



ACTIVITE T1.1

Identification des risques difficiles à prévoir liés aux actions atmosphériques et marines

Livrable T1.1.3

Procédure commune de prévision des risques naturels difficiles à détecter: composante marine

Partenaire responsable:

Consorzio LaMMA

date de début et de fin:

01/04/2019 – 31/10/2022

Version finale, 7/12/2022



Ce rapport a été rédigé par:

Carlo Brandini

Alessio Innocenti

(Consorzio LaMMA)

Tania Del Giudice

(ARPAL)

Joseph Jenkins

Adeline Paiement

Yann Ourmieres

Jacques Verron

Hervé Glotin

(Université de Toulon)



1	Résumé	
2	Introduction	5
3	Partie 1	6
3.1	Actions marines difficilement prévisibles	6
3.1.1	Vagues scélérates	6
3.2	Modèles de prédiction des vagues	7
3.2.1	Modèles phase averaging: WW3	8
3.2.2	Indices de probabilité des vagues scélérates	8
3.3	Observations	9
3.4	Modèles phase-resolving à haute résolution	9
3.5	Reconstruction de la probabilité des vagues scélérates	11
3.6	Application de la méthodologie à la réanalyse des incidents	12
3.6.1	Cas d'étude	13
3.6.2	Résultats	15
3.6.2.1	Conditions synoptiques	15
3.6.2.2	Analyse statistique	17
3.7	Conclusions	20
3.8	References	21
4	Partie 2	24
4.1	Introduction	24
4.2	Modèle de prévision maritime en Méditerranée	25
1.		25
4.2.1	Modélisation atmosphérique	25
4.2.2	Modélisation météo-marine	25
4.2.3	Calibrage du modèle	27
4.2.4	Storm Power dans les données historiques de la Ligurie	28
4.2.5	Storm Power Index (SPI), Coastal Vulnerability Index (CVI) e Risk Index (RI)	29
4.3	References	34
5	Partie 3	35
5.1	Contexte	35
5.1.1	Le contexte du Projet	35
5.1.2	Problème de formulation	36
5.2	Méthodologies actuelles	38
5.3	Projet de recherche	39

5.4	Tests préliminaires	42
5.5	Introduction	44
5.6	Travaux antérieurs	46
5.7	Données	47
5.7.1	Champs de vélocité des courants de surface	47
5.7.2	Trajectoires de particules	48
5.7.2.1	Advection de particules	49
5.7.2.2	Trajectoires probabilistes	49
5.7.2.3	Instantanés de trajectoires probabilistes	50
5.7.3	Déploiement des trajectoires probabilistes	50
5.8	Méthodologie	51
5.8.1	Cartes de densité de probabilité pour la représentation de l'incertitude	51
5.8.2	Réseau de neurones	52
5.8.2.1	Architecture	52
5.8.2.2	Fonction de perte	52
5.9	Expériences	54
5.9.1	Comparaison des fonctions de perte	54
5.9.2	Généralisation à différents dynamiques	55
5.9.3	Application à différentes représentations des courbes de surface	58
5.10	Discussion sur les trajectoires probabilistes	59
5.10.1	Caractéristiques	59
5.10.2	Prévisions préliminaires	60
5.11	Conclusions	61
5.12	References	64



2 Introduction

L'objectif de ce rapport est de fournir des procédures pour identifier les dangers difficilement prévisibles, en mettant l'accent sur la composante marine. La modélisation utilisée de manière opérationnelle dans les centres météorologiques pour soutenir la navigation, la surveillance de l'environnement et toutes les activités liées à l'écosystème marin, nécessite des procédures communes et facilement reproductibles pour la prévision et l'identification des dangers qui, bien que rares et difficiles à prévoir, peuvent avoir de graves conséquences. À ce jour, les modèles de mouvement des vagues et les modèles océanographiques se concentrent sur la prédiction de quantités physiques "moyennes", telles que la hauteur significative des vagues, alors qu'ils ne traitent pas, ou très approximativement, les événements extrêmes, qui sont plus rares et donc aussi plus difficiles à reproduire. Dans l'optique, donc, d'identifier des outils partageables capables de combler ces lacunes dans les modèles de prévision, ce rapport présente les résultats obtenus par les partenaires LaMMA, ARPAL et Université de Toulon sur différents aspects et facettes des aléas difficiles à prévoir concernant la composante marine. En particulier, le produit est divisé en trois parties :

- la première partie rapporte la contribution du Consortium LaMMA sur la prédiction des vagues scélérates avec des modèles à haute résolution ;
- la deuxième partie rapporte la contribution d'ARPAL sur l'implémentation, la validation et l'optimisation d'un indice de risque côtier dans un modèle de vagues ;
- la troisième partie rapporte la contribution de l'Université de Toulon sur l'utilisation de méthodologies d'apprentissage automatique pour le suivi d'objets dérivants ou la dispersion de polluants suite à des déversements.

Ces éléments, liés à des événements météorologiques rares ou au facteur humain, présentent des risques pour la navigation, l'environnement marin et côtier et les infrastructures, et nécessitent des procédures spécifiques en fonction du contexte. Dans les trois parties suivantes du rapport, les procédures et les résultats pour ces trois types de risques sont décrits.



3 Partie 1

Cette première partie se concentre sur le travail effectué par le Consortium LaMMA en collaboration avec l'Université de Turin sur l'utilisation de modèles à haute résolution pour une approche statistique de la prédiction des vagues scélérates.

3.1 Actions marines difficilement prévisibles

3.1.1 Vagues scélérates

Les vagues scélérates, également appelées "rogue waves", "freak waves" ou plus simplement "extreme waves" dans la littérature scientifique internationale, sont des vagues beaucoup plus hautes que la hauteur moyenne des vagues voisines et peuvent être extrêmement dangereuses pour la navigation. Parler de vagues extrêmes, ou vagues scélérates, nécessite une approche statistique en raison de la nature stochastique des vagues de mer en général et de la rareté de ce type d'événement. Les vagues de mer dans un état de mer donné peuvent être prédites en termes de probabilité d'occurrence d'une vague ou d'une hauteur de crête donnée. Des conditions de mer différentes peuvent donner lieu à des descriptions statistiques différentes, avec des distributions de probabilité qui peuvent être prédites théoriquement ou adaptées avec des représentations ad hoc. Si le champ des vagues est considéré comme une somme infinie des vagues élémentaires de fréquences différentes et de phases aléatoires, dans l'approximation linéaire, les hauteurs des ondes et des crêtes sont décrites par la distribution de Rayleigh. En particulier, l'adoption de la définition des vagues scélérates $H > 2,2H_s$ donne une probabilité de dépassement égale à $PR(H > 2,2H_s) = 6,25 \times 10^{-5}$, tandis que l'adoption du critère des crêtes $\eta_c > 1,25H_s$ donne une probabilité de dépassement égale à $PR(\eta_c > 1,25H_s) = 3,73 \times 10^{-6}$. La plupart des modèles traditionnels utilisent des distributions de probabilité "standard" dérivées de la théorie, telles que la distribution de Rayleigh ou d'autres distributions plus précises qui prennent également en compte les non-linéarités. Toutefois, en fonction des conditions de mer, on peut avoir des distributions différentes et donc avoir, pour des mers de même hauteur significative, une probabilité encore très différente de rencontrer des vagues scélérates. En fait, des mesures sur le terrain [1, 2] et des expériences en laboratoire [3] ont montré que la distribution de probabilité peut être fortement non gaussienne, notamment en ce qui concerne les crêtes des vagues. Par conséquent, des conditions apparemment similaires peuvent en réalité être associées à des niveaux de risque très différents.



Ces vagues plus extrêmes d'une distribution, d'un point de vue physique, peuvent se former pour diverses raisons. Le plus souvent, elles sont générées par superposition linéaire, c'est-à-dire que des vagues provenant de directions différentes se focalisent en un même point de l'espace au même moment, générant une seule vagues de plus grande hauteur (focalisation directionnelle), ou cette superposition peut être provoquée par des vagues de longueurs d'onde différentes se déplaçant à des vitesses différentes, se superposant ainsi à un instant donné et dans une position donnée. Ces mécanismes sont pris en compte par les modèles traditionnels avec lesquels il est possible d'avoir une première approximation de la fréquence des ondes scélérates, ou au-delà d'un certain seuil. Cependant, il existe des mécanismes plus complexes qui sont difficiles à prendre en compte dans les modèles. Par exemple, les vagues interagissent avec leur environnement, comme la bathymétrie (c'est-à-dire la profondeur et la morphologie des fonds marins), les courants marins et le vent, des phénomènes qui peuvent entraîner la concentration des vagues en un même point et ainsi augmenter leur hauteur. Enfin, les vagues interagissent entre elles de manière non linéaire, transférant l'énergie d'une fréquence à l'autre, avec la possibilité de l'émergence de mécanismes tels que l'instabilité modulatoire, ou instabilité de Benjamin-Feir, dans laquelle un train d'ondes caractérisé par une porteuse haute fréquence et modulé ou perturbé par une onde de fréquence beaucoup plus basse peut devenir instable et provoquer la concentration de l'énergie en un point donné.

La prévisibilité de ces phénomènes est très faible et ne peut être traitée que de manière probabiliste.

3.2 Modèles de prédiction des vagues

Les modèles de prévision des vagues utilisés par les centres météorologiques tels que ARPAL et LaMMA ont des résolutions typiques de quelques km, avec une plage de prévision de 2 à 3 jours. Les modèles à maillage non structuré peuvent fournir une résolution variable, avec des cellules de calcul plus petites près de la côte et plus grandes au large; dans ce cas, il est possible d'arriver à des résolutions localement même inférieures à 1 km près de la côte. Pour pouvoir fournir une prévision en temps opérationnel, la seule possibilité actuelle est d'utiliser des modèles phase-averaging, c'est-à-dire "moyennés dans la phase", qui ne traitent pas les vagues individuellement mais utilisent le spectre comme variable pronostique. D'autre part, les modèles à résolution de phase sont capables de représenter des vagues individuelles, mais en raison de la haute résolution numérique qu'ils requièrent (de l'ordre de 10m), ils ne peuvent pas être utilisés pour des résultats opérationnels dans de grandes zones maritimes. Toutefois, ces modèles peuvent être appliqués à l'analyse de cas spécifiques liés à des conditions



météorologiques et maritimes particulièrement énergiques afin d'obtenir des informations utiles concernant l'apparition de vagues scélérates.

3.2.1 Modèles phase averaging: WW3

Dans ces modèles, l'équation d'évolution est une équation d'équilibre pour l'action des vagues N , définie comme le rapport entre l'énergie spectrale E et la fréquence relative σ , qui est liée à la fréquence absolue ω par la dépendance du vecteur nombre d'onde k et du vecteur U représentant la vitesse du courant (si présent). L'équation d'équilibre est donc la suivante :

$$\frac{\partial N}{\partial t} + \nabla \cdot (\dot{x}N) + \frac{\partial}{\partial k} \dot{k}N + \frac{\partial}{\partial \theta} \dot{\theta}N = \frac{S}{\omega}$$

Les différents termes représentent la variation temporelle, la propagation spatiale (x,y) et la propagation intra-spectrale (k, θ). Les termes à droite des égaux sont les termes dits sources qui prennent en compte l'interaction vague-vent et donc la génération des vagues, les interactions non linéaires des vagues qui transfèrent l'énergie entre les différentes parties du spectre, et les phénomènes de dissipation dus au whitecapping ou à la friction au fond. Contrairement à l'énergie, l'action des vagues est conservée en présence de courants.

Plusieurs schémas numériques sont disponibles pour résoudre l'équation de conservation de l'action des vagues et, récemment, plusieurs efforts ont été déployés pour améliorer la représentation des termes sources.

Les centres météorologiques (tels que LAMMA et ARPAL) exécutent des modèles de mouvement des vagues en utilisant des codes tels que WW3, SWAN et MIKE-SW. Les chaînes de prévision adoptées par LAMMA et ARPAL ont été validées.

3.2.2 Indices de probabilité des vagues scélérates

Dans la littérature scientifique, il existe différents indices qui peuvent être associés à une fréquence accrue des vagues scélérates. L'indice le plus largement utilisé est l'indice de Benjamin-Feir (BFI) [4], qui est donné par le rapport entre la pente de l'onde et la largeur de bande spectrale, pondérant ainsi la non-linéarité par rapport à la dispersion. Des valeurs élevées du BFI, proches de l'unité ou supérieures à celle-ci, indiquent une probabilité plus élevée de vagues extrêmes liées au mécanisme de formation de



l'instabilité de Benjamin-Feir (ou instabilité modulatoire). Cependant, ce type d'indice peut être efficace dans des conditions où le spectre a une faible dispersion angulaire, c'est-à-dire des vagues quasi-unidirectionnelles, mais a peu de pouvoir prédictif dans les cas où le spectre est plus multidirectionnel. Pour cette raison, il ne peut fournir que des informations indicatives dans de nombreux cas réels.

Des variantes du BFI ont également été proposées, comme le BFI effectif [5], qui l'étend en tenant également compte de la directionnalité du spectre, mais elles n'ont pas été mises en œuvre dans des modèles opérationnels et, en tout état de cause, leur efficacité n'a pas été démontrée de manière approfondie dans des conditions réelles liées aux spectres de vagues en mer.

Il convient également de souligner que le calcul du BFI nécessite l'estimation de plusieurs paramètres tels que la largeur de bande spectrale et la fréquence de crête et peut donc conduire à des résultats différents selon la méthode de calcul utilisée.

D'autres méthodes ont récemment été proposées dans le cadre de modèles de prévision tels que WW3 [6], qui fondent la prédiction des "extrêmes spatio-temporels" sur des distributions théoriques, et qui sont capables de fournir, par exemple, des cartes de la hauteur maximale attendue en chaque point du domaine. Ce type d'estimation, bien que plus raffiné que le BFI, est toujours basé sur des distributions théoriques paramétriques et peut donc au mieux prendre en compte les moments spectraux, plus simplement la pente et l'amplitude du spectre, perdant ainsi des informations importantes sur la forme du spectre (par exemple, mono ou multi-crête).

3.3 Observations

Les observations du mouvement des vagues sont principalement basées sur des bouées capables d'enregistrer des séries temporelles de la hauteur des vagues à partir de capteurs accélérométriques. A partir des séries temporelles, des paramètres synthétiques tels que la hauteur significative, la période de pointe, la période moyenne et la direction moyenne sont obtenus, tandis que dans certains cas, des spectres d'ondes réels sont reconstruits. Ces outils sont extrêmement utiles pour surveiller les zones d'intérêt particulier et pour valider les modèles de mouvement des vagues.

3.4 Modèles phase-resolving à haute résolution

Les modèles de mouvement des vagues phase-resolving, c'est-à-dire sans moyenne de phase, sont les modèles qui offrent une représentation vague par vague. On peut distinguer deux grandes classes de modèles de ce type: les modèles dédiés à l'environnement côtier, qui exploitent typiquement la simplification des eaux peu profondes, en faisant ainsi la moyenne sur la verticale, et qui sont utilisés pour des études de processus ou des études relatives, par exemple, à l'impact des vagues de tempête sur la côte, en tenant compte des phénomènes de réfraction, de diffraction, de shoaling et de rupture. Il s'agit, par exemple, de MIKE 21BW et de X-Beach. Un autre type de modèles est le modèle spectral d'ordre élevé (High Order Spectral, HOS), qui simule l'évolution temporelle de la surface libre η et du potentiel de vitesse ϕ . Sous les hypothèses d'un écoulement incompressible, inviscide et irrotationnel, l'équation de continuité peut être exprimée en termes de potentiel par l'équation de Laplace $\Delta\phi=0$. Les équations dynamiques pour l'élévation de la surface libre et le potentiel de vitesse sont obtenues à partir des conditions limites non linéaires de la surface libre, qui imposent un écoulement à vitesse nulle à la surface libre (condition limite cinématique) et une continuité de pression (condition limite dynamique), ce qui donne:

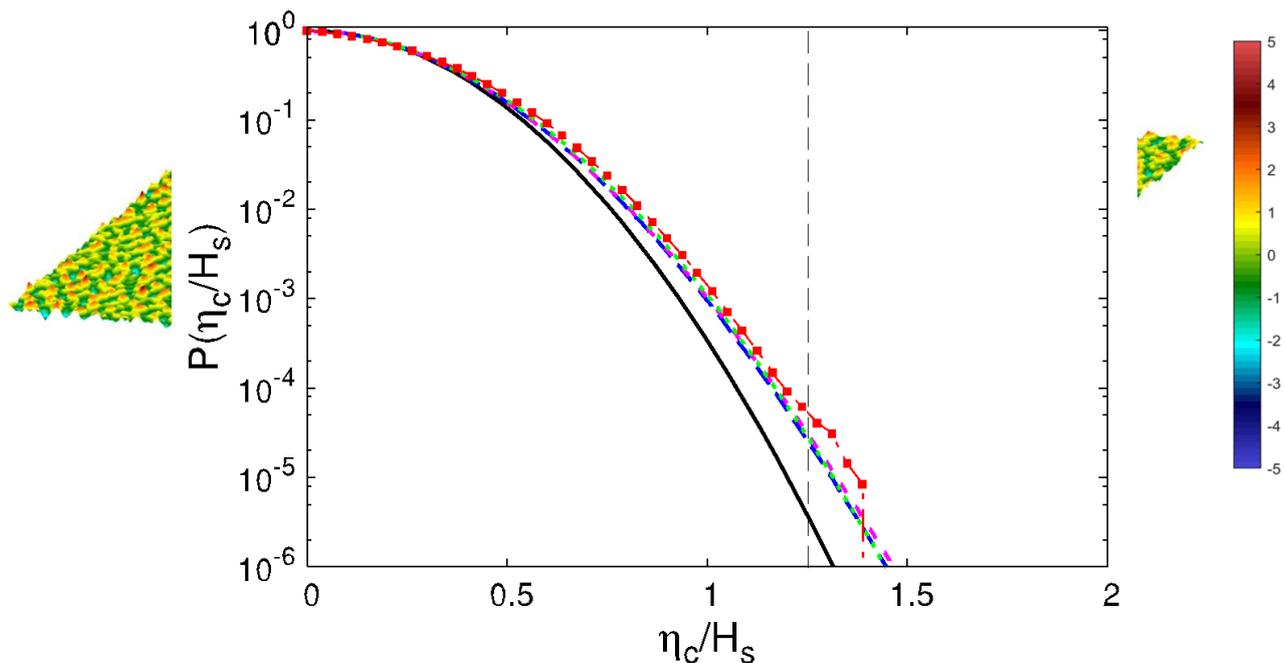
$$\frac{\partial \phi_s}{\partial t} = -g\eta - \frac{1}{2}|\nabla\phi_s|^2 + \frac{1}{2}(1 + |\nabla\eta|^2) \left(\frac{\partial\phi}{\partial z}\right)^2,$$

$$\frac{\partial\eta}{\partial t} = (1 + |\nabla\eta|^2) \frac{\partial\phi}{\partial z} - \nabla\phi_s \cdot \nabla\eta.$$

tous deux évalués à $z = \eta(x,y,t)$. Le potentiel de vitesse est exprimé comme une quantité surfacique, c'est-à-dire en fonction du potentiel de surface $\phi_s(x,y,t) = \phi(x,y,\eta,t)$. La quantité $W = \partial\phi/\partial z$ est évaluée de manière itérative jusqu'à un certain ordre non linéaire M et est ensuite utilisée pour la progression temporelle des équations. Cette méthode a été initialement proposée indépendamment par Dommermuth et Yue [7] et West et al. [8] et a été largement appliquée pour l'étude des interactions non linéaires des vagues et la formation des vagues scélérates [9, 10, 11]. Dans le présent travail, nous avons utilisé le code open source HOS-ocean [12, 13, 14] (<https://github.com/LHEEA/HOSocean>), qui résout les équations décrites ci-dessus dans un domaine doublement périodique, typiquement d'une longueur égale à quelques dizaines de longueurs d'onde crête. À l'instant initial, la surface libre est initialisée à partir d'un spectre de vague avec des phases aléatoires, après quoi l'évolution libre est simulée, sans introduire de forçage externe supplémentaire, mais en évaluant simplement la dynamique des vagues à partir d'un état initial. Typiquement, cette évolution est simulée pour environ 250 périodes de pointe T_p , c'est-à-dire pour des durées allant de 40 minutes à l'ordre d'une heure. La Figure 1 montre un exemple d'un instantané de l'élévation de la surface dans une simulation utilisant la méthode HOS.



3.5 Reconstruction de la probabilité des vagues scélérates



A partir des modèles à résolution de phase, il est possible de faire des analyses statistiques utiles pour reconstruire en termes probabilistes l'occurrence de vagues scélérates dans certaines conditions marines. Ayant l'élévation de la surface η comme variable pronostique, et déterminant les maxima et minima à chaque instant présents sur l'ensemble du domaine, on peut calculer le nombre de crêtes et de lignes pour chaque élévation afin de reconstruire des distributions de probabilité. De même, une fois qu'une méthode unique pour déterminer la hauteur de la vague a été définie, il est possible de construire des distributions similaires pour la hauteur de la vague dérivée de la différence entre la crête et le creux. Cependant, la reconstruction de la hauteur des vagues à partir d'une analyse spatiale peut être complexe dans les mers multidirectionnelles et une analyse temporelle effectuée en des points fixes du domaine ou des sondes virtuelles est plus simple. En définissant un seuil pour identifier les vagues scélérates, par exemple $H/H_s > 2$ ou $\eta_c/H_s > 1,25$, nous pouvons associer à chaque cas une certaine valeur de probabilité de vagues scélérates. La Figure 2 montre diverses distributions de probabilité pour un cas d'étude intéressant (voir le chapitre sur les cas d'étude), en comparant diverses distributions "théoriques" où sont pris en compte au maximum les paramètres de moyenne spectrale tels que la raideur $\varepsilon = k_p H_s/4$, et l'étalement angulaire σ_θ , avec la distribution obtenue à partir de l'analyse statistique de la simulation avec le modèle HOS.

L'un des objectifs de l'analyse que nous avons effectuée était de comprendre comment la probabilité de vagues extrêmes dépend des paramètres de la moyenne spectrale, afin que des estimations puissent également être obtenues dans les modèles de vagues opérationnels.

D'autres indices statistiques très intéressants sont l'asymétrie, λ_3 , et l'aplatissement, λ_4 , de l'élévation de la surface libre, c'est-à-dire les moments statistiques de troisième et quatrième ordre. Ils sont associés à l'asymétrie du PDF et au poids de la queue, respectivement, et donc indirectement à l'apparition de vagues extrêmes. Il a également été démontré, tant sur le plan théorique que par des simulations numériques de cas réels, que l'aplatissement dépend de la pente de la vague, ce qui peut donc être considéré comme un signal utile pour la détection de conditions des vagues scélérates. Une forte dispersion angulaire du spectre, au contraire, a l'effet inverse, bien que ce paramètre ne tienne pas compte de la topologie du spectre (par exemple, il ne différencie pas les spectres avec un seul pic des spectres avec plusieurs pics), qui pourrait équilibrer ce comportement ou même l'inverser.

3.6 Application de la méthodologie à la réanalyse des incidents

Une approche hybride "en cascade", consistant à utiliser les spectres de vagues obtenus à partir de prévisions a posteriori du modèle de vagues à moyenne de phase pour initialiser les modèles à résolution de phase, a été utilisée pour la réanalyse des accidents dus à des vagues extrêmes. Cette méthodologie a été appliquée à une série des cas d'étude d'accidents dans le nord/nord-ouest de la Méditerranée causés par des vagues extrêmement hautes.

L'hindcast des vagues a été réalisée avec le modèle Wave Watch III (WW3, version 6.07) avec une configuration nichée : une grande grille couvrant toute la mer Méditerranée avec une résolution de $0,1^\circ$ (environ 10 km), et une grille nichée sur la mer Méditerranée du Nord-Ouest avec une résolution de $0,03^\circ$ (environ 3 km), au nord de la latitude $39,6^\circ\text{N}$. La bathymétrie provient de la version 2018 d'EMODnet (<https://www.emodnet-bathymetry.eu>) avec une résolution spatiale de 1/16 de minute d'arc. Dans l'espace spectral, la résolution de WW3 est de 36 directions (intervalles de 10°) et 40 fréquences, allant de 0,0418Hz à 0,585Hz. Le forçage atmosphérique est produit par une analyse du modèle Weather Research and Forecasting (WRF) [15] avec une résolution spatiale de 3 km. L'analyse WRF est forcée à la limite toutes les 6 heures, tandis que le champ de vent produit est utilisé pour forcer le modèle WW3 à intervalles d'une heure.

Les simulations d'hindcast des vagues ont été effectuées avec un calcul exact de l'intégrale de Boltzmann, dans le but d'avoir la reconstruction la plus précise possible des spectres des vagues passées. Ce calcul exact, qui nécessite un temps de calcul énorme, a également été comparé à la méthode approximative de l'interaction discrète (DIA), généralement utilisée dans les prévisions opérationnelles. Le forçage du vent et la dissipation due au déferlement des vagues ont été paramétrés avec le paquet de termes sources ST4.

Aux positions approximatives des incidents, on a extrait des spectres des vagues qui ont été utilisés pour initialiser le champ de mouvement dans les simulations à résolution de phase utilisant la méthode HOS. Pour chaque cas d'étude, plusieurs simulations ont été effectuées, en réinitialisant les phases de façon aléatoire à chaque fois, afin d'avoir une plus grande signification statistique. Les simulations ont été effectuées dans un domaine limité de taille 40 fois supérieure à la longueur d'onde maximale dans les directions x et y.

3.6.1 Cas d'étude

Nous avons examiné cinq études de cas d'accidents de navires dans le nord-ouest de la Méditerranée couvrant les 20 dernières années. Quatre d'entre eux ont été rapportés dans les catalogues de Didenkulova et al. [16, 17, 18] tandis que le cinquième, concernant l'accident d'Euro-cargo Venezia en 2011, n'a été rapporté que par les journaux et médias locaux..

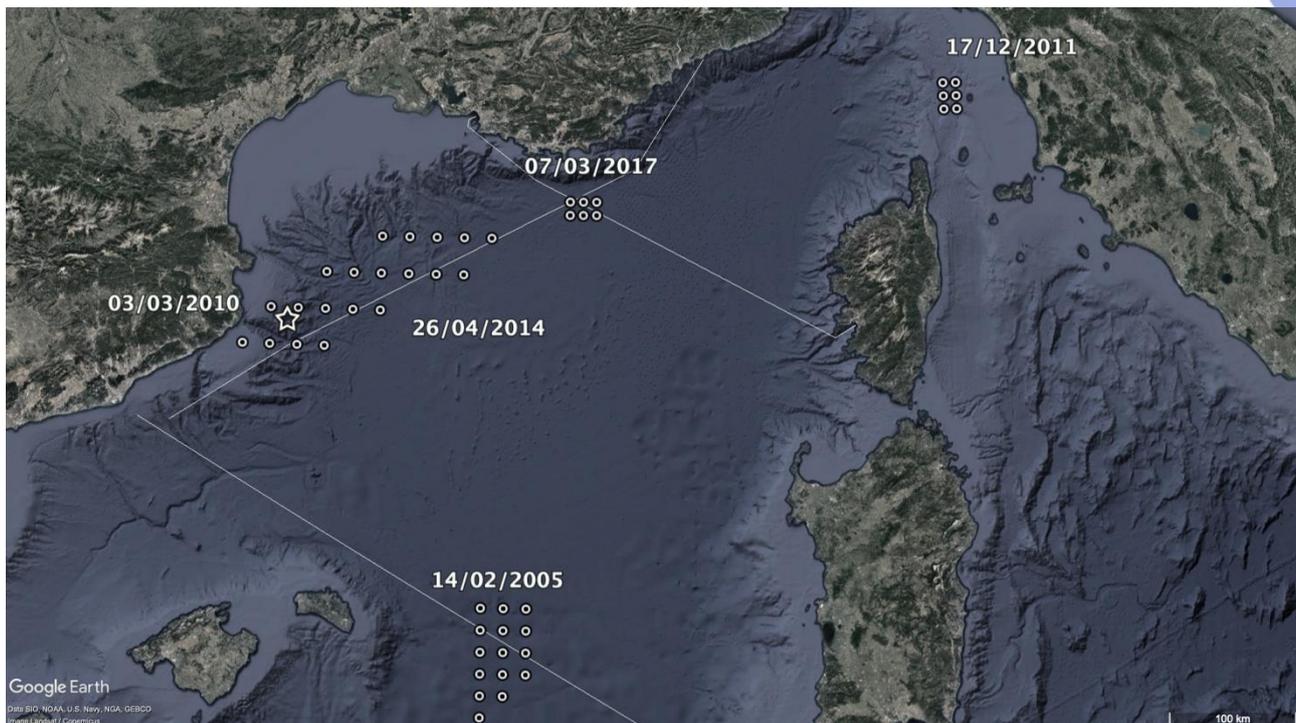


Figure 3: Lieux des accidents. Les lignes représentent la route idéale des navires, tandis que les séries de points sont les positions analysées couvrant la zone potentielle des incidents. Le cas n° 2 est représenté par une seule étoile, puisque la position exacte est connue. La carte a été créée avec Google Earth

La Figure 3 mostra le posizioni approssimative degli incidenti, mentre di seguito riassumiamo brevemente i rapporti sugli incidenti:

1. 14/2/2005: Le navire de croisière "Grand Voyager", qui faisait route de Tunis à Barcelone, a été frappé par une énorme vague pendant une tempête, à environ 60 miles au large de Minorque, causant des dommages aux équipements électroniques et aux moteurs [18]. Les rapports sur le web (<http://www.cruisejunkie.com/events2005.html>, 2006) indiquent que l'heure de l'incident se situe aux alentours de 08 UTC, tandis que le lieu exact est inconnu. Les rapports parlent d'une série de vagues d'environ 14 m de hauteur. La hauteur significative de la vague, H_s , estimée dans [18] est $H_s = 2,4$ m. Le hindcast avec lequel nous avons analysé l'incident prévoit $H_s = 5-6$ m, montrant une variabilité significative.
2. 3/3/2010: le navire de croisière "Louis-Majesty", faisant route de Barcelone à Gênes dans des conditions de mer tempétueuse, a été frappé par une vague qui s'est abattue sur le pont 5, à 16,7 m au-dessus de la ligne de flottaison, à 24 miles au large du cap Begur, en Espagne, vers 15 UTC, causant des dommages au navire, des blessures et deux décès [16]. La hauteur significative des vagues était supérieure à 4 m, [16] rapportant $H_s \approx 4$ m, [19] $H_s \approx 4,8$ m. La bouée Begur, située à proximité à Puertos del Estado [20], a enregistré $H_s \approx 4,8$ m au moment de l'incident.

3. 26/4/2014: le navire de croisière "Crystal Serenity", en route d'Espagne vers Monte Carlo, a été frappé par une vague vers 1:30 UTC, causant des dommages à la salle à manger principale du navire [17]
(<http://freaquewaves.blogspot.com/2014/05/encountered-by-cruise-ship-crystal.html>). La localisation du navire est incertaine, mais d'après les rapports des blogs, nous avons estimé qu'il se trouvait au sud du Golfe du Lion, le long d'une route idéale entre l'Espagne et Monte Carlo. L'hindcast que nous avons utilisée prévoit que Hs \approx 2-3 m dans cette zone.
4. 7/3/2017: un ferry faisant route d'Ajaccio à Marseille a été endommagé par une vague qui a brisé la fenêtre du cockpit (environ 20 m au-dessus de la ligne de flottaison) vers 21h30 UTC, à 40 km au sud des îles Levante [17]
(<https://www.varmatin.com/vie-locale/endommagement-par-une-vague-de-15m-un-ferry-est-obligé-d'accoster-plus-tôt-que-prévu-a-toulon-119787>). La hauteur significative des vagues prédite par ce hindcast dans la zone est approximativement Hs \approx 5m. Les mesures de la bouée voisine de Porquerolles provenant de la base de données CANDHIS du Cerema (<https://candhis.cerema.fr>) ne sont pas disponibles pour la période de l'accident.
5. 17/12/2011: Le cargo "Eurocargo Venezia", qui faisait route de Catane à Gênes dans la nuit du 16 au 17 décembre, a perdu deux semi-remorques contenant des fûts de produits toxiques en raison de la tempête. L'accident s'est produit dans un laps de temps d'environ trois heures, alors que le navire naviguait près de l'île de Gorgona, dans l'archipel toscan. La reconstruction actuelle des prévisions rétrospectives indique une hauteur de vague significative d'environ Hs \approx 5 m dans cette zone. Ce dernier incident ne fait pas partie des catalogues de [16, 17, 18], mais a été ajouté pour élargir le panorama des incidents, même s'il n'était pas clair s'il était causé par une vague scélérate ou une manœuvre erronée.

3.6.2 Résultats

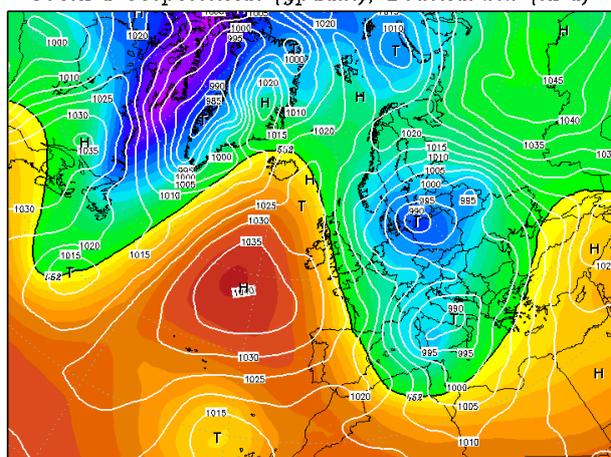
3.6.2.1 Conditions synoptiques

Dans 3 des 5 cas, ceux relatifs aux accidents de 2005, 2011 et 2017, les conditions peuvent être associées à un bloc de haute pression sur l'Atlantique et à un bloc de basse pression sur l'Europe du Nord, entraînant des flux de nord/nord-ouest. Autour de la zone

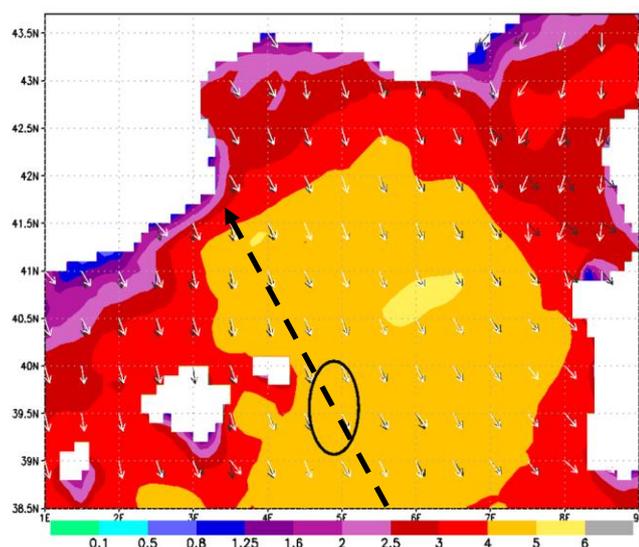


et de l'heure des incidents respectifs, il y a eu une rotation des vents vers le nord ou l'ouest associée à un approfondissement de la dépression sur la péninsule italienne. Dans certains cas, la rotation du vent s'est produite avant le moment de l'accident et est donc également visible dans les spectres extraits du modèle de vague, qui représentent une conformation avec deux pics près des deux directions principales du vent, avant et après la rotation. Dans ces cas, la hauteur significative est de l'ordre de 5 à 6 mètres. La Figure 4 montre la carte du géopotential et la carte de la hauteur significative des vagues et de la direction du vent pour le cas de l'accident de 2005.

14FEB2005 12Z
500hPa Geopotential (gdam), Bodendruck (hPa)



Daten: CFS Reanalysis
 (C) Wetterzentrale
 www.wetterzentrale.de



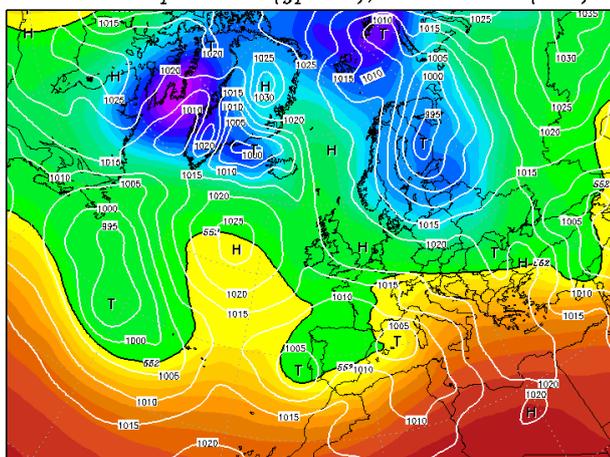
Le cas relatif à l'accident de 2010 est associé au passage d'un minimum dépressionnaire autour de la zone de l'accident, qui a entraîné localement une variation rapide de l'intensité et de la direction du vent, comme on peut le voir sur la Figure 5. En particulier, dans les heures précédant l'accident, le minimum s'est déplacé des Baléares vers le nord-est en direction de la zone de l'accident et les vents, initialement de sud-est/est, ont eu tendance à tourner vers le sud/sud-ouest, bien que dans le golfe du Lion la composante d'est soit restée prédominante. Dans ce cas, le spectre extrait à l'instant de l'incident montre une très grande dispersion angulaire avec des contributions énergétiques



provenant de toute la section Est / Sud-Est. Dans ce cas, la hauteur significative des vagues est d'environ 4,8 mètres.

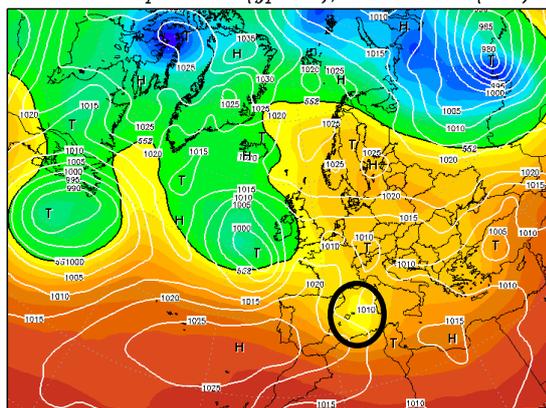
Enfin, dans le cas de l'incident de 2014, il n'y a pas de conditions particulièrement énergétiques en termes d'intensité du vent et de hauteur significative des vagues, mais il est associé au passage d'une poche d'air froid qui a pu provoquer des orages avec de fortes rafales de vent.

03MAR2010 12Z
500hPa Geopotential (gdam), Bodendruck (hPa)



Daten: CFS Reanalysis
(C) Wetterzentrale
www.wetterzentrale.de

25APR2014 12Z
500hPa Geopotential (gdam), Bodendruck (hPa)



Daten: CFS Reanalysis
(C) Wetterzentrale
www.wetterzentrale.de

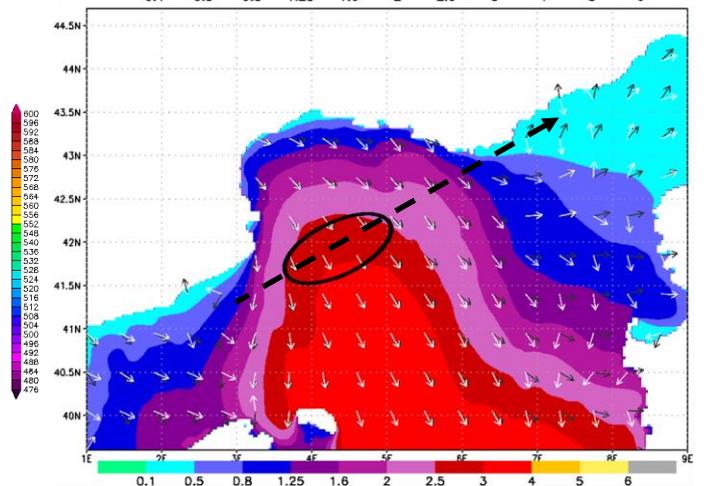
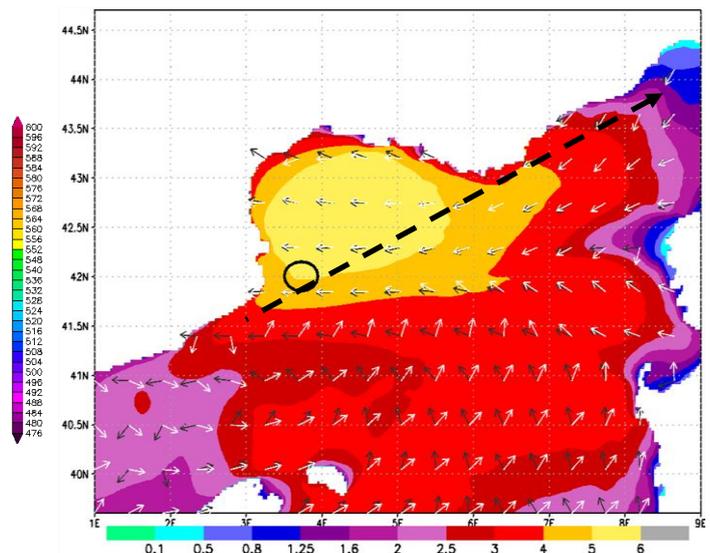


Figure SEQ Figure * ARABIC 6: Carte géopotentielle (à gauche) et carte de la hauteur significative des vagues, de la direction du vent et de la vague moyenne (à droite), relatives à l'étude de cas de 2014. La carte du géopotential fait référence à 12 UTC le 25/04/2014, la carte de la hauteur des vagues à 01 UTC le 26/04/2014.

36.2.2 Analyse statistique

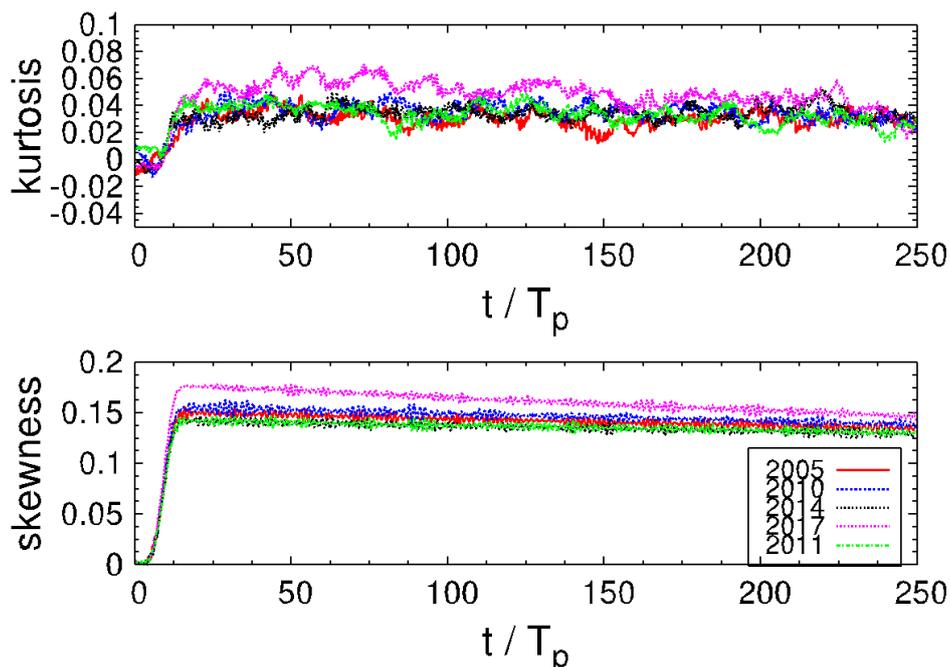


Le Tableau 1 présente les paramètres spectraux et les moyennes statistiques pour les cinq cas d'étude.

date	ε	σ_θ	λ_3	λ_4
14/2/2005	0.048	0.38	0.147	0.034
3/3/2010	0.053	0.45	0.152	0.038
26/4/2014	0.048	0.46	0.140	0.034
7/3/2017	0.058	0.33	0.171	0.057
17/12/2011	0.047	0.31	0.140	0.036

Tableau 1: paramètres spectraux et statistiques moyens pour les différentes études de cas

La pente de vague la plus forte est observée dans le cas de 2017, qui est également celui où l'asymétrie et l'aplatissement sont les plus élevés, tandis que le cas de 2005, bien

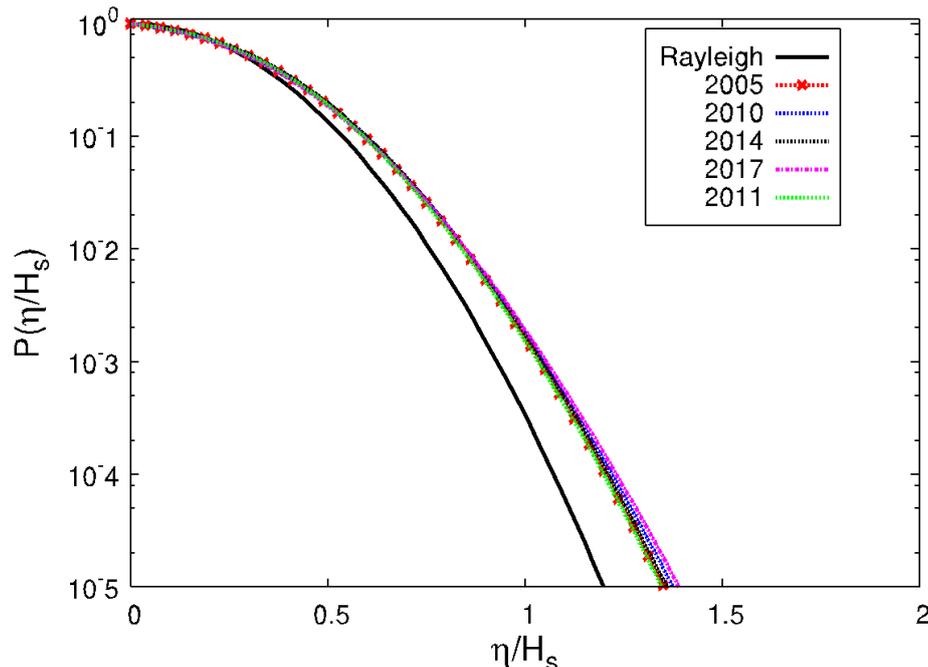


qu'ayant une forme spectrale très similaire, donne lieu à des valeurs plus faibles de ces statistiques. La Figure 7 montre également l'évolution temporelle de l'asymétrie et de l'aplatissement pour les cinq cas après la moyenne sur 20 réalisations.

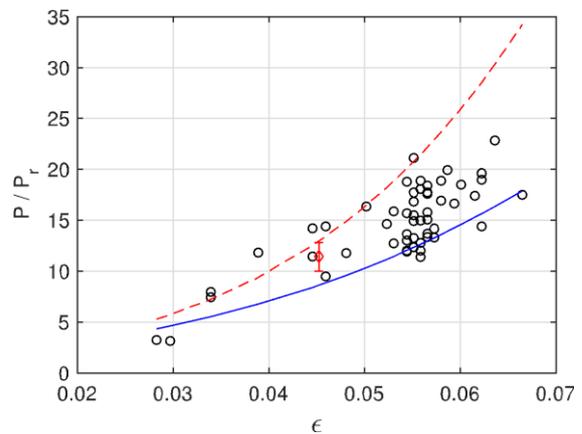
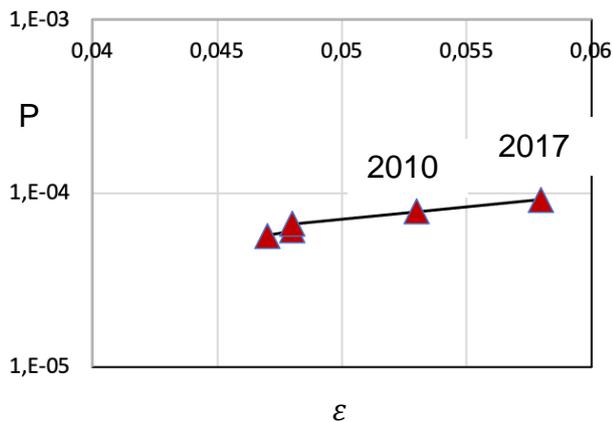
La tendance est la même que celle indiquée par les valeurs moyennes du Tableau 1: le seul cas présentant une évolution légèrement plus élevée de l'asymétrie et de l'aplatissement est celui de 2017, tandis que tous les autres cas ont un comportement très similaire. Ceci est également confirmé par la distribution des hauteurs de crête de la Figure 8, obtenue à partir d'une analyse statistique spatio-temporelle de toutes les crêtes. Dans tous les cas, la distribution de probabilité de la hauteur de la crête est nettement



supérieure à la distribution de probabilité de Rayleigh, comme on peut s'y attendre pour des vagues de mer réelles [21, 11], et par conséquent les valeurs d'aplatissement et d'asymétrie sont supérieures à zéro. La déviation la plus prononcée par rapport à la distribution de Rayleigh est à nouveau notée pour le cas d'étude 2017, tandis qu'une déviation moins prononcée, mais toujours pertinente par rapport aux autres cas, est également notée pour le cas 2010.



Enfin, la Figure 9 montre la probabilité de vagues scélérates en fonction de la raideur de la vague pour les 5 cas, qui, bien que dans une gamme limitée de valeurs, montrent une tendance bien définie, c'est-à-dire des probabilités plus élevées pour des valeurs plus élevées de la raideur. Cette dépendance a également été vérifiée par rapport à une plus grande base de données de cas d'étude, et les résultats sont présentés dans le panneau de droite de la figure. Dans ce cas, une étude plus approfondie a été faite de tous les événements à haute énergie qui se sont produits à l'endroit de l'accident de 2010 (c'est-à-dire près de la bouée de Begur) pendant toute l'année 2010, soit 54 cas au total. Là encore, l'augmentation de la probabilité en fonction de la pente est bien définie, les valeurs se situant approximativement dans une fourchette délimitée par les deux courbes obtenues à partir des distributions théoriques d'ordre 2 et 3.



3.7 Conclusions

L'utilisation d'une approche de modélisation en cascade phase-averaging / phase-resolving nous a permis d'analyser en détail certains cas d'étude intéressants afin d'évaluer précisément si certains incidents survenus dans le nord-ouest de la Méditerranée en raison de l'impact des vagues présentaient des conditions particulièrement favorables à la formation de vagues scélérates. La prévisibilité de tels événements est extrêmement difficile en raison de leur rareté et seules des analyses statistiques et des analyses "aux ingrédients" peuvent être effectuées. L'analyse des conditions synoptiques permet de comprendre la présence de conditions communes à l'occurrence de tels phénomènes qui peuvent constituer des indicateurs de risque. Dans le cadre de ces scénarios, afin d'avoir une indication plus précise et quantitative, il est nécessaire de recourir aux résultats de l'analyse statistique, qui montrent que les vagues à forte pente sont associées à des probabilités plus élevées de vagues scélérates. En outre, une autre conclusion importante est fournie par la forme du spectre: les conditions de mer croisées, donc associées à des spectres comportant plus d'un pic, peuvent être associées à une probabilité plus élevée de vagues scélérates par rapport à des spectres unimodaux dans des conditions d'énergie similaires.

3.8 References

- [1] N. Mori, P. Liu e T. Yasuda, «Analysis of freak wave measurements in the Sea of Japan,» *Ocean Engineering*, vol. 29, pp. 1399-1414, 2002.
- [2] F. Fedele e M. Tayfun, «On nonlinear wave groups and crest statistics,» *Journal of Fluid Mechanics*, vol. 620, 2009.
- [3] M. Onorato, A. Osborne, M. Serio, L. Cavaleri, C. Brandini e C. Stansberg, «Observation of strongly non-Gaussian statistics for random sea surface gravity waves in wave flume experiments,» *Physical Review E*, vol. 70, 2004.
- [4] P. A. E. M. Janssen, «Nonlinear Four-Wave Interactions and Freak Waves,» *Journal of Physical Oceanography*, vol. 33, pp. 863-884, 2003.
- [5] T. Waseda, T. Kinoshita e H. Tamura, «Evolution of a Random Directional Wave and Freak Wave Occurrence,» *Journal of Physical Oceanography*, vol. 39, pp. 621-639, 2009.
- [6] F. Barbariol, J. H. G. Alves, A. Benetazzo, F. Bergamasco, L. Bertotti, S. Carniel, L. Cavaleri, Y. Y. Chao, A. Chawla, A. Ricchi, M. Sclavo e H. Tolman, «Numerical modeling of space-time wave extremes using WAVEWATCH III,» *Ocean Dynamics*, vol. 67, pp. 535-549, 2017.
- [7] D. G. Dommermuth e D. K. Yue, «A high-order spectral method for the study of nonlinear gravity waves,» *Journal of Fluid Mechanics*, vol. 184, pp. 267-288, 1987.
- [8] B. J. West, K. A. Brueckner, R. S. Janda, D. M. Milder e R. L. Milton, «A new numerical method for surface hydrodynamics,» *Journal of Geophysical Research: Oceans*, vol. 92, pp. 11803-11824, 1987.
- [9] M. Tanaka, «A method of studying nonlinear random field of surface gravity waves by direct numerical simulation,» *Fluid Dynamics Research*, vol. 28, p. 41, 2001.
- [10] A. Toffoli, E. Bitner-Gregersen, A. R. Osborne, M. Serio, J. Monbaliu e M. Onorato, «Extreme waves in random crossing seas: Laboratory experiments and numerical simulations,» *Geophysical Research Letters*, vol. 38, 2011.

- [11] W. Xiao, Y. Liu, G. Wu e D. K. Yue, «Rogue wave occurrence and dynamics by direct simulations of nonlinear wave-field evolution,» *Journal of Fluid Mechanics*, vol. 720, pp. 357-392, 2013.
- [12] G. Ducrozet, F. Bonnefoy, D. Le Touzé e P. Ferrant, «Hos-ocean: Open-source solver for nonlinear waves in open ocean based on high-order spectral method,» *Computer Physics Communications*, vol. 203, pp. 245-254, 2016.
- [13] G. Ducrozet, F. Bonnefoy, D. L. Touzé e P. Ferrant, «3-d hos simulations of extreme waves in open seas,» *Natural Hazards and Earth System Sciences*, vol. 7, pp. 109-122, 2007.
- [14] F. Bonnefoy, G. Ducrozet, D. Le Touzé e P. Ferrant, «Time domain simulation of nonlinear water waves using spectral methods,» *In Advances in Numerical Simulation of Nonlinear Water Waves. World Scientific*, pp. 129-164.
- [15] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang e J. G. Powers, «A description of the advanced research wrf version 3. ncar technical note-475+ str,» 2008.
- [16] I. Nikolkina e I. Didenkulova, «Rogue waves in 2006-2010,» *Natural Hazards and Earth System Sciences*, pp. 2913-2924, 2011.
- [17] E. Didenkulova, «Catalogue of rogue waves occurred in the world ocean from 2011 to 2018 reported by mass media sources,» *Ocean & Coastal Management*, vol. 188, p. 105076, 2020.
- [18] I. Didenkulova, A. Slunyaev, E. Pelinovsky e C. Kharif, «Freak waves in 2005,» *Natural Hazards and Earth System Sciences*, pp. 1007-1015, 2006.
- [19] L. Cavaleri, L. Bertotti, L. Torrisi, E. Bitner-Gregersen, M. Serio e M. Onorato, «Crossing sea state and rogue wave probability during the prestige accident,» *Journal of Geophysical Research: Oceans*, vol. 120, pp. 7113-7136, 2012.
- [20] Puerto del Estado, 2015. [Online]. Available: <https://bancodatos.puertos.es/bd/informes/int2.pdf>. [Consultato il giorno 5 10 2022].
- [21] F. Fedele, J. Brennan, S. DeLeon, J. Dudley e F. Dias, «Real world ocean rogue waves explained without the modulational instability,» *Scientific reports*, vol. 6, pp. 1-11, 2016.
- [22] N. Mori e P. A. Janssen, «On kurtosis and occurrence probability of freak waves,» *Journal of Physical Oceanography*, vol. 36, pp. 1471-1483, 2006.



4 Partie 2

Dans le cadre de l'activité T1.1 du projet GIAS "Identification des risques difficilement prévisibles liés aux actions atmosphériques et marines", le produit T1.1.3 "Procédure commune pour la prédiction des risques naturels difficilement identifiables: composante marine" a parmi ses objectifs la mise en œuvre de méthodologies pour la prédiction de phénomènes atmosphériques dangereux difficilement prévisibles pour la navigation en haute mer. Pour cette activité, ARPAL a activé une collaboration technico-scientifique avec le Département d'ingénierie civile, chimique et environnementale de l'Université de Gênes, pour la mise en œuvre, la validation et l'optimisation de l'indice de risque côtier au sein de la chaîne de prévision du mouvement des vagues WW3 d'ARPAL, dont le rapport technique est joint.

4.1 Introduction

Ces dernières années, la côte ligure a été victime d'intenses tempêtes de mer, qui ont causé des dommages considérables aux biens et aux infrastructures le long du littoral. La figure montre deux clichés des tempêtes de mer survenues à des dates différentes en Ligurie: octobre 2018 à gauche et novembre 2019 à droite. En ce qui concerne le premier événement, les dommages estimés s'élèvent à environ 180 millions d'euros (Chambre de commerce de Gênes).



Dans ce contexte, il est apparu nécessaire de mettre en place un système de prévision efficace, qui permette de réguler les activités côtières suffisamment à l'avance par rapport aux ondes de tempête prévues et, dans des cas exceptionnels, de déplacer les biens exposés (petits bateaux, etc.), lorsque cela est possible. C'est

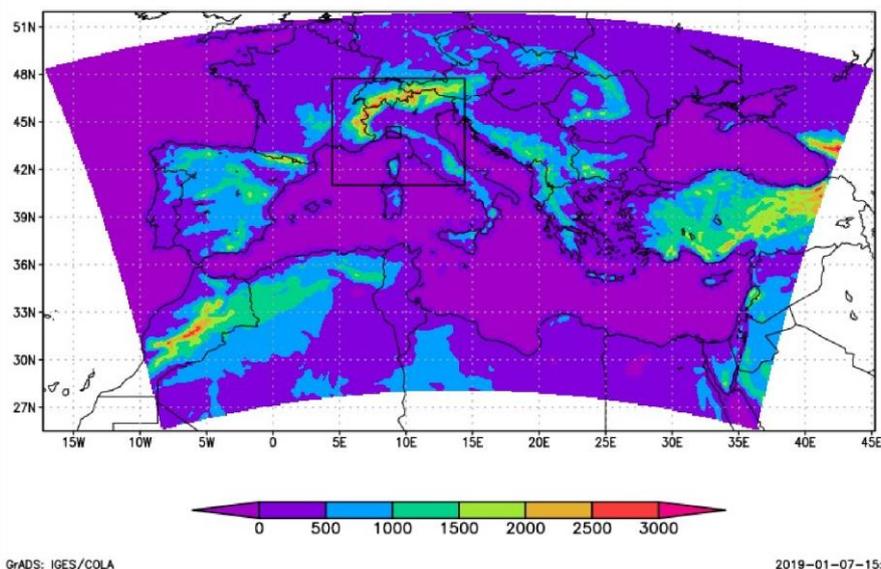
la raison pour laquelle ce projet et la collaboration technico-scientifique pour le développement d'un indice de risque côtier ont été développés, visant à définir un Système d'Alerte Précoce sur la base des prévisions de vagues développées par la chaîne de modélisation en place au Centre Météo-Hydrologique Fonctionnel de la Protection Civile de Ligurie.

4.2 Modèle de prévision maritime en Méditerranée

1.

4.2.1 Modélisation atmosphérique

Les simulations effectuées avec wviii utilisent les champs de vent hautement détaillés générés par WRF (Skamarock, 2008) sur le bassin étudié, mis en évidence dans la figure ci-dessous. Le WRF est initialisé par des simulations du modèle à l'échelle mondiale Global Forecast System GFS.



4.2.2 Modélisation météo-marine

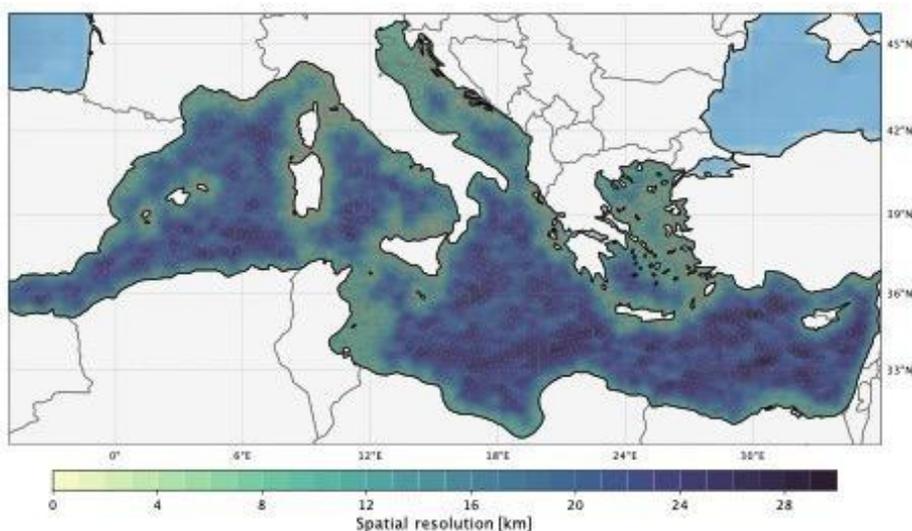
La chaîne de modélisation de la prévision des vagues utilise les informations introduites précédemment sur le forçage atmosphérique afin de prévoir le climat des vagues au large de la côte de la mer Tyrrhénienne supérieure.

Les sorties du modèle météorologique sont donc nichées dans le modèle de vagues wviii, un modèle de génération et de propagation des vagues de mer qui permet de simuler les variations spatio-temporelles de l'élévation de la surface de la mer, du courant de surface, de l'oscillation de surface due au vent et des effets bathymétriques (Tolman, 2009). Les sorties du modèle comprennent, entre autres, la

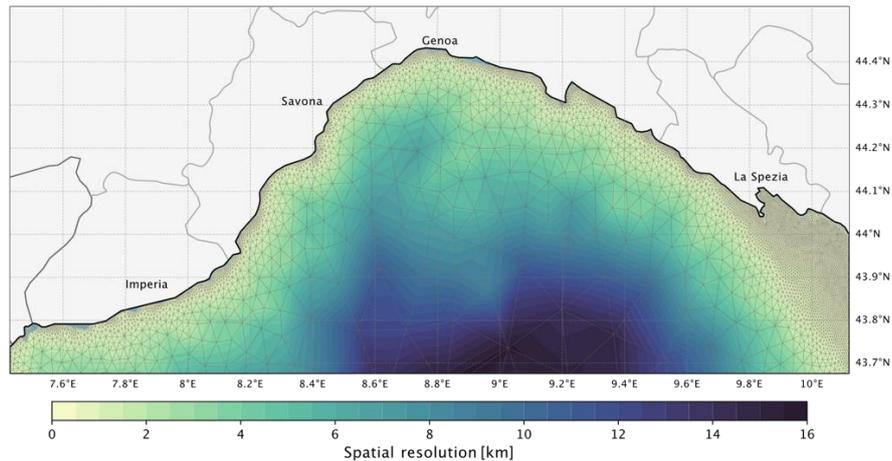
hauteur significative des vagues avec sa période et sa direction associées (valeurs moyennes et de pointe), définies sur les intervalles de temps souhaités.

La résolution de la grille de calcul sur laquelle les paramètres des vagues doivent être résolus a été améliorée en traitant une grille de calcul non structurée avec une résolution variable depuis le large, où un degré de détail moindre est requis, vers la côte, où les effets bathymétriques du fond deviennent prédominants et peuvent être résolus avec précision au moyen d'une plus grande résolution du maillage de calcul. La résolution a été augmentée (~10 km) pour atteindre une résolution de 300 m sur le littoral de la zone d'étude. Inversement, lorsque les conditions du fond marin sont telles que les conditions des vagues ne changent pas, la résolution a été réduite à ~25 km. En dehors de la zone d'intérêt, en revanche, une résolution de ~10 km a été maintenue sur le littoral, en augmentant progressivement la distance jusqu'à ~25 km dans des conditions de "profondeur infinie".

Le maillage de calcul résultant et un zoom du maillage dans la zone d'étude sont présentés dans la figure ci-dessous. Sur cette figure, on peut également constater que de nombreuses îles situées en dehors de la zone du projet ont été supprimées afin de rationaliser la chaîne de calcul et d'augmenter les performances du modèle (par exemple, de nombreuses îles de l'archipel des Cyclades). Dans tous les cas, les îles retirées sont suffisamment éloignées de la zone de la Haute Tyrrhénienne, et il est donc raisonnable de supposer que la qualité des simulations dans cette zone n'est pas compromise (comme le prouvent les comparaisons avec les données des bouées, présentées plus loin). La maille est caractérisée par un nombre total de nœuds de 65392.



Grille de calcul non structurée sur la mer Méditerranée



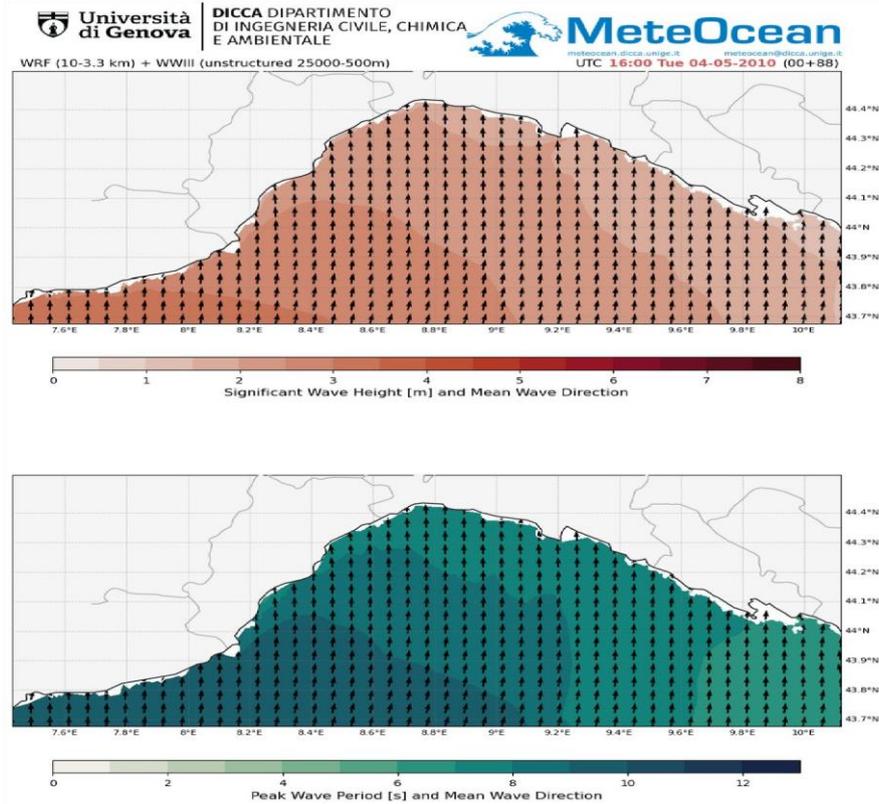
Zoom de la grille de calcul dans la zone d'étude

4.2.3 Calibrage du modèle

Une fois la grille de calcul produite et les temps de calcul vérifiés selon les paramètres du modèle, le modèle a été calibré sur les ondes de tempête historiques survenues dans la zone du projet. Le but de la calibration est de définir le réglage optimal de w_{iii} afin de maximiser les comparaisons entre les données observées et simulées. À cette fin, on a utilisé des données de vagues échantillonnées à partir de bouées appartenant au réseau EMODNET (<https://www.emodnet.eu/en>), et donc considérées comme d'une certaine fiabilité.

Enfin, les paramètres jugés optimaux pour les deux termes sources différents ont été utilisés sur la même grille de calcul à l'aide de la suite de codes UOST (Mentaschi et al., 2019), implémentée pour prendre en compte les obstacles non résolus par le maillage (c'est-à-dire les îles supprimées pour rationaliser les temps de calcul).

La figure présente les résultats de la houle de 2010 :



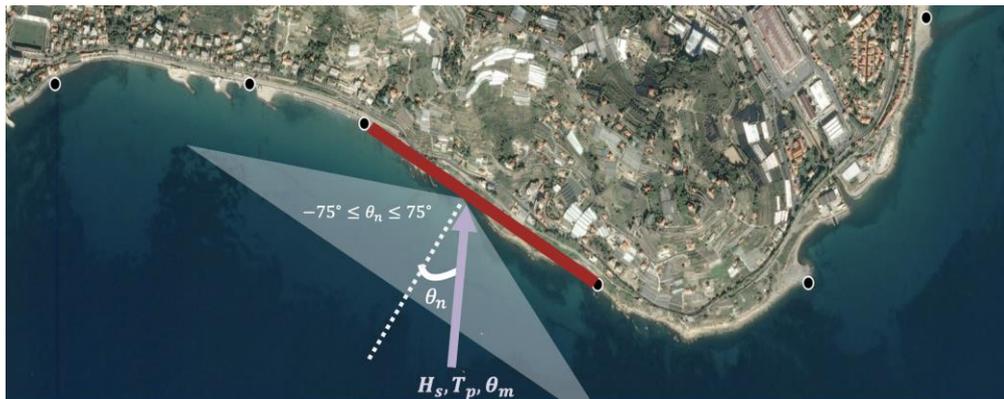
Hauteur et période des vagues de la houle 2010.

4.2.4 Storm Power dans les données historiques de la Ligurie

Ce travail propose l'utilisation du flux d'énergie des vagues, ou puissance des vagues par unité de longueur de front de vague (P), comme mesure instantanée de la houle qui mesure l'interdépendance de la hauteur significative des vagues H_s , de la période de pic, T_p , et de la direction moyenne d'incidence par rapport à la direction moyenne du trait de côte:

$$P = \int_0^d E(t) \cdot c_g(t) dt = \int_0^d \left(\frac{1}{16} \rho g H_s(t)^2 \right) \frac{g T_p(t)}{4\pi} \cos \cos \theta_n dt$$

où g est l'accélération gravitationnelle et ρ est la densité de l'eau. Différents points ont été sélectionnés tous les 2 km environ le long de la côte ligure et l'orientation moyenne de la côte a été dérivée de la procédure sur 10 km environ. La puissance a été calculée en utilisant la série temporelle de rétrodiffusion la plus proche du littoral.



La puissance des tempêtes est évaluée pour chaque état de mer dans la simulation rétrospective à l'aide d'une fenêtre mobile de $d=18$ heures, ce qui permet d'obtenir une série chronologique P de même longueur que la rétrospective et sa fonction de distribution cumulative empirique (ECDF) correspondante.

4.2.5 Storm Power Index (SPI), Coastal Vulnerability Index (CVI) e Risk Index (RI)

Ce travail utilise une classification de SPI selon la probabilité d'occurrence de l'énergie houlomotrice, définissant cinq classes allant de très faible ($SPI = 1$) à très élevée ($SPI = 5$) comme présenté dans le tableau:

Probability	≤ 0.4	$(0.4 - 0.75]$	$(0.75 - 0.9]$	$(0.9 - 0.99]$	$(0.99 - 1]$
SPI	1 <i>very low</i>	2 <i>low</i>	3 <i>medium</i>	4 <i>high</i>	5 <i>very high</i>

L'indice de vulnérabilité côtière (CVI) est défini comme la sensibilité de la zone côtière au risque de vagues et une mesure de sa capacité à faire face et à résister aux événements de tempête et à leurs impacts (Gilard et Givone, 1996 ; Weichselgartner, 2001). Comme l'indice de puissance des tempêtes, le CVI est divisé en cinq classes ($CVI = 1-5$) par une analyse qualitative de la région côtière étudiée.

CVI	Descrizione
1 - Molto bassa	Quando vi sia cliff/scogliera o comunque costa a livello del mare ma in assenza di infrastrutture viarie (strade/ferrovie), e/o abitazioni e stabilimenti balneari
2 - Bassa	Quando vi siano prospicienti alla costa infrastrutture viarie protette, o con opere di protezione (dighe soffolte/revetments etc.) o per la morfologia della costa. Ovvero, infrastrutture sufficientemente elevate rispetto al livello medio mare o all'interno. Ad ogni modo, questo livello implica che NON vi siano abitazioni e/o stabilimenti balneari.
3 - Media	Quando vi siano prospicienti alla costa infrastrutture viarie NON protette. Di nuovo, questo livello implica che NON vi siano abitazioni e/o stabilimenti balneari
4 - Alta	Livello analogo al livello 2 ma in presenza di case e/o stabilimenti balneari
5 - Molto Alta	. Livello analogo al livello 3 ma in presenza di case e/o stabilimenti balneari

L'évaluation des risques comprend la combinaison du danger et de la vulnérabilité, définis comme suit,

$$RI = SPI \cdot CVI$$

Puisque *SPI* et *CVI* sont définis dans une plage comprise entre 1 et 5, des valeurs *RI* comprises entre 1 et 25 sont fournies. Pour des raisons de cohérence, le *RI* est également divisé en cinq classes, allant de *RI* = 1 (très faible) pour *SPI*·*CVI*=1-3 à *RI*=5 (très élevé) pour *SPI*·*CVI*=16-25 :

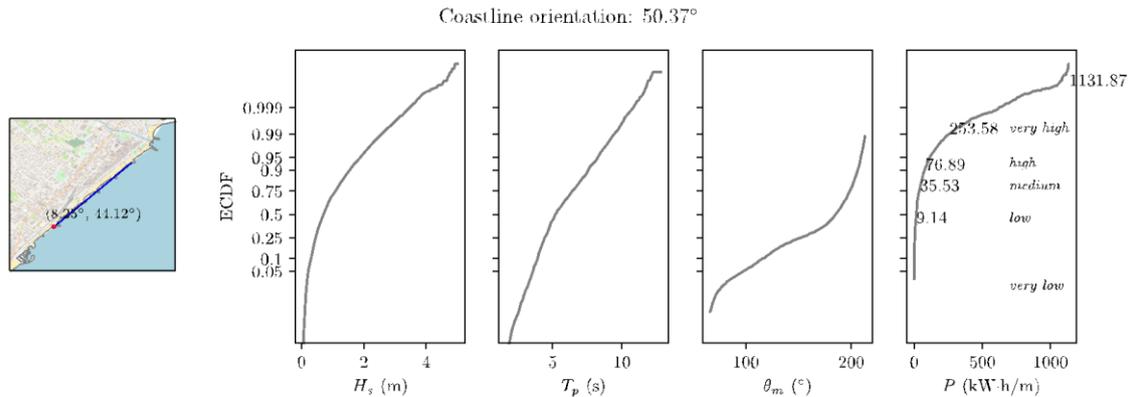
Risk Index (<i>RI</i>)	<i>SPI</i>						
	1	2	3	4	5		
<i>CVI</i>	1	1	2	3	4	5	<i>very low</i>
	2	2	4	6	8	10	<i>low</i>
	3	3	6	9	12	15	<i>medium</i>
	4	4	8	12	16	20	<i>high</i>
	5	5	10	15	20	25	<i>very high</i>

L'évaluation du risque et du risque de vagues est effectuée sur la côte ligurienne. La côte ligurienne comprend un mélange de côtes rocheuses et de plaines et se caractérise par un niveau élevé d'urbanisation et d'infrastructures portuaires, ainsi que par des autoroutes côtières et des chemins de fer.

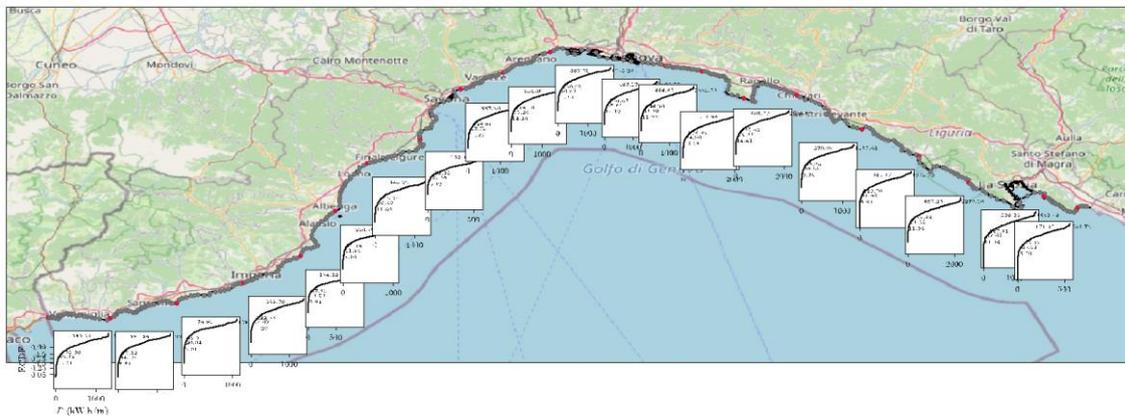
Près de 400 points ont été définis le long des 350 km de côtes de la Ligurie sur la base des caractéristiques du littoral et des infrastructures existantes. Les points définis délimitent des tronçons de côte de ≈0,15-2 km pour lesquels la puissance des vagues, *P* a été calculée en utilisant une durée de *d* = 18 h, en raison du fait que les tempêtes dans la mer Tyrrhénienne résultent généralement du passage de systèmes de basse pression qui expirent en moins d'une journée, et en utilisant l'orientation côtière moyenne de chaque tronçon côtier comme référence pour θ_n . Les conditions de houle ont été obtenues à partir de simulations rétrospectives du 1er janvier 1979 au 31 décembre 2020 en utilisant le nœud situé à une distance de 10 km de chaque point. Ensuite, la série chronologique de la puissance des vagues, *P*, est calculée et la fonction de distribution cumulative empirique (ECDF) est obtenue, ce qui fournit des informations sur la probabilité que *P* soit inférieure ou égale à une valeur donnée de la puissance des vagues de la tempête et permet une définition efficace de l'indice de puissance des tempêtes.

La définition *SPI* de chaque point est obtenue en suivant différentes valeurs de probabilité. La figure présente, pour trois emplacements et en considérant leur

orientation, les ECDF de H_s , T_p , θ_m , P et SPI comme exemple des résultats obtenus pour l'ensemble du littoral ligure.



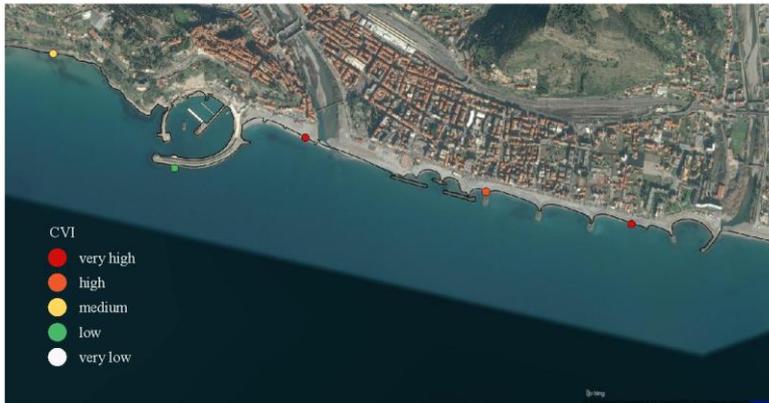
La figure présente l'ECDF de P et la définition de SPI pour différents points de la côte ligure. Le tableau présente les valeurs minimales, moyennes et maximales de P au sein des points analysés pour différentes valeurs de probabilité. Le long du littoral ligure, des valeurs moyennes de SPI sont obtenues pour P pour (52,38-122,74) kW h/m, des valeurs élevées de SPI pour (122,74-400,67) kW h/m et des valeurs très élevées pour (400,67-1535,82) kW h/m.



Probability	Minimum	Mean	Maximum
100	365.94	1535.82	2567.59
99	132.84	400.67	678.13
90	35.46	122.74	200.90
75	14.45	52.38	85.44

Pour chaque lieu, un indice de vulnérabilité côtière (CVI) correspondant aux cinq classes définies en fonction des caractéristiques du littoral et de la présence

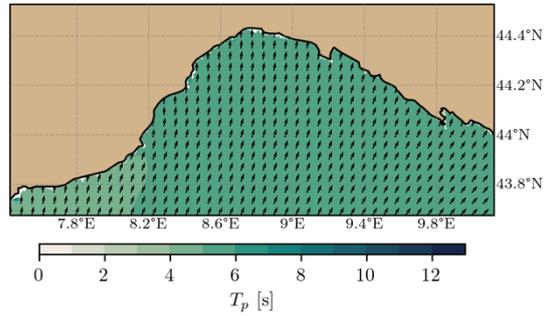
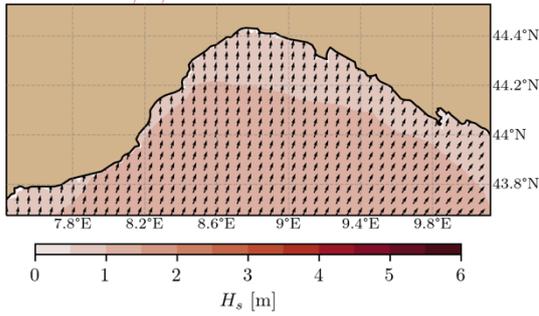
d'infrastructures a été défini de manière qualitative. La figure présente des exemples de *CVI* définis pour différents endroits de la côte ligurienne.



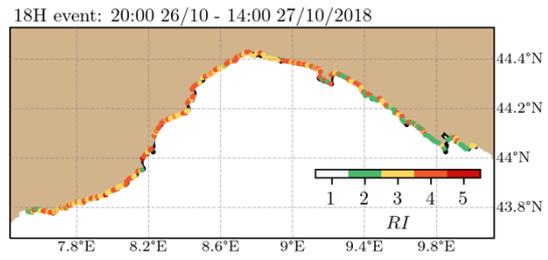
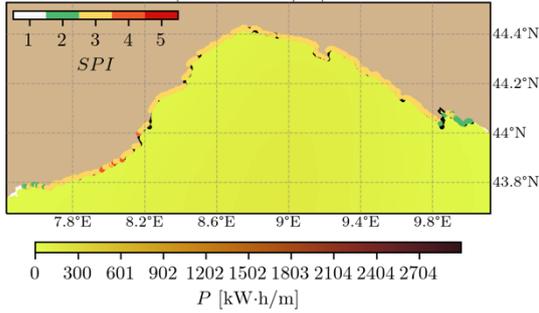
En suivant la définition de *SPI* varie en fonction de la puissance de tempête des vagues P pour chaque emplacement et de son *CVI*, nous pouvons calculer la série chronologique prévisionnelle de P et l'indice de puissance de tempête correspondant, *SPI*, et l'indice de risque de tempête, *RI*.

À titre d'exemple, les figures ci-dessous présentent les résultats de la prévision de la hauteur significative des vagues H_s , de la période de pointe T_p , de la puissance de tempête des vagues P et de l'indice de puissance de tempête *SPI* et de l'indice de risque *RI* pour deux moments de tempête en octobre 2018.

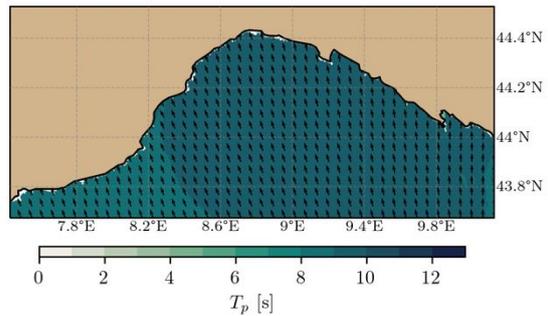
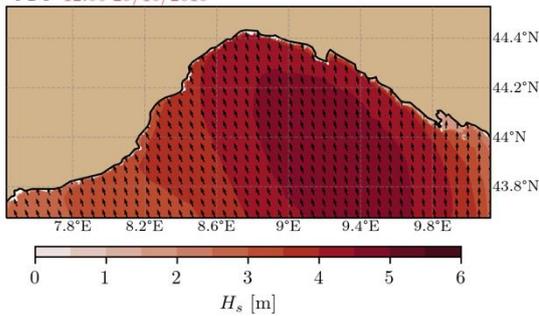
UTC 20:00 26/10/2018



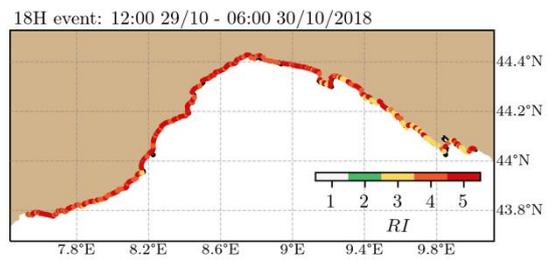
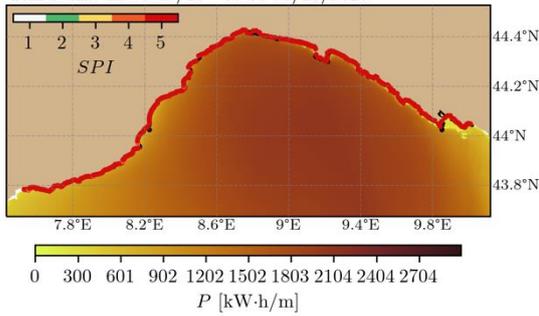
18H event: 20:00 26/10 - 14:00 27/10/2018



UTC 12:00 29/10/2018



18H event: 12:00 29/10 - 06:00 30/10/2018



4.3 References

Camera di Commercio di Genova

Gilard, O., Givone, P.-E., 1996. Flood risk management : new concepts and methods for objective negotiations. In: Destructive Water Conference AISH Conference,

Anaheim, USA. United States, p. 13.

Mentaschi, L., Vousdoukas, M., Besio, G., Feyen, L., 2019. alphaBetaLab: Automatic estimation of subscale transparencies for the Unresolved Obstacles Source Term in ocean wave modelling. *SoftwareX* 9, 1–6.
<http://dx.doi.org/10.1016/j.softx.2018.11.006>.

Skamarock, W.C., 2008. A description of the advanced research WRF version 3. Tech. Note 1–96.

The WAVEWATCH III® Development Group, 2019. User Manual and Documentation WAVEWATCH III® v6.07. Technical Report.

Weichselgartner, J., 2001. Disaster mitigation: the concept of vulnerability revisited. *Disaster Prev. Manage.: Int. J.* 10 (2), 85–95. <http://dx.doi.org/10.1108/09653560110388609>.



5 Partie 3

5.1 Contexte

5.1.1 Le contexte du Projet

L'objectif du projet GIAS est d'accroître la sécurité de la navigation dans la région de Pelagos, située au nord-ouest de la Méditerranée (Figure 1). Il s'agit d'atténuer le risque de collision pendant la navigation en augmentant la sensibilisation aux objets à la dérive situés à proximité. En outre, la dispersion des polluants et des gros objets doit être connue pour permettre une intervention de nettoyage appropriée. En définitive, pour atteindre l'objectif du projet, il faut développer un modèle qui améliore la tâche de suivi des objets à la dérive.



Figure 1: The Pelagos region

Le projet vise à atteindre cet objectif en tirant parti des progrès récents de l'apprentissage automatique. Au cours de la dernière décennie, l'apprentissage automatique - et plus particulièrement l'apprentissage profond - a démontré sa capacité à résoudre de nombreux types de tâches dans de nombreux types de domaines grâce à l'augmentation substantielle de la puissance de calcul et de la disponibilité des données [1]. Au fil du temps, la capacité de l'apprentissage automatique s'est améliorée de façon continue et rapide, et l'objectif autrefois



ambitieux de dépasser les performances humaines devient de plus en plus courant pour de nombreuses tâches. La capacité d'apprendre des informations et des compétences qui transcendent nos connaissances et nos limites actuelles est un développement clé qui peut nous aider à faire progresser de manière significative nos outils et modèles scientifiques.

Cependant, l'application réussie de l'apprentissage automatique à un domaine spécialisé nécessite encore beaucoup d'expertise interdisciplinaire. Avec le récent regain d'intérêt pour l'apprentissage automatique, son application à des problèmes scientifiques est principalement restée dans la phase expérimentale initiale. Cela a souvent conduit les chercheurs à avoir un biais de connaissances vers l'une des deux disciplines, où un manque de connaissances de l'autre entraîne une diminution des chances de succès. Pour éviter cette limitation, ce projet sera mené par une équipe de chercheurs de différents groupes d'expertise, assurant une contribution égale des connaissances en apprentissage automatique et en océanographie physique. Le projet sera réalisé par les trois groupes de recherche suivants :

LIS, équipe DYNI (UTLN) - Experts en apprentissage automatique et en science des données appliquées aux données scientifiques.

MIO, équipe OPCL (UTLN) - Experts en physique de la dynamique des océans.

Ocean Next - Experts dans le travail sur les données géoscientifiques en relation avec l'océanographie et l'hydrologie.

5.1.2 Problème de formulation

Le suivi d'objets à la dérive est nécessaire pour de nombreuses applications, notamment : la recherche et le sauvetage, le suivi de polluants tels que les marées noires ou les déchets plastiques, et le suivi d'objets de grande taille tels que les conteneurs de marchandises ou les baleines mortes. Les applications peuvent relever des trois scénarios opérationnels possibles que sont les prévisions rétrospectives, les prévisions immédiates et les prévisions, qui visent à décrire l'état de la mer dans le passé, le présent et le futur, respectivement.

À l'aide d'instruments d'observation, la mer peut être mesurée pour comprendre son comportement dans l'état actuel. Par exemple, les dériveurs GPS fournissent des informations sur le comportement des objets à la dérive en mesurant des données très précises de vitesse et de localisation à un moment donné. En général, les données d'observation seules ne peuvent pas être utilisées pour des applications de

prévision a posteriori ou de prévision, à l'exception peut-être de l'utilisation d'analyses rétrospectives à grande échelle pour découvrir des tendances récurrentes à long terme. Même ainsi, l'utilisation exclusive de données d'observation dans tout scénario opérationnel est rarement pratique. La nature passive des bouées dérivantes, combinée à la nature chaotique de l'océan, rend leurs trajectoires hautement imprévisibles. De plus, le déploiement des bouées en vrac est à la fois extrêmement coûteux et long, et la quantité de couverture spatiale et temporelle est en fin de compte un facteur limitatif important pour la plupart des cas d'utilisation.

Les modèles numériques fournissent aux océanographes physiques un substitut très pratique aux données d'observation. Ils mettent en œuvre les principes physiques qui régissent l'océan, tels que les équations de Navier-Stokes, ce qui permet de réaliser des simulations dans toutes les conditions souhaitées. Ils sont peu coûteux et presque instantanés par rapport au temps réel, ce qui constitue un moyen efficace de produire des expériences à grande échelle pour l'analyse. La liberté de mouvement le long de la dimension temporelle les rend également capables d'être utilisés pour des applications de prévision a posteriori et de prévision.

Cependant, le suivi fiable d'objets dérivants à l'aide de modèles numériques est un problème intrinsèquement difficile en raison de l'accumulation massive d'erreurs. Les modèles numériques doivent être discrétisés sur une grille d'une résolution déterminée, ce qui limite par conséquent l'échelle des caractéristiques qui peuvent être modélisées. Les modèles de prévision, en particulier, doivent sacrifier une partie importante de leur résolution, ce qui limite leur capacité à prévoir à court terme en raison de leur manque de connaissance de la dynamique à petite échelle. En outre, la physique sous-jacente qui régit la dynamique de l'océan est très complexe en raison de sa nature turbulente, ce qui signifie que la résolution de la dynamique à petite échelle ne consiste pas seulement à améliorer la résolution, mais aussi à améliorer notre compréhension de la dynamique de l'océan. Le manque de compréhension nuancée de la physique, la nature discrète et finie du calcul, et les approximations faites en faveur d'une complexité de calcul réduite contribuent tous à des erreurs dans les modèles numériques pour les applications à courte échelle de temps. L'effet de ces erreurs est encore amplifié par les incertitudes liées à des paramètres tels que la position et les caractéristiques initiales d'un objet, ainsi que les conditions atmosphériques, météorologiques et océaniques. Par exemple, le mouvement des objets à la dérive dépend de la façon dont les vents, les courants et les vagues agissent sur l'objet, et il est donc impératif de connaître ces variables. Les modèles de prévision opérationnels nous donnent une estimation de ces variables mais sont limités par leur résolution et l'échelle de temps à laquelle ils restent précis.



5.2 Méthodologies actuelles

Plusieurs travaux ont développé des méthodologies pour prendre en compte les incertitudes dans la modélisation des objets à la dérive. Différents types d'objets ont des incertitudes différentes, et donc chaque méthodologie se concentre généralement sur une application spécifique. Des exemples de telles applications sont la recherche et le sauvetage [2-5], les marées noires [3, 6, 7], les conteneurs de fret [8], les carcasses [9, 10] et les particules virtuelles [11, 12].

Les incertitudes peuvent être prises en compte par la modélisation stochastique, où les paramètres inconnus sont considérés comme des variables aléatoires décrites par une fonction de densité de probabilité (PDF). En simulant plusieurs trajectoires, chacune avec une variation des paramètres, l'emplacement des objets par rapport au temps peut être décrit par une fonction de vraisemblance. Cependant, il n'est pas évident de définir des PDF pour chaque paramètre, ce qui fait que la fonction de vraisemblance est mal définie. Dans [12], ils supposent que la PDF est modélisée par une certaine distribution antérieure et estiment ensuite ses paramètres en maximisant la probabilité que la PDF décrive un ensemble de trajectoires observées localement. Une approche alternative est utilisée dans [3] qui ne nécessite pas de données d'observation ; au lieu de cela, ils créent synthétiquement les trajectoires utilisées pour estimer la PDF en introduisant du hasard dans la composante turbulente du modèle de trajectoire. Pour des raisons de simplicité, les deux méthodes supposent que la PDF est modélisée par une distribution gaussienne.

Viser à réduire les incertitudes plutôt qu'à les expliquer est un problème beaucoup plus difficile, puisque les incertitudes existantes sont le résultat des limites inhérentes aux progrès actuels de la science et de la technologie. Il faut combiner l'amélioration de trois obstacles fondamentaux : nos ressources informatiques, le développement des instruments et la compréhension de la physique. Le travail de [13] est un exemple de la façon dont l'apprentissage automatique pourrait être utilisé pour surmonter ces obstacles. Ils ont cherché à savoir s'il était possible d'utiliser les données des dériveurs GPS pour apprendre des termes inconnus de la dynamique des océans par le biais d'un réseau neuronal afin de faciliter le développement d'un modèle de prévision des objets à la dérive. Leurs expériences ont montré des preuves prometteuses que la physique précédemment inconnue avait effectivement été incorporée dans leur modèle. Cependant, les expériences ont également révélé une suradaptation radicale à leurs données d'entraînement, ce qui a rendu le modèle inutile pour un déploiement dans le monde réel. Néanmoins, la capacité de saisir certaines des subtilités de la dynamique des océans à petite échelle à partir de données est une étape clé pour surmonter les obstacles susmentionnés :

1. Les modèles capturés à partir des données ne sont pas limités par notre compréhension actuelle de la physique.
2. Contrairement aux instruments à faible résolution actuellement utilisés dans les modèles de prévision, les dériveurs GPS sont spatialement continus et ont une très haute résolution temporelle.
3. Les réseaux neuronaux peuvent potentiellement apprendre une représentation plus efficace des équations de la physique, réduisant ainsi la charge de calcul.

Comme mentionné précédemment, l'utilisation des dériveurs GPS n'est pas très pratique en raison de leur couverture effective limitée. Les auteurs de l'étude [13] ont bien montré comment cela a des répercussions importantes sur la capacité d'un modèle entraîné à se généraliser au monde réel. Actuellement, à notre connaissance, seuls deux travaux [13, 14] ont tenté d'utiliser l'apprentissage automatique pour le suivi d'objets à la dérive. Chaque méthode repose sur l'utilisation exclusive de techniques traditionnelles d'apprentissage supervisé pour apprendre la distribution des données réelles. Puisque la distribution des dériveurs GPS est si limitée, et représente donc un sous-ensemble trivial de la distribution du monde réel régie par les complexités de la dynamique océanique, cette approche constitue une limitation considérable de ces travaux car un modèle formé de cette manière ne peut jamais se généraliser au monde réel.

5.3 Projet de recherche

Par l'utilisation de l'apprentissage automatique, nous visons à améliorer l'état de l'art des modèles existants pour le suivi des objets à la dérive en apprenant à

1. Mieux prendre en compte les incertitudes dans la distribution des paramètres.
2. Mieux prendre en compte le manque de compréhension physique et de capacité de modélisation de la dynamique océanique à petite échelle.

En plus de l'entraînement de notre modèle sur des données réelles de dériveurs, nous utiliserons d'autres sources de données pour augmenter la quantité d'informations que notre modèle peut apprendre. Cela contribuera à rendre le modèle plus adapté au déploiement en réduisant le risque de suradaptation à ses conditions d'entraînement. Tout d'abord, l'utilisation de dériveurs synthétiques nous permettra de créer une distribution d'échantillonnage beaucoup plus efficace et représentative des variations possibles des objets à la dérive. Cela ira de pair avec l'utilisation de modèles numériques pour modéliser les scénarios réels et synthétiques. En outre, nous pourrions également utiliser d'autres sources de données d'observation telles

que la hauteur et la température de la surface de la mer provenant de satellites, ou les courants de surface de la mer provenant de radars HF.

Avec l'accès à un vaste éventail de données, il existe de nombreuses méthodologies possibles qui peuvent être utilisées pour extraire des informations non encore décrites par les modèles physiques. Par exemple, les modèles opérationnels nous fournissent des modèles précis assimilés à des données, tandis que les modèles non opérationnels nous fournissent une grille à plus haute résolution. En utilisant les deux modèles, nous pourrions chercher à apprendre une cartographie qui combine les avantages des deux, produisant ainsi un modèle opérationnel à haute résolution à utiliser pour les prévisions. En effet, le modèle apprendra indirectement à mettre à l'échelle les données assimilées par le biais de la cartographie.

Toutefois, ce processus est assez coûteux, car il n'existe pas de données d'observation de base à haute résolution à partir desquelles apprendre une cartographie, et l'ajustement linéaire des techniques d'assimilation des données ne permet pas de saisir les non-linéarités des données. L'émergence récente des réseaux neuronaux informés par la physique peut aider à surmonter ce problème, car nous pourrions intégrer des connaissances physiques dans le modèle pour tenir compte de l'incomplétude des données de base. Dans ce cas, l'ajout de connaissances physiques servira de couche supplémentaire de vérité de base afin de limiter l'extrapolation du modèle assimilé à une résolution plus élevée. En fait, l'intégration de connaissances en physique dans nos modèles apprenants s'avérera une technique très utile, car elle permettra d'améliorer la convergence lors de l'apprentissage de dynamiques complexes régies par des équations différentielles partielles non linéaires [15]. La figure 2 montre un exemple de la manière dont les connaissances en physique peuvent être intégrées dans un réseau neuronal.

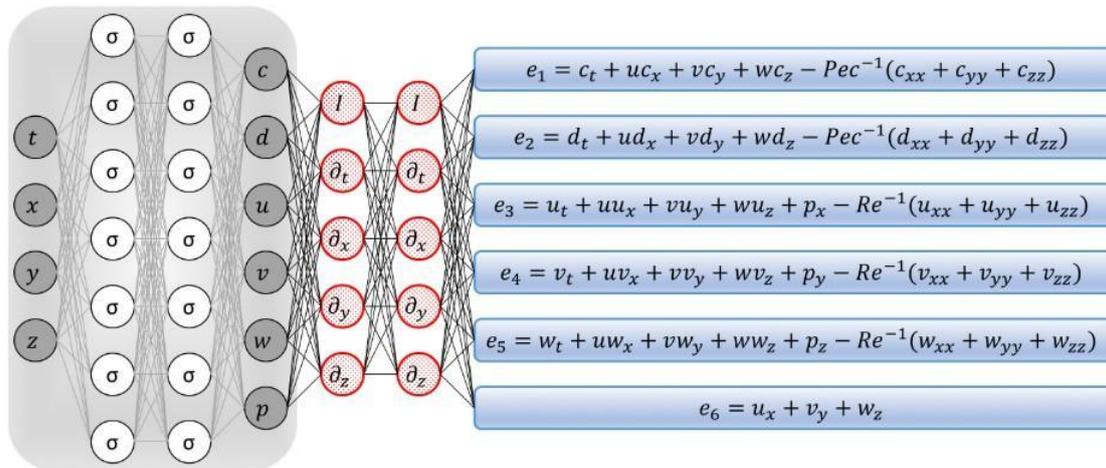


Figure 2 : Réseau neuronal de Navier-Stokes informé par la physique [16]. Étant donné l'entrée t, x, y, z , le modèle s'assure que ses sorties c, d, u, v, w, p sont contraintes par les termes e_1 - e_6 des équations de Navier-Stokes pendant l'apprentissage.

Nous pouvons également utiliser des moteurs physiques existants tels que le moteur océanique NEMO [17] pour encoder des connaissances physiques plus complexes dans un réseau neuronal. Des scénarios synthétiques peuvent être créés de telle sorte qu'un modèle puisse être entraîné sur un sur-ensemble de n'importe quel ensemble de paramètres du monde réel. Cela garantit qu'un modèle est capable de produire des estimations raisonnables pour n'importe quel scénario, même si de nouveaux scénarios du monde réel sont introduits au fil du temps (ce que l'on appelle la dérive des concepts). Cependant, un tel modèle souffrira de l'écart de réalité qui existe entre la distribution simulée et la distribution du monde réel. Il est courant de traiter ce problème par l'application de techniques d'adaptation au domaine, qui visent à transférer un modèle appris dans un domaine source à un domaine cible connexe. En utilisant des données d'observation réelles, une telle technique peut être utilisée pour réajuster la distribution simulée à une distribution du monde réel (par exemple, une région particulière de l'océan comme Pelagos). Puisque la tâche de réajuster une distribution est beaucoup plus simple que d'apprendre la distribution à partir de zéro, un modèle peut apprendre à généraliser même lorsqu'il est entraîné avec un ensemble limité d'exemples réels.

L'utilisation de sources de données secondaires provenant de satellites ou de radars HF peut également aider à surmonter l'accès limité aux exemples réels.

L'utilisation de sources de données secondaires provenant de satellites ou de radars HF peut également aider à surmonter l'accès limité aux exemples réels, même si ces sources de données ne sont pas utilisées comme données d'entrée dans un cadre

opérationnel. Il a été démontré que l'ajout de tâches auxiliaires peut améliorer l'apprentissage d'une tâche principale par l'exploitation d'informations connexes [18]. De même, des tâches auxiliaires simples peuvent également être entraînées sur des données factices générées automatiquement en augmentant la source de données primaire. Il a été démontré que l'ajout de ces tâches simples peut réduire considérablement le nombre d'exemples supervisés nécessaires [19]. Les techniques d'apprentissage non supervisé peuvent également être utilisées pour extraire des informations de données non étiquetées, et il a été démontré qu'elles fonctionnent bien pour prédire l'écoulement de la glace [20]. Une telle technique peut également être utile pour prédire la dérive d'objets dans le flux océanique.

En résumé, notre objectif est d'utiliser une variété de sources de données et de techniques d'apprentissage automatique combinées à des connaissances en physique pour prédire la dérive des objets combinées à des connaissances en physique afin de maximiser la quantité d'informations représentées dans notre modèle. Cela nous aidera à former un modèle plus représentatif des conditions du monde réel, malgré la couverture très limitée des objets réels à la dérive.

5.4 Tests préliminaires

Nous avons effectué quelques simulations préliminaires de particules dérivantes à l'aide du paquet Ocean- Parcels1, en utilisant les sorties de courants de la configuration GLAZURE64 (64° / 1 km de résolution) du moteur océanique NEMO. La figure 3 présente un exemple de trajectoires de particules moyennées dans le temps sur une période de 12 jours et montre l'émergence de structures cohérentes lagrangiennes. Les trajectoires ont été superposées à la hauteur de la surface de la mer (SSH) pour montrer la corrélation entre les deux phénomènes. Cette corrélation justifie l'utilisation des données satellitaires SSH comme entrée secondaire pour l'apprentissage de nos modèles d'apprentissage automatique.

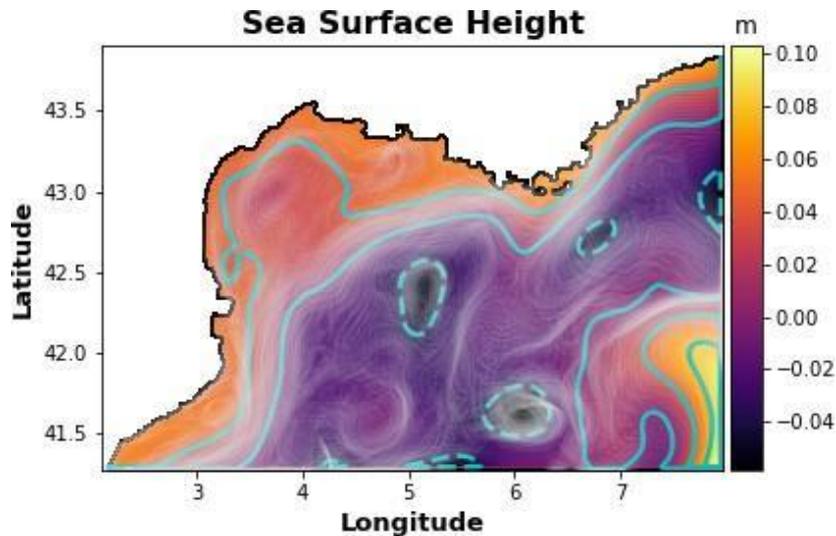


Figure 3 : Densité de 10 000 simulations de particules sur 12 jours (en blanc) superposée à la hauteur moyenne de la surface de la mer (SSH) en mètres. Les contours cyan montrent clairement la corrélation entre le changement de SSH et le flux de particules. La région représentée correspond à une région du nord-ouest de la mer Méditerranée.

La figure 4 montre un exemple d'analyse plus poussée des trajectoires sur 12 jours. Les figures 4a et 4b mettent en évidence les zones dans lesquelles les particules parcourent la plus grande distance, ce qui est donc un indicateur de la vitesse des courants. La figure 4c montre la différence entre la distance totale et la distance en ligne droite des trajectoires, ce qui met en évidence les zones de courbure maximale. L'analyse de ce type de métrique peut nous aider à comprendre le comportement de certaines zones. Du point de vue de l'apprentissage automatique, ces informations nous permettront d'expliquer pourquoi nos modèles sont peu performants pour certains sous-ensembles de nos données.

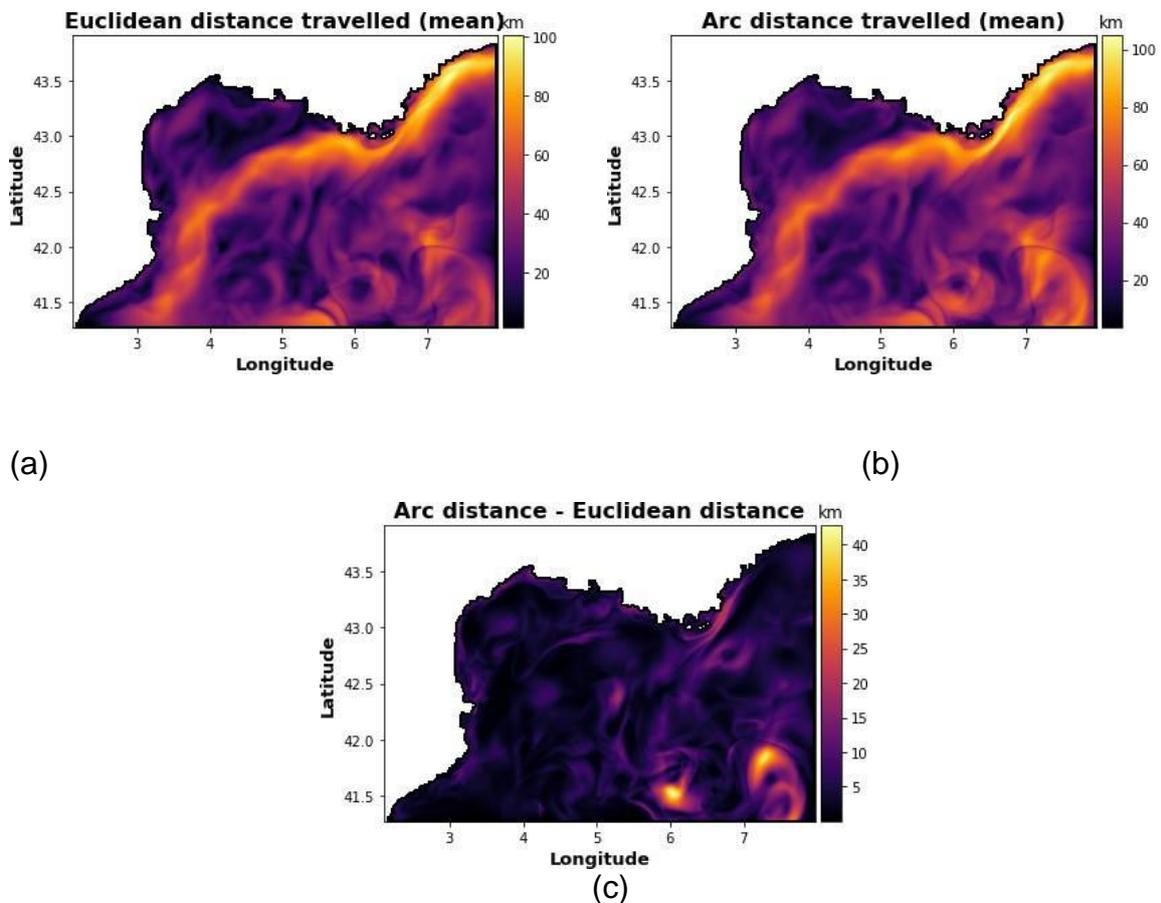


Figure 4 : Distance moyenne parcourue en kilomètres par les particules déployées sur 12 jours. a) Distance en ligne droite entre la position de départ et la position d'arrivée. b) Distance totale le long des trajectoires des particules. c) Différence entre les deux mesures de distance.

5.5 Introduction

Nous présentons un cadre piloté par les données pour apprendre la dérive lagrangienne sur un pas de temps donné (par exemple un jour dans nos expériences) en présence d'incertitude. L'incertitude survient lorsque les systèmes dynamiques ne peuvent être parfaitement décrits en raison d'une modélisation imparfaite de la dynamique ou de l'état du système. Les capacités de modélisation sont limitées par la disponibilité des ordinateurs, la connaissance des phénomènes physiques sous-jacents (en particulier à petite échelle) et la résolution des capteurs. Dans le même temps, les systèmes chaotiques font que des variations d'état mineures peuvent avoir des influences significatives et imprévisibles sur le comportement observé. Ainsi, la modélisation de la dérive avec incertitude est critique pour de nombreuses applications dont la dynamique est chaotique ou dont les entrées (par exemple le

champ de vitesse) ne résolvent pas suffisamment la dynamique nécessaire. Cela inclut des applications telles que la dynamique des océans et de l'atmosphère [21].

L'incertitude est généralement modélisée par des équations différentielles stochastiques (SDE) pour tenir compte des phénomènes physiques non capturés à de petites échelles [22]. Si les phénomènes peuvent être capturés de telle sorte que la principale source d'incertitude provient des effets du chaos, la prise en compte de l'incertitude peut être simplifiée par un échantillonnage déterministe des dérivées des particules avec des variations aléatoires des conditions initiales, par exemple [23]. Cependant, dans la pratique, les applications de prévision sont limitées par des modèles à basse résolution qui ne parviennent pas à capturer les phénomènes physiques nécessaires et doivent donc être pris en compte par exemple par le biais de SDEs.

Nous suivons une approche différente pour prendre en compte l'incertitude de la dérive qui ne repose pas sur la résolution d'équations physiques. Nous utilisons un réseau neuronal profond (DNN) pour modéliser la dérive (section 8.2), et nous proposons une représentation probabiliste de l'emplacement des particules pour représenter l'incertitude (section 8.1). En utilisant cette représentation, notre DNN inclut intrinsèquement le concept d'incertitude dans sa modélisation interne. Bien que nous démontrions notre approche en utilisant des dérivées simulées où l'incertitude est produite par échantillonnage par rapport à la position initiale des particules comme dans [23], notre cadre est plus général et peut être entraîné avec des dérivées non simulées ou différentes manières de produire l'incertitude à l'avenir.

Nos simulations sont effectuées sur des courbes de surface océaniques réalistes, à haute résolution, représentatives des états océaniques passés du monde réel dans le nord-ouest de la Méditerranée. Ainsi, nous démontrons notre cadre en apprenant un modèle de dérive d'objets flottants en mer. Comme notre cadre est entièrement piloté par les données, il permet une grande flexibilité dans ce qu'il peut prendre en entrée. Toute information représentative des courants de surface, comme les champs de vitesse ou les mesures de la hauteur de la surface de la mer (SSH), peut être utilisée. Même si ces informations ne rendent pas compte de certains phénomènes physiques, un modèle piloté par les données peut être capable de déduire les phénomènes manquants, à condition qu'ils soient pris en compte dans les exemples d'entraînement. Nous pensons qu'il s'agit là d'avantages considérables par rapport aux approches de modélisation traditionnelles basées sur les équations de la physique, qui manquent de flexibilité pour extraire des informations de différentes quantités physiques et résolutions.

Nos contributions peuvent être résumées comme suit : 1) Nous proposons une nouvelle approche de modélisation de la dérive lagrangienne avec incertitude basée sur l'apprentissage profond. Notre cadre de travail est applicable aux observations représentatives de toute source (simulée ou non) ou mesure d'incertitude. 2) Cette

approche est soutenue par une nouvelle représentation statistique de l'emplacement des particules pour la modélisation de la dérive incertaine. A notre connaissance, aucun cadre précédent n'a utilisé une localisation probabiliste dans la modélisation de la dérive lagrangienne. 3) Nous démontrons notre méthode sur un nouvel ensemble de données de dérives simulées d'objets flottants en mer caractérisées par l'incertitude des positions initiales des objets, qui sera publié en même temps que cet article.

Le reste de cet article est organisé comme suit. La section 6 passe en revue les travaux antérieurs sur la modélisation des dérives et des trajectoires incertaines. La section 7 introduit notre jeu de données et la section 8 présente notre méthode. Les résultats expérimentaux sont discutés dans la section 9. La section 11 conclut l'article.

5.6 Travaux antérieurs

L'incertitude de la dérive lagrangienne est généralement modélisée à l'aide de trajectoires stochastiques par le biais d'EDS. La stochasticité peut être utilisée pour paramétrer la physique non résolue à des échelles inférieures à la maille, soit en formulant l'EDS comme une équation de Fokker-Planck [24], soit en adaptant une EDS à des trajectoires stochastiques simulées [25]. Pour l'application des courants de surface de la mer, les exemples de physique non résolue sont les mouvements des tourbillons, des vagues ou de la turbulence à petite échelle. Pour une revue de l'utilisation des SDEs pour rendre compte des phénomènes océaniques, voir [22]. Les trajectoires stochastiques peuvent être simulées en faisant varier aléatoirement le déplacement, la vitesse ou l'accélération d'une particule. Contrairement à une approche purement axée sur les données, les EDD peuvent ne pas être en mesure de décrire des sources d'incertitude arbitraires. De plus, ils sont limités dans leur dépendance à l'égard de facteurs physiques spécifiques.

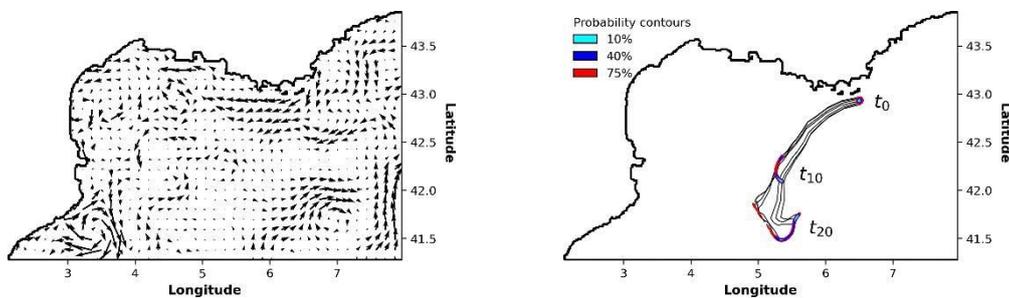


Figure 5 : Vue d'ensemble des données. Gauche) Exemple de champ de vitesse. Pour des raisons de visualisation, nous réduisons considérablement la résolution perçue du champ. Droite) Superposition de 3 instantanés de densité de probabilité à t_0 , t_{10} et t_{20} . Nous affichons les instantanés sous forme de contours de probabilité

remplis. Des exemples de trajectoires utilisées pour générer les instantanés de densité de probabilité de t_0 à t_{20} sont représentés en noir. Les travaux précédents utilisant l'apprentissage automatique pour prédire la dérive lagrangienne visent à modéliser la dérive de manière déterministe plutôt que probabiliste. [Les travaux précédents utilisant l'apprentissage automatique pour prédire la dérive lagrangienne visent à modéliser la dérive de manière déterministe plutôt que probabiliste. Au lieu d'utiliser des simulations, [13, 14] apprennent à partir d'observations passées de dériveurs en mer. [14] ont utilisé un réseau neuronal pour prédire le déplacement des dériveurs à partir du vent et de la vitesse de l'écoulement, tandis que [13] résolvent les équations physiques avec un réseau neuronal mettant en œuvre un terme supplémentaire pour apprendre les phénomènes physiques inconnus. Ce faisant, [13] a démontré sa capacité à modéliser le comportement de la dérive qui n'était pas décrit par un modèle physique de base. Cependant, ces capacités de modélisation diminuaient lorsque les conditions spatio-temporelles s'écartaient de l'ensemble d'apprentissage, indiquant un manque de généralisation à différents flux. En raison de la disponibilité limitée des observations passées, combinée à la nature passive de leur dérive dans les environnements du monde réel, la couverture des conditions qui en résulte est généralement insuffisante pour représenter de manière significative la véritable distribution. [28] considère le cas de l'apprentissage de la modélisation de la dérive d'un champ scalaire par opposition aux particules individuelles. Cependant, leur méthodologie est étroitement intégrée à l'équation d'advection des champs de concentration, ce qui est fondamentalement différent de l'advection des particules avec incertitude.

5.7 Données

Nous présentons un nouvel ensemble de données composé 1) des champs de vitesse des courants de surface de la mer (figure 5 gauche et section 7.1) et 2) des instantanés probabilistes 2D des positions des particules dérivantes (figure 5 droite et section 7.2). Nos instantanés sont produits à partir de simulations de dérive lagrangienne sur nos champs de vitesse, qui sont à leur tour produits à partir de modèles numériques réalistes à haute résolution validés par des observations réelles.

Dans cette étude, nous modélisons la dérive sur un pas de temps d'un jour, nous préparons donc nos champs de vitesse et nos instantanés pour avoir un pas de temps d'un jour. Ce pas de temps est motivé par la résolution spatiale de nos champs de vitesse et la dynamique de la région considérée dans notre étude. Nous observons une dérive notable mais suffisamment faible entre les pas de temps, ce qui est important pour fournir un critère d'apprentissage significatif.

5.7.1 Champs de vélocité des courants de surface

Nous utilisons les courants de surface du modèle de circulation générale océanique (OGCM) GLAZUR64 [29] qui est basé sur l'OGCM NEMO [17]. Le modèle a été validé avec des données d'observation réelles (courantomètres et température/hauteur de la surface de la mer) [29, 30] afin de fournir des instantanés réalistes à haute résolution des états océaniques passés au sein de la mer Méditerranée nord-ouest (lon 2-8° E, lat 41,3-43,9° N). Dans cette étude, nous considérons que la surface est bidimensionnelle en utilisant uniquement la couche supérieure. Dans la communauté de modélisation des océans, il est courant d'approximer la dérive des objets flottants en ignorant les informations sur la profondeur [31]. La résolution bidimensionnelle est de $1/64^\circ$, ce qui signifie que chaque cellule de la grille (pixel)² est représentative de 1,3 1,3 km. GLAZUR64 produit une sortie toutes les minutes, nous faisons donc la moyenne de ses sorties sur une période d'un jour afin de réaliser notre scénario d'un jour.

Les courants de surface sont composés de deux composantes de vitesse, chacune représentée par une matrice sur le domaine spatial : U (composante zonale du flux) et V (composante méridionale). NEMO utilise une grille décalée pour représenter les composantes de vitesse de sorte que U et V sont décalés d'une demi-cellule de grille vers la droite et vers le bas, respectivement. En préparation de la section 8.2 où nous fournissons les composantes de vitesse en entrée d'un CNN, nous alignons les composantes sur les centres des pixels en utilisant une interpolation linéaire. Nous remplaçons également les valeurs NaN des pixels terrestres par 0, ce qui fournit une interprétation naturelle du flux nul au CNN. Un exemple du champ de vitesse final est illustré à la figure 5 gauche.

Dans la section 7.2, nous simulons les trajectoires des particules en utilisant les champs de vitesse pour les années 2018 et 2016. Nous échantillonnons des années complètes afin de capturer de manière représentative les variances saisonnières des courants. Dans la section 9, nous utilisons les données de 2018 pour l'entraînement et les tests. Par conséquent, les champs de vitesse d'une même saison, voire d'un même jour, peuvent être utilisés à la fois pour la formation et les tests³. Les données de 2016 sont exclusivement réservées aux tests afin de fournir une évaluation plus approfondie de la généralisation de notre modèle aux vitesses et dérives non vues. La dynamique des courants de surface de la mer n'étant pas corrélée d'une année à l'autre, nous pensons que l'évaluation de 2016 sera représentative des performances des autres années.

5.7.2 Trajectoires de particules

En utilisant les champs de vitesse quotidiens des courants de surface de GLAZUR64 (section 7.1), nous simulons des trajectoires de 30 jours de particules flottantes et les

divisons en instantanés quotidiens. Nous travaillons avec des trajectoires probabilistes, ce qui signifie que pour chaque instantané, la position de la particule n'est pas représentée par un point 2D déterministe, mais plutôt par une distribution de probabilité 2D (voir Figure 5 à droite). Nous définissons la représentation de cette distribution dans la section 8.1.

Bien que la prémisse méthodologique de notre travail soit de prédire la dérive lagrangienne sur un seul pas de temps, nous générons de longues trajectoires, dont nous extrayons des instantanés, dans le but d'introduire de la variance dans le niveau d'incertitude des instantanés. Au fur et à mesure qu'une particule se déplace dans le temps, l'incertitude de sa position augmentera, car le potentiel d'emprunter différents chemins augmente (voir la figure 5, à droite).

57.2.1 Advection de particules

Nous utilisons la bibliothèque OceanParcels [32] pour advecter des particules flottantes sans masse sur nos champs de vitesse en utilisant un schéma d'intégration Runge-Kutta de 4ème ordre, où nous mettons à jour l'état des particules toutes les six heures. Le positionnement des particules étant continu, OceanParcels effectue une interpolation spatio-temporelle des champs de vitesse discrétisés. Nous advectons les particules jusqu'à 30 jours et sauvegardons leurs positions quotidiennement. Les particules n'effectuent pas toujours une trajectoire complète de 30 jours en raison des interactions aux frontières. Il existe deux types de frontières : la frontière entre l'océan et la terre et la frontière ouverte (voir Figure 5). La frontière ouverte est appelée ainsi parce qu'elle est causée par la coupure de notre couverture de données, de sorte que les valeurs océaniques voisines sont inconnues. Nous définissons deux conditions pour l'arrêt prématuré des particules : 1) une étape d'advection a fait en sorte qu'une particule s'échappe de la frontière océan-terre ou de la frontière ouverte, ou 2) une particule est entrée en contact avec une cellule océanique (pixel) à la frontière ouverte. La deuxième condition existe pour empêcher les particules de rester coincées et de s'accumuler à la frontière ouverte.

57.2.2 Trajectoires probabilistes

Comme nous l'avons vu dans la section 6, l'introduction d'un comportement aléatoire dans la modélisation de la dérive lagrangienne permet de prendre en compte l'incertitude des données ou du processus de dérive. Ainsi, nous générons des

trajectoires dont les dérives sont définies de manière probabiliste par une source d'aléa choisie. Nos trajectoires probabilistes représentent la position d'une particule comme une distribution de probabilité plutôt que comme un point unique. Nous approchons cette distribution en advectant des particules NP, où nous décidons empiriquement que $NP = 10\ 000$ est suffisant pour approcher la distribution. Pour démontrer notre cadre, nous perturbons aléatoirement la position initiale de chaque particule dans un rayon de 5 km. Ce choix de l'aléatoire est motivé par le fait qu'il est simpliste et efficace, car il permet au processus d'advection de rester complètement déterministe. En pratique, notre méthodologie pourrait être appliquée à n'importe quelle source d'aléa, telle que les perturbations de la vitesse ou de la position d'une particule à la dérive.

57.2.3 Instantanés de trajectoires probabilistes

Pour préparer l'entraînement de notre CNN à prédire la dérive lagrangienne sur un pas de temps donné (un jour dans nos expériences), nous divisons les trajectoires probabilistes en instantanés. Chaque instantané représente une distribution de probabilité de la position d'une particule dans l'espace. Comme nous nous approchons de cette distribution en advectant les particules NP, nos instantanés sont initialement représentés comme un groupe de particules, avant d'être convertis en cartes de densité de probabilité dans la section 8.1. Comme nous l'avons mentionné précédemment, l'advection des particules peut être empêchée aux frontières, de sorte que la somme de la distribution d'un instantané peut être inférieure à un parce que le nombre de particules dans un instantané est inférieur à NP.

5.7.3 Déploiement des trajectoires probabilistes

Pour l'année 2018, nous déployons NT trajectoires probabilistes de 30 jours dont les positions initiales sont échantillonnées aléatoirement sur le domaine spatio-temporel de l'océan. Nous choisissons $NT = 10\ 000$ pour encourager une large couverture spatio-temporelle de nos trajectoires simulées. Pour s'assurer qu'une trajectoire complète de 30 jours puisse être réalisée, nous avons fixé la dernière date d'échantillonnage au 1er décembre. Pour évaluer notre modèle sur une année différente, nous déployons $0,15NT$

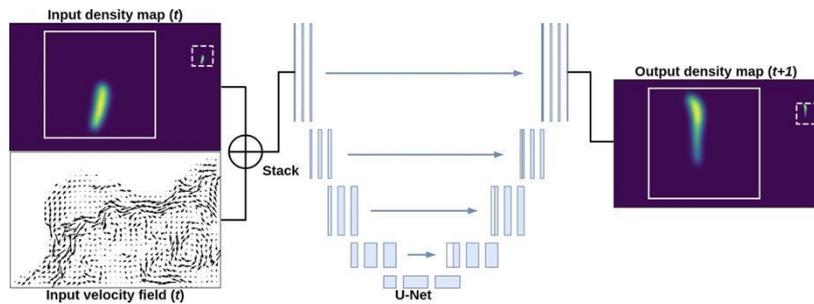


Figure 6 : Vue d'ensemble de la méthodologie. La carte de densité d'entrée $DM \times N$ est empilée avec le champ de vitesse correspondant $U \times M \times N$ pour obtenir la matrice d'entrée $X \times M \times N$, qui est la suivante t introduite dans une architecture U-Net pour produire l'instantané successif de la carte de densité $DM \times N$. Nous montrons des superpositions zoomées (carré blanc plein) de la région de premier plan des instantanés de carte de densité (carré blanc pointillé) instantanés de la carte de densité (carré blanc pointillé). Notez que le littoral n'est pas visible dans les cartes de densité en raison du paramétrage des valeurs terrestres pour avoir des valeurs de densité de probabilité de 0. trajectoires pour l'année 2016 avec les mêmes caractéristiques d'échantillonnage que pour l'année 2018.

5.8 Méthodologie

Nous formons un CNN (section 8.2) pour prédire l'évolution sur une journée de la dérive des particules avec incertitude. Il prend en entrée un champ de vitesse (section 7.1) et une carte de densité de probabilité (section 8.1) de l'emplacement des particules. Une vue d'ensemble de ce processus est présentée à la figure 6. Nous entraînons le CNN en régressant les valeurs de densité de probabilité par pixel entre les cartes de densité réelles et prédites à $t+1$ (un jour). Nous proposons et comparons trois fonctions de perte différentes dans les sections 8.2 et 9.1.

5.8.1 Cartes de densité de probabilité pour la représentation de l'incertitude

Dans la section 7.2, nous approximations la distribution de la dérive incertaine des particules en advectant un grand nombre de particules. Ici, nous choisissons de représenter la distribution de probabilité 2D de la localisation des particules sous la

forme de cartes de densité de probabilité. Ce choix est motivé par la possibilité d'homogénéiser les représentations entre la distribution de probabilité et le champ de vitesse, qui sont tous deux donnés en entrée d'un CNN dans la section 8.2. Pour créer nos cartes de densité, nous produisons une première approximation de la distribution de probabilité en calculant un histogramme 2D des emplacements des N_p particules de l'instantané (voir section 7.2) par rapport à la grille de nos champs de vitesse. Pour tenir compte du fait que nous ne disposons que d'un nombre fini de particules, nous produisons une approximation plus plausible de la distribution en appliquant un filtre gaussien ($\sigma = 1$) à l'histogramme afin d'atténuer ses gradients spatiaux anormalement nets. Nous notons que la somme de nos cartes de densité peut ne pas toujours être égale à un en raison de la possibilité que des particules s'échappent de la région, comme expliqué dans la section 7.2.1.

5.8.2 Réseau de neurones

58.2.1 Architecture

Notre CNN prend en entrée une matrice à 3 canaux constituée d'un champ de vitesse à 2 canaux (U et V) (Section 7.1) empilé avec une carte de densité de probabilité à 1 canal (Section 8.1) de la localisation des particules. Il régresse la représentation de la carte de densité de probabilité définie dans la section 8.1. Pour démontrer notre méthodologie, nous utilisons l'architecture populaire U-Net [33]. L'architecture codeur-décodeur de U-Net est bien adaptée à la régression pixel par pixel grâce à ses couches de déconvolution qui renvoient les codages de caractéristiques de faible dimension dans l'espace de l'image originale. Les connexions de saut sont également utilisées pour préparer des représentations de plus haute dimension de l'information spatiale à différentes échelles. En pratique, d'autres architectures, par exemple conçues pour la segmentation, peuvent être appliquées à notre configuration de régression pixel par pixel, par exemple [34] et [35].

58.2.2 Fonction de perte

Nous considérons trois fonctions de perte pour évaluer la qualité de la carte de densité \hat{D} with par rapport à la carte de densité de référence D_{t+1} . L_{position} minimise

l'EQM entre les cartes de densité de probabilité de position prédite et de vérité du sol (Eq. (1)). L_{drift} évalue l'EQM sur la dérive entre t et $t + 1$ (Eq. (2)). $L_{\text{threshold}}$ est similaire à L_{position} mais évalue l'EQM sur les pixels de premier plan (i.e.

non-zero) of \hat{D} and D_{t+1} (Eq. (3)).

As we know that particles can only exist within

the set of ocean pixels O , we ignore any land pixels in the output, including when computing the loss.

$$L_{position} = \frac{1}{|O|} \sum_{x \in O} (D_{t+1} - \hat{D}_x)^2 \quad (1)$$

$$L_{drift} = \frac{1}{|O|} \sum_{x \in O} (R - \hat{R}_x)^2 \quad (2)$$

$$L_{threshold} = \frac{1}{|O_T|} \sum_{x \in O_T} (D_{t+1} - \hat{D}_x)^2 \quad (3)$$

where $R = D_{t+1} - D_t$, $\hat{R} = \hat{D} - \hat{D}^t$, and $O_T = \{x \in O \mid (D_{t+1} > 0) \wedge (\hat{D}_x > 0)\}$.

Comme nous cherchons à apprendre le changement subi par une carte de densité à cause d'une dérive, la perte de dérive L_{drift} fournit un critère d'apprentissage qui correspond directement à ces objectifs. Cela permet de normaliser les critères d'apprentissage pour différents niveaux de gradients de densité, car $L_{position}$ peut encourager le modèle à apprendre la fonction d'identité en réponse à de petits gradients. Pendant l'inférence, nous ajoutons la prédiction de la carte de dérive \hat{R} à l'entrée D_t pour récupérer la carte de densité \hat{D} .

La nature éparses de nos cartes de densité motive l'utilisation de la perte seuillée $L_{threshold}$ pour concentrer l'apprentissage sur les rares informations de premier plan par opposition à l'arrière-plan abondant. En utilisant un seuil dur de 0 pour identifier le premier plan, le modèle est encouragé à discriminer implicitement entre l'arrière-plan et le premier plan en prédisant des valeurs négatives pour l'arrière-plan. Pendant l'inférence, nous coupons les valeurs négatives à zéro⁴. Grâce à l'opération de

lissage effectuée sur nos cartes de densité (voir section 8.1), la transition entre le premier plan et l'arrière-plan se fait en douceur.

	Identity	$L_{position}$	L_{drift}	$L_{threshold}$
2018	$5 \cdot 10^{-3}$	2.76 (1.30)	2.52 (1.17)	5.57 (1.06)
2016	$5 \cdot 10^{-3}$	4.38 (0.16)	4.47 (0.13)	5.67 (1.10)
# of epochs	–	43.3 (20.1)	37.3 (13.3)	11.3 (2.38)

Tableau 1 : évaluation de l'EQM des fonctions de perte (à la magnitude 10^{-8}), et nombre d'époques d'apprentissage, en tant que 'moyenne (std)' sur 4 exécutions. Pour référence, nous montrons l'EQM de D_t et D_{t+1} (identité).

Arrière-plan. Par conséquent, le seuillage n'introduit pas de gradients importants dans les cartes de densité.

5.9 Expériences

Pour chaque expérience, nous entraînons quatre modèles avec des graines différentes (0-3) et présentons les résultats sous forme de "moyenne (std)". Nous utilisons l'optimiseur Adam [36] avec des bêtas (0,9, 0,999), sans décroissance de poids, et un taux d'apprentissage de 0,0001. Nous diminuons le taux d'apprentissage d'un facteur 10 lorsque la perte de validation ne s'est pas améliorée pendant 3 époques. Nous arrêtons la formation avant la deuxième décroissance du taux d'apprentissage. Nous utilisons une taille de lot de 16. Avant d'extraire des instantanés de nos trajectoires probabilistes, nous divisons aléatoirement les trajectoires en ensembles de formation, de validation et de test avec un ratio de 70/15/15.

5.9.1 Comparaison des fonctions de perte

Les résultats pour les trois fonctions de perte décrites dans la section 8 sont donnés dans le tableau 1. En comparant L_{drift} à $L_{position}$, nous constatons peu de différence dans la performance moyenne, bien qu'une diminution de 14% du temps de

convergence moyen soit observée pour Ldrift. Cette formation plus rapide est conforme aux résultats de travaux antérieurs (par exemple [37]) selon lesquels l'apprentissage des résidus améliore la convergence. Lthreshold est moins performant et arrête l'apprentissage beaucoup plus tôt, ce qui indique que la fonction de perte a du mal à converger. Dans des travaux futurs, ce problème pourrait être résolu en ajustant le schéma d'apprentissage. En raison de ses bonnes performances et de sa convergence plus rapide, nous recommandons Ldrift comme perte de choix dans les travaux futurs. Pour des raisons de simplicité, et puisque Lposition obtient des résultats tout aussi bons, nous l'utilisons pour les prochaines expériences de ce document.

5.9.2 Généralisation à différents dynamiques

Nous évaluons la capacité de nos modèles (formés sur les données de 2018) à se généraliser à différentes situations dynamiques en les testant sur des données complètement inédites d'une autre année (2016). Comme nous utilisons un nombre limité de champs de vitesse (365) pour générer un grand nombre d'instantanés (300k), les champs de vitesse d'entrée de 2018 sont les mêmes dans les ensembles de données d'entraînement, de validation et de test. Par conséquent, les résultats présentés pour 2016 sont un meilleur indicateur des performances de notre modèle, car il n'y a pas de biais lié à l'utilisation des mêmes champs de vitesse entre les ensembles de formation et de test. Les courants de surface ne présentent pas de tendances annuelles, nous nous attendons donc à ce que les performances sur 2016 soient représentatives des performances sur d'autres années inédites.

En comparant les résultats entre les différentes années, nous observons que les performances de Lthreshold restent constamment inférieures à celles des autres. Lposition et Ldrift sont en moyenne moins performants pour l'ensemble de données de 2016, bien que l'écart soit beaucoup plus faible. Cependant, dans le tableau 1 (Identité), nous vérifions que leurs dérives prédites ont toujours une plus grande valeur ajoutée que la simple approximation de \hat{D} comme D_t . En outre, après inspection (voir la figure 7), ces moins bons résultats pour 2016 proviennent en partie d'une plus grande dispersion des distributions de probabilité, signe d'une plus grande incertitude. Cependant, les dérives prédites tendent à aller dans la bonne direction, et sont visuellement proches de celles de la vérité terrain (même si les erreurs sont plus nombreuses que pour 2018). Cela ouvre des perspectives prometteuses de généralisation à différents lieux et moments.

Ces résultats suggèrent que les champs de vitesse sont bien pris en compte dans la détermination de la dérive. Ceci est vérifié par une petite expérience dans la figure

8 où le champ de vitesse d'un échantillon d'entraînement aléatoire de 2018 a été inversé. La dérive s'est produite de manière cohérente vers la direction supérieure gauche, au lieu de la direction initiale vers le bas et la droite. Ainsi, le champ de vitesse d'entrée joue un rôle essentiel dans la prédiction de la dérive. Néanmoins, l'incertitude plus élevée pour l'année d'essai 2016 indique qu'un modèle interne de dérive pour l'emplacement spatio-temporel peut également jouer un rôle (conflictuel) dans les résultats. Les travaux futurs devront aborder l'équilibre entre ces deux influences, afin d'améliorer la généralisation à d'autres situations dynamiques.

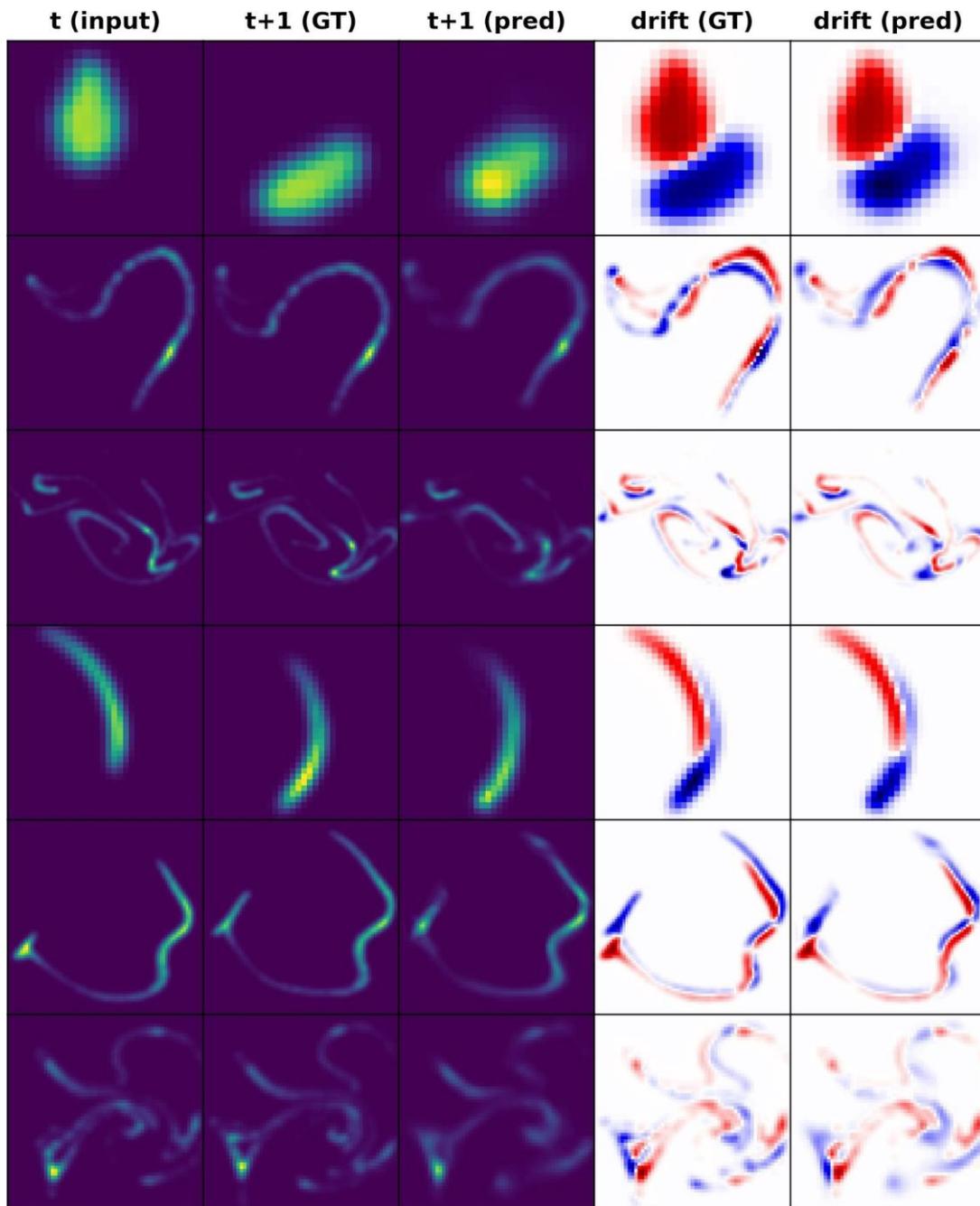


Figure 7 : Exemples de résultats pour Lposition pour différents niveaux d'incertitude et différentes années. De gauche à droite : D_t , D_{t+1} , \hat{D} , dérive de la vérité du sol R , et dérive prédite \hat{R} . Trois rangées supérieures : échantillons de 2016, trois rangées inférieures : échantillons de 2018. Les tracés sont zoomés sur les régions de premier plan et couvrent donc différentes échelles. La direction de la dérive dans les colonnes 4-5 est indiquée du rouge au bleu.

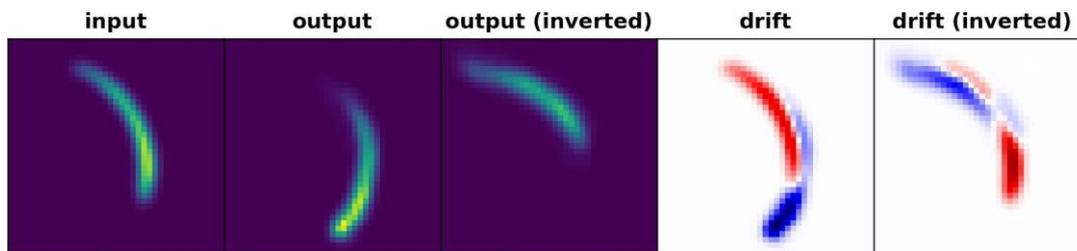


Figure 8 : Résultat de l'inversion du champ de vitesse. De gauche à droite : D_t , D^* avec les vitesses d'origine et inversées, dérive prédite \hat{R} avec les vitesses d'origine et inversées. La position L est utilisée. Notez que les prédictions reflètent le changement de direction du champ de vitesse.

	Low-res velocity	SSH
2018	3.45 (1.09)	2.69 (0.79)
2016	4.33 (0.16)	7.27 (0.37)
# of epochs	34.5 (17.2)	46.8 (8.53)

Tableau 2 : évaluation de l'EQM (à la magnitude 10^{-8}) de différentes représentations des courants de surface en utilisant Lposition, et le nombre d'époques d'apprentissage, comme 'moyenne (std)' sur 4 passages.

5.9.3 Application à différentes représentations des courbes de surface

Pour démontrer la capacité de notre cadre à s'adapter à différentes représentations des courants de surface, nous présentons les résultats de l'entraînement avec des champs de vitesse de résolution réduite (filtre gaussien de $\sigma = 1$) et des cartes SSH (obtenues à partir de GLAZUR64 en plus de nos champs de vitesse) en utilisant Lposition. Les champs de vitesse utilisés pour la prévision des applications en mer ont une résolution beaucoup plus faible par rapport aux modèles de recherche tels que celui utilisé dans cette étude (voir section 7.1). De son côté, la SSH est représentative des courants de surface de la mer à grande échelle mais ne permet pas de saisir les phénomènes à petite échelle. Cependant, sa disponibilité à l'échelle mondiale grâce aux observations par satellite en fait une source d'information très utile pour la modélisation de l'océan.

Comme l'illustre la figure 9, nous observons peu de différence entre l'utilisation de notre champ de vitesse à haute résolution et les deux autres représentations des courants de surface. Dans le tableau 2, en comparant avec le tableau 1, nous observons que l'erreur MSE n'augmente que marginalement pour 2018 lors de l'utilisation des deux représentations à basse résolution. Cela indique que le CNN est

capable d'inférer les phénomènes physiques manquants qui ne sont pas présents dans les données d'entrée, peut-être en puisant dans son modèle interne de dérive dont nous avons parlé à la fin de la section précédente.

Pour 2016, les champs de vitesse à basse résolution obtiennent également des résultats cohérents, tandis que le CNN semble avoir plus de mal avec les SSH, avec une erreur MSE passant de 4,38(0,16) à 7,27(0,37). Cela peut être interprété comme l'inférence de phénomènes manquants de SSH étant plus dépendante de l'emplacement et du temps donnés. Par conséquent, bien que les résultats semblent toujours visuellement plausibles, les futurs utilisateurs devront peut-être garder à l'esprit que la généralisation à de nouvelles situations dynamiques est plus difficile pour les sciences humaines.

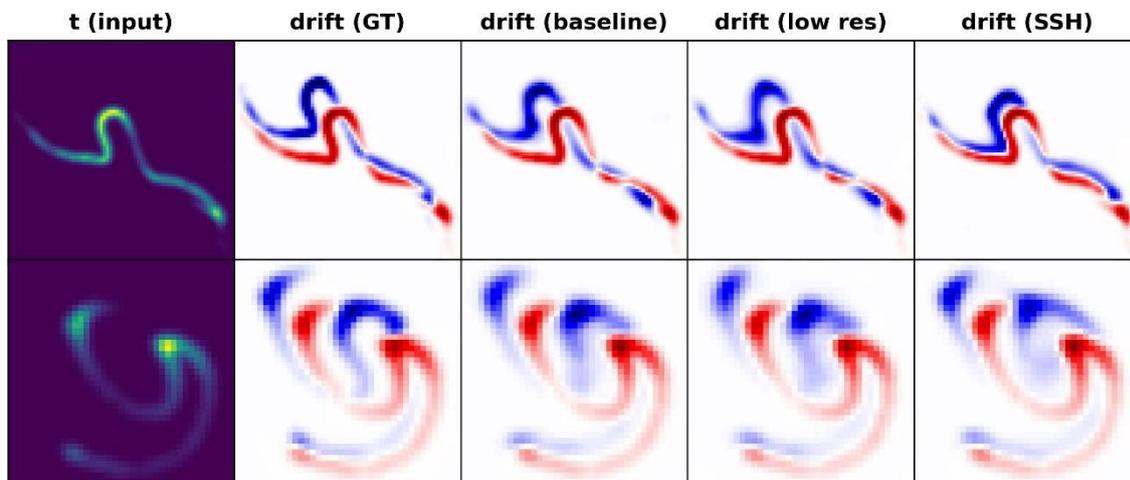


Figure 9 : Exemples de résultats pour Lposition entraînée sur différentes représentations des courants de surface. courants de surface. De gauche à droite : Dt , R , \hat{R} (champs de vitesse originaux), \hat{R} (champs de vitesse à faible rés. de vitesse), et \hat{R} (SSH). Rangée du haut : échantillon aléatoire de 2016, rangée du bas : échantillon aléatoire de 2018.

5.10 Discussion sur les trajectoires probabilistes

5.10.1 Caractéristiques

Nous montrons des exemples d'animations de trajectoires probabilistes complètes à http://sabiod.lis-lab.fr/pub/gias/probabilistic_trajectories. Comme nos trajectoires probabilistes sont simulées à l'aide de courants de surface modélisés à partir d'états océaniques passés réels, les modèles que nous formons sont capables d'être déployés et utilisés pour des applications dans notre région donnée (nord-ouest de la

Méditerranée). Par exemple, nous observons les phénomènes suivants qui sont typiques de la dynamique de la région :

- 0.gif - Nous voyons que les particules sont contenues dans un tourbillon relativement statique, correspondant à un mouvement en spirale avec peu de translation spatiale.
- 1.gif - La position initiale correspond à un emplacement répulsif, ce qui entraîne deux séparations distinctes du flux. Nous voyons cet effet dans nos trajectoires probabilistes où le chemin est incertain entre deux directions.
- 4.gif - La position initiale se trouve dans le courant du nord, qui est le courant principal et le plus fort de la région. Nous voyons les particules s'écouler le long du courant d'est en ouest à un rythme rapide.
- 5.gif - Les particules sont piégées dans un tourbillon (en haut à droite), ce qui fait que la distribution spatiale correspondante de l'emplacement des particules reste statique dans le temps.
- 9.gif - Nous voyons la distribution être attirée par le courant du nord, qui la transporte ensuite vers la région du sud-ouest.

5.10.2 Prévisions préliminaires

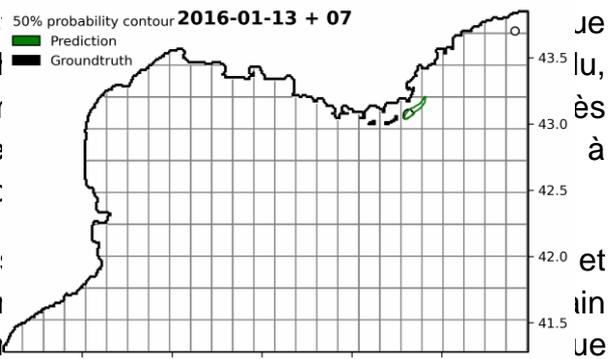
En utilisant notre modèle de prédiction du lendemain, nous montrons des résultats très préliminaires de prévisions d'une semaine visualisées avec des contours de probabilité de 50% à l'adresse :

http://sabiody.lis-lab.fr/pub/gias/probabilistic_trajectories/preliminary_forecasts.

Étant donné la carte de densité de probabilité à t_0 , nous obtenons des prédictions de séquence en alimentant récursivement les prédictions du modèle comme entrée pour le pas de temps suivant. Ce scénario est beaucoup plus difficile car toute erreur dans les prédictions s'accumule au fil du temps. Malgré cela, nos résultats préliminaires sont prometteurs si l'on considère que notre modèle n'a pas été entraîné à corriger les erreurs de prédiction des séquences.

Comme prévu, nous observons que la distribution prédite des trajectoires est souvent mal alignée avec la vérité du terrain en raison de l'accumulation d'erreurs. Le tableau 4.gif en est l'exemple le plus parlant, où les contours de 50% prédits et de la vérité du terrain sont complètement séparés au jour 7. Nous pouvons également voir, par exemple dans 0.gif et 2.gif, que le contour prédit peut être plus grand que la vérité terrain, indiquant ainsi un plus grand niveau d'incertitude dans les prédictions. La figure 3.gif montre un cas intéressant de division de la distribution de la vérité du terrain en deux, où l'occurrence de la division est modélisée de manière correcte,

mais pas l'alignement de la deuxième dis-
 les prédictions sont raisonnablement col-
 même après des prévisions sur 7 jour
 prometteurs et montrent qu'il est possible
 condition de l'entraîner de manière approx



La figure 10 montre la métrique IOU (inter-
 confirme le fait que les prévisions sont gé-
 degré d'alignement. Il est intéressant de

l'IOU diminue de façon monotone au fil du temps, mais nous observons au contraire
 des pics (plus particulièrement pour 0.gif au pas de temps 4). Cela est probablement
 dû aux effets de l'incertitude : à mesure que la distribution s'étend, les contours de
 probabilité atteignent un point où leur concentration diminue. Ainsi, lors de la
 prévision, un compromis apparaît entre la réduction de la concentration (proportion
 correcte plus grande) et l'accumulation d'erreurs (proportion correcte plus petite).

5.11 Conclusions

Nous avons abordé le problème de la modélisation de la dérive lagrangienne sous l'influence de l'incertitude en tirant parti de la flexibilité des CNN. Nous avons démontré notre cadre en considérant une application des courants de surface de la mer dans laquelle nous avons modélisé la dérive incertaine d'objets flottants. Nous avons généré des données d'entraînement en simulant un grand nombre de trajectoires afin d'approximer l'évolution temporelle de la distribution de probabilité d'une particule à la dérive dans un espace 2D. Nos simulations ont été effectuées sur des représentations de champs de vitesse des courants de surface produits par un modèle océanique réaliste à haute résolution. Nous avons constaté que les dérives prédites par notre modèle entraîné étaient en bon accord avec les simulations, et nous avons trouvé que notre modèle se généralisait raisonnablement bien aux différentes situations dynamiques d'une année différente. Nous avons également évalué la capacité de notre cadre à compenser les phénomènes physiques manquants dans les données d'entrée en effectuant des tests sur les courants de surface d'un champ de vitesse à faible résolution et des cartes SSH. Sur ce scénario plus difficile, le CNN a été capable d'inférer certaines des informations manquantes à partir de sa modélisation interne de la dérive, et même d'appliquer ce modèle à différentes situations dynamiques avec un succès raisonnable. Dans des travaux futurs, notre cadre peut être étendu pour modéliser des trajectoires de dérive complètes plutôt que des pas de temps uniques.

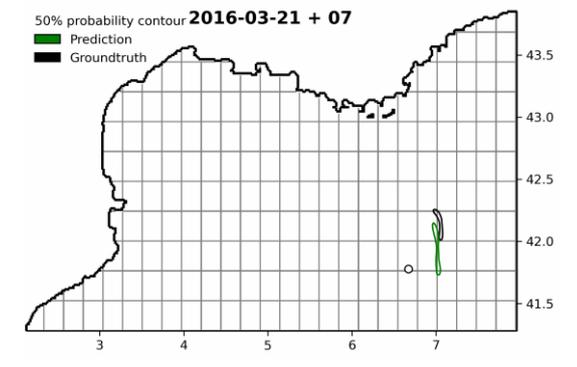
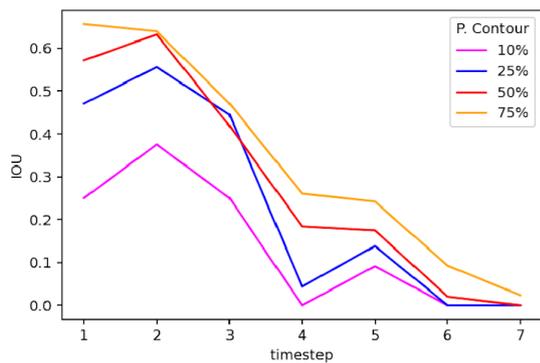
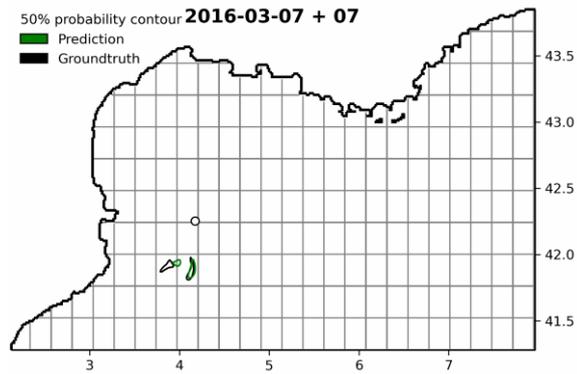
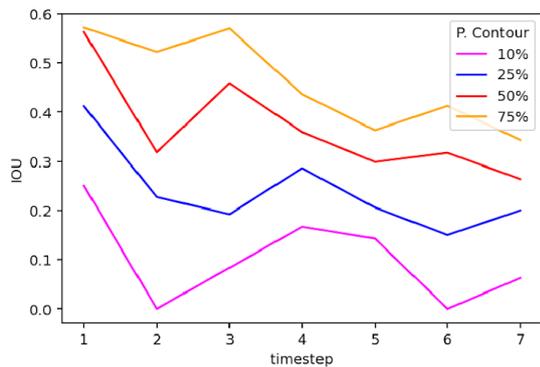
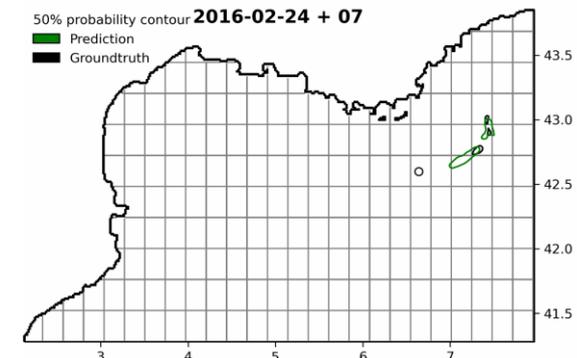
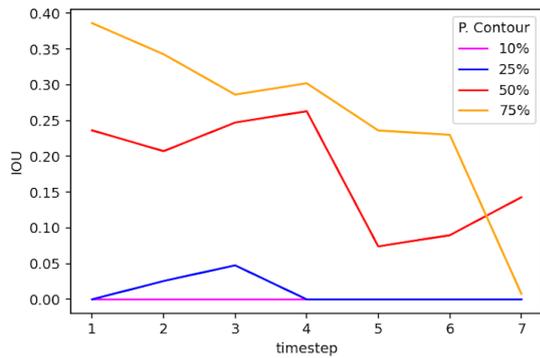
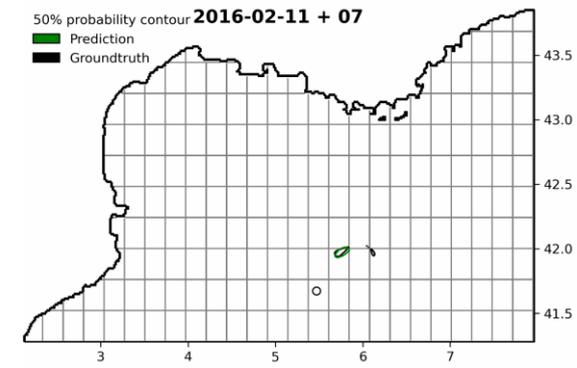
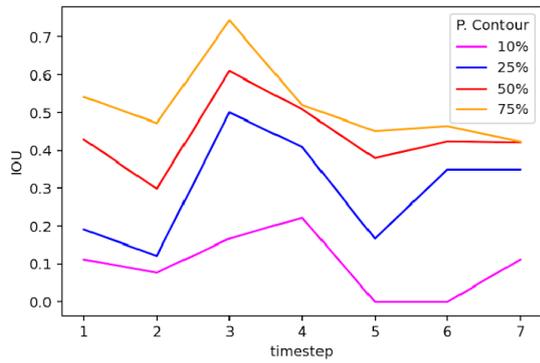
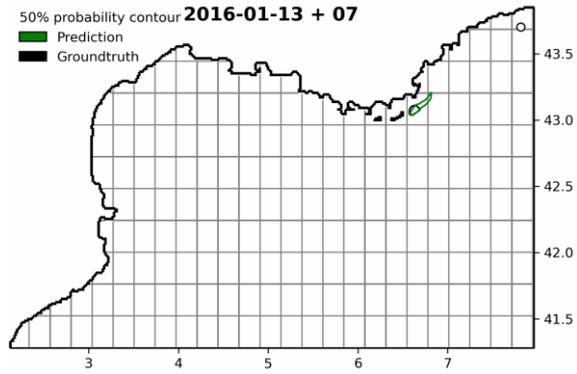
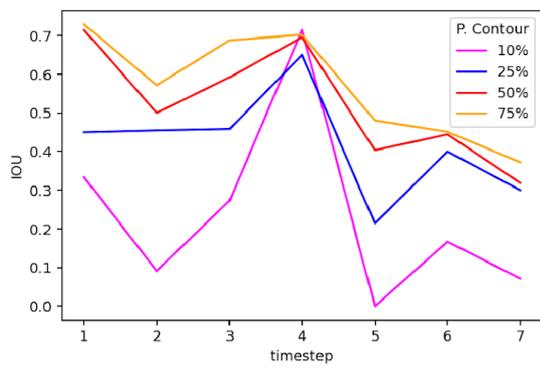


Figure 10: Left) IOU of groundtruth and predicted probability contours. Right) Final day (7) of forecast. Columns: 0–4.gif.



5.12 References

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 1. MIT press Cambridge, 2016.
- [2] Ø. Breivik and A. A. Allen, “An operational search and rescue model for the norwegian sea and the north sea,” *Journal of Marine Systems*, vol. 69, no. 1-2, pp. 99–113, 2008.
- [3] J. M. Sayol, A. Orfila, G. Simarro, D. Conti, L. Renault, and A. Molcard, “A Lagrangian model for tracking surface spills and SaR operations in the ocean,” *Environmental modelling & software*, vol. 52, pp. 74–82, 2014.
- [4] S.-z. Wang, H.-b. Nie, and C.-j. Shi, “A drifting trajectory prediction model based on object shape and stochastic motion features,” *Journal of Hydrodynamics, Ser. B*, vol. 26, no. 6, pp. 951–959, 2015.
- [5] J. Zhang, A. P. Teixeira, C. G. Soares, and X. Yan, “Probabilistic modelling of the drifting trajectory of an object under the effect of wind and current for maritime search and rescue,” *Ocean Engineering*, vol. 129, pp. 253–264, 2017.
- [6] P. Sebastião and C. G. Soares, “Uncertainty in predictions of oil spill trajectories in open sea,” *Ocean Engineering*, vol. 34, no. 3-4, pp. 576–584, 2007.
- [7] A. J. Abascal, S. Castanedo, F. J. Mendez, R. Medina, and I. J. Losada, “Calibration of a Lagrangian transport model using drifting buoys deployed during the Prestige oil spill,” *Journal of Coastal Research*, vol. 25, no. 1 (251), pp. 80–90, 2009.
- [8] P. Daniel, G. Jan, F. Cabioc’h, Y. Landau, and E. Loiseau, “Drift modeling of cargo containers,” *Spill Science & Technology Bulletin*, vol. 7, no. 5-6, pp. 279–288, 2002.
- [9] H. Peltier, W. Dabin, P. Daniel, O. Van Canneyt, G. Dorémus, M. Huon, and V. Ridoux, “The significance of stranding data as indicators of cetacean populations at sea: modelling the drift of cetacean carcasses,” *Ecological Indicators*, vol. 18, pp. 278–290, 2012.
- [10] B. S. Santos, D. M. Kaplan, M. A. Friedrichs, S. G. Barco, K. L. Mansfield, and J. P. Manning, “Consequences of drift and carcass decomposition for estimating sea turtle mortality hotspots,” *Ecological Indicators*, vol. 84, pp. 319–336, 2018.
- [11] M. Rixen and E. Ferreira-Coelho, “Operational surface drift prediction using linear and non-linear hyper-ensemble statistics on atmospheric and ocean models,” *Journal of Marine Systems*, vol. 65, no. 1-4, pp. 105–121, 2007.
- [12] R. Minguez, A. Abascal, S. Castanedo, and R. Medina, “Stochastic Lagrangian trajectory model for drifting objects in the ocean,” *Stochastic environmental research and risk assessment*, vol. 26, no. 8, pp. 1081–1093, 2012.
- [13] N. O. Aksamit, T. Sapsis, and G. Haller, “Machine-learning mesoscale and sub-mesoscale surface dynamics from lagrangian ocean drifter trajectories,” *Journal of Physical Oceanography*, vol. 50, no. 5, pp. 1179–1196, 2020.





- [14] Y.-W. Nam, H.-Y. Cho, D.-Y. Kim, S.-H. Moon, and Y.-H. Kim, “An improvement on estimated drifter tracking through machine learning and evolutionary search,” *Applied Sciences*, vol. 10, no. 22, p. 8123, 2020.
- [15] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- [16] M. Raissi, A. Yazdani, and G. E. Karniadakis, “Hidden fluid mechanics: A navier-stokes informed deep learning framework for assimilating flow visualization data,” *arXiv preprint arXiv:1808.04327*, 2018.
- [17] G. Madec, R. Bourdall’e-Badie, P.-A. Bouttier, C. Bricaud, D. Bruciaferri, D. Calvert, J. Chanut, E. Clementi, A. Coward, D. Delrosso, *et al.*, “Nemo ocean engine,” 2017.
- [18] S. Ruder, “An overview of multi-task learning in deep neural networks,” *arXiv preprint arXiv:1706.05098*, 2017.
- [19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.
- [20] Y. Min, S. K. Mukkavilli, and Y. Bengio, “Predicting ice flow using machine learning,” *arXiv preprint arXiv:1910.08922*, 2019.
- [21] A. E. Gill, *Atmosphere-ocean dynamics*, vol. 30. Academic press, 1982.
- [22] E. Van Sebille, S. M. Griffies, R. Abernathey, T. P. Adams, P. Berloff, A. Bias-toch, B. Blanke, E. P. Chassignet, Y. Cheng, C. J. Cotter, *et al.*, “Lagrangian ocean analysis: Fundamentals and practices,” *Ocean Modelling*, vol. 121, pp. 49–75, 2018.
- [23] I. M. Koszalka, T. W. Haine, and M. G. Magaldi, “Fates and travel times of denmark strait overflow water in the irminger basin,” *Journal of Physical Oceanography*, vol. 43, no. 12, pp. 2611–2628, 2013.
- [24] A. W. Visser, “Lagrangian modelling of plankton motion: From deceptively simple random walks to fokker–planck and back again,” *Journal of Marine Systems*, vol. 70, no. 3-4, pp. 287–299, 2008.
- [25] J. LaCasce, “Statistics from lagrangian observations,” *Progress in Oceanography*, vol. 77, no. 1, pp. 1–29, 2008.
- [26] M. Han, S. Sane, and C. R. Johnson, “Exploratory lagrangian-based particle tracing using deep learning,” *arXiv preprint arXiv:2110.08338*, 2021.
- [27] M. D. Grossi, M. Kubat, and T. M. Özgökmen, “Predicting particle trajectories in oceanic flows using artificial neural networks,” *Ocean Modelling*, vol. 156, p. 101707, 2020.





- [28] J. Zhuang, D. Kochkov, Y. Bar-Sinai, M. P. Brenner, and S. Hoyer, “Learned discretizations for passive scalar advection in a two-dimensional turbulent flow,” *Physical Review Fluids*, vol. 6, no. 6, p. 064605, 2021.
- [29] Y. Ourmières, B. Zakardjian, K. Béranger, and C. Langlais, “Assessment of a nemo-based downscaling experiment for the north-western mediterranean region: Impacts on the northern current and comparison with adcp data and altimetry products,” *Ocean Modelling*, vol. 39, no. 3-4, pp. 386–404, 2011.
- [30] K. Guihou, J. Marmain, Y. Ourmieres, A. Molcard, B. Zakardjian, and P. Forget, “A case study of the mesoscale dynamics in the north-western mediterranean sea: a combined data–model approach,” *Ocean Dynamics*, vol. 63, no. 7, pp. 793–808, 2013.
- [31] J. Mansui, A. Molcard, and Y. Ourmières, “Modelling the transport and accumulation of floating marine debris in the mediterranean basin,” *Marine pollution bulletin*, vol. 91, no. 1, pp. 249–257, 2015.
- [32] P. Delandmeter and E. v. Sebille, “The parcels v2.0 lagrangian framework: new field interpolation schemes,” *Geoscientific Model Development*, vol. 12, no. 8, pp. 3571–3584, 2019.
- [33] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [34] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- [35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [36] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015.