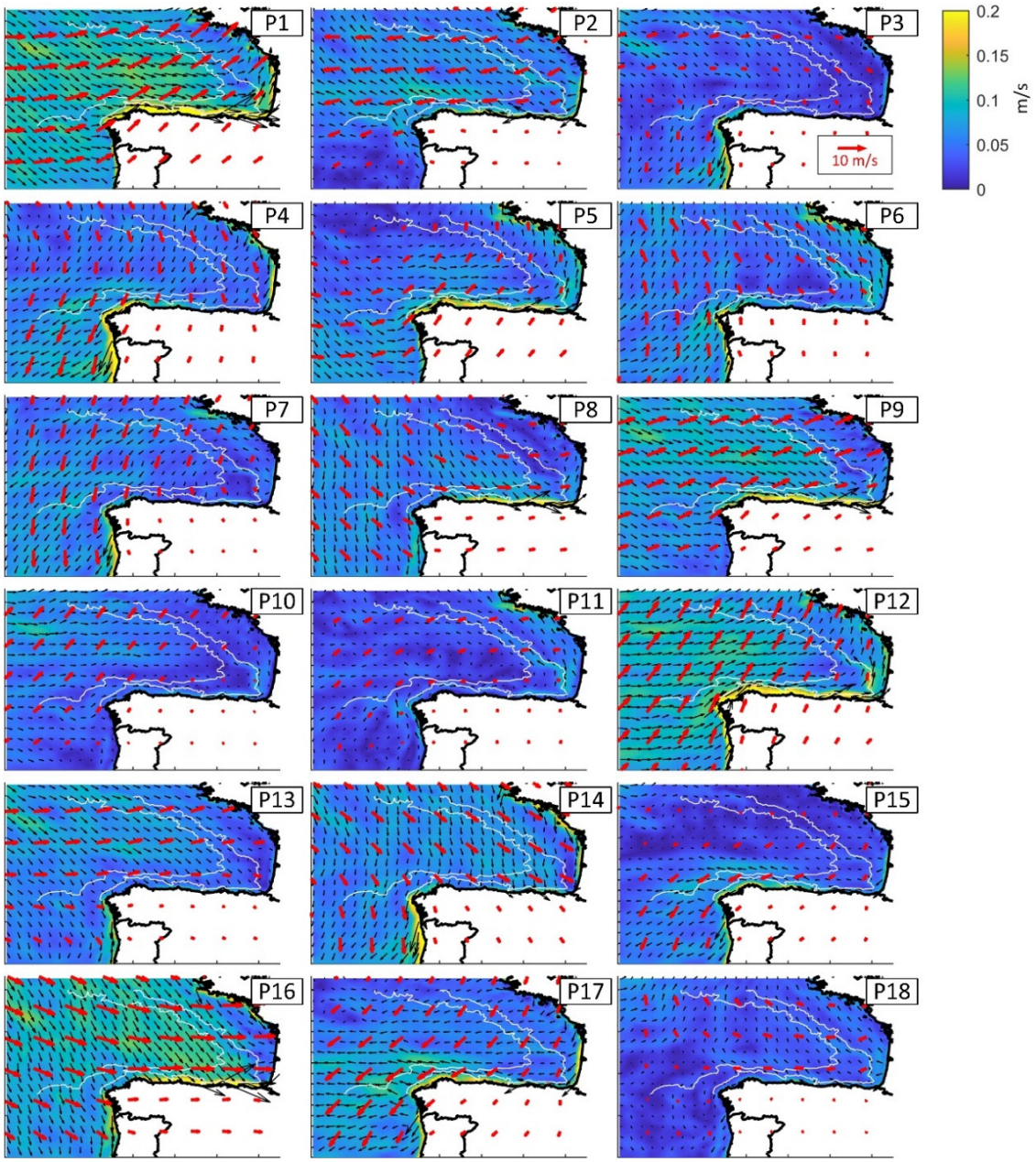

Figure 16. K-means extracted daily patterns of wind and surface currents. Red vectors represent wind, while black vectors and background color maps refer to surface currents. Distances among vectors do not represent the actual resolution of the data: for clarity's sake only the data of one of every two wind grid points and of one for every six current grid points are represented.

3.5.2 Logistic Regression model

In order to build the LR model for the simulation of wind and current patterns over the Bay of Biscay, the potential effect of several variables were tested: sea salinity, sea surface temperature (SST), sea level pressure (SLP), SLP anomalies, SLP gradients, all of them considered in the study area and in a wider domain including the whole North Atlantic, several climate indices and seasonality. All these variables were used individually and in all possible combinations as predictors of a LR model. As mentioned in Section 3.2.1.2, the most efficient way to select among a group of LR models is to cross-validate them and choose the one which performs better. In this case the *k*-fold cross-validation technique was used (Gutiérrez et al., 2013). The historical 24- year series was split into 24 one-year segments, or folds. Iteratively, one fold was held out for validation while data from the remaining 23 years were used to train the model. For each of the 24 iterations the *F*-score metric was calculated and the average of these 24 scores represents the performance ability of each model. Table 5 shows the average *F*-score obtained for some of the considered predictors or combination of predictors. As shown, very low values of the average *F*-score were achieved when considering as predictors any of SST, salinity (both over the study domain and over the whole North Atlantic), seasonality and some of the climate indices with a higher impact in the study area (cases 1 to 7). This results suggest that these features do not have a major influence on the model's fit. A significantly higher average *F*-score (0.34) was achieved for case 7 when the SLP over the North Atlantic is considered as predictor of the LR model, meaning that the variability of the considered pattern evolution is partly related to the variability in this variable. The average *F*-score further increases when using the local Bay of Biscay's SLP as

predictor (0.48, case 9), suggesting that local SLP variability has a more direct implication in the evolution of the considered patterns, compared with the larger scale SLP effect. To further explore the local SLP potential for the LR model, the SLP anomalies with respect to daily SLP mean values and spatial gradients were considered, with the result that the SLP gradients increased the model average *F*-score to 0.51 (case 11). The scores obtained in cases 12 and 13 show that the performance of the LR model does not improve when considering the local salinity (case 12) or the EA index (case13) together with the SLP gradients in the study domain. The highest average *F*-score (0.52) was obtained with the combination of predictors including the SLP gradients over the Bay of Biscay and seasonality (case 14). This combination was then used to force the LR model in order to predict future patterns of wind and surface currents.

Table 5. Evaluation of the performance of different predictors and predictor combinations in the LR model of wind and current patterns.

Case	**Predictor variables**	**Average *F*-score**
1	North Atlantic SST	0.06
2	North Atlantic Salinity	0.08
3	Bay of Biscay SST	0.07
4	Bay of Biscay Salinity	0.10
5	Seasonality	0.06
6	NAO index	0.02
7	EA index	0.03
8	North Atlantic SLP	0.34
9	Bay of Biscay SLP	0.48
10	Bay of Biscay SLP Anomalies	0.38
11	Bay of Biscay SLP Gradients	0.51
12	Bay of Biscay SLP Gradients + Bay of Biscay Salinity	0.50
13	Bay of Biscay SLP Gradients + EA index	0.50
14	Bay of Biscay SLP Gradients + Seasonality	0.52

In order to assess the general performance of the defined model, the confusion matrix obtained by comparing the actual pattern series (the BMU series) and the series of most probable patterns simulated with the defined LR model is represented (Figure 17). When read row-wise, this matrix indicates how many times the model correctly predicts a pattern (principal diagonal) and which pattern is predicted when the true one is missed. For example, reading the first row, the confusion matrix indicates that pattern P1 is correctly predicted by the model 50 % of the times (with respect to the total instances in which pattern P1 occurs in the historical series). Among the remaining 50 % of the times in which the model doesn't predict pattern P1, it wrongly predicts pattern P9 20 % of the times, and P5, P8 and P12 10 % of the times, respectively. Thus, as shown by the matrix in Figure 17, the model performs better when simulating pattern P2, and has a lower precision when dealing with patterns P3, P7, P11 and P18, that represent calm situations. This could indicate that the LR model performs better when not dealing with calm conditions. Despite this, the model's capabilities are in general good since the higher percentages lay always on the principal diagonal. The uncertainties found may be related to large scale effects or other predictors not included in the model. It should be highlighted that when the model fails to predict a true pattern, in general the wrongly predicted pattern tends to be similar to the real one. For example, the first row shows that when the model misses pattern P1, most of the times it predicts P9 or P12. As commented in Section 3.4.1, these three spatial patterns represent conditions characteristic of the winter season with south-westerly winds and an intense eastward current over the Spanish shelf. At this point it is logical to think that if fewer patterns were extracted, representing hence more distinctive conditions, the LR model performance would improve. Even if this seems to be the case, as mentioned in Section 3.4.1, it is also important to use a

number of patterns large enough to gather principal variations in the main local environmental conditions. Indeed, these slight differences among similar patterns can produce substantial differences in the final trajectory calculations.

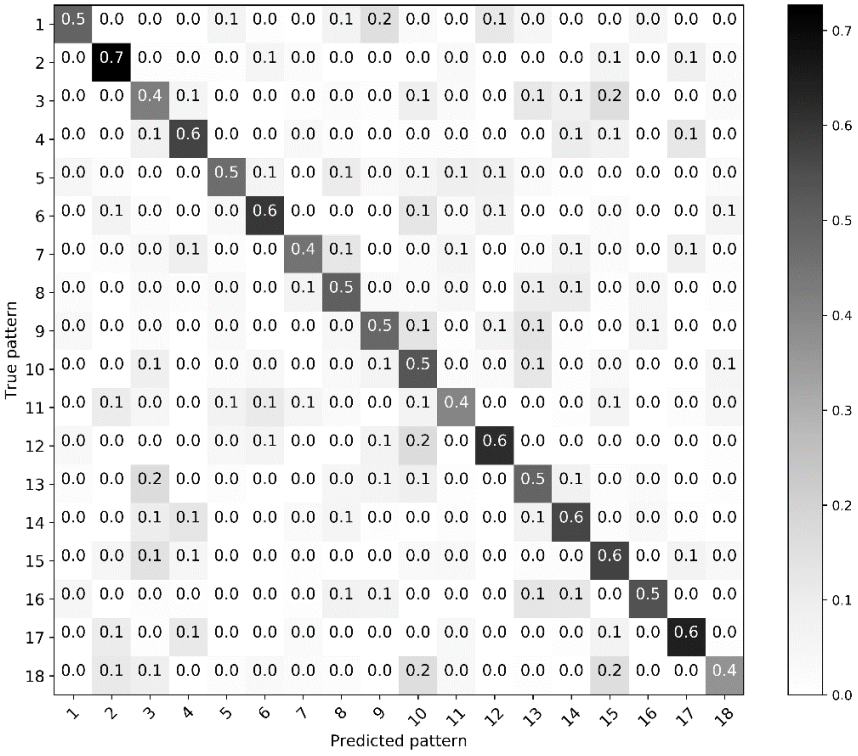

Figure 17. Confusion matrix of the defined Logistic Regression model for wind and current patter simulations.

Once the LR model has been fitted and its capability to simulate the historical pattern series is proved, it can be used to simulate future values. As shown in Table 4, the first buoy was deployed on December 19th, 2002 and the last one was collected on May 19th, 2003, for an overall buoy drifting period of 5 months. We used the defined LR model to forecast daily wind and current patterns for these 5 months. This period lays inside the overall period used

to build and fit the LR model (1993-2016). To avoid this lack of independence between the forecast period and the calibration data, we recalibrated the model by using only the response and predictor data outside this time frame.

To consider the probabilistic nature of the LR model, 100 independent simulations of the model were executed. As a result, 100 equally likely predictions of daily wind and currents were generated, in terms of the 18 considered patterns.

As an example, in Figure 18, results of the 100 LR model simulations are shown as histograms for the first 9 days of prediction. For each day, the actual pattern is indicated in brackets in the subplot title and also highlighted in bold in the X-axis. The Y-axis shows the probabilities predicted for each pattern. As shown, during the first day of simulation (December 19th, 2002) the actual conditions in the Bay of Biscay were identified by pattern P6 (Figure 16). Considering the 100 independent simulations for this day, in over 50% of them, the predicted pattern was indeed P6. The day after, the conditions changed to P12, with intense south-westerly winds and an intensification of the shelf current. In this case, the pattern most commonly simulated was P10, which also corresponds to south-westerly winds, though less intense. A high probability, however, corresponded to P12. During the next six days, until December 26th, conditions remained similar, with a persistent P12, typical of winter. The model's predictions correctly selected pattern P12 as the most likely until December 25th. On December 26th in approximately 50 of the 100 simulations the predicted pattern was P1 which also represents south-westerly winds and eastward currents, but more intense (Figure 16). On December 27th the conditions became calmer, but still with some residual south-westerly winds and a less intense poleward current, identified by pattern P5. As shown in Figure 18, the LR prediction was able to capture this change. An *F*-score of 0.47

was obtained for the total predicted 5-month period, quite a positive value considering the high number of categories included in the analysis (18).

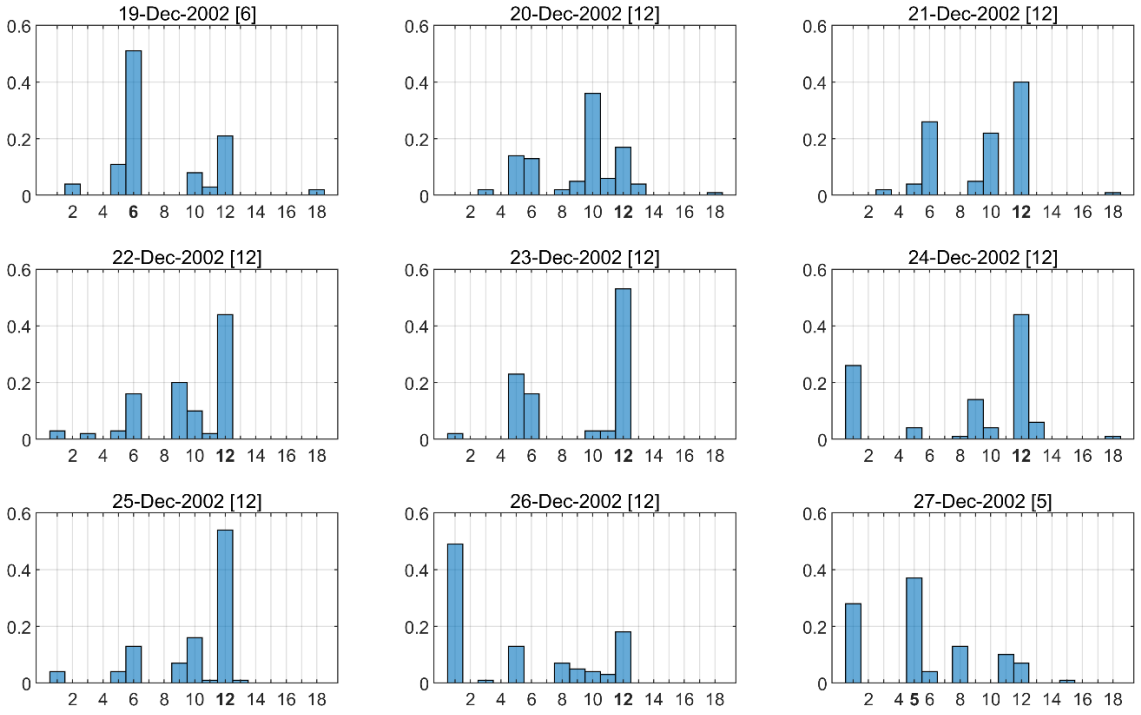

Figure 18. Histograms of pattern prediction based on 100 LR model simulations for the first 9 days of the 5-month forecast period. The true pattern is indicated in brackets in each plot title and in bold in the X-axes.

3.6 Probabilistic trajectory forecasts

At this point in the methodology, the statistical prediction of wind and surface current patterns was available for the 5-month period from December 19th, 2002 to May 19th, 2003. Next, the oil spill model, TESEO, was used to simulate the trajectories of the drifters in Table 4. Starting from the initial position of each drifter and using the predicted met-ocean patterns, TESEO was run to simulate drift trajectories for the drifting period of each buoy. Note that

due to the proximity of their initial positions and the fact that their drifting periods started on the same date, the two groups of drifters 16751, 16752, 16753, 16754 and 23258, 23348, respectively, were simulated with individual TESEO runs. The TESEO diffusion coefficient D was set to 100 m^2/s, since it is an open sea application, and the wind drag coefficient C_D to 0.025 (Abascal et al., 2009b). The numerical map used for probability calculations has a 0.25º resolution, resulting from a compromise between computational cost and the quality of results visualization.

The cumulative probability maps calculated for all the buoys are shown in Figure 19. In each panel of Figure 19 the actual drifter paths are plotted as black continuous lines. Although it is not accurate to compare a simulated probabilistic trajectory map with an observed one, such comparison is still valuable and necessary to evaluate the methodology's results. In general, drifter trajectories were in good agreement with the probabilistic maps obtained. As visible in all panels, the actual drifter trajectories lay inside the simulated trajectory maps. Only a very minor fraction of actual tracks in panels *b*, *d* and *h* lays outside the maps. In many cases the observational track lay near the core of the simulated maps, where the estimated probability is highest (panels *a*, *d*, *e*, *f* and *g*). Moreover, the only two cases in which more than one drifter was deployed at the same time and location, shown in panels *d* and *f*, demonstrate how the spread of the computed maps is adequate to take into account the separation among trajectories expectable in a real situation. Regarding these two simulations, it must also be highlighted that, starting so near the coast, the calculated probabilities are affected by the number of particles that stuck to the coast during the first phase of the trajectory simulations. The probabilistic trajectory map represented in panel *c*, correspond to a trajectory simulation period of about 4.5 months (Table 4). During this

period, the deployed buoy drifted following a highly variable path, covering a long distance and a wide portion of the Bay of Biscay. Such an irregular and variable situation is reflected in the probability map that presents a particularly wide area as the likely region crossed by the trajectory. In spite of this complicated setting, the actual path lays inside the contours of the obtained map. Similar situations but for shorter time horizons are shown in panels *b* and *h*, where again the highly irregular drifter track is located inside the forecasted probabilistic map.

To show some results of instantaneous trajectory maps, Figure 20 shows six maps corresponding to six different instants, each taken on a 4 day interval basis, during the drifting period of buoy 16291. In this case, the results are represented as search areas based on the obtained probabilities. The contour plots enclose those cells of the probability grid containing the percentage of numerical particles represented by the contour line labels. In order to consider the results of the 100 equally likely simulations, the values shown in the diagrams refer to the percentage of particles averaged over the 100 computed simulations. The actual trajectory of the drifter is shown as a black continuous line, the initial point is shown as a white triangle and the location of the drifter at each considered time is represented as a red circle. As shown, during the whole simulation period (24 days) the drifter's position is always inside the computed search areas. After an approximately 10-day drift towards the Northeast (panels *a*, *b* and *c*), the buoy made a turn of approximately 90º to its right heading towards the Eastern coast of Spain (panels *d*, *e* and *f*). Note that the simulated search area also shifts accordingly, containing in all cases the drifter's observed position. In order to quantify the capability of the model to provide probability areas that actually contain the observed drifter locations, the lagrangian separation between the actual drifter locations and the modelled

probabilistic maps was analyzed for each drifter in Table 4. An index widely used in the literature for the evaluation of numerical trajectory model results is the skill score (*ss*) proposed by Liu and Weisberg (2011) (e.g., Janeiro et al., 2014; Phillipson and Toumi, 2017; Röhrs et al., 2012). According to Liu and Weisberg (2011) a non-dimensional separation distance *s* can be defined as:

$$s = \frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} l_{0i}} \tag{11}$$

where d_i is the separation distance between the observed endpoints and the simulated positions, corresponding to drifter trajectory segment *i*, *N* is the number of drifter trajectory segments and l_{0i} is the length of the *i*-th trajectory segment. In this case, the separation distance d_i was calculated between drifter locations and the corresponding simulated search areas, so it was considered equal to zero if the drifter location lay inside the modelled area. The simulation skill score is then defined as:

$$ss = \begin{cases} 1 - \frac{s}{n}, & (s \leq n) \\ 0, & (s > n) \end{cases} \tag{12}$$

Where n is a tolerance threshold. In this work, as suggested by Liu and Weisberg (2011), *n* was set equal to 1, i.e. the normalized separation distance should not be larger than the considered trajectory length. Table 6 presents the obtained *ss* for each drifter used in the study, considering the simulated areas corresponding to the 0.1% and 2% contour lines (Figure 19). As shown, the model performs well for all the 12 considered buoys. With respect to the 0.1% contour line the *ss* was greater than or equal to 0.9 for all cases. When considering the 2% contour plot, the *ss* still depicts an overall good performance of the model, being in

all cases larger than 0.5. The lowest values of the *ss* score (between 0.57 and 0.69) were obtained for the group of four drifters that were deployed on December 19th, 2002 in front of the western Galician coast. This could be related to the curve suffered by these drifter trajectories toward the central Spanish coast, which wasn't properly reflected by the model.

Table 6. Liu and Weisberg (2011) ss score considering the distance between the drifter's trajectories and the simulated areas with probability 0.1% and 2%.

Drifter	**ss 0.1%**	**ss 2%**
16291	1	0,96
16651	0,98	0,87
16735	0,95	0,82
16751	0,94	0,60
16752	0,90	0,57
16753	0,98	0,64
16754	0,99	0,69
23249	1	0,99
23258	0,97	0,83
23259	0,99	0,92
23282	0,96	0,88
23348	1	0,99

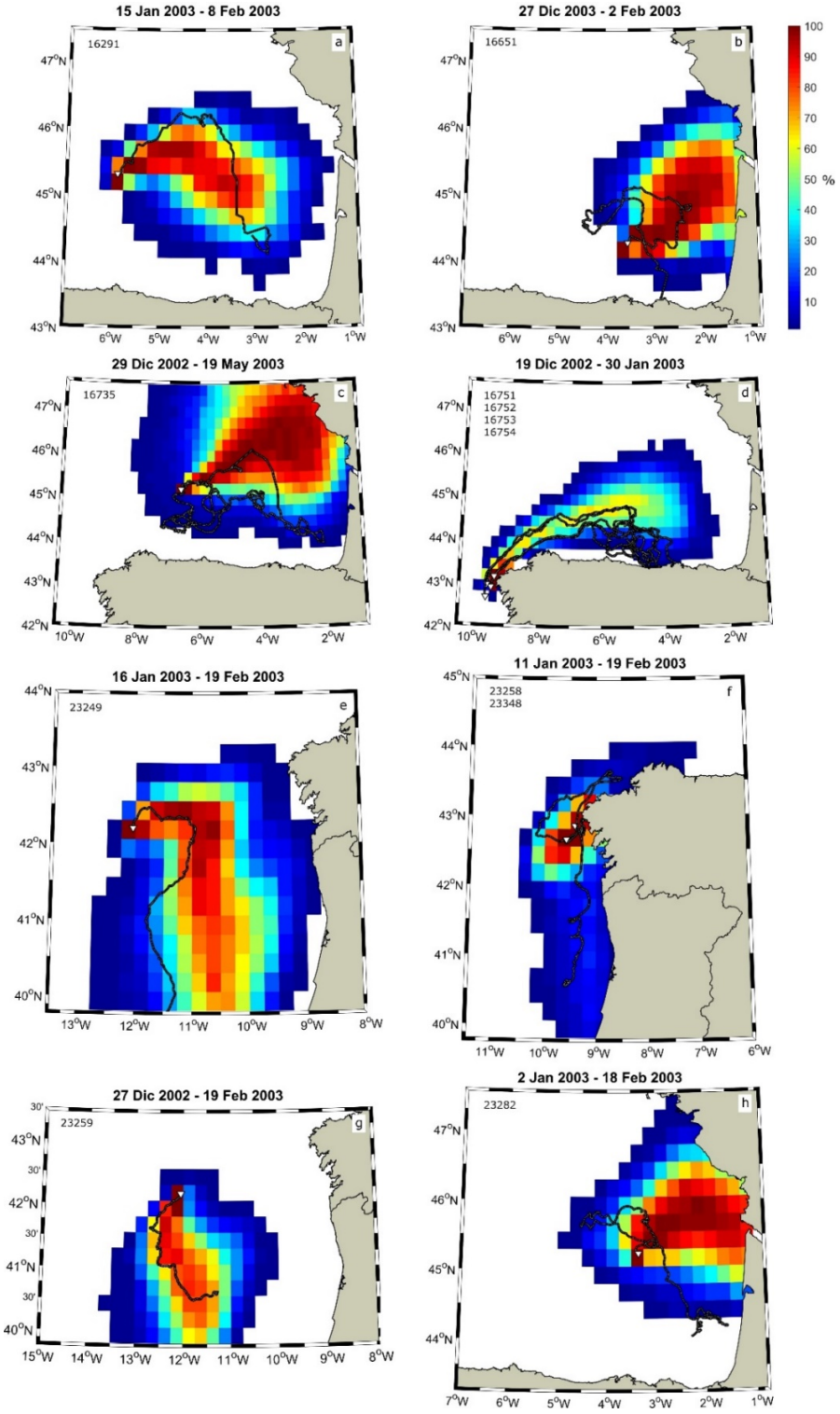

Figure 19. Cumulative probability maps compared to actual drifter trajectories.

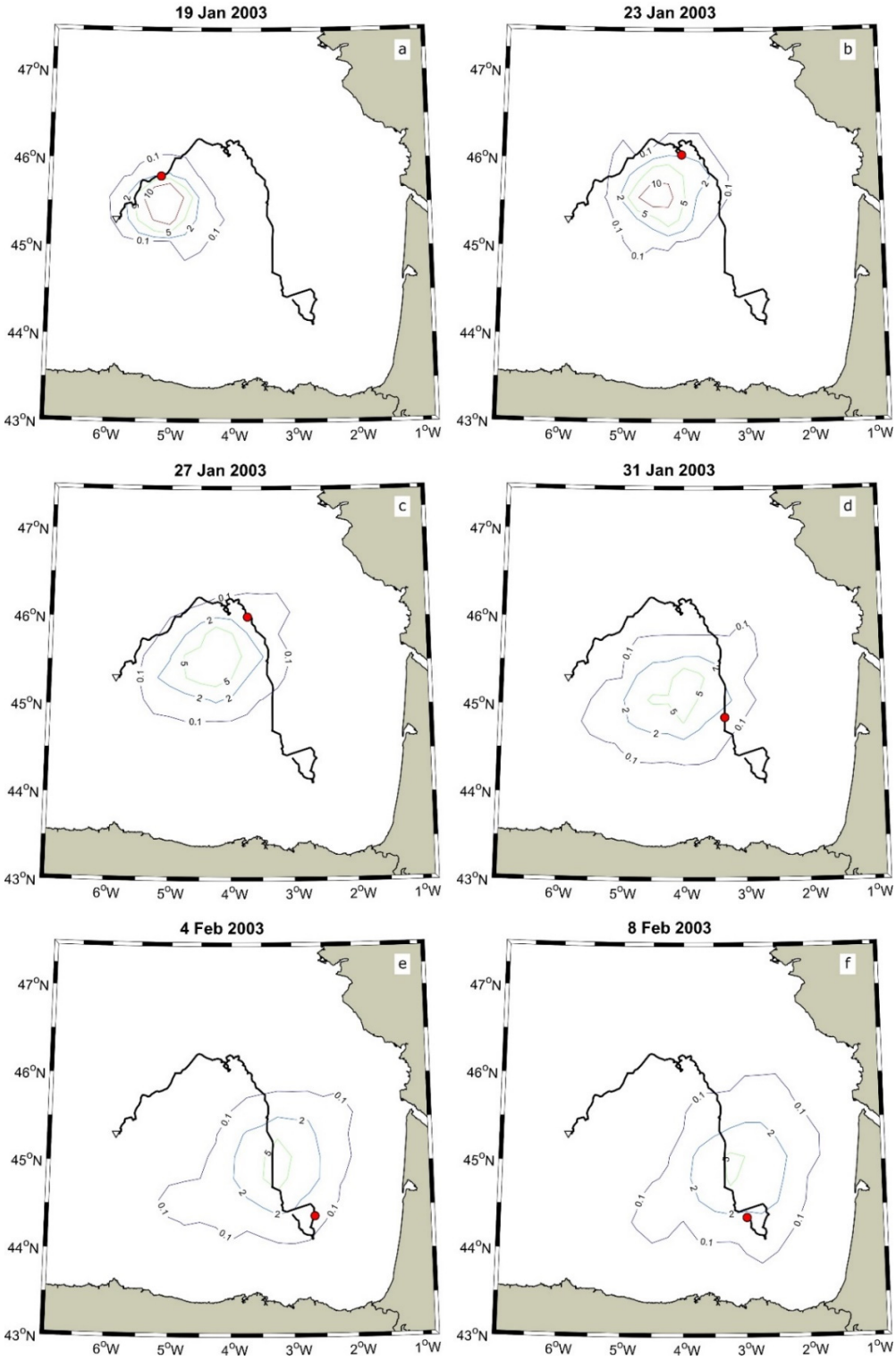

Figure 20. Instantaneous probability maps compared with drifter 16291's track. The contour plots shown represent the density (as percentages) of the simulated particles at each time interval.

3.7 Discussion

The statistical simulation of oil trajectories has been a common approach in the oil spill modelling field during the last decades. In particular, the rise of publicly available numerical reanalyses of currents and wind has enhanced the number of studies that, taking advantage of the large amount of historical met-ocean information, assess oil spill related hazard using a statistical approach, frequently focusing on mid-long term simulations. Thus, it is relevant to analyze in detail differences and benefits of the proposed methodology with respect to the existing capabilities of the state-of-the-art approaches. In order to directly compare our methodology with other statistical techniques for oil spill trajectory simulations, we focused here on previous studies centered in the Bay of Biscay. Abascal et al. (2010) used a Monte Carlo approach to select initial dates from a 44-year long met-ocean reanalysis. Considering winter and summer seasons separately, 200 initial dates were randomly selected for each season. Starting from those dates, the authors considered the environmental evolution over the following 30 days based on the available historical databases. These 400 met-ocean scenarios (200 for each of the two considered seasons) randomly extracted from the original reanalysis databases were used to force a lagrangian oil spill model in order to evaluate the oil spill hazard over the whole Bay of Biscay considering several hypothetical spill sites. Note that this approach, with minor variations, has been traditionally used in many other oil spill hazard studies (e.g., Lee and Jung, 2015; Liubartseva et al., 2015; Romero et al., 2013). In order to check the performance of their proposed model, Abascal et al. (2010) compared their results with actual drifter trajectories. Five of the considered drifter trajectories matched some of the ones we used in our work, i.e. drifters 16751, 16752, 16753,

16754 and 23258. This allowed for a direct comparison between the performance of a traditionally applied technique for the oil spill hazard assessments and of our mid-long term prediction methodology. In Abascal et al. (2010), since these five buoys were deployed during winter months, all of them were simulated using the 200 met-ocean scenarios corresponding to the winter season. Moreover, having been deployed approximately at the same location, a unique initial point was considered for the lagrangian simulations. After 30 days of simulation, the authors achieved a probabilistic map showing the probability of contamination over the study domain. The region with a higher risk was clearly identified around the western coast of the North Iberian Peninsula, corresponding to the area covered by the observed trajectory of drifter 23258. However, the region of the southern Bay of Biscay, where four out of the five considered drifters went, was estimated to have a low probability of contamination (less than 5%). On the contrary, as shown in Figure 19d, the probability obtained while simulating the trajectories of the same drifters 16751, 16752, 16753 and 16754 with the methodology proposed in this chapter is in a very good agreement with the observed trajectories. The main movement of the buoys towards the inner Bay of Biscay was effectively reproduced by our model's results, with the four trajectories laying inside a region with a predicted probability higher than 50%. Then, after the drifters turned to the southeast and started to move erratically towards the Spanish coast, they are still inside the estimated area although the probability varies between 40% and 10%. It is important to emphasize however, that the goal of Abascal et al. (2010), similarly to any other oil spill hazard assessment, was to provide general pollution probabilities related to representative met-ocean conditions in their considered study area. Conversely, our methodology aimed at predicting the contamination probabilities specifically related to those met-ocean conditions

expected for the considered simulation period and hence it better represents the evolution of the spill during this specific period, as shown by the comparison with the actual drifter's trajectories.

In order to highlight the relevance of the simulation's specific time frame in our methodology and to show how the proposed technique could be useful for an oil spill responder, a numerical experiment was conducted simulating two spills occurring at the same location but at different dates. As an example, the deploying location of drifter 16291 was considered as spill site, while the periods from January 15th to February 8th, 2002 and January 15th to February 8th, 2005 were established as simulation time frames (i.e. the period corresponding to drifter's 16291 trajectory but in different years). Results show (Figure 21) how in the first case (panel *a*) the expected trajectory would likely head northeast towards the central-northern coast of France while in the second case (panel *b*), the probabilistic trajectory map after 24 days of simulation followed a clockwise rotation heading towards the western coast of Spain, in a totally opposite direction than the one shown in panel *a*. As shown, in case of a real spill, the results provided by the proposed methodology would provide a more valuable information for a mid-long term spill response, compared to the one commonly provided by traditional oil spill hazard systems.

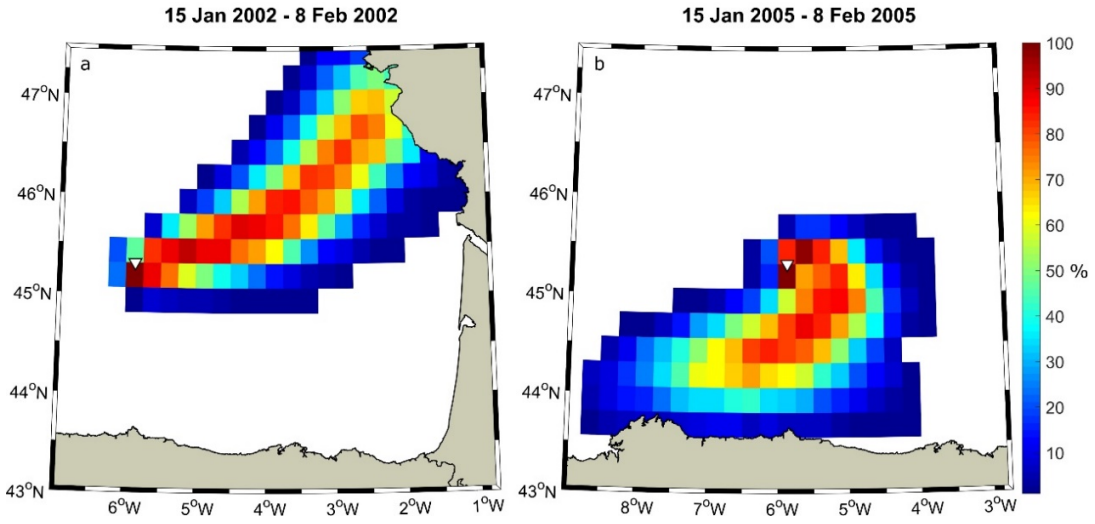

Figure 21. Cumulative probability maps obtained using drifter's16291 deploying position (white triangle) as initial location and two different simulation time frames: January 15th to February 8th, 2002 (panel a) and January 15th to February 8th, 2005 (panel b).

3.8 Conclusions

In this chapter a methodology for the probabilistic mid-long term oil spill forecast has been presented. The methodology, based on the statistical prediction of met-ocean patterns in the mid-long term by means of logistic regression modeling, has been applied in the Bay of Biscay. The obtained results have been compared with the trajectories of several drifting buoys deployed in the area in the aftermath of the *Prestige* accident, in 2002. The comparison highlight the capacities of the methodology and the benefits it could bring to the fight against marine pollution. In this chapter, the specific objectives 1 and 2 of the thesis are achieved.

CHAPTER IV.

DEEP OIL SPILL HAZARD ASSESSMENT USING STOCHASTIC MODELING BASED ON MET-OCEAN SPATIO-TEMPORAL PATTERNS. APPLICATION TO THE NORTH SEA.[3]

4.1 Introduction

Chapter 2 and 3 were centered on the development of new methodologies for the improvement of real time oil spill response systems, considering the mid-long term prediction of oil spills. In this chapter, the focus moves to the field of oil spill risk assessment.

[3] The findings from this chapter are going to be submitted to Mar. Pollut. Bull. as: Chiri, H., Abascal, A.J., Castanedo, S., Medina, R.. Deep oil spill hazard assessment based on spatio-temporal met-ocean patterns.

Risk assessment is a powerful tool used to minimize consequences of oil spills and represents a key element in many oil spill contingency plans. An important piece of risk assessments is represented by stochastic oil spill modeling, aimed to probabilistically assess the likely consequences of a spill by considering representative environmental conditions of the considered area.

In this chapter a novel methodology for stochastic oil spill modeling is proposed, based on the statistical selection of met-ocean temporal evolution patterns of the considered study area. The benefits of this methodology with respect to existing methods are: 1) it can be applied to deep spills, considering the subsurface and the surface oil transport; 2) it is based on classification techniques for the optimum selection of met-ocean patterns to represent the main environmental variability of a considered study area; 3) it allows an optimization of the computational costs, a limiting factor in this kind of studies. The methodology is general and worldwide applicable, since it only relies on numerical reanalysis databases.

The proposed methodology is applied to analyze the probability of pollution due to a deep oil spill in the North Sea, one of the regions with highest level of offshore oil activities. In this case the methodology integrated both the surface and the subsurface processes affecting the evolution of a spill in the sea. However, a similar methodology, considering only surface processes and met-ocean forcings can be used in case of considering a surface spill (Figure 22). With this chapter, the third specific objective of the thesis is achieved.

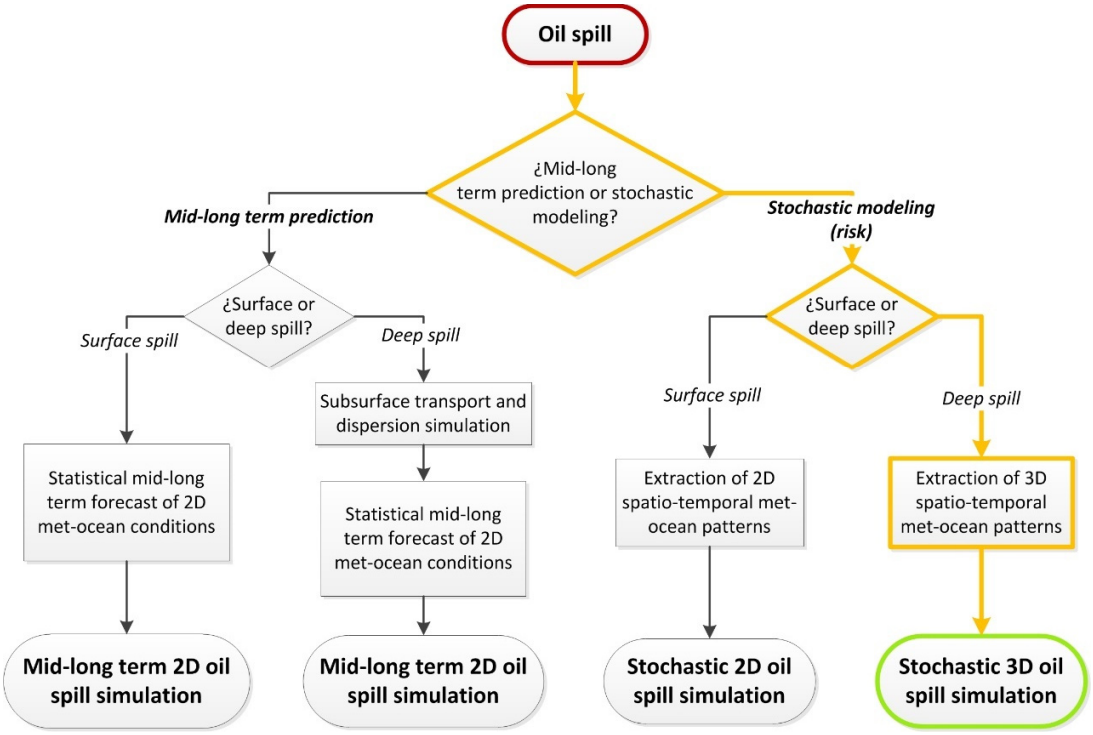

Figure 22. Flowchart of the methodology case study presented in this chapter in the framework of the general methodology pursued in the thesis.

4.2 General methodological framework

The proposed methodology allows for the estimation of the pollution hazard at the sea surface due to a deep oil spill. The main steps constituting the presented method are schematically shown in Figure 23.

The methodology can be applied worldwide, provided that the following hypothesis are accomplished:

I. The oil droplets' rise velocity driven by their buoyancy is significantly higher than the oil droplets' horizontal velocity in the water column.

II. There is not a significant vertical gradient in the average water column current velocity.

III. Wave-induced Stokes drift is neglected

The accomplishment of the first hypothesis means that the spilled oil experiments a quick rise towards the sea surface due to the buoyance force. Spending a limited amount of time underwater, the main horizontal dispersion of the spill occurs once at the surface (Figure 24). By assuming the second hypothesis, a depth-averaged value of the subsurface currents can be considered as representative of the whole water column. The third assumption is generally applicable in wind-dominated circulation areas and if the main focus of the study is not in a coastal area, where the effect of the waves on the oil transport become important (Guo et al., 2014; Janeiro et al., 2014; Weisberg et al., 2017).

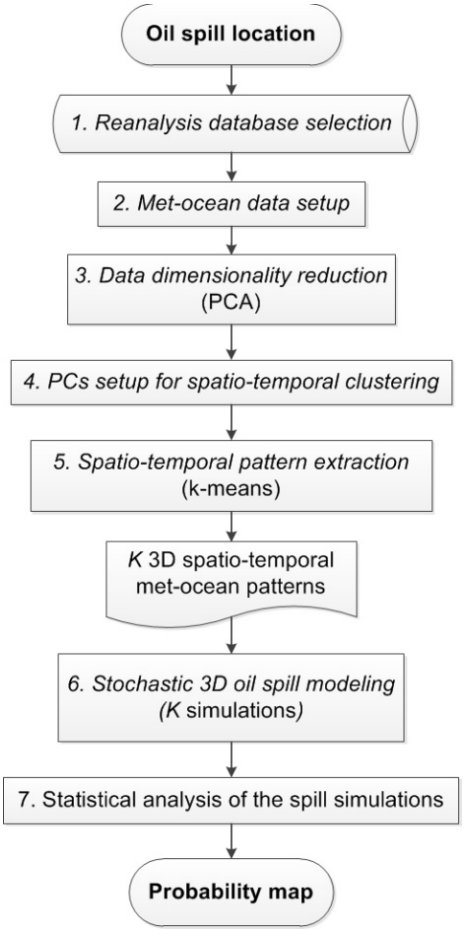

Figure 23. Flow chart of the proposed methodology.

4.2.1 Reanalysis database selection

The first step of the methodology is the selection of those met-ocean reanalyses most suitable for the study scope. Besides reliability, the main requirement for a numerical reanalysis to be used in an oil spill hazard assessment is its temporal extent. Since the aim of a hazard study is to simulate likely slick trajectories under typical met-ocean conditions in a study area, it is crucial that the considered databases cover enough time so that the patterns extractable from them will not be conditioned by the specific period covered by the database

but, on the contrary, they can be considered generic and representative of the considered application domain. Besides this aspect, another essential features of a numerical reanalysis is its spatial resolution. The resolution should be fine enough to properly represent the main spatial features of the study domain. As an example, if the water circulation is highly affected by the flow between two islands or by a narrow continental shelf, the ocean current reanalysis to be used should have a spatial resolution adequate to properly represent such features.

4.2.2 Met-ocean data setup

The assumption of hypothesis *I* allows to consider subsurface current data only from the current grid cell containing the release location. Then, based on hypothesis *II*, a depth-averaged value of this subsurface current data is calculated and will be representative of the subsurface currents for the oil spill simulations. Subsequently, wind and currents datasets has to be arranged in order to uniform their temporal scales. In first place, the time frame extents of the two reanalysis are matched by cutting the data to the common time period covered by both series. Then, the time frequencies of both datasets are adjusted in order to have both data series with exactly the same time frame.

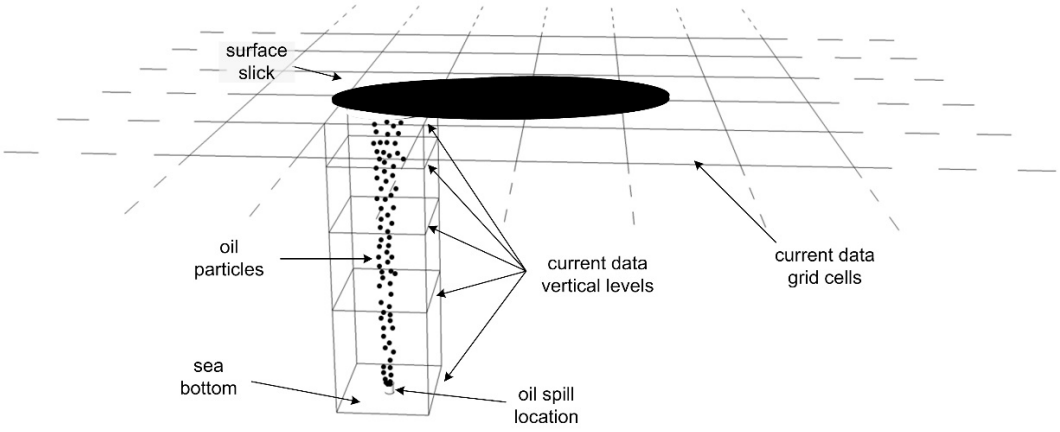

Figure 24. Sketch showing that different degrees of horizontal dispersion between the water column and the sea surface.

4.2.3 Data dimensionality reduction (PCA)

In order to facilitate the subsequent steps of the methodology and to enhance the k-means algorithm performance, the dimensionality of the considered met-ocean data must be reduced using Principal Component Analysis (PCA). PCA is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. It accomplishes this reduction by identifying directions, called empirical orthogonal functions (EOFs), along which the variation in the data is maximal (Jolliffe, 2002). The temporal series of the data in these new directions are called principal components (PCs). By keeping those PCs accounting for a certain percentage of the original data variance, generally 95% (e.g. Antolínez et al., 2016; Camus et al., 2016), the dimension is reduced to the number of considered PCs. If n_c and n_w are the number of numerical nodes of the considered current and wind grids, respectively, and T is the common number of time steps of both series, then the value of the u and v components of surface currents and wind can be expressed as Uc_i^t, Vc_i^t,

Uw_j^t and Vw_j^t, respectively, where i = 1,…, n_c, j = 1,…, n_w and t = 1,…, T, and the u and v components of the depth-averaged currents can be expressed as Ud^t and Vd^t, respectively. The matrix *M0* to be built for the PCA application will have dimensions $\left(T \times (2n_c + 2n_w + 2)\right)$ (Table 7). In order to give the same relative weight to the different considered met-ocean variables, each column of *M0* must be standardized before applying PCA. As a result of applying PCA to matrix *M0*, if *N* is the number of PCs obtained to explain 95% of the original data variance, then *M0* is reduced to a new matrix *M1* of dimensions $(T \times N)$ (Table 8) (step 3).

Table 7. 3D current and wind data setup for PCA application (M0).

U components of wind	V components of wind	U components of surface currents	V components of surface currents	U depth-averaged current	V depth-averaged current
Uw_1^1 Uw_2^1 … $Uw_{n_w}^1$	Vw_1^1 Vw_2^1 … $Vw_{n_w}^1$	Uc_1^1 Uc_2^1 … $Uc_{n_c}^1$	Vc_1^1 Vc_2^1 … $Vc_{n_c}^1$	Ud^1	Vd^1
Uw_1^2 Uw_2^2 … $Uw_{n_w}^2$	Vw_1^2 Vw_2^2 … $Vw_{n_w}^2$	Uc_1^2 Uc_2^2 … $Uc_{n_c}^2$	Vc_1^2 Vc_2^2 … $Vc_{n_c}^2$	Ud^2	Vd^2
⋮ ⋮ ⋮ ⋮	⋮ ⋮ ⋮ ⋮	⋮ ⋮ ⋮ ⋮	⋮ ⋮ ⋮ ⋮	⋮	⋮
Uw_1^T Uw_2^T … $Uw_{n_w}^T$	Vw_1^T Vw_2^T … $Vw_{n_w}^T$	Uc_1^T Uc_2^T … $Uc_{n_c}^T$	Vc_1^T Vc_2^T … $Vc_{n_c}^T$	Ud^T	Vd^T

Table 8. Matrix M1 obtained as a result of applying PCA to M0.

PC_1^1	PC_2^1	…	PC_N^1
PC_1^2	PC_2^2	…	PC_N^2
⋮	⋮	⋮	⋮
PC_1^T	PC_2^T	…	PC_N^T

4.2.4 PCs setup for spatio-temporal clustering

The fourth step is to set up the data in order to extract patterns of evolutions within the study domain. To do so, if the temporal evolution of the spatio-temporal patterns that

want to be obtained corresponds to D time steps, then a new matrix *M2* must be create by rearranging *M1* to the dimensions $((T-D)\times(N\times D))$ (Table 9).

Table 9. Matrix M2 obtained by reshaping M0 to consider the temporal evolution of the PCs.

PC_1^1	PC_1^2	…	PC_1^D	PC_2^1	PC_2^2	…	PC_2^D	…	PC_N^1	PC_N^2	…	PC_N^D
PC_1^2	PC_1^3	…	PC_1^{D+1}	PC_2^2	PC_2^3	…	PC_2^{D+1}	…	PC_N^2	PC_N^3	…	PC_N^{D+1}
PC_1^3	PC_1^4	…	PC_1^{D+2}	PC_2^3	PC_2^4	…	PC_2^{D+2}	…	PC_N^3	PC_N^4	…	PC_N^{D+2}
…	…	…	…	…	…	…	…	…	…	…	…	…
PC_1^{T-D}	PC_1^{T-D+1}	…	PC_1^T	PC_2^{T-D}	PC_2^{T-D+1}	…	PC_2^T	…	PC_N^{T-D}	PC_N^{T-D+1}	…	PC_N^T

4.2.5 Spatio-temporal pattern extraction (k-means)

The next step is the extraction of the patterns of simultaneous evolutions of wind, surface currents and water column currents with k-means algorithm. In order to apply this clustering technique, it is required to define the number of patterns *K* that have to be extracted. Following Bárcena et al. (2015) and Núñez et al. (2019), an index is used to assess the performance of k-means for a variable number of extracted patterns. Willmott's "index of agreement", *d*, (Willmott, 1981) is a valuable index to compare a modelled series versus the real one (Núñez et al., 2019). The number of patterns to extract *K* can then be defined as that one above which the increase of *d* is not significant. Once *K* is defined, k-means is apply to extract the *K* most relevant 3D met-ocean spatio-temporal patterns from the available dataset. However, it is worth noting that the obtained patterns are expressed in term of the (standardized) PCs from matrix *M2*. It is required, then, to reconstruct these patterns to the original space by multiplying them for the EOFs found in step 3 and by unstandardizing the results. By doing so, the obtained pattern are reconstructed in the original space of matrix

M0. The original data series can be reconstructed in terms of the obtained patterns by considering in each case the closest pattern or best matching unit (BMU), and hence a probability of occurrence $P_{o,k}$ can be calculated for each of the *K* patterns.

4.2.6 Stochastic 3D oil spill modeling

At this point, the *K* most representative spatio-temporal patterns of wind, surface currents and depth-averaged currents are available. Subsequently, a state of-the-art 3D lagrangian oil spill model, TESEO (Abascal et al., 2007), is used to stochastically modeling the evolution of the considered spill. A description of TESEO model is provided in the appendix at the end of the document. *K* simulations have to be run, each with *K* patterns of wind and 3D currents evolution identified.

4.2.7 Statistical analysis

For each lagrangian simulation, as also done in the traditional oil spill hazard assessment approach, an auxiliary grid is used to count the simulation particles at the surface. The probability of oil contamination $P_{cont,i}$ in the *i*-th grid cell is obtained as follow:

$$P_{cont,i} = \sum_{k=1}^{K} \left(P_{o,k} \cdot PA_i \right) \tag{13}$$

where, PA_i is a binary variable indicating the presence or absence of any particle within the cell i and $P_{o,k}$ is the probability of occurrence of the k-th met-ocean pattern, with k=1,…,K:

$$PA_i = \begin{cases} 0, & \text{if no particle lays inside the } i-\text{th cell} \\ 1, & \text{if any particle lays inside the } i-\text{th cell} \end{cases} \quad (14)$$

The hazard map can be finally represented based on the probabilities P_{cont} of each cell grid.

4.3 Application to a hypothetical deep spill in the North Sea

In order to facilitate the understanding of the proposed methodology and to check for its potential, it is applied to a hypothetical deep spill in the North Sea, one of the most active areas in the world in offshore oil and gas producing and exploring operations. Due to the high volume of oil and gas operations, the area is continuously under an unavoidable treat of oil spill accidents and represents an interesting case study to test the capabilities of the proposed methodology. The methodology was applied to assess the stochastic oil trajectory evolution of an accidental oil deep release from one of the currently active oil production platforms of the North Sea. A platform located in the central part of the North Sea was arbitrary selected as the accident location and a 30-day period was selected as the stochastic modeling time horizon.

4.3.1 Study area

The North Sea is a shelf sea bounded by the British Isles, Norway and the European Continent and reaching approximately the Shetland Islands latitude to the North (Figure 25). It has a mean depth of 74 m, with two thirds of the region having depths shallower than 100 m (ICES, 1983). The deeper area is the Norwegian Trench with a typical depth of 300 m, increasing to 700 m in the Skagerrak.

In the 1960s and 1970s, advances in oil and gas technologies encouraged oil companies to explore new areas in search of reserves. The North Sea immediately showed its potential. Over the years, this area emerged as an important oil producing region and, thanks to the proximity to some of the European major consumer markets and the region's political stability, it has become an alternative to the traditional exporting countries (De Groot, 1996; Noreng, 2016). Almost 8000 wells have been drilled to date, delivering 44 billion barrels of crude, only considering the United Kingdom continental shelf (OGA, 2018a). Such an intense activity has clearly not been free of accidents. In 2017, for example, 253 accidental oil releases occurred on the UK continental shelf, with a total of 23 tonnes of crude oil spilled to the marine environment (OGA, 2018b). For this application case, a hypothetical deep spill location was considered in the middle of the North Sea (3.55 ºN, 56.187ºE) at a depth of 71 m (Figure 25).

Regarding the North Sea met-ocean dynamics, the warmer and more salty water from the Atlantic Ocean enters the area through the English Channel or the northern boundary, around the Shetland Islands or through the Norwegian Trench (Winther and Johannessen, 2006). The Atlantic water tends to flow southward while is mixed with fresher water from

the British Islands and the European continent. This flow moves eastward reaching the Skagerrak and entering the Baltic Sea, where the interaction with the brackish water makes the flow to turn back following the Norwegian coast westward originating the Norwegian Coastal Current (Winther and Johannessen, 2006). Winds have shown to significantly control circulation of the North Sea with four identified main patterns of wind-driven currents (Sündermann, 2003). Prevailing westerly winds are generally associated with an intense cyclonic circulation, which in case or more occasional easterly winds tends to reverse. For north-westerly and south-easterly winds states of stagnation appear. Tides, mainly semidiurnal, also play an important role on water circulation in the North Sea (Huthnance, 1991).

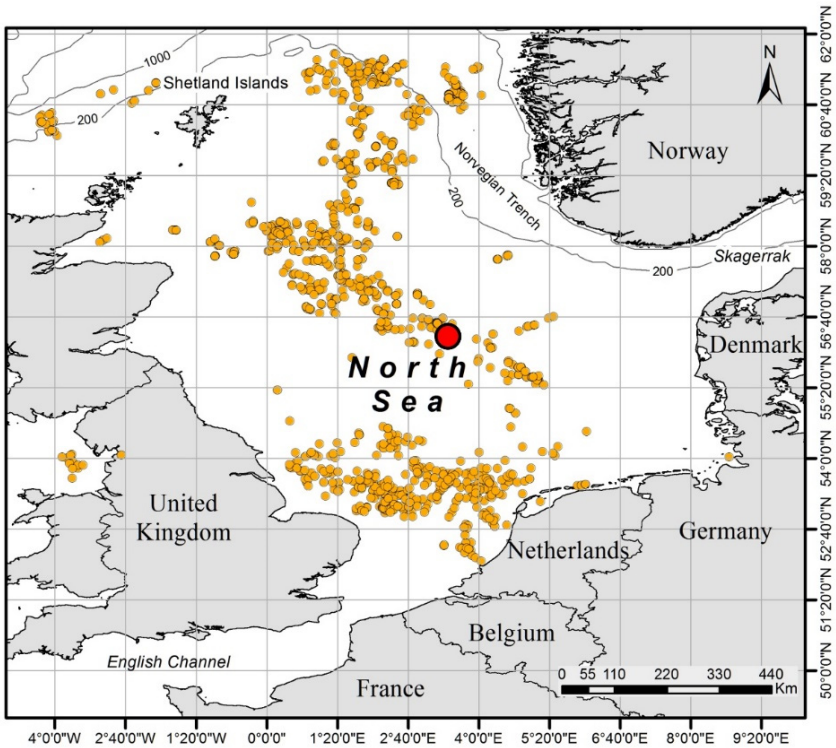

Figure 25. Study area map. Also shown the locations of offshore platforms in the area (orange circles) (OSPAR, 2015). The red circle represents the considered spill location.

4.3.2 Reanalysis database selection

4.3.2.1 Currents

As 3D current data source the U.S. Naval Research Laboratory (NRL)'s Global Ocean Forecasting System (GOFS) 3.1 Reanalysis was selected. This ocean reanalysis covers 22 years (1994-2015), a period of time long enough to comprehend the general variability of the area, with a 3-hourly frequency. The spatial resolution of the data is 1/12º, approximately 9 km, counting with approximately 10 grid cells in the most narrow portion of the English Channel, and 39 vertical levels. It was obtained as a result of the Hybrid Coordinate Ocean Model (HYCOM, Bleck (2002)) and the Navy Coupled ocean Data Assimilation (NCODA, Cummings, 2005). HYCOM is a comprehensive, three-dimensional hydrodynamic model. It uses a hybrid coordinate system that is isopycnal in the open, stratified ocean, but smoothly reverts to a terrain-following (sigma) coordinate in shallow coastal regions, and to *z*-level coordinates in the mixed layer and/or unstratified regions. The surface fluxes of momentum, heat and freshwater are derived from the Climate Forecast System Reanalysis (CFSR) products. River "runoff" is added to the precipitation field at a single grid point and smoothed over surrounding grid points. The data assimilated by NCODA include remotely sensed sea surface height (SSH), sea surface temperature (SST) and sea ice concentration as well as in-situ surface and subsurface observations of temperature and salinity. The GOFS 3.1 reanalysis variables considered in this study were the 3D current velocities. As said in Section 4.2, only currents of the subsurface layers from the grid cell containing the spill location (3.55 ºN, 56.187 ºE) were considered. Surface data were cut to the study domain [-4º, 10º] longitude and [51º, 61º] latitude. Given the high resolution of the database and considering

the spatial scale of the study domain, one grid point every three in latitude and longitude was considered, for a total of 4814 (83 x 58) GOFS 3.1 Reanalysis grid points. Moreover, as said in Section 4.2, also the water column current values of the grid cell containing the simulated spill location were used to calculate the depth-averaged currents.

4.3.2.2 Wind

Wind data used in this study came from National Centers for Environmental Prediction (NCEP)'s Climate Forecast System Reanalysis (CFSR). CFSR covers 41 years (1979-present). Wind data are provided with hourly frequency and a spatial resolution of 0.3º until 2010 and 0.2º starting from 2011, when a new Climate Forecast System version2 (CFSv2) was implemented. This reanalysis uses a coupled atmosphere-ocean-land surface-sea ice system and a 3D-VAR data assimilation system (Saha et al., 2010). The atmospheric model has 64 vertical levels from the surface to 0.26 hPa. Model physics include solar radiation following Hou et al., (1996), the cumulus convection scheme of Hong and Pan (1998), gravity wave drag (Kim and Arakawa, 1995) and cloud water and ice (Zhao and Carr, 1997). CFSv2 incorporates advanced parametrization schemes with latest version of model components. Similarly to Shinoda et al. (2019), to ensure the consistency of the used data, only wind data up to 2010, previous to the implementation of CFSv2 model, were considered. The same domain limits as the current data were considered for a total of 1408 (44 x 32) grid points.

4.3.3 Met-ocean data setup

Considering the fine resolution of the GOFS 3.1 currents and in order to facilitate the methodology application, data of one every three grid nodes was considered for the surface layer currents. For the subsurface currents, once identified the reanalysis grid cell containing the horizontal position of the spill site, the depth-average currents were obtained by calculating the arithmetic mean of the currents from the 17 subsurface model layers available for that location (for both *x* and *y* current components). Both wind and current datasets were then cut the overlapping temporal extent covered by both datasets, being the 17-year period 1994 to 2010. Note that 17 years is a sufficiently long period of historical met-ocean information for the purpose of the present study. To completely match the time frames of the wind and current data sets, both series were rescaled to a common time frequency. In this application case, a temporal frequency of 1 day was used to uniform the data. Finally, the met-ocean data were merged in a matrix *M0* having dimensions 6208 x 12446, where the number of rows reflects the time steps (number of days in the 17-year considered period) and the number of columns corresponds to the considered data features (1408 wind data grids + 4814 surface current data grids + 1 depth average current feature, replicated for both *x* and *y* velocity components).

4.3.4 Data dimensionality reduction (PCA)

In order to reduce the dimensionality of the data, PCA was applied to this matrix. Given the different nature of the considered variables, hence their characteristic magnitudes, all the columns of the matrix were previously standardized in order to enhance the

performance of PCA. As a result of applying this technique, 51 PCs were obtained that explained 95% of the original data variance. So, a new, lighter matrix *M1* was obtained with dimensions 6208 x 51, containing these obtained 51 PCs, and representing the data in the new EOF space. Such an important reduction of the met-ocean data dimensionality (having sacrificed just 5% of their variance) highlights that a high correlation was present in the considered original data set.

4.3.5 PCs setup for spatio-temporal clustering

As previously mentioned, in this application case, the oil pollution hazard within a period of 30 days from the beginning of the spill was to be assessed. To achieve lagrangian oil spill simulations at this time horizon, temporal evolutions of wind and 3D currents fields during 30 days were required. To do so, before the application of k-means, the data (*M1*) was rearranged in a new matrix, *M2*, having size 6178 (6280-30) x 1530 (51 PCs x 30 days). The groups extracted using this matrix as input data of k-means algorithm will represent patterns of evolutions of the data at a 30-day horizon.

4.3.6 Spatio-temporal pattern extraction (k-means)

The results of the sensitivity analysis aiming to determine the number of patterns to be extracted are shown in Figure 26. In this figure, the evolution of *d* index is represented for a varying number of patterns from 0 to 400. Note that the number of considered spatio-temporal met-ocean patterns equals the number of oil spill numerical simulations to be

computed, and then the higher the number of patterns the higher the computational costs required for the stochastic oil spill modeling. As shown in Figure 26, the increasing in *d* starts losing strength starting from 125 patterns, indicating that the benefit of using a higher number of patterns may not be significant. Willmott et al. (1985) suggest that values of *d* larger than 0.5 represent good model performance. Considering 125 patterns *d* equals 0.65, indicating the proper performance of the clustering technique. As an example, in Figure 27 one of the 125 obtained patterns is shown, after reconstruction. Each snapshot corresponds to one day. Black arrows and background colors refer to surface currents and red arrows to wind. *x* and *y* components of the depth-averaged currents at the spill location grid cell are shown as a text in the lower left corner of each snapshot.

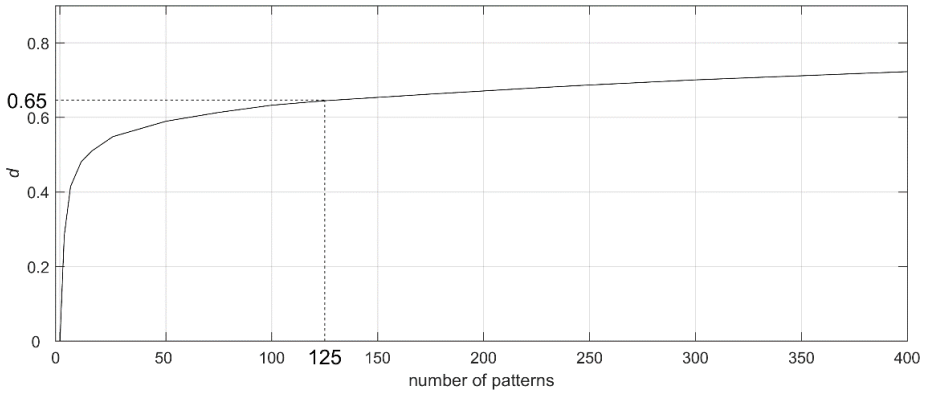

Figure 26. Evolution of Willmott's index of agreement d with the numbers of patterns considered.

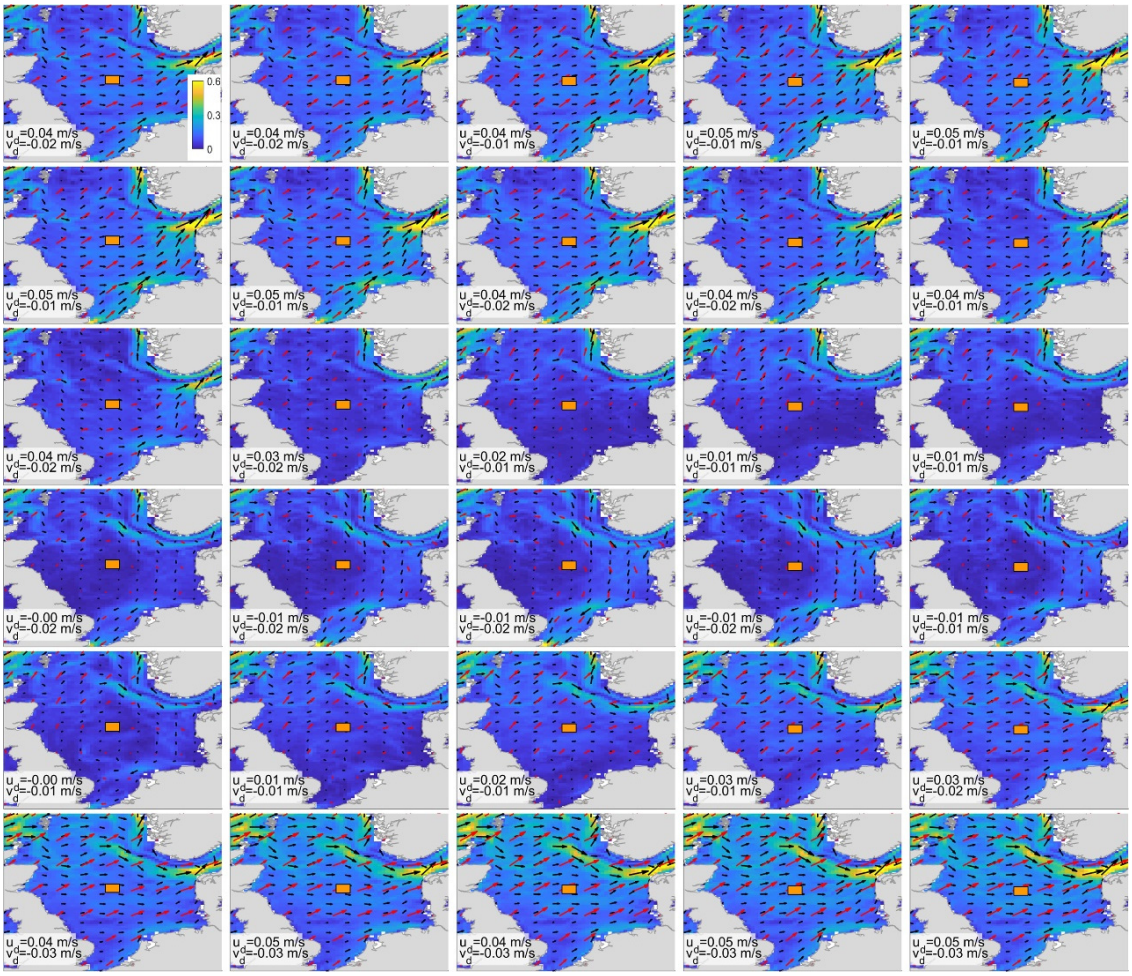

Figure 27. Representation of one of the 125 spatio-temporal patterns of 30-day evolution of wind and 3D currents at daily frequency in the North Sea. Each snapshot corresponds to one day. Black arrows with background color maps (in m/s) represent surface currents, red arrows refer to wind. Subsurface current information is shown as a text in the lower left corner in each snapshot. The location of the current grid cell to which subsurface currents information refers (corresponding to the spill location) is indicated as a filled orange rectangle. Distances among vectors do not represent the actual resolution of the data: for clarity's sake only the data of one of every five grid points, for both wind and surface currents, are represented.

4.3.7 Spill numerical simulation

For each of the 125 lagrangian spill simulations computed, all the simulation parameters were maintained identical, except the environmental forcings. The spill was located at 3.455ºE, 56.176ºN at a depth of 71 m. A continuous release was considered during 48 hours with a spill rate of 20,000 bbl/day. A product with density of 811.6 kg/m3 at 15ºC and viscosity of 6.89 cSt at 12ºC, characteristic of the study area reservoirs, was considered (Flaherty and Murray, 2001). For each simulation, a total amount of 100.000 particles were released. An example of the oil spill simulations is shown in Figure 28, where the 3D trajectory evolution, considering one of the 125 identified met-ocean scenarios, after 7 hours from the beginning of the spill is represented. In panel *b* the same particle distribution from panel *a* is shown, but with a different level of zoom. Note how the subsurface trajectory of the particles is essentially vertical, showing a minimum horizontal transport due to subsurface currents. In order to estimate the contamination probability, an auxiliary regular grid was considered at the surface with cells of size 0.25º x 0.25º.

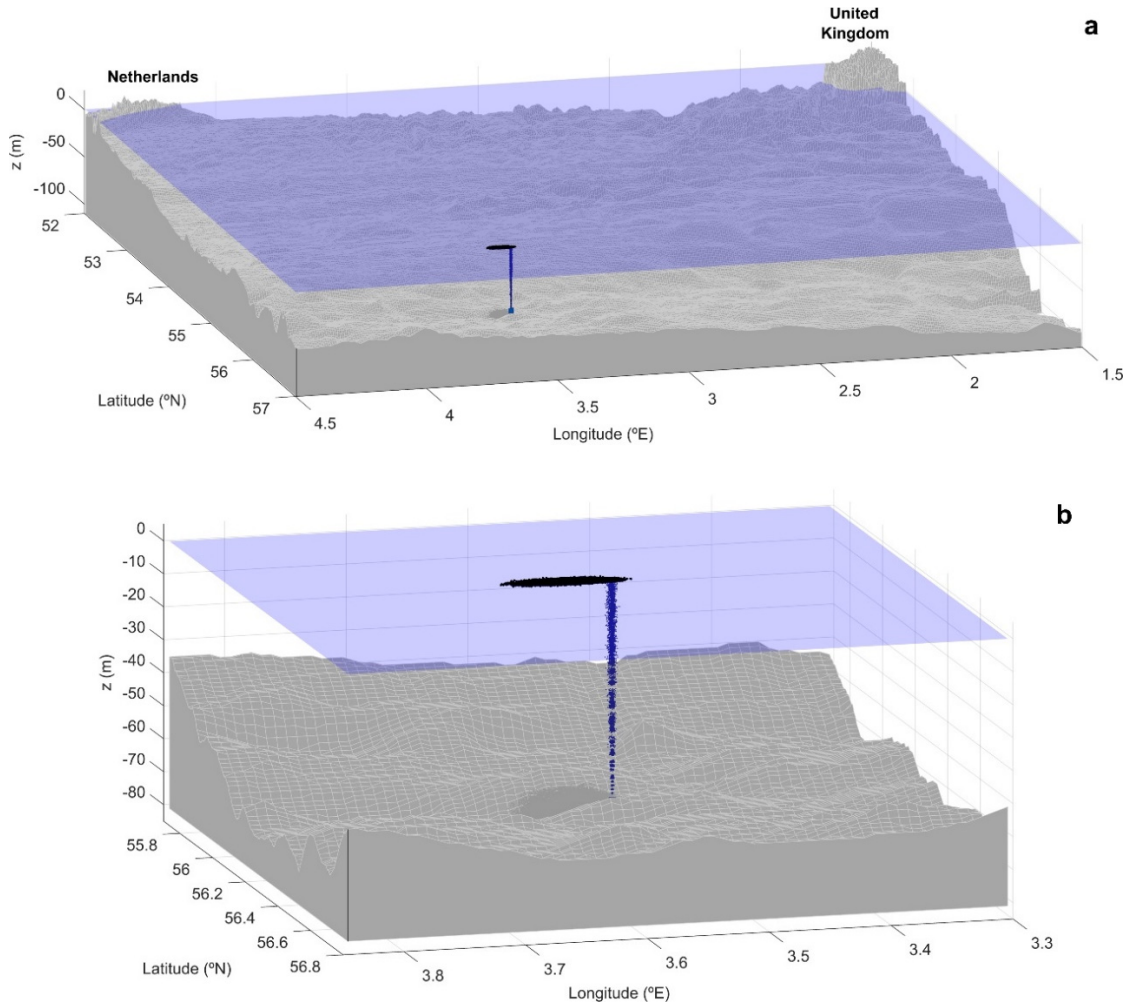

Figure 28. A snapshot from one on the spill simulations with two levels of zoom.

4.3.8 Results

Figure 29 shows the results obtained for the cumulative probability of contamination at the sea surface, for the 0.25º auxiliary grid cells, after the 30-day simulation period. The color bar shows the probability of pollution as percentage. For instance, the deep red color area has probability of contamination between 90% and 100%, within 30 days. The obtained results show that most of the North Sea surface could be affected in case of occurring the

considered spill. Higher probabilities are found in the mid-eastern part of the basin, indicating that most of the simulations followed an eastward trajectory. Denmark, Norway, Netherland and Germany coastlines resulted to be the more exposed among the countries surrounding the North Sea. The highest coastal contamination probability is obtained in Denmark, being approximately 15-20% around the Skagerrak facing side. This suggests that the water flow towards the Baltic Sea through the southern section of the Skagerrak may play an important role in the oil spill transport in the area. A probability of about 5-10% is found in the northeastern section of the Netherland coastline and at the southern side of Norway. A small probability (less than 5%) is found, however, through the whole coast of Norway, as well as at the northern German coast and, although very localized, in England and Scotland.

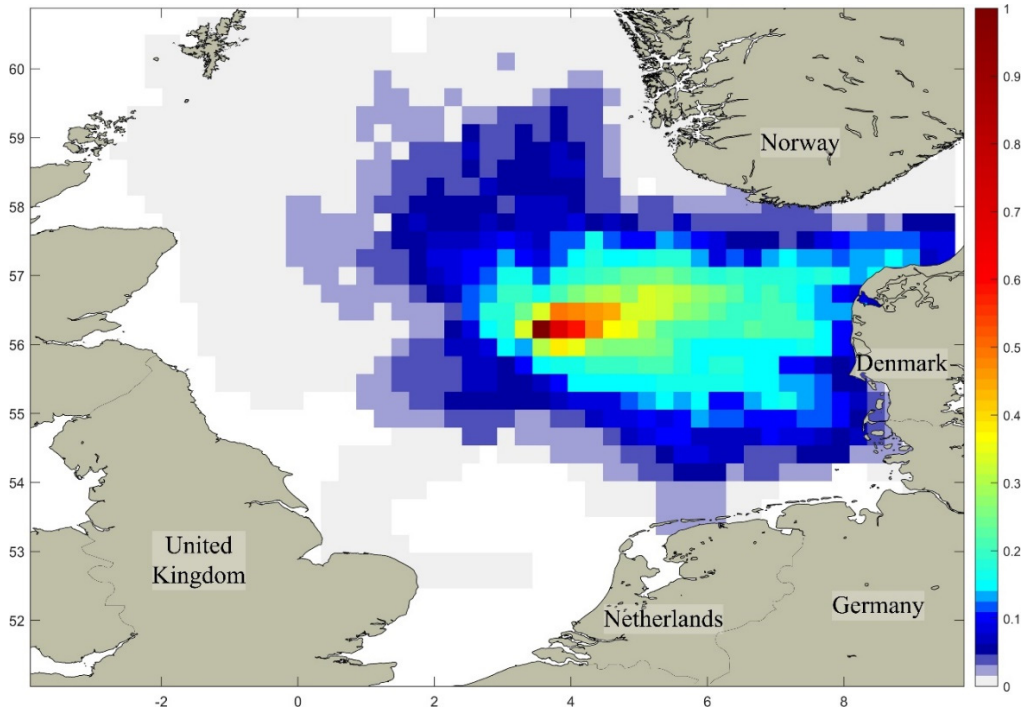

Figure 29. Probability, expressed as a percentage, of oil contamination in 0.25° x 0.25° gridded cells over the 30 day period following the beginning of the release, based on the proposed methodology.

4.3.9 Comparison with traditional stochastic modeling technique

In order to compare the obtained hazard map with the one achievable using a traditional Monte Carlo approach, the same oil spill scenario was considered to assess the contamination hazard when randomly selecting 1000 30-day wind and currents evolution from the same met-ocean databases. Using 1000 simulations to assess the oil spill hazard is consistent with various studies of similar nature (Abascal et al., 2010; Canu et al., 2015; Elliott, 2004; Liubartseva et al., 2015). After running 1000 simulations with the same 3D oil spill model and considering the exact same parameters, the hazard map shown in Figure 30 was achieved. Compared with the map of Figure 29, it is evident how the probability gradients are, in this case, more smoothed. This is expectable when using such a large number of simulations. However, the main features shown in this map are similar to the ones from Figure 29. Also in this case, minimum degree of contamination probability are found for most of the North Sea domain. Similarly to what obtained with the proposed methodology, the higher probabilities are found in the mid-eastern part of it. The coastline with higher exposure corresponds to Denmark (approximately 15%), having Norway, Germany and Netherlands a probability of about 5%, similarly to what shown in Figure 29. In order to quantify the comparison between the results obtained with the two methods, in Figure 31 a map of the differences between the probabilities estimated with both methodologies are presented. Positive values (warmer colors) refer to overestimation and negative values (colder colors) refer to underestimation of the proposed methodology results respect to the ones obtained using the Monte Carlo approach. As shown, the highest differences are obtained in the proximity of the oil spill surfacing point, to the North, West and South sides

of it, but not to the East side where, as shown, most of the trajectories headed. In these areas the proposed methodology results underestimated the probabilities obtained with the Monte Carlo method of about 15%. This may be due to some of the 1000 randomly extracted met-ocean scenarios of the Monte Carlo method that were not taken into account in the proposed methodology, which focused more in those met-ocean scenarios producing the predominant eastward transport. The differences between the two methodologies' results decrease moving away from the surfacing spill area. The variations in the estimated pollution probabilities in the coast do not go behind a 3%-4% in the southern Norway and Denmark coasts. The overall average difference is -2%, reflecting the high degree of similarity in the results achieved using the two methodologies.

This comparison highlight the benefits of the presented stochastic oil spill methodology in the framework of an oil spill risk analysis. Indeed, by focusing on the analysis of extensive met-ocean databases with the described data mining techniques, leading to the specific selection of the most relevant scenarios (in this case, 125), the proposed methodology leads substantially to the same oil contamination probabilities than those achievable with a traditional Monte Carlo method, often requiring the inclusion of hundreds or thousands randomly selected met-ocean scenarios (in this case, 1000). Such a reduction in the number of considered met-ocean scenarios is reflected in a reduction of the required oil spill numerical simulations, which, eventually, produce an important computational costs and system memory save. This reduction in computational resources is particularly valuable when including the oil subsurface processes in oil spill risk assessments for offshore installations, as suggested by the international requirements (IPIECA and IOGP, 2013).

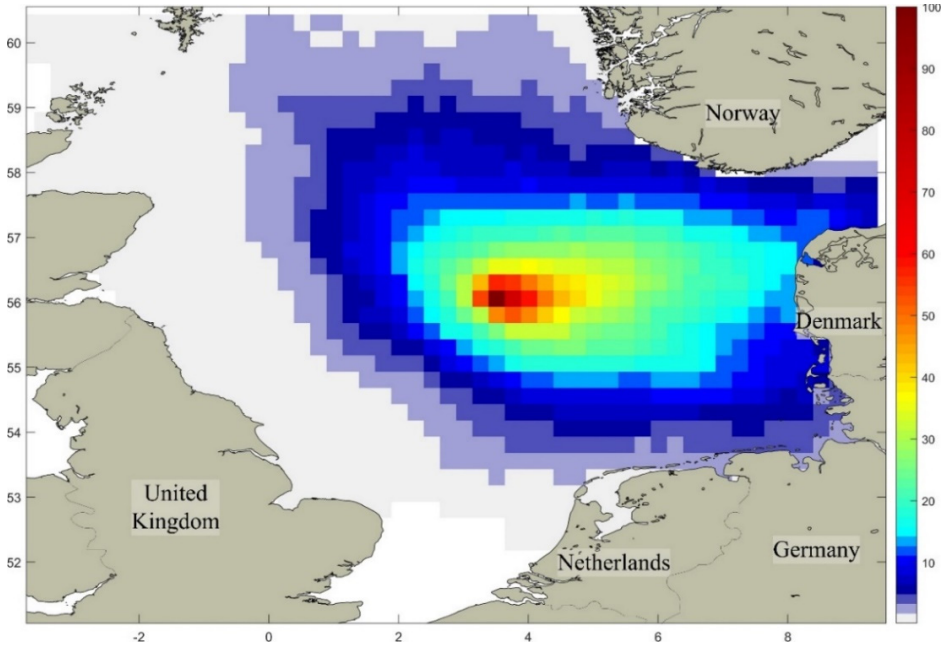

Figure 30. Probability, expressed as a percentage, of oil contamination in 0.25° gridded cells over the 30 day period following the beginning of the release, based on 1000 random met-ocean scenarios.

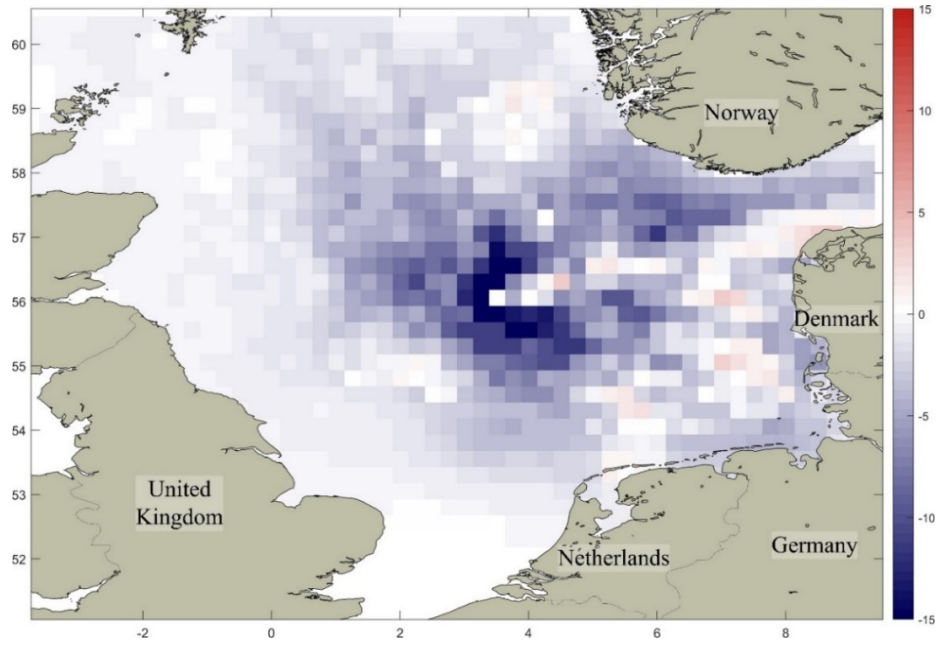

Figure 31. Map of the probability difference between the probabilities of contamination obtained with the proposed methodology and the ones obtained with the Monte Carlo approach considering 1000 simulations. Positive values refer to overestimation and negative values refer to underestimation of the proposed methodology results respect to the ones obtained using the Monte Carlo approach.

4.4 Conclusions

In this chapter a methodology for stochastic oil spill modeling, based on the selection of spatio-temporal met-ocean patterns, is presented. This methodology to be used in oil spill risk assessments, present the following advantages with respect to existing state-of-the-art techniques: 1) it can be applied to subsurface spills, 2) it is based on spatio-temporal met-ocean pattern extraction and 3) it allows to reduce computational costs of a risk assessment, a critical issue in this kind of studies. Obtained results in an application to a deep spill in the North Sea highlight the capabilities of the proposed method and the computational cost reduction with respect to traditional oil spill stochastic modeling methods. With this chapter, the third specific objective of the thesis is achieved.

CHAPTER V.

CONCLUSIONS AND FUTURE RESEARCH.

5.1 Conclusions

The overall objective of this thesis was the definition and development of a general framework for the simulation at the mid-long term of oil spill, taking advantage of different statistical techniques. The definition of such a goal was the result of a thorough literature review. Two main problematics were identified. The first one was the need of methodologies for the prediction of oil spill evolution beyond the typical temporal extent of approximatively 5 days, imposed by met-ocean forecast coverage. The second one was the lack of efficient techniques for the stochastic modeling, a tool required in the oil spill risk assessment framework, in case of considering subsurface oil releases. In order to fill the first gap, the first and second specific objectives of the thesis were pursued in Chapter 2 and Chapter 0, respectively, while the third specific objective was considered in Chapter 4.

Following, the main conclusions of each of these three chapters are summarized.

Chapter II. Statistical simulation of current patterns using a logistic regression model. Application case in the Gulf of Mexico.

- Statistical simulation of current patterns can be achieved using logistic regression.
- A logistic regression model was set up for the simulation of surface currents in the Gulf of Mexico, based on a likelihood ratio statistic test for the selection of relevant covariates.
- The model showed good capacity of reproducing both the inter-annual and the intra-annual variability of current patterns in the Gulf of Mexico.
- The effectiveness of the model is confirmed by the obtained values of Sensitivity, Precision, F-score and Accuracy being, in all cases, greater than 0.75.
- Statistical simulation of current patterns represent the first step towards the mid-long term prediction of oil spills in areas dominated by general circulation.

Chapter III. Mid-long term prediction of oil spills based on logistic regression modeling of met-ocean forcings. Application case in the Bay of Biscay.

- A methodology for probabilistic mid-long term oil spill trajectory predictions is presented.

- As a first part of this methodology, a mid-long term prediction of met-ocean patterns is required. The obtained predictions represent the environmental forcing inputs of the subsequent oil spill numerical simulations.
- The final results of the methodology are provided as probability trajectory maps, preserving the probabilistic nature of the statistical model.
- The proposed methodology allowed to produce mid-long term spill simulations in the Bay of Biscay.
- The comparison of the obtained results with drifter trajectory data collected during the *Prestige* emergency showed the potential of the proposed methodology.
- The methodology may fill a gap in present-day operational oil spill response systems, allowing to obtain information about the evolution of a spill in the mid-long term.

Chapter IV. Deep oil spill hazard assessment using stochastic modeling based on met-ocean spatio-temporal patterns. Application to the North Sea.

- A methodology for the stochastic modeling of deep oil spills, integrating water column and surface processes is proposed.
- In the methodology, the inclusion of the overall met-ocean variability, required to obtain the likely consequence of a hypothetical spill, is achieved by the extraction of different patterns of evolutions of multiple met-ocean variables, each with a probability of occurrence.

- By simulating the considered spill under the selected met-ocean spatio-temporal patterns and considering the probability of occurrence of each scenario, an oil pollution probability map is generated.
- The methodology capacity is tested with a hypothetical deep spill in the North Sea, where 125 patterns of 1-month evolution of wind, surface and subsurface currents are selected to represent the local environmental variability. The resulting probability pollution map is similar to the one obtained by applying a traditional approach involving 1000 oil spill simulations, versus the 125 used with the proposed methodology.
- The benefits of the proposed methodology are twofold: 1) it allows to consider the integrated subsurface and surface evolution of a spill and 2) it reduced the computational costs required to perform a stochastic oil spill simulation with respect to the traditional techniques.
- The methodology represents an improvement in the field of oil spill risk assessment and could help to enhance contingency plans, especially for offshore oil platform activities.

Although in this thesis the proposed methodologies are applied alternatively to 2D or 3D cases, they are general and can be applied both for surface and subsurface accidents. As an overall conclusion of this thesis, responding to the general objective suggested, in Figure 32 the general framework achieved, detailing the different procedures presented in each chapter, is graphically represented. As shown, either for mid-long term oil spill predictions

or for stochastic modeling, a similar methodological flow is proposed for both surface and subsurface release conditions. In case of considering blowouts, an extra step is required consisting in the numerical modeling of the near field processes affecting the first spill stage.

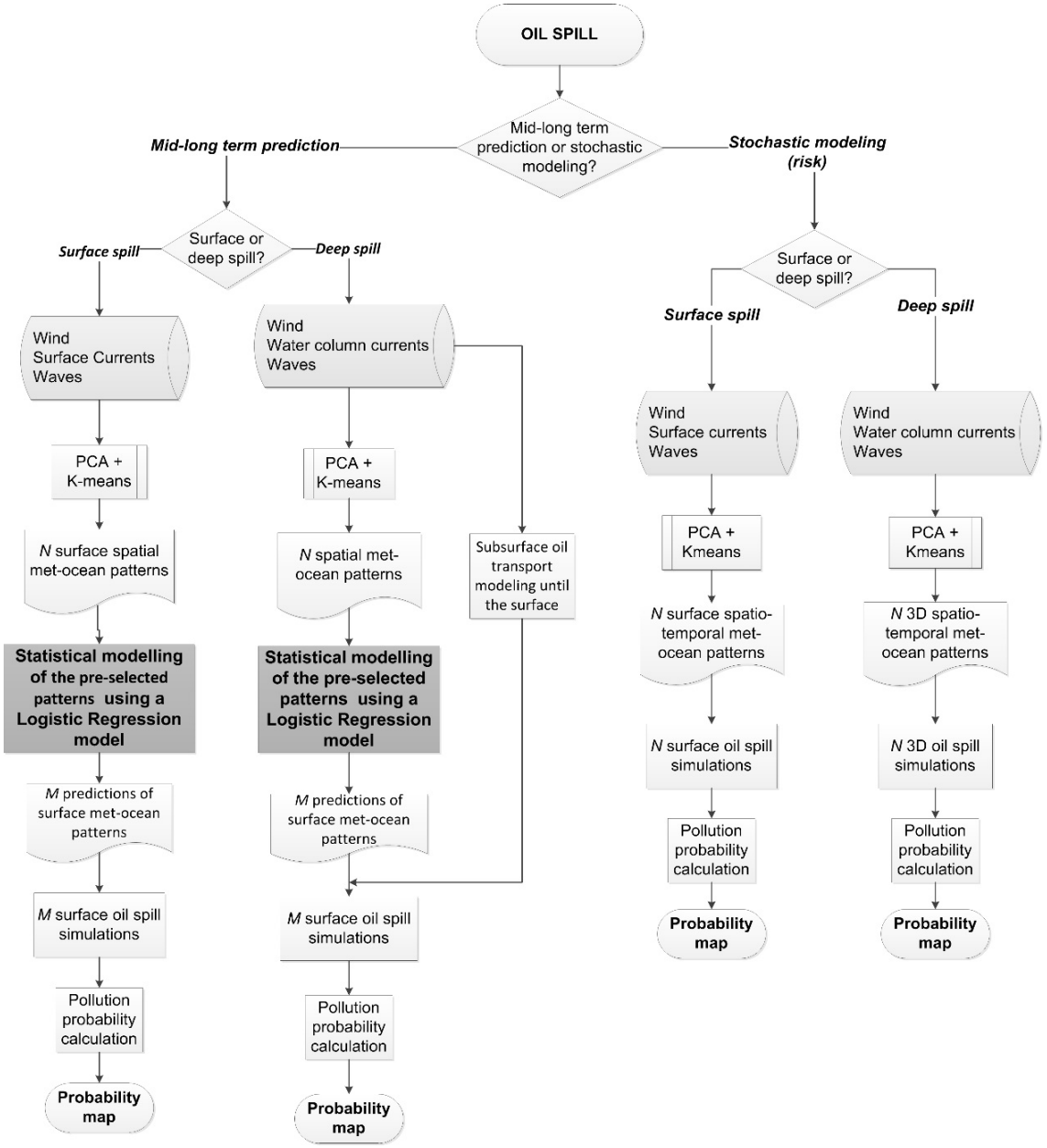

Figure 32. Overall methodological framework proposed in this thesis for the mid-long term statistical simulation, considering both mid-long term predictions and stochastic modeling for oil spill risk assessments.

5.2 Future research

During the development of this thesis, some aspects emerged that would require further investigation and some of the obtained results suggested interesting ideas that could be analyzed in the future:

Regarding the mid-long term oil spill prediction methodology:

- In the application of a logistic regression model, the selection of the number of patterns constituting the response variables plays a key role. So, a method for an objective selection of the number of patterns to be considered for the statistical met-ocean simulation would be beneficial for the proposed methodology.
- The predictive capacity of logistic regression models in the met-ocean field has been shown to be promising. However, further application cases, especially focused on 3D environmental variables, would provide more detailed insides about the capabilities of this technique.
- The methodology for the statistical simulation of met-ocean patterns could be explored when dealing with 3D met-ocean variables (e.g., wind, 3D currents, etc.). This would enhance the possibility of more detailed mid-long term 3D oil spill predictions.
- In line with the previous point, in case of considering blowout accidents, the methodology should be able to include near field modeling and, hence, the inclusions of related relevant variables (e.g. water salinity and temperature).

- The methodology developed for mid-long term forecast of oil spill could be taken as a basis to develop an analogous method for the mid-long term forecasts of marine litters since the main environmental forcings affecting the drift oil slicks and of marine litters are the same.
- The results of the proposed methodology, as seen, are probabilistic. Although in the case of the mid long-term prediction in the Bay of Biscay, the results were compared with a few observed trajectories, a comprehensive validation analysis should be carried on with the deployment and tracking of a considerable amount of drifters. Such an analysis would provide a more detailed source of evaluation of the proposed statistical methods.

Regarding the oil spill stochastic modeling methodology:

- When considering blowout-related oil spills, a blowout model should be implemented to properly simulate the near field processes. To do so, relevant variables for the near field modeling should be integrated in the methodology for environmental pattern extraction.
- In this thesis, the assumption of a small vertical gradient in the subsurface horizontal current velocities is assumed, so to simplify the application of the proposed procedure. For a more realistic subsurface modeling and to further generalized the methodology, subsurface currents from different vertical layers may be included in the methodology.

- Others environmental variables, such as waves and density, could be included in the pattern extraction process if relevant in certain application domain.
- Given the probabilistic nature of the stochastic modeling results, similarly to what proposed for the prediction methodology, disposing of observed trajectory data from deployment exercises involving an high number of drifting buoy would represent a valuable mean to better assess the capabilities of the proposed methodology.

References

Aamo, O.M., Reed, M., Daling, P.S., 1993. A laboratory based weathering model: PC version for coupling to transport models, in: 16th Arctic and Marine Oil Spill Program Technical Seminar. pp. 617–626.

Aamo, O.M., Reed, M., Downing, K., 1997. Oil spill contingency and response (OSCAR) model system: sensitivity studies. Int. Oil Spill Conf. Proc. 1997, 429–438. https://doi.org/10.7901/2169-3358-1997-1-429

Abascal, A.J., Castanedo, S., Gutierrez, A.D., Comerma, E., Medina, R., Losada, I.J., 2007. Teseo , an Operational System for Simulating Oil Spills Trajectories and fate processes. Proc. Int. Offshore Polar Eng. Conf. 1751–1758.

Abascal, A.J., Castanedo, S., Medina, R., Liste, M., 2010. Analysis of the reliability of a statistical oil spill response model. Mar. Pollut. Bull. 60, 2099–2110. https://doi.org/10.1016/j.marpolbul.2010.07.008

Abascal, A.J., Castanedo, S., Medina, R., Losada, I.J., Alvarez-Fanjul, E., 2009a. Application of HF radar currents to oil spill modelling. Mar. Pollut. Bull. 58, 238–248. https://doi.org/10.1016/j.marpolbul.2008.09.020

Abascal, A.J., Castanedo, S., Mendez, F.J., Medina, R., Losada, I.J., 2009b. Calibration of a Lagrangian Transport Model Using Drifting Buoys Deployed during the Prestige Oil Spill. J. Coast. Res. 251, 80–90. https://doi.org/10.2112/07-0849.1

Abascal, A.J., Castanedo, S., Núñez, P., Mellor, A., Clements, A., Pérez, B., Cárdenas, M., Chiri, H., Medina, R., 2016. A high-resolution operational forecast system for oil spill response in Belfast Lough. Mar. Pollut. Bull. 114, 302–314.

https://doi.org/10.1016/j.marpolbul.2016.09.042

Abascal, A.J., Sanchez, J., Chiri, H., Ferrer, M.I., Cárdenas, M., Gallego, A., Castanedo, S., Medina, R., Alonso-Martirena, A., Berx, B., Turrell, W.R., Hughes, S.L., 2017. Operational oil spill trajectory modelling using HF radar currents: A northwest European continental shelf case study. Mar. Pollut. Bull. 119, 336–350. https://doi.org/10.1016/j.marpolbul.2017.04.010

Alvera-Azcárate, A., Barth, A., Weisberg, R.H., 2009. The Surface Circulation of the Caribbean Sea and the Gulf of Mexico as Inferred from Satellite Altimetry. J. Phys. Oceanogr. 39, 640–657. https://doi.org/10.1175/2008JPO3765.1

Antolínez, J.A.A., Méndez, F.J., Camus, P., Vitousek, S., González, E.M., Ruggiero, P., Barnard, P., 2016. A multiscale climate emulator for long-term morphodynamics (MUSCLE-morpho). J. Geophys. Res. Ocean. 121, 775–791. https://doi.org/10.1002/2015JC011107

Aravamudan, K., Raj, P., Ostlund, J., Newman, E., Tucker, W., 1982. Break up of oil on rough seas--simplified models and step-by-step calculation.

ASCE, 1996. State-of-the-Art Review of Modeling Transport and Fate of Oil Spills. J. Hydraul. Eng. 122, 594–609. https://doi.org/10.1061/(ASCE)0733-9429(1996)122:11(594)

Azevedo, A., 2010. Sistema Integrado de Modelação para Apoio à Prevenção e Mitigação de Acidentes de Hidrocarbonetos em Estuários e Orla Costeira. Lisboa, Portugal: Faculdade de Ciências da Universidade de Lisboa.

Azevedo, A., Oliveira, A., Fortunato, A.B., Zhang, J., Baptista, A.M., 2014. A cross-scale numerical modeling system for management support of oil spill accidents. Mar. Pollut. Bull. 80, 132–147. https://doi.org/10.1016/j.marpolbul.2014.01.028

Bandara, U.C., Yapa, P.D., 2011. Bubble Sizes, Breakup, and Coalescence in Deepwater Gas/Oil Plumes. J. Hydraul. Eng. 137, 729–738. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000380

Bárcena, J.F., Camus, P., García, A., Álvarez, C., 2015. Selecting model scenarios of real hydrodynamic forcings on mesotidal and macrotidal estuaries influenced by river discharges using K-means clustering. Environ. Model. Softw. 68, 70–82. https://doi.org/10.1016/j.envsoft.2015.02.007

Barker, C.H., 2011. A Statistical Outlook for the Deepwater Horizon Oil Spill. pp. 237–244. https://doi.org/10.1029/2011GM001129

Berry, A., Dabrowski, T., Lyons, K., 2012. The oil spill model OILTRANS and its application to the Celtic Sea. Mar. Pollut. Bull. 64, 2489–2501. https://doi.org/10.1016/j.marpolbul.2012.07.036

Blaikley, D.R., Dietzel, G.F.L., Glass, A.W., van Kleef, P.J., 1977. "SLIKTRAK"—A computer simulation of offshore oil spills, cleanup, effects and associated costs. Int. Oil Spill Conf. Proc. 1977, 45–52. https://doi.org/10.7901/2169-3358-1977-1-45

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Model. 4, 55–88. https://doi.org/10.1016/S1463-5003(01)00012-9

Borja, A., Amouroux, D., Anschutz, P., Gómez-Gesteira, M., Uyarra, M.C., Valdés, L., 2019. The Bay of Biscay, in: World Seas: An Environmental Evaluation. Elsevier, pp. 113–152. https://doi.org/10.1016/B978-0-12-805068-2.00006-1

Brandvik, P.J., Johansen, Ø., Farooq, U., Glen, A., Leirvik, F., 2014. Subsurface oil releases – Experimental study of droplet distributions and different dispersant injection techniques Version 2. Trondheim Norway.

Brandvik, P.J., Johansen, Ø., Leirvik, F., Farooq, U., Daling, P.S., 2013. Droplet breakup in subsurface oil releases – Part 1: Experimental study of droplet breakup and effectiveness of dispersant injection. Mar. Pollut. Bull. 73, 319–326. https://doi.org/10.1016/j.marpolbul.2013.05.020

Bureau of Ocean Energy Management, 2016. 2016 Update of Occurrence Rates for Offshore Oil Spills.

Caballero, A., Espino, M., Sagarminaga, Y., Ferrer, L., Uriarte, a., González, M., 2008. Simulating the migration of drifters deployed in the Bay of Biscay, during the Prestige crisis. Mar. Pollut. Bull. 56, 475–482. https://doi.org/10.1016/j.marpolbul.2007.11.005

Camus, P., Mendez, F.J., Medina, R., Cofiño, A.S., 2011. Analysis of clustering and selection algorithms for the study of multivariate wave climate. Coast. Eng. 58, 453–462. https://doi.org/10.1016/j.coastaleng.2011.02.003

Camus, P., Menéndez, M., Méndez, F.J., Izaguirre, C., Espejo, A., Cánovas, V., Pérez, J., Rueda, A., Losada, I.J., Medina, R., 2014. A weather-type statistical downscaling framework for ocean wave climate. J. Geophys. Res. Ocean. 119, 7389–7405. https://doi.org/10.1002/2014JC010141

Camus, P., Rueda, A., Méndez, F.J., Losada, I.J., 2016. An atmospheric-to-marine synoptic classification for statistical downscaling marine climate. Ocean Dyn. 66, 1589–1601. https://doi.org/10.1007/s10236-016-1004-5

Camus, P., Herrera, S., Gutiérrez, J.M., Losada, I.J., 2019. Statistical downscaling of seasonal wave forecasts. Ocean Model. 138, 1–12. https://doi.org/10.1016/j.ocemod.2019.04.001

Canu, D.M., Solidoro, C., Bandelj, V., Quattrocchi, G., Sorgente, R., Olita, A., Fazioli, L., Cucco, A., 2015. Assessment of oil slick hazard and risk at vulnerable coastal sites. Mar. Pollut. Bull. 94, 84–95. https://doi.org/10.1016/j.marpolbul.2015.03.006

Cardenas, M., Abascal, A.J., Castanedo, S., Chiri, H., Núñez, P., Antolinez, J.A.A., Mellor, A., Clements, A., Perez-Diaz, B., Medina, R., 2017. Short-term and medium-term forecast of oil spill trajectories: Application to local and regionals scales. Int. Oil Spill Conf. Proc. 2017, 1890–1910. https://doi.org/10.7901/2169-3358-2017.1.1890

Casas-Prat, M., Wang, X.L., Sierra, J.P., 2014. A physical-based statistical method for modeling ocean wave heights. Ocean Model. 73, 59–75. https://doi.org/10.1016/j.ocemod.2013.10.008

Castanedo, S., Medina, R., Losada, I.J., Vidal, C., Méndez, F.J., Osorio, A., Juanes, J.A.,

Puente, A., 2006. 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. 226, 1474–1489. https://doi.org/10.2112/04-0364.1

Charria, G., Lazure, P., Le Cann, B., Serpette, A., Reverdin, G., Louazel, S., Batifoulier, F., Dumas, F., Pichon, A., Morel, Y., 2013. Surface layer circulation derived from Lagrangian drifters in the Bay of Biscay. J. Mar. Syst. 109–110, S60–S76. https://doi.org/10.1016/j.jmarsys.2011.09.015

Chen, F., Yapa, P.D., 2003. A model for simulating deep water oil and gas blowouts - Part II: Comparison of numerical simulations with "Deepspill" field experiments. J. Hydraul. Res. 41, 353–365. https://doi.org/10.1080/00221680309499981

Cid, A., Camus, P., Castanedo, S., Méndez, F.J., Medina, R., 2017. Global reconstructed daily surge levels from the 20th Century Reanalysis (1871–2010). Glob. Planet. Change 148, 9–21. https://doi.org/10.1016/j.gloplacha.2016.11.006

Conil, S., Hall, A., 2006. Local Regimes of Atmospheric Variability: A Case Study of Southern California. J. Clim. 19, 4308–4325. https://doi.org/10.1175/JCLI3837.1

Coppini, G., De Dominicis, M., Zodiatis, G., Lardner, R., Pinardi, N., Santoleri, R., Colella, S., Bignami, F., Hayes, D.R., Soloviev, D., Georgiou, G., Kallos, G., 2011. Hindcast of oil-spill pollution during the Lebanon crisis in the Eastern Mediterranean, July–August 2006. Mar. Pollut. Bull. 62, 140–153. https://doi.org/10.1016/j.marpolbul.2010.08.021

Crowley, D., Mendelsohn, D., Mulanaphy, N.W., Li, Z., Spaulding, M., 2014. Modeling Subsurface Dispersant Applications for Response Planning and Preparation. Int. Oil Spill Conf. Proc. 2014, 933–948. https://doi.org/10.7901/2169-3358-2014.1.933

Cummings, J.A., 2005. Operational multivariate ocean data assimilation. Q. J. R. Meteorol. Soc. 131, 3583–3604. https://doi.org/10.1256/qj.05.105

Daling, P.S., Aamo, O.M., Lewis, A., Strøm-Kristiansen, T., 1997. SINTEF/IKU oil-

weathering model: predicting oils' properties at sea. Int. Oil Spill Conf. Proc. 1997, 297–307. https://doi.org/10.7901/2169-3358-1997-1-297

Daniel, P., Josse, P., Dandin, P., 2004. Forecasting the Prestige oil spills. Proc. Interspill Conf. 402.

Daniel, P., Marty, F., Josse, P., Skandrani, C., Benshila, R., 2003. Improvement of Drift Calculation in Mothy Operational Oil Spill Prediction System. Int. Oil Spill Conf. Proc. 2003, 1067–1072. https://doi.org/10.7901/2169-3358-2003-1-1067

Dasanayaka, L.K., Yapa, P.D., 2009. Role of plume dynamics phase in a deepwater oil and gas release model. J. Hydro-environment Res. 2, 243–253. https://doi.org/10.1016/j.jher.2009.01.004

De Dominicis, M., Pinardi, N., Zodiatis, G., Lardner, R., Archetti, R., 2013. MEDSLIK-II, a Lagrangian marine surface oil spill model for short-term forecasting – Part 1: Theory. Geosci. Model Dev. 6, 1851–1869. https://doi.org/10.5194/gmd-6-1851-2013

De Groot, S., 1996. Quantitative assessment of the development of the offshore oil and gas industry in the North Sea. ICES J. Mar. Sci. 53, 1045–1050. https://doi.org/10.1006/jmsc.1996.0130

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828

Delvigne, G.A.L., 1993. Natural dispersion of oil by different sources of turbulence. Int. Oil Spill Conf. Proc. 1993, 415–419. https://doi.org/10.7901/2169-3358-1993-1-415

Delvigne, G.A.L., Hulsen, J.M., 1994. Simplified laboratory measurement of oil dispersion

coefficient - application in computations of natural oil dispersion, in: Proceedings of the Seventeenth Arctic and Marine Oil Spill Program (AMOP) Technical Seminar. Vancouver, B. C., Environmental Protection Service, Environment Canada, pp. 173–187.

Delvigne, G.A.L., Sweeney, C.E., 1988. Natural dispersion of oil. Oil Chem. Pollut. 4, 281–310. https://doi.org/10.1016/S0269-8579(88)80003-0

Diez-Sierra, J., del Jesus, M., 2019. Subdaily Rainfall Estimation through Daily Rainfall Downscaling Using Random Forests in Spain. Water 11, 125. https://doi.org/10.3390/w11010125

Diez-Sierra, J., del Jesus, M., 2017. A rainfall analysis and forecasting tool. Environ. Model. Softw. 97, 243–258. https://doi.org/10.1016/j.envsoft.2017.08.011

Eckle, P., Burgherr, P., Michaux, E., 2012. Risk of Large Oil Spills: A Statistical Analysis in the Aftermath of Deepwater Horizon. Environ. Sci. Technol. 46, 13002–13008. https://doi.org/10.1021/es3029523

Elliott, A.J., 2004. A probabilistic description of the wind over Liverpool Bay with application to oil spill simulations. Estuar. Coast. Shelf Sci. 61, 569–581. https://doi.org/10.1016/j.ecss.2004.06.022

Fannelop, T., Sjoen, K., 1980. Hydrodynamics of Underwater Blowouts, in: 18th Aerospace Sciences Meeting. American Institute of Aeronautics and Astronautics, Reston, Virigina. https://doi.org/10.2514/6.1980-219

Fay, J.A., 1971. Physical processes in the spread of oil on a water surface. Int. Oil Spill Conf. Proc. 1971, 463–467. https://doi.org/10.7901/2169-3358-1971-1-463

Fay, J.A., 1969. The Spread of Oil Slicks on a Calm Sea, in: Oil on the Sea. Springer US, Boston, MA, pp. 53–63. https://doi.org/10.1007/978-1-4684-9019-0_5

Fingas, M., 2017. Oil Spill Science and Technology. Elsevier. https://doi.org/10.1016/C2015-0-04851-1

Fingas, M., 1996. The evaporation of oil spills: Prediction of equations using distillation data. Spill Sci. Technol. Bull. 3, 191–192. https://doi.org/10.1016/S1353-2561(97)00009-1

Fingas, M.F., 1999. The Evaporation of Oil Spills: Development and Implementation of New Prediction Methodology. Int. Oil Spill Conf. Proc. 1999, 281–287. https://doi.org/10.7901/2169-3358-1999-1-281

Flaherty, G., Murray, D., 2001. Condensate density penalty modification study. Report Prepared for the Canadian Association of Petroleum Producers.

Fogg, P.G.T., Gerrar, W., 1990. Solubility of Gases in Liquids, Chemie Ingenieur Technik. New York. https://doi.org/10.1002/cite.330631231

French-McCay, D., Crowley, D., Rowe, J.J., Bock, M., Robinson, H., Wenning, R., Walker, A.H., Joeckel, J., Nedwed, T.J., Parkerton, T.F., 2018. Comparative Risk Assessment of spill response options for a deepwater oil well blowout: Part 1. Oil spill modeling. Mar. Pollut. Bull. 133, 1001–1015. https://doi.org/10.1016/j.marpolbul.2018.05.042

French-McCay, D.P., 2004. Oil spill impact modeling: Development and validation. Environ. Toxicol. Chem. 23, 2441–2456. https://doi.org/10.1897/03-382

French-McCay, D.P., Jayko, K., Li, Z., Horn, M., Kim, Y., Isaji, T., Crowley, D., Spaulding, M., Decker, L., Turner, C., Zamorki, S., Fontenault, J., Shmookler, R., Rowe, J.J., 2015. Technical Reports for Deepwater Horizon Water Column Injury Assessment – WC_TR14: Modeling Oil Fate and Exposure Concentrations in the Deepwater Plume and Cone of Rising Oil Resulting From the Deepwater Horizon Oil Spill. South Kingstown, RI, USA.

French, D.P., Rines, H.M., 1997. Validation and use of spill impact modeling for impact assessment. Int. Oil Spill Conf. Proc. 1997, 829–834. https://doi.org/10.7901/2169-3358-1997-1-829

García-Ladona, E., Font, J., Río, E. del, Julià, A., Salat, J., Chic, O., Orfila, A., Alvarez, A.,

Basterretxea, G., Vizoso, G., Piro, O., Tintoré, J., Castanedo, S., Coto, M.G., Herrera, J.L., 2005. The use of surface drifting floats in the monitoring of oil spills. The Prestige case. Int. Oil Spill Conf. Proc. 2005, 613–617. https://doi.org/10.7901/2169-3358-2005-1-613

Girin, M., Cabióch, F., Peigné, G., Gouriou, V., Abascal, A.J., Castanedo, S., Parthiot, F., 2006. Synthese sur la contribution des produits et services satellitaires a la prevision de derive de nappes d'hydrocarbures pour la lutte en mer et a l'information de la terre. Sécrétariat Général pour les Affaires Régionales. INTERREG III B-Post Prestige.

Girin, M., Cabióch, F., Peigné, G., Uribe, J., Menéndez, J., Pozo, R., González, M., Uriarte, A., Castanedo, S., 2004. Exchanging Information for Improved Response and Public Communication in a Transboundary Oil Spill: The Prestige Experience. Interspill 2004 403.

Goerlandt, F., Kujala, P., 2011. Traffic simulation based ship collision probability modeling. Reliab. Eng. Syst. Saf. 96, 91–107. https://doi.org/10.1016/j.ress.2010.09.003

González, J.J., Viñas, L., Franco, M.A., Fumega, J., Soriano, J.A., Grueiro, G., Muniategui, S., López-Mahía, P., Prada, D., Bayona, J.M., Alzaga, R., Albaigés, J., 2006. Spatial and temporal distribution of dissolved/dispersed aromatic hydrocarbons in seawater in the area affected by the Prestige oil spill. Mar. Pollut. Bull. 53, 250–259. https://doi.org/10.1016/j.marpolbul.2005.09.039

Gouriou, V., Cabon, R., Gieu, V., Girardot, A., Lepetit, A., 2004. Data management system concerning both pollution monitoring and response operations during an oil spill : the Prestige experience. Interspill 2004 405.

Grace, J.R., Wairegi, T., Brophy, J., 1978. Break-up of drops and bubbles in stagnant media. Can. J. Chem. Eng. 56, 3–8. https://doi.org/10.1002/cjce.5450560101

Guanche, Y., Mínguez, R., Méndez, F.J., 2014. Autoregressive logistic regression applied to atmospheric circulation patterns. Clim. Dyn. 42, 537–552.

https://doi.org/10.1007/s00382-013-1690-3

Guo, W., Hao, Y., Zhang, L., Xu, T., Ren, X., Cao, F., Wang, S., 2014. Development and application of an oil spill model with wave–current interactions in coastal areas. Mar. Pollut. Bull. 84, 213–224. https://doi.org/10.1016/j.marpolbul.2014.05.009

Gutiérrez, J.M., San-Martín, D., Brands, S., Manzanas, R., Herrera, S., 2013. Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions. J. Clim. 26, 171–188. https://doi.org/10.1175/JCLI-D-11-00687.1

Hastie, T., Friedman, J., Tibshirani, R., 2001. Model Assessment and Selection. pp. 193–224. https://doi.org/10.1007/978-0-387-21606-5_7

Hong, S.-Y., Pan, H.-L., 1998. Convective Trigger Function for a Mass-Flux Cumulus Parameterization Scheme. Mon. Weather Rev. 126, 2599–2620. https://doi.org/10.1175/1520-0493(1998)126<2599:CTFFAM>2.0.CO;2

Hosmer, D.W., Lemeshow, S., 2000. Applied logistic regression (Second edition). John Wiley & Sons, Canada.

Hou, Y.T. , Campana, K.A., Yang, S.K., 1996. Shortwave radiation calculations in the NCEP's global model. International Radiation Symposium, IRS-96, August 19-24, Fairbanks, AL.

Hsieh, C.H.H., Chiu, T.S., Shih, C.T., 2004. Copepod diversity and composition as indicators of intrusion of the Kuroshio Branch Current into the Northern Taiwan Strait in Spring 2000. Zool. Stud. 43. https://doi.org/http://zoolstud.sinica.edu.tw/Journals/43.2/393.pdf

Huang, J.C., 2005. A review of the state-of-the-art of oil spill fate/behavior models. 2005 Int. Oil Spill Conf. IOSC 2005 7275.

Hunter, J.R., Craig, P.D., Phillips, H.E., 1993. On the use of random walk models with spatially variable diffusivity. J. Comput. Phys. 106, 366–376. https://doi.org/10.1016/S0021-9991(83)71114-9

Huthnance, J., 1991. Physical oceanography of the North Sea. Ocean Shorel. Manag. 16, 199–231. https://doi.org/10.1016/0951-8312(91)90005-M

ICES, 1983. Flushing times of the North Sea. Copenaghen.

IPIECA, IOGP, 2013. Oil spill risk assessment and response planning for offshore installations.

ITOPF, 2011a. TIP 16: Contingency Planning for Marine Oil Spills. ITOPF Tech. Inf. Pap. 12. https://doi.org/10.1021/acs.biochem.6b00280

ITOPF, 2011b. Fate of marine oil spills.

Janeiro, J., Fernandes, E., Martins, F., Fernandes, R., 2008. Wind and freshwater influence over hydrocarbon dispersal on Patos Lagoon, Brazil. Mar. Pollut. Bull. 56, 650–665. https://doi.org/10.1016/j.marpolbul.2008.01.011

Janeiro, J., Zacharioudaki, A., Sarhadi, E., Neves, A., Martins, F., 2014. Enhancing the management response to oil spills in the Tuscany Archipelago through operational modelling. Mar. Pollut. Bull. 85, 574–589. https://doi.org/10.1016/j.marpolbul.2014.03.021

Ji, Z.-G., Johnson, W.R., Li, Z., 2011. Oil Spill Risk Analysis Model and Its Application to the Deepwater Horizon Oil Spill Using Historical Current and Wind Data. pp. 227–236. https://doi.org/10.1029/2011GM001117

Jirka, G.H., 2004. Integral Model for Turbulent Buoyant Jets in Unbounded Stratified Flows. Part I: Single Round Jet. Environ. Fluid Mech. 4, 1–56.

Johansen, Ø., 2003. Development and verification of deep-water blowout models. Mar. Pollut. Bull. 47, 360–368. https://doi.org/10.1016/S0025-326X(03)00202-9

Johansen, Ø., 2000. DeepBlow – a Lagrangian Plume Model for Deep Water Blowouts. Spill Sci. Technol. Bull. 6, 103–111. https://doi.org/10.1016/S1353-2561(00)00042-6

Johansen, Ø., Brandvik, P.J., Farooq, U., 2013. Droplet breakup in subsea oil releases – Part 2: Predictions of droplet size distributions with and without injection of chemical

dispersants. Mar. Pollut. Bull. 73, 327–335. https://doi.org/10.1016/j.marpolbul.2013.04.012

Jolliffe, I.T., 2002. Principal component analysis, second edition. New York.

Jones, R.K., 1997. A simplified pseudo-component oil evaporation model, in: 20. Arctic and Marine Oil Spill Program Technical Seminar. Ottawa, Ontario, Canada, pp. 43–61.

Kersalé, M., Marié, L., Le Cann, B., Serpette, A., Lathuilière, C., Le Boyer, A., Rubio, A., Lazure, P., 2016. Poleward along-shore current pulses on the inner shelf of the Bay of Biscay. Estuar. Coast. Shelf Sci. 179, 155–171. https://doi.org/10.1016/j.ecss.2015.11.018

Khelifa, A., Charron, D.M., Tong, T., Roshan Singh, N., 2011. Effects of chemical dispersants on oil-brine interfacial tension and droplet formation: new results using new high resolution imaging setup and weathered oils, in: 34. AMOP Technical Seminar on Environmental Contamination and Response. Ottawa, Ontario, Canada.

Khelifa, A., So, L.L.C., 2009. Effects of chemical dispersants on oil-brine interfacial tension and droplet formation, in: 32 AMOP Technical Seminar on Environmental Contamination and Response. Environment Canada, Ottawa, ON, p. 940.

Kim, T.-S., Park, K.-A., Li, X., Lee, M., Hong, S., Lyu, S.J., Nam, S., 2015. Detection of the Hebei Spirit oil spill on SAR imagery and its temporal evolution in a coastal region of the Yellow Sea. Adv. Sp. Res. 56, 1079–1093. https://doi.org/10.1016/j.asr.2015.05.040

Kim, Y.-J., Arakawa, A., 1995. Improvement of Orographic Gravity Wave Parameterization Using a Mesoscale Gravity Wave Model. J. Atmos. Sci. 52, 1875–1902. https://doi.org/10.1175/1520-0469(1995)052

Kolpack, R.L., Plutchak, N.B., Stearns, R.W., 1977. Fate of oil in a water environment—Phase II, a dynamic model of the mass balance for released oil. Washington, DC.

Lazure, P., Jegou, A.-M., 1998. 3D modelling of seasonal evolution of Loire and Gironde plumes on Biscay Bay continental shelf. Oceanol. Acta 21, 165–177. https://doi.org/10.1016/S0399-1784(98)80006-6

Lee, M., Jung, J.-Y., 2015. Pollution risk assessment of oil spill accidents in Garorim Bay of Korea. Mar. Pollut. Bull. 100, 297–303. https://doi.org/10.1016/j.marpolbul.2015.08.037

Lefebvre, A.H., McDonell, V.G., 2017. Atomization and Sprays. CRC Press. https://doi.org/10.1201/9781315120911

Lehr, W., Jones, R., Evans, M., Simecek-Beatty, D., Overstreet, R., 2002. Revisions of the ADIOS oil spill model. Environ. Model. Softw. 17, 189–197. https://doi.org/10.1016/S1364-8152(01)00064-0

Lehr, W.J., Fraga, R.J., Belen, M.S., Cekirge, H.M., 1984. A new technique to estimate initial spill size using a modified fay-type spreading formula. Mar. Pollut. Bull. 15, 326–329. https://doi.org/10.1016/0025-326X(84)90488-0

Lehr, W.J., Overstreet, R., Jones, R., Watabayashi, G., 1992. ADIOS-Automatic Data Inquire for Oil Spill, in: 15th Arctic Marine Oilspill Program, Technical Seminar. Ottawa, Ontario, pp. 31–45.

Lehr, W.J., Wesley, D., Simecek-Beatty, D., Jones, R., Kachook, G., Lankford, J., 2000. Algorithm and interface modifications of the NOAA oil spill behavior model, in: 23rd Arctic and Marine Oil Spill Program (AMOP) Technical Seminar. Vancouver, BC, pp. 525–539.

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., Benkiran, M., Testut, C.-E., Bourdalle-Badie, R., Gasparin, F., Hernandez, O., Levier, B., Drillet, Y., Remy, E., Le Traon, P.-Y., 2018. Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1⁄12° high-resolution system. Ocean Sci. 14, 1093–1126. https://doi.org/10.5194/os-14-1093-2018

Li, Z., Bird, A., Payne, J.R., Vinhateiro, N., Kim, Y., Davis, C., Loomis, N., 2015. Technical Reports for Deepwater Horizon Water Column Injury Assessment: Oil Particle Data from the Deepwater Horizon Oil Spill. South Kingstown, RI, USA.

Li, Z., Spaulding, M., French McCay, D., Crowley, D., Payne, J.R., 2017a. Development of a unified oil droplet size distribution model with application to surface breaking waves and subsea blowout releases considering dispersant effects. Mar. Pollut. Bull. 114, 247–257. https://doi.org/10.1016/j.marpolbul.2016.09.008

Li, Z., Spaulding, M.L., French-McCay, D., 2017b. An algorithm for modeling entrainment and naturally and chemically dispersed oil droplet size distribution under surface breaking wave conditions. Mar. Pollut. Bull. 119, 145–152. https://doi.org/10.1016/j.marpolbul.2017.03.048

Liu, Y., Weisberg, R.H., 2011. Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. J. Geophys. Res. 116, C09013. https://doi.org/10.1029/2010JC006837

Liu, Y., Weisberg, R.H., Hu, C., Zheng, L., 2011. Tracking the Deepwater Horizon Oil Spill: A Modeling Perspective. Eos, Trans. Am. Geophys. Union 92, 45–46. https://doi.org/10.1029/2011EO060001

Liu, Y., Weisberg, R.H., Vignudelli, S., Mitchum, G.T., 2016. Patterns of the loop current system and regions of sea surface height variability in the eastern Gulf of Mexico revealed by the self□organizing maps. J. Geophys. Res. Ocean. 121, 2347–2366. https://doi.org/10.1002/2015JC011493

Liu, Y., Weisberg, R.H., Yuan, Y., 2008. Patterns of upper layer circulation variability in the South China Sea from satellite altimetry using the Self-Organizing Map. Acta Oceanol. Sin. 27, 129–144.

Liubartseva, S., De Dominicis, M., Oddo, P., Coppini, G., Pinardi, N., Greggio, N., 2015. Oil spill hazard from dispersal of oil along shipping lanes in the Southern Adriatic and Northern Ionian Seas. Mar. Pollut. Bull. 90, 259–272.

https://doi.org/10.1016/j.marpolbul.2014.10.039

Lugo-Fernández, A., 2007. Is the Loop Current a Chaotic Oscillator? J. Phys. Oceanogr. 37, 1455–1469. https://doi.org/10.1175/JPO3066.1

Lynch, D.R., Greenberg, D.A., Bilgili, A., McGillicuddy, J., Manning, J.P., Aretxabaleta, A.L., 2015. Particles in the Coastal Ocean Theory and Applications. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107449336

Mackay, D., Buist, I.A., Mascarenhas, R., Paterson, S., 1980. Oil spill processes and models. Oil spill processes and models.

Mackay, D., Trudel, K., S., P., 1980. A Mathematical Model of Oil Spill Behaviour. Canada. Environmental Emergency Branch. Research and Development Division, Canada. Environmental Impact Control Directorate.

Mackay, D., Zagorski, W., 1982. Studies of water-in-oil emulsions. Ottawa, Ontario, Canada.

Martins, F., Leitao, P., Silva, A., Neves, R., 2001. 3D modelling in the Sado estuary using a new generic vertical discretization approach. Oceanol. Acta 24, 51–62. https://doi.org/10.1016/S0399-1784(01)00092-5

Mcdougall, T.J., 1978. Bubble plumes in stratified environments. J. Fluid Mech. 85, 655–672. https://doi.org/10.1017/S0022112078000841

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2006. North American Regional Reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360. https://doi.org/10.1175/BAMS-87-3-343

Mishra, N., Soni, H.K., Sharma, S., Upadhyay, A.K., 2017. A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction. J. ICT Res. Appl. 11, 168. https://doi.org/10.5614/itbj.ict.res.appl.2017.11.2.4

Montero, P., Blanco, J., Cabanas, J., Maneiro, J., Pazos, Y., Moroño, A., Balseiro, C.F., Carracedo, P., Gomez, B., Penabad, E., Pérez-Muñuzuri, V., Braunschwieg, F., Fernandes, R., Leitao, P.C., Neves, R., 2003. Oil Spill Monitoring and Forecasting on the Prestige-Nassau Accident, in: Proceedings 26th Artic and Marine Oil Spill Program (AMOP) Technical Seminar, Environment Canada, Ottawa, Canada, Vol. 2. pp. 1013–1029.

Muller-Karger, F.E., Smith, J.P., Werner, S., Chen, R., Roffer, M., Liu, Y., Muhling, B., Lindo-Atichati, D., Lamkin, J., Cerdeira-Estrada, S., Enfield, D.B., 2015. Natural variability of surface oceanographic conditions in the offshore Gulf of Mexico. Prog. Oceanogr. 134, 54–76. https://doi.org/10.1016/j.pocean.2014.12.007

NOAA, 1994. ADIOS, Automated data inquiry for oil spills, User's Manual. Seattle, Washington, DC.

Noreng, O., 2016. The Oil Industry and Government Strategy in the North Sea. Routledge.

Núñez, P., García, A., Mazarrasa, I., Juanes, J.A., Abascal, A.J., Méndez, F., Castanedo, S., Medina, R., 2019. A methodology to assess the probability of marine litter accumulation in estuaries. Mar. Pollut. Bull. 144, 309–324. https://doi.org/10.1016/j.marpolbul.2019.04.077

Oey, L.-Y., Lee, H.-C., Schmitz, W.J., 2003. Effects of winds and Caribbean eddies on the frequency of Loop Current eddy shedding: A numerical model study. J. Geophys. Res. 108, 3324. https://doi.org/10.1029/2002JC001698

OGA, 2018a. UK Oil & Gas Authority, Wells Insight Report 2018. Available at: https://www.ogauthority.co.uk/media/5107/oga_wells_insight_report_2018.pdf

OGA, 2018b. UK Oil & Gas Authority, Environment Report 2018. Available at: https://oilandgasuk.co.uk/wp-content/uploads/2019/05/OGUK-Environment-Report-2018.pdf

Omrani, H., Arsouze, T., Béranger, K., Boukthir, M., Drobinski, P., Lebeaupin-Brossier, C., Mairech, H., 2016. Sensitivity of the sea circulation to the atmospheric forcing in

the Sicily Channel. Prog. Oceanogr. 140, 54–68. https://doi.org/10.1016/j.pocean.2015.10.007

OSPAR, 2015. Update of the inventory of oil and gas Offshore Installations in the OSPAR Maritime Area. OSPAR Commission, Offshore Industry Series, London, UK.

Palomar, P., Lara, J.L., Losada, I.J., Tarrade, L., 2013. Numerical modeling of brine discharge: commercial models, MEDVSA online simulation tools and advanced computational fluid dynamics. Desalin. Water Treat. 51, 543–559. https://doi.org/10.1080/19443994.2012.714625

Payne, J.R., Driskell, W.B., 2015. Offshore Adaptive Sampling Strategies. Encinitas, CA, USA.

Payne, J.R., Kirstein, B.E., McNabb, G.D., Lambach, J.L., de Oliveira, C., Jordan, R.E., Hom, W., 1983. Multivariate analysis of petroleum hydrocarbon weathering in the subarctic marine environment 1. Int. Oil Spill Conf. Proc. 1983, 423–434. https://doi.org/10.7901/2169-3358-1983-1-423

Pedragosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dunborg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Cuchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://doi.org/10.1145/2786984.2786995

Perez, J., Menendez, M., Camus, P., Mendez, F.J., Losada, I.J., 2015. Statistical multi-model climate projections of surface ocean waves in Europe. Ocean Model. 96, 161–170. https://doi.org/10.1016/j.ocemod.2015.06.001

Peterson, C.H., 2003. Long-Term Ecosystem Response to the Exxon Valdez Oil Spill. Science (80-.). 302, 2082–2086. https://doi.org/10.1126/science.1084282

Phillipson, L., Toumi, R., 2017. Impact of data assimilation on ocean current forecasts in the Angola Basin. Ocean Model. 114, 45–58. https://doi.org/10.1016/j.ocemod.2017.04.006

Picou, J.S., Formichella, C., Marshall, B.K., Arata, C., 2009. Community Impacts of the Exxon Valdez Oil Spill : A Synthesis and Elaboration of Social Science Research. Synth. Three Decad. Res. Socioecon. Eff. Relat. to Offshore Pet. Dev. Coast. Alaska 21.

Pingree, R.D., 1993. Flow of surface waters to the west of the British Isles and in the Bay of Biscay. Deep Sea Res. Part II Top. Stud. Oceanogr. 40, 369–388. https://doi.org/10.1016/0967-0645(93)90022-F

Pingree, R.D., Le Cann, B., 1992a. Anticyclonic eddy X91 in the southern Bay of Biscay, May 1991 to February 1992. J. Geophys. Res. 97, 14353. https://doi.org/10.1029/92JC01181

Pingree, R.D., Le Cann, B., 1992b. Three anticyclonic slope water oceanic eDDIES (SWODDIES) in the Southern Bay of Biscay in 1990. Deep Sea Res. Part A. Oceanogr. Res. Pap. 39, 1147–1175. https://doi.org/10.1016/0198-0149(92)90062-X

Pingree, R.D., Le Cann, B., 1990. Structure, strength and seasonality of the slope currents in the Bay of Biscay region. J. Mar. Biol. Assoc. United Kingdom 70, 857. https://doi.org/10.1017/S0025315400059117

Pingree, R.D., Le Cann, B., 1989. Celtic and Armorican slope and shelf residual currents. Prog. Oceanogr. 23, 303–338. https://doi.org/10.1016/0079-6611(89)90003-7

Powers, D.M.., 2011. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. J. Mach. Learn. Technol. 2, 37–63.

Prasad, K., Dash, S.K., Mohanty, U.C., 2009. A logistic regression approach for monthly rainfall forecasts in meteorological subdivisions of India based on DEMETER retrospective forecasts. Int. J. Climatol. n/a-n/a. https://doi.org/10.1002/joc.2019

Proctor, R., Flather, R.A., Elliott, A.J., 1994. Modelling tides and surface drift in the Arabian Gulf—application to the Gulf oil spill. Cont. Shelf Res. 14, 531–545. https://doi.org/10.1016/0278-4343(94)90102-3

Reed, M., Aamo, O.M., Daling, P.S., 1995. Quantitative analysis of alternate oil spill response strategies using OSCAR. Spill Sci. Technol. Bull. 2, 67–74. https://doi.org/10.1016/1353-2561(95)00020-5

Reed, M., Daling, P.S., Brakstad, O.G., Singsaas, I., Faksness, L.G., Hetland, B., Ekrol, N., 2000. OSCAR2000 : a multi-component 3-dimensional oil spill contingency and response model, in: 23. Arctic and Marine Oilspill Program (AMOP) Technical Seminar.

Reed, M., Johansen, Ø., Leirvik, F., Brors, B., 2009. Numerical Algorithm to Compute the Effects of Breaking Waves on Surface Oil Spilled at Sea. Trondheim Norway.

Reed, M., Rye, H., 1995. A THREE-DIMENSIONAL OIL AND CHEMICAL SPILL MODEL FOR ENVIRONMENTAL IMPACT ASSESSMENT. Int. Oil Spill Conf. Proc. 1995, 61–66. https://doi.org/10.7901/2169-3358-1995-1-61

Röhrs, J., Christensen, K.H., Hole, L.R., Broström, G., Drivdal, M., Sundby, S., 2012. Observation-based evaluation of surface wave effects on currents and trajectory forecasts. Ocean Dyn. 62, 1519–1533. https://doi.org/10.1007/s10236-012-0576-y

Romero, A.F., Abessa, D.M.S., Fontes, R.F.C., Silva, G.H., 2013. Integrated assessment for establishing an oil environmental vulnerability map: Case study for the Santos Basin region, Brazil. Mar. Pollut. Bull. 74, 156–164. https://doi.org/10.1016/j.marpolbul.2013.07.012

Rueda, A., Vitousek, S., Camus, P., Tomás, A., Espejo, A., Losada, I.J., Barnard, P.L., Erikson, L.H., Ruggiero, P., Reguero, B.G., Mendez, F.J., 2017. A global classification of coastal flood hazard climates associated with large-scale oceanographic forcing. Sci. Rep. 7, 5038. https://doi.org/10.1038/s41598-017-05090-w

Rye, H., 1994. Model for calculation of underwater blow-out plume.

Rye, H., Reed, M., Brandvik, P.J., 1996. Subsurface oil release field experiment - observations and modelling of subsurface plume behaviour, in: Nineteenth Arctic and

Marine Oilspill Program (AMOP) Technical Seminar. Ottawa, Ontario, Canada.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, Jiande, Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, Jun, Hou, Y.-T., Chuang, H., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G., Goldberg, M., 2010. The NCEP Climate Forecast System Reanalysis. Bull. Am. Meteorol. Soc. 91, 1015–1058. https://doi.org/10.1175/2010BAMS3001.1

Shinoda, T., Zamudio, L., Guo, Y., Metzger, E.J., Fairall, C.W., 2019. Ocean variability and air-sea fluxes produced by atmospheric rivers. Sci. Rep. 9, 2152. https://doi.org/10.1038/s41598-019-38562-2

Sloan, E.D., Dekker, J.M., 1998. Clathrate Hydrates of Natural Gases, 2nd Ed Revised and Expanded, Energy & Fuels. New York. https://doi.org/10.1021/ef000056e

Smith, L.C., Smith, M., Ashcroft, P., 2010. Analysis of Environmental and Economic Damages from British Petroleum's Deepwater Horizon Oil Spill. SSRN Electron. J. https://doi.org/10.2139/ssrn.1653078

Socolofsky, S.A., Adams, E.E., 2002. Multi-phase plumes in uniform and stratified crossflow. J. Hydraul. Res. 40, 661–672. https://doi.org/10.1080/00221680209499913

Socolofsky, S.A., Adams, E.E., Boufadel, M.C., Aman, Z.M., Johansen, Ø., Konkel, W.J., Lindo, D., Madsen, M.N., North, E.W., Paris, C.B., Rasmussen, D., Reed, M., Rønningen, P., Sim, L.H., Uhrenholdt, T., Anderson, K.G., Cooper, C., Nedwed, T.J., 2015. Intercomparison of oil spill prediction models for accidental blowout scenarios with and without subsea chemical dispersant injection. Mar. Pollut. Bull. 96, 110–126. https://doi.org/10.1016/j.marpolbul.2015.05.039

Solabarrieta, L., Rubio, A., Cárdenas, M., Castanedo, S., Esnaola, G., Méndez, F.J.,

Medina, R., Ferrer, L., 2015. Probabilistic relationships between wind and surface water circulation patterns in the SE Bay of Biscay. Ocean Dyn. 65, 1289–1303. https://doi.org/10.1007/s10236-015-0871-5

Solidoro, C., Bandelj, V., Barbieri, P., Cossarini, G., Fonda Umani, S., 2007. Understanding dynamic of biogeochemical properties in the northern Adriatic Sea by using self-organizing maps and k-means clustering. J. Geophys. Res. 112, C07S90. https://doi.org/10.1029/2006JC003553

Sönmez, İ., Kömüşcü, A.Ü., 2011. Reclassification of rainfall regions of Turkey by K-means methodology and their temporal variability in relation to North Atlantic Oscillation (NAO). Theor. Appl. Climatol. 106, 499–510. https://doi.org/10.1007/s00704-011-0449-1

Sotillo, M.G., Fanjul, E.A., Castanedo, S., Abascal, A.J., Menendez, J., Emelianov, M., Olivella, R., García-Ladona, E., Ruiz-Villarreal, M., Conde, J., Gómez, M., Conde, P., Gutierrez, A.D., Medina, R., 2008. Towards an operational system for oil-spill forecast over Spanish waters: Initial developments and implementation test. Mar. Pollut. Bull. 56, 686–703. https://doi.org/10.1016/j.marpolbul.2007.12.021

Spaulding, M.L., 2017. State of the art review and future directions in oil spill modeling. Mar. Pollut. Bull. 115, 7–19. https://doi.org/10.1016/j.marpolbul.2017.01.001

Spaulding, M.L., 1988. A state-of-the-art review of oil spill trajectory and fate modeling. Oil Chem. Pollut. 4, 39–55. https://doi.org/10.1016/S0269-8579(88)80009-1

Spaulding, M.L., Bishnoi, P.R., Anderson, E., Isaji, T., 2000. An Integrated Model for Prediction of Oil Transport from a Deep Water Blowout, in: 23rd, Arctic and Marine Oilspill Program. Environment Canada, Vancouver, Canada, pp. 611–636.

Spaulding, M.L., Howlett, E., Anderson, E., Jayko, K., 1992. OILMAP: a global approach to spill modeling, in: 15th Arctic and Marine Oil Spill Program, Technical Seminar. Edmonton, Alberta, Canada, pp. 15–1.

Spaulding, M.L., Saila, S.B., Lorda, E., Walker, H., Anderson, E., Swanson, J.C., 1983.

Oil-spill fishery impact assessment model: Application to selected Georges Bank fish species. Estuar. Coast. Shelf Sci. 16, 511–541. https://doi.org/10.1016/0272-7714(83)90083-5

Stiver, W., Mackay, D., 1984. Evaporation rate of spills of hydrocarbons and petroleum mixtures. Environ. Sci. Technol. 18, 834–840. https://doi.org/10.1021/es00129a006

Stolzenbach, K.T., Madsen, O.S., Adams, E.E., Pollack, A.M., Cooper, C., 1977. A review and evaluation of basic techniques for predicting the behavior of surface oil slicks. Cambridge.

Sun, C., Mao, D., Zhao, T., Shang, X., Wang, Y., Duan, M., 2015. Investigate Deepwater Pipeline Oil Spill Emergency Repair Methods. Aquat. Procedia 3, 191–196. https://doi.org/10.1016/j.aqpro.2015.02.210

Sündermann, J., 2003. The changing North Sea: Knowledge, speculation and new challenges. Oceanologia 45, 247–259.

Surís-Regueiro, J.C., Garza-Gil, M.D., Varela-Lafuente, M.M., 2007. The Prestige oil spill and its economic impact on the Galician fishing sector. Disasters 31, 201–215. https://doi.org/10.1111/j.1467-7717.2007.01004.x

Swann, C., 2007. The History of Oilfield Diving: An Industrial Adventure.

Tang, L., 2004. Cylindrical liquid-liquid jet instability. University of Hawai'i.

Taylor, A.H., 1987. Modelling contaminants in the North Sea. Sci. Total Environ. 63, 45–67. https://doi.org/10.1016/0048-9697(87)90035-0

Uppala, S.M., K□llberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Berg, L. Van De, Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.P., Mahfouf, J.-F., Morcrette, J.-

J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 re-analysis. Q. J. R. Meteorol. Soc. 131, 2961–3012. https://doi.org/10.1256/qj.04.176

Venkataraman, P., Tang, J., Frenkel, E., McPherson, G.L., He, J., Raghavan, S.R., Kolesnichenko, V., Bose, A., John, V.T., 2013. Attachment of a Hydrophobically Modified Biopolymer at the Oil–Water Interface in the Treatment of Oil Spills. ACS Appl. Mater. Interfaces 5, 3572–3580. https://doi.org/10.1021/am303000v

Vignudelli, S., Kostianoy, A.G., Cipollini, P., Benveniste, J., 2011. Coastal altimetry, First edit. ed. Springer, New York.

Wang, X.L., Feng, Y., Swail, V.R., 2012. North Atlantic wave height trends as reconstructed from the 20th century reanalysis. Geophys. Res. Lett. 39. https://doi.org/10.1029/2012GL053381

Weisberg, R.H., Lianyuan, Z., Liu, Y., 2017. On the movement of Deepwater Horizon Oil to northern Gulf beaches. Ocean Model. 111, 81–97. https://doi.org/10.1016/j.ocemod.2017.02.002

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194. https://doi.org/10.1080/02723646.1981.10642213

Winther, N.G., Johannessen, J.A., 2006. North Sea circulation: Atlantic inflow and its destination. J. Geophys. Res. 111, C12018. https://doi.org/10.1029/2005JC003310

Yapa, P.D., Li, Z., 1997. Simulation of oil spills from underwater accidents I: Model development. J. Hydraul. Res. 35, 673–688. https://doi.org/10.1080/00221689709498401

Yapa, P.D., Zheng, L., Chen, F., 2001. A Model for Deepwater Oil/Gas Blowouts. Mar. Pollut. Bull. 43, 234–241. https://doi.org/10.1016/S0025-326X(01)00086-8

Yi, J., Du, Y., Wang, X., He, Z., Zhou, C., 2013. A clustering analysis of eddies' spatial distribution in the South China Sea. Ocean Sci. 9, 171–182. https://doi.org/10.5194/os-

9-171-2013

Zafirakou, A., 2019. Oil Spill Dispersion Forecasting Models, in: Monitoring of Marine Pollution. IntechOpen. https://doi.org/10.5772/intechopen.81764

Zelenke, B., O'Connor, C., Barker, C., Beegle-Krause, C.J., 2012. General NOAA Operational Modeling Environment (GNOME) Technical Documentation. Seattle, WA.

Zeng, X., Li, Y., He, R., Yin, Y., 2015. Clustering of Loop Current patterns based on the satellite-observed sea surface height and self-organizing map. Remote Sens. Lett. 6, 11–19. https://doi.org/10.1080/2150704X.2014.998347

Zhao, L., Boufadel, M.C., Socolofsky, S.A., Adams, E., King, T., Lee, K., 2014. Evolution of droplets in subsea oil and gas blowouts: Development and validation of the numerical model VDROP-J. Mar. Pollut. Bull. 83, 58–69. https://doi.org/10.1016/j.marpolbul.2014.04.020

Zhao, Q., Carr, F.H., 1997. A Prognostic Cloud Scheme for Operational NWP Models. Mon. Weather Rev. 125, 1931–1953. https://doi.org/10.1175/1520-0493(1997)125

Zheng, L., Yapa, P.D., 2002. Modeling gas dissolution in deepwater oil/gas spills. J. Mar. Syst. 31, 299–309. https://doi.org/10.1016/S0924-7963(01)00067-7

Zheng, L., Yapa, P.D., Chen, F., 2003. A model for simulating deepwater oil and gas blowouts - Part I: Theory and model formulation. J. Hydraul. Res. 41, 339–351. https://doi.org/10.1080/00221680309499980

A APPENDIX.

IMPROVEMENTS AND DEVELOPMENTS OF NUMERICAL TOOLS FOR OIL SPILL SIMULATION.

A1.Introduction

Based on the oil spill numerical modeling state of the art presented in Chapter 1, we present here the developments achieved with this thesis in oil spill numerical modeling. The final aim of these advances is to allow more accurate numerical simulations of oil spills in the water column and at the sea surface, including those spills originated by underwater oil wells, i.e. blowouts. The achieved developments may be organized in two different main components of oil spill modeling: the far field modeling, including weathering processes and the 3D oil transport due to environmental forcings and oil properties, and the near field modeling, required in case of blowout simulations to solve those specific processes occurring in the first phase of a deep spill, mainly driven by the spilling flux momentum and its properties. Concerning the advances in far field oil spill modeling, the oil spill model TESEO

(Abascal et al., 2007) has been updated so to allow, in first place, the simulation of the 3D evolution of a spill through the whole water column and, in second place, to include new relevant oil spill processes aiming to achieve more realistic spill simulations. Regarding near field modeling, a first-stage blowout model has been developed to simulate the main features of a jet plume hydrodynamics and the oil droplet size distribution resulting from the oil flux breakup during a blowout, including the effect of chemical dispersants in such distribution. It is worth pointing out that all the numerical developments presented in this chapter are not the results of investigation carried out during the thesis but, as mentioned through the text, are based on different studies from the literature.

A2.Advances in the TESEO oil spill model (far field)

TESEO is a state of the art transport and weathering oil spill model, whose capabilities have been validated in several studies (Abascal et al., 2017, 2016, 2009b). Until now, however, this model allowed to simulate uniquely surface spills, not taking into account any type of 3D features in the simulation process. As presented Chapter 1, a trend exists in the field consisting in expanding the capacities of these models to simulate the evolution of oil spills through the whole water column, including those spills originated by subsurface blowouts (in that case associated, if possible, with near field models). For that, in this thesis we addressed the updating of TESEO model, so to include all major 3D processes occurring during a spill evolution in the sea, and hence allowing a more accurate numerical simulations at the surface and in the water column. The 3D processes included in TESEO model are: a)

the 3D movement through the water column and the surface, b) the sedimentation of oil particles to the sea bed and c) the vertical natural dispersion of oil due to breaking waves (also called entrainment).

A2.1 3D oil transport

The movement of a spill is simulated by means of the lagrangian transport of several numerical particles, equivalent to the slicks. At each time step, the new position of each particle is computed following:

$$\frac{dx}{dt} = u(x,t) + u_d(x,t) \tag{15}$$

$$\frac{dy}{dt} = v_a(y,t) + u_d(y,t) \tag{16}$$

$$\frac{dz}{dt} = w_a(z,t) + w_f(z,t) + w_d(z,t) \tag{17}$$

where u_a, v_a and w_a refer to the advective velocity components, u_d, v_d and w_d to the diffusive velocity components and w_f to the droplets' rise velocity.

For a particle below the sea surface, the advective velocity is given by the subsurface currents, while for a particle at the surface wind and surface currents (and wave-induced Stokes drift if relevant) are taken into account. In any case, the values of the met-ocean variables are interpolated to the particle's position.

The turbulent diffusive velocity is obtained using a Monte Carlo sampling in the range of velocities $\left[-\overrightarrow{U_{dH}}, \overrightarrow{U_{dH}}\right]$ for the horizontal components, $\left[-\overrightarrow{U_{dV}}, \overrightarrow{U_{dV}}\right]$ for the vertical component, and it is assumed proportional to the diffusion coefficient (Hunter et al., 1993). The velocity fluctuations for each time step are defined as follows (Taylor, 1987):

$$\left|\overrightarrow{U_{dH}}\right| = \sqrt{\frac{6D_H}{\Delta t}} \tag{18}$$

$$\left|\overrightarrow{U_{dV}}\right| = \sqrt{\frac{6D_V}{\Delta t}} \tag{19}$$

where D_H is the horizontal diffusion coefficient, typically in the range of 1-100 m^2/s (ASCE, 1996). D_V is the vertical diffusion coefficient. Following De Dominicis et al. (2013b), D_V is set to 0.01 m^2/s in the mixed layer, assumed to be 30 m deep, and 0.0001 m^2/s below it. However, this values should be calibrated for each simulation case.

The rise velocity, w_f, depends on oil and seawater densities and on the droplet size and it is defined as Proctor et al. (1994):

$$w_f = \begin{cases} \dfrac{gd^2(1-\rho_o/\rho_w)}{18\nu} & \text{for } d \le d_c \\ [8/3gd(1-\rho_o/\rho_w)]^{1/2} & \text{for } d > d_c \end{cases} \tag{20}$$

where g is the gravity acceleration, d is the oil droplet diameter, ρ_o the oil density, ρ_w the sea water density, ν the dynamic viscosity of seawater and d_c is a critic droplet diameter given by Aravamudan et al. (1982):

$$d_c = \frac{9.52\nu^{2/3}}{g^{2/3}(1-\rho_o/\rho_w)^{1/3}} \quad (21)$$

A2.2 Sedimentation

As shown in Eq. (20), when oil density ρ_o is greater than the water density ρ_w, the droplet rise velocity w_f becomes negative. This means that if a downward w_f is not compensate by an equal or greater upward velocity $w_a + w_d$ (17), this droplet inverts its vertical movement heading towards the seabed. All particles that reaches the sea bed are considered sedimented oil and are not allow to move anymore. The sedimented oil mass is calculated at each time step by considering the mass of all the sedimented particles and is taken into account in the oil mass balance calculation.

A2.3 Natural dispersion

One of the most relevant processes included in the latest version of TESEO is the entrainment due to breaking waves. Entrainment has been implemented in the model following Delvigne and Sweeney (1988) and Delvigne and Hulsen (1994). Based on these studies, the entrainment rate due to breaking waves, Q_d (kg/m^{2}s), for a particle with diameter d, can be expressed as:

$$Q_d = C^* D_d^{0.57} SF d^{0.7} \Delta d \quad (22)$$

where C^* is an empirical entrainment constant that depends on oil type and weathering state, D_d is the dissipated breaking wave energy per unit surface area (J/m^2), S is the fraction of sea surface covered by oil, assumed to 0.75 (Berry et al., 2012), F is the fraction of sea surface hit by breaking waves and Δd is the oil droplets interval diameter. Based on French-McCay (2004), depending on the kinematic viscosity of the product, C^* can be expressed as:

$$C^* = \begin{cases} exp(-0.1023\ ln\ \nu + 7.572), & \text{for } \nu < 132\ cst \\ exp(-1.8927\ ln\ \nu + 16.313), & \text{for } \nu \geq 132\ cst \end{cases} \quad (23)$$

and the dissipated breaking wave energy per unit surface area D_d (J/m^2) is obtained by:

$$D_d = 0.0034 \rho_w g H^2 \quad (24)$$

where g is the acceleration due to gravity and H is the mean breaking wave height.

The fraction of sea surface hit by breaking waves is parametrized depending on the intensity of the wind:

$$F = \begin{cases} 3 \cdot 10^{-6} U_W^{3.5} / T_W & \text{for } U_W \leq U_{th} \\ 0.032 (U_W - U_{th}) / T_W & \text{for } U_W > U_{th} \end{cases} \quad (25)$$

where U_W is the wind speed at 10 m above the sea surface, U_{th} is a threshold wind speed for onset of breaking waves assumed as 6 m/s (French-McCay, 2004) and T_W is the significant wave period.

Δd is the oil droplet interval diameter (m), equally spaced between d_{max} and d_{min}, the maximum and minimum droplet diameters, respectively, defined as $d_{min} = 0.1\ d_{50}$ and $d_{max} = 2\ d_{50}$, where d_{50} is the mean droplet diameter of entrained oil, in μm, defined as

Delvigne and Sweeney, (1988). So, for each time step, the total mass entrained due to breaking waves M_E (kg) is given by:

$$M_E = A_s \sum_{i=1}^{10} Q_{d,i} \Delta t \tag{26}$$

where A_s represent the area of the surface slick.

A3.Development of a blowout model (near field)

Different phases occur during the evolution of an oil spill originated by an underwater blowout. The first phase, believed to occur very close to the spill source, consists in the initial jet breakup into oil droplets and, often, also gas bubbles (Socolofsky et al., 2015). This phenomenon is modeled by means of the estimation of a droplet size distribution. Two different approaches exist for this estimation: the approach based on empirical equilibrium equations (e.g., Johansen, Brandvik, and Farooq 2013; Li et al. 2017) and the approach based on dynamic population evolution models (e.g., Bandara & Yapa, 2011; Zhao et al., 2014). Following the breakup region, the spill enters the jet phase, characterized by the flow momentum as main driving force and by the entrainment of ambient water with the consequence plume section enlargement. Finally, when the oil droplet rise velocity due to the relative density oil-sea water is predominant in respect to the velocity due to the jet momentum, the spill enters the far field and is modeled by the 3D lagrangian oil spill model (TESEO).

A3.1 Jet model

As suggested by (Socolofsky et al., 2015), the jet phase can be efficiently modeled using a buoyant jet integral model, similarly to sewage outfall plumes or brine discharges (Jirka, 2004). In this work the model used for the blowout modeling is, indeed, mainly based on a preexisting model of underwater brine discharge, namely BrIHne-Jet (Palomar et al., 2013). BrIHne-Jet is a eulerian model for the simulation of underwater brine discharges as individual jets. The model is based on the integration of the Navier-Stokes equations and the transport equation in the transversal section of the jet, calculating the evolution of the output variables along the jet centerline, assuming an unbounded environmental ambient and self-similarity between cross-sections. The model considers the same numerical approach and equations proposed by (Jirka, 2004). As a result, it calculates, among others variables, the trajectory of the jet, the evolution of its radius and the velocity at the centerline. In order to adapt this model to simulate oil blowout jet, a simple modification to consider positive buoyancy of the flux, in contrast with the negative buoyancy of a brine, was applied. Although this modified eularian model can be applied to calculate the geometry and the centerline velocity of an oil jet originated from a blowout, as suggested by (Johansen, 2003; Zheng et al., 2003), results of a blowout model may be affected if the considered flux is not pure oil, but a mix of oil and gas. Since for the inclusion of the gas effect in a blowout model is recommended to use a lagrangian approach (Spaulding, 2017), we opted to keep a first-stage eulerian blowout model that does not takes into account the effect of the gaseous phase. The model takes into account the effect of the sea current that may affect the position of the

plume. The output of the jet model are the 3D trajectory of the centerline, the radius of the plume and the velocity at the centerline.

A3.2 Oil droplet size distribution module

A.3.2.1 Droplet size distribution

As previously mentioned, the modeling of the size distribution of the oil droplets formed during a blowout is of great relevance since the rise velocity of the oil droplets highly depends, among others variables, on their size. Then, a specific module for the simulation of the oil breakup was developed, mainly based on the recent work by Li et al. (2017a). These authors proposed an equilibrium-based algorithm to simulate oil droplet formation that, unlike previous methods (e.g., Delvigne & Sweeney, 1988), accounts for the change of oil-water interfacial tension which represents a key factor when dealing with the application of chemical dispersants.

Following Li et al. (2017a), a maximum stable droplet size, d_0, is defined as the smaller between the blowout release nozzle diameter, d_n, and the maximum diameter based on Rayleight-Taylor instability (Grace et al., 1978):

$$d_o = \min\left(4 * \left[\frac{\sigma_{o-w}}{\Delta\rho * g}\right]^{0.5}, d_n\right) \quad (27)$$

where $\Delta\rho = \rho_w - \rho_o$ is the density difference between seawater (ρ_w) and oil (ρ_o) ($kg \cdot m^{-3}$); g is gravity acceleration rate ($m \cdot s^{-2}$); and σ_{o-w} is the oil-water interfacial tension ($N \cdot m^{-1}$).

The ratio of disruptive momentum forces, due to the hydrodynamics, and the restorative forces, represented by the oil-water interfacial tension, indicated as the Weber number *We*, is defined as:

$$We = \frac{\rho_w * U^2 * d_o}{\sigma_{o-w}} \tag{28}$$

where U is the exit velocity of the oil ($m \cdot s^{-1}$). The Ohnesorge number, *Oh*, is another dimensionless number used to relate the viscous forces to inertial and surface tension forces (Lefebvre and McDonell, 2017). In this case it can be written as:

$$Oh = \frac{\mu_o}{\sqrt{\rho_o \sigma_{o-w} d_o}} \tag{29}$$

where μ_o is the dynamic viscosity of oil ($Pa \cdot s$ or $N \cdot m^{-2} \cdot s$ or $kg \cdot m^{-1} \cdot s^{-1}$).

Based on these variables, Li et al. (2017a) proposed the following equation to calculate a characteristic droplet diameter:

$$\frac{d_{50}}{d_0} = r * (1 + 10 * Oh)^p * We^q \tag{30}$$

The exponential coefficient p and q with base *Oh* and *We*, respectively, and the scaling coefficient r were empirically determined, based on data from Delvigne and Hulsen (1994), Reed et al. (2009) and the Deep Spill experiment (Brandvik et al., 2014), to be r = 14.05, p = 0.46 and q = -0.518.

Typically, in equilibrium-based oil droplet distribution models, Rossin-Rammler or lognormal distribution are used to evaluate the whole distribution once the characteristic diameter is known. In this case, Li et al. (2017a) suggested to use a lognormal distribution for its simplicity with respect to the Rossin-Rammler distribution. *N* equally spaced droplet

diameter intervals of size Δd are then defined between 0 m and the maximum droplet diameter previously obtained, d_0. The number of oil droplets within the *i*-th interval, f_i, with $I = 1,\ldots, N$, can be defined as:

$$f_i = \frac{1}{\sqrt{2\pi} d_{p,i} ln\sigma_g} exp\left(-\frac{\left(lnd_{p,i} - lnd_{50}\right)^2}{2\left(ln\sigma_g\right)^2}\right)\Delta d \quad (31)$$

where $ln\sigma_g$, the standard deviation, assumed 0.5 (Li et al., 2017a) and $ln\mathrm{d}_{50}$ represent the two distribution parameters, $d_{p,i}$ is the mean droplet diameter of the *i*-th interval.

A.3.2.2 Use of chemical dispersant

Laboratory studies about the injection of chemical dispersants in the proximity of a deep spill showed how these can reduce the size of the oil droplets produced during the breakup of various order of magnitude (Brandvik et al., 2013; Tang, 2004). The first application of chemical dispersants to a real spill was during the Deepwater Horizon accident (Li et al., 2015; Payne and Driskell, 2015).

Following (Crowley et al., 2014), the effect of applying chemical dispersant was implemented in our model by means of a reduction of the oil-water interfacial tension. The model also allows to evaluate the interfacial tension reduction as a function of the dispersant-to-oil ratio (DOR) considered. The DOR is defined as $Q_{dispersant}/Q_{oil}$, where $Q_{dispersant}$ is the volumetric rate of dispersant and Q_{oil} the volumetric rate of oil. Several works exist that relates the oil-water interfacial tension to DOR (Johansen et al., 2013; Khelifa et al., 2011; Khelifa and So, 2009; Venkataraman et al., 2013). In the present work, achievements

published by (Venkataraman et al., 2013) are considered. The authors carried out several experiments using Louisiana sweet crude oil and Corexit 9500, one of the more commonly used chemical dispersant. The empirical relation between DOR and interfacial tension (IFT) found by (Venkataraman et al., 2013) and, implemented in the oil droplet distribution model presented here, is shown in Figure 33.

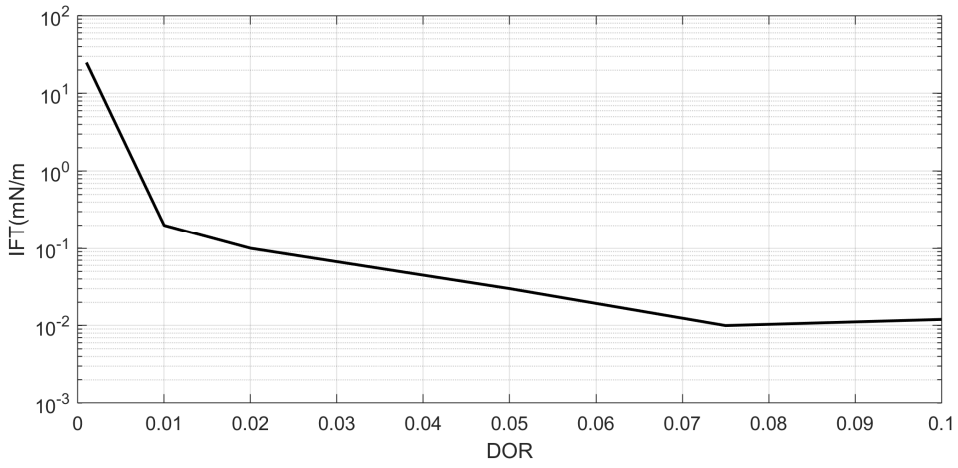

Figure 33. DOR-IFT relationship as obtained by (Venkataraman et al., 2013).

A4. Near field – far field coupling

In case of simulating a spill produced by a blowout, a link is required to connect the two modeling components previously presented: the near field and the far field models. As said, the output of the blowout models are, on the one side, the geometry of the plume, in particular the trajectory of the centerline and the radius, and, on the other, the oil droplet size distribution. In order to start running a far field model, such as TESEO, it is required to define

the initial oil source location. So, as shown by previous studies (Rye et al., 1996; Yapa and Li, 1997), the problem of linking the two modeling components consists, substantially, in the definition of that point in which the plume dynamics becomes negligible respect with the movement of the oil droplets caused by their own buoyant velocities. This transition point is commonly referred to as plume terminal level (Socolofsky and Adams, 2002). (Dasanayaka and Yapa, 2009) proposed a criterion for the definition of the plume terminal level based on the argument that an oil droplet rises with its own buoyant velocity once the plume velocity drops to such velocity. Based on this criterion, the terminal level of the near field plume is defined as the point in which the plume centerline velocity equals the median rise velocity of the oil droplets (Dasanayaka and Yapa, 2009). At this point occurs the coupling with the far field model. The oil particles, with their size distribution obtained in the near field modeling, are uniformly distributed at the estimated terminal level, within a circle with the same radius as the plume radius at the terminal level. From this point, the evolution of the oil particles, as well as the weathering processes once at the surface, are simulated by the far field TESEO oil spill model.