Learning-based forecasting of metocean variables : a path to maintenance operations optimization for offshore wind energy

Robin Marcille

COLLEGES BRETAGNE LOIRE
SCIENCES POUR L'INGENIEUR ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE
MINES-TÉLÉCOM ATLANTIQUE BRETAGNE
PAYS DE LA LOIRE – IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 648
Sciences pour l'Ingénieur et le Numérique
Spécialité : *Mathématiques et Sciences et Technologies de l'Information et de la Communication*

Par
Robin MARCILLE

Learning-based forecasting of metocean variables

A path to maintenance operations optimization for offshore wind energy

Thèse présentée et soutenue à IMT Atlantique, Plouzané, France, le 13 Septembre 2024
Unité de recherche : Lab-STICC, UMR 6285
Thèse N° : 2024IMTA0416

Rapporteur•ice•s avant soutenance :

Jethro BROWELL Professeur, Université de Glasgow
Valérie MONBET Professeure, Université de Rennes 1

Composition du jury :

Président :	Georges KARINIOTAKIS	Professeur, Mines Paris Tech
Rapporteur•ice•s :	Jethro BROWELL	Professeur, Université de Glasgow
	Valérie MONBET	Professeure, Université de Rennes 1
Examinateur•ice•s :	Laure RAYNAUD	Chercheuse, Météo-France
	François ROUSSEAU	Professeur, IMT Atlantique
	Pierre TANDEO	Maître de Conférence, IMT Atlantique
Dir. de thèse :	Ronan FABLET	Professeur, IMT Atlantique
Co-dir. de thèse :	Pierre PINSON	Professeur, Imperial College London

Invités :

Jean-François FILIPOT Directeur scientifique, France Énergies Marines
Maxime THIEBAUT Chercheur, France Énergies Marines

This work was supported by France Énergies Marines and the French government, managed by the Agence Nationale de la Recherche under the Investissements d'Avenir program, with the reference ANR-10-IEED-0006-34. This work was carried out in the framework of the FOWRCE SEA, POWSEIDOM and FLOWTOM projects.

Abstract

Offshore wind energy maintenance operations are highly sensitive to environmental conditions. To be safely executed, an ensemble of limiting parameters are to be below operability limits for the duration of the operation. This makes maintenance operations decision-making problems under forecast uncertainty. Current numerical weather and wave predictions are limited for the full estimation of forecast uncertainty due to their computational costs. Learning-based methods on the other hand are computationally inexpensive for inference, and can provide an interesting framework for multivariate probabilistic forecasting. Their low computational cost furthermore permits the integration of recent in-situ measurements for metocean forecast emulation.

The importance of offshore measurements is highlighted by the thesis results, and a learning-based method for designing optimal sensors networks is proposed. It is shown that unsupervised clustering can be used efficiently for the sparse sampling of offshore wind compared to state-of-the-art methods. This method has been used by the French weather service to define the deployment of floating LIDAR on future offshore wind energy tender areas.

This thesis explores deep learning models for the joint probabilistic forecasting of offshore wind energy operations limiting parameters. The importance of taking as input both numerical weather predictions and in-situ measurements is shown, and the interest of deep learning model for assimilating a large amount of input data is demonstrated. Different probabilistic framework are proposed for the multivariate probabilistic forecast emulation. A Gaussian posterior assumption, particularly relevant for bi-variate wind prediction, is compared to a non-parametric generative approach using normalizing flows. It is shown that the use of normalizing flows can relax any assumption on the shape of posterior distribution while maintaining sampling and likelihood computation capabilities.

A real case study dataset is built using co-located buoys and LIDAR measurements on a relevant area for offshore wind energy development. The probabilistic models are adapted for the joint wind and wave forecasting, for which the non-Gaussian properties of the normalizing flows is beneficial for forecast reliability. Several approaches based on Gaussian copulas are proposed to emulate temporal dependency between lead times.

Normalizing flows-based temporal correlation embedding are also proposed. The use of normalizing flows for scenario generation proves promising and should be further investigated. Eventually, the different models are evaluated in a new operational value framework that considers the economic impact of the decision-making based on the forecast. As an addition to the literature on the topic, a new metric that includes risk taken when operating under dangerous conditions is proposed. It is shown that the search for an economic optimum in the probabilistic decision-making leads to higher risk during operations, and this should be taken into account for forecast selection and evaluation.

Remerciements

La thèse est un voyage. On part à la voile, vent arrière d'abord, en direction du large, sous une légère brise. Les vents tournent, la mer se lève, se déchaine parfois, tant et si bien qu'il peut être dur de tenir la barre. Au cours du voyage, on prend pied sur différents continents. On découvre des contrées qu'on avait idéalisées, d'autres que l'on n'attendait pas. De ces surprises, de ces détours, il faut faire des opportunités. Être ouvert tout en gardant le cap. Le voyage est fait de moments d'euphorie, de galère, de doute, de joie, de combat. A l'heure de coucher ces mots sur mon fidèle ordinateur de bord, je finis la traversée retour. Un léger souffle au portant me pousse vers la rade de Brest, où tout a commencé. Lessivé mais serein, je regarde en arrière et souhaite remercier toutes les personnes qui m'ont accompagné durant ces 4 années de vadrouille.

Je souhaite en premier lieu remercier mes encadrants : Pierre P., Ronan, Pierre T. et Maxime pour leur soutien constant. Votre bienveillance, votre pertinence et vos idées m'ont guidé durant ces 4 années en eaux troubles, et tout ce qui est écrit sur ce manuscrit je vous le dois. Je n'aurais pas pu rêver meilleur encadrement et je mesure la chance d'avoir pu travailler dans ces conditions. Un bon capitaine de bateau dirige, rassure, partage et forme. Si je retourne à Brest à la barre de mon propre bateau, c'est parce que j'ai eu le privilège d'apprendre à naviguer sous votre direction. Du fond du cœur, merci.

Un immense merci à mon jury de thèse qui a accepté de juger mon travail. Échanger avec vous qui avez nourri ma réflexion à travers vos travaux est une très grande chance. Sincere acknowledgements to my Ph.D. jury who took the time to read and evaluate my work. Discuss my work with you, who've inspired me through your publications is a great opportunity. Thanks for your time and implication.

Merci ensuite aux personnes qui m'ont convaincu de faire cette thèse. Elle est née de discussions avec Jean-François, et a été construite grâce à la pugnacité de Florence. Merci d'avoir toujours cru en moi et de m'avoir guidé pour démarrer ce voyage. Merci à toutes les personnes avec qui j'ai pu travailler de près ou de loin durant ce voyage. Flo-

rence, Laure, Youen, Oscar, Maxime, Matteo, Arthur, Amédée, Emmanuel, Rui, Florian, Jean-François, Florent, les membres du projet FLOWTOM. Vous m'avez tous aidé d'une manière ou d'une autre à tracer un morceau de chemin.

Bien sûr lorsque l'on planifie un itinéraire tout semble simple et linéaire sur le papier. Une fois sur le terrain, la réalité nous oblige à nous adapter en permanence. Merci à toutes les collègues qui au quotidien ont nourri mes réflexions, ont tempéré mes doutes, ont conforté mes opinions. Pour votre bienveillance, votre expertise et votre amitié, merci Florence, Laure, Sophie, Georges, Vincent, Arnaud, Linta, Florent et Maxime. Merci aux collègues de France Energies Marines pour ces années enrichissantes, pour les discussions pluridisciplinaires qui m'ont toujours sorti de ma zone de confort. La grande richesse de mon voyage a été de marcher avec vous dans la même direction. Merci Laetitia, Youssra, Paul, Oscar, Alizée, Julien, Youen, Amédée, Romain, Adrien, Lydie, Emma, Karine, Rui, Fabien, François, Florian D.R., Mélusine, Ronan R., Anne-Sophie, Herveline, et je ne peux citer tout le monde ici. Un voyage comme celui-ci ne peut avoir lieu sans toutes les personnes œuvrant dans l'ombre. Merci à toi Laurène pour ton infinie patience et ta bienveillance. Merci Stevenn d'avoir toujours pris mes problèmes au sérieux aussi mineurs soient-ils. Vous avez été mon support météo et logistique au long de ce périple. Merci.

Lors de chacune des épreuves de mon voyage, j'ai pu compter sur le soutien continu de mes amis. Merci aux Bigoudingues, c'est avec vous tout a commencé, c'est vous qui avez lancé la bouteille de champagne sur la coque du bateau, et vous m'avez accompagné tout au long de ce voyage. Merci à mes compagnons de toujours, Arthur, Arthur, Gabriel, Vincent, Julie, Lucas, Hélène, Chloé, Yannis, Lucas, Pierre. Lorsqu'on est au milieu de l'océan, c'est bon de penser à la maison. C'est comme un phare qui indique la direction. Merci à mes camarades d'aventures, Manon, Lulu, Toto, Mathias, Cécile, Antoine, Scott, Alice, Antoine, Kim, Gauthier, Marthe, tous les amis qui ont accepté de me suivre dans mes sommets au bord du chemin, Jean-Côme, Charlotte, Victor, Mathilde, Thomas Rh., Sophie et Aurel, et je ne peux citer tout le monde ici. Merci aux Corsican Wild Rovers pour cette ultime respiration salvatrice, Alec, Adélie, Edgar, Capu, Scott, Toto, Lulu, Mathias, Cameçon, Manon, Antoine. Parfois, faire des sommets au bord du chemin permet de lever les yeux, d'appréhender la suite du parcours, et d'avoir une vue plus globale de la topo environnante. Merci pour ces moments de clarté. Une pensée émue pour toutes les personnes qui ont nourri ma réflexion politique au long de ce parcours, merci pour vos

oreilles attentives Julie, Morgane, Jean, Noé, Gauthier, Kim, Toto, Lulu, Manon. Merci Val et les Coco-dines pour la bande son de la fin de ma thèse. Enfin merci à tous ceux que je ne peux citer ici dans ma métaphore filée mais dont l'amitié m'apporte de la joie au quotidien Etienne, Mathias T., la coloc brestoise, Stan, Marion, Pierre et Mélanie, Thomas, la team Beluga, la team Yoann Huget de Copenhague. J'espère que vous pouvez compter sur moi comme je peux compter sur vous.

Perdu au milieu de la pampa pendant des mois, sous une météo changeante et avec peu de vivres, j'ai toujours trouvé un soutien indéfectible et une confiance aveugle auprès de ma famille. Notre liaison constante par téléphone satellite a été un fil d'Ariane, une bouée de sauvetage, une bouteille d'oxygène dans les moments les plus difficiles. J'aurais beau faire toutes les métaphores du monde je ne pourrais vous dire à quel point je vous aime, je suis fier de vous, et je suis fier que vous soyez fiers de moi aujourd'hui. Vous avoir est ma plus grande chance et ma plus grande force. Je laisse Ben Mazué et Jacques Higelin décrire ce que je ressens.

A mes sœurs, «*C'est toi qui me porte, c'est toi qui décide, c'est toi qui apporte du vent dans mes cimes, c'est toi qui me change, et tes discours raisonnent, tes yeux derrière ta frange m'ont convaincu mille fois d'être un homme*».

A ma mère, «*Le petit bonhomme, libre et fier comme un vagabond, Maman, est-ce que tu le revois de temps en temps? Celui qui te faisait sourire et rêver au bord de l'eau, toujours à courir devant, dès les premiers beaux jours du printemps. A part toi personne, ne me parle jamais de lui, Maman, à part toi qui m'a donné la vie*».

A mon père, tu m'as fait lever les yeux vers les cimes. «*Faut que je marche, parce que je comprends quand je marche. Faut que je marche, parce que j'apprends quand je marche. Faut que je marche, parce que je pense quand je marche, parce que j'avance quand je marche, parce que je rêve quand je marche*».

Daddy, je pense à toi. T'imaginer fier de moi a été ma principale pensée quand je voulais tout balancer. «*Et puis de te parler, ça j'ai pas oublié. Et puis de te penser, ça j'ai pas arrêté. Tous les jours c'est facile, t'es partout quand j'avance. Dans mon âme tu es l'île des souvenirs de l'enfance*»

Enfin, un mot pour toi ma partner in crime. Tu accompagnes mes pas au quotidien avec une attention et une bienveillance constante. Sans toi j'aurais pris un avion pour écourter mon voyage. Tu l'as compris, cette métaphore c'est un clin d'œil à l'aventure

dans laquelle on se lance. Avec toi tout est limpide et les possibilités sont infinies. Merci d'être là, merci d'être toi, merci d'être nous. «*Et ça fait 10 ans que t'es passée devant mes yeux, 10 ans que j'me dit que c'est c'qui m'est arrivé de mieux.* »

Je vous laisse avec ces vers de Clara Luciani, qui n'est pas étrangère à la complétion de ce manuscrit.

«*Quand je suis coincée dans un corps trop étroit, quand je sens l'univers se refermer sur moi, je pense aux fleurs, qui sont parfaites, qui n'ont pas d'autre rôle que de l'être. Je pense aux fleurs et, c'est bête, mais j'envie leur beauté muette.*»

Et ces vers de Neil Young qui résument les convictions qui m'ont porté pendant ce voyage.

«*Respect Mother Earth and her giving ways, or trade away our children's days.*»

Table of Contents

Résumé en français

Contexte

Dans un contexte global de lutte contre le changement climatique [1], la France fait face à un défi immense de décarbonation de ses systèmes énergétiques. Cette transition énergétique historique va faire appel à l'ensemble des énergies bas carbone disponibles, et l'éolien en mer y tient une place majeure [2]. De par sa position privilégiée et ses diverses façades maritimes, la France est amenée à jouer un rôle majeur dans le développement de l'éolien en mer. Les premières fermes éoliennes utilisant des technologies posées ont été mises en service en 2022 en France, avec un retard conséquent sur d'autres pays européens comme le Danemark, l'Allemagne, les Pays-Bas et le Royaume-Uni. Pour l'éolien flottant en revanche, la plus faible maturité technologique et l'immensité du potentiel de capacité installée en font un enjeu de recherche et développement majeur.

Dans ce cadre, France Énergies Marines, l'institut de transition énergétique français dédié aux énergies marines renouvelables (EMR) mène avec ses partenaires industriels et académiques et avec le soutien de l'Agence Nationale de la Recherche des projets de recherche collaboratifs pour augmenter le niveau de maturité des technologies EMR. Les projets de recherche émanent directement de problématiques industrielles, et ont pour objet des sujets aussi divers que l'intégration environnementale, la caractérisation de sites, l'optimisation de fermes et le design et suivi en service des systèmes.

Un des freins majeurs au développement de l'éolien flottant à travers le monde est la difficulté et le surcout associés aux opérations de maintenance. A la différence des éoliennes posées, les opérations de maintenance de l'éolien flottant nécessitent de faire des manœuvres de flottant à flottant. Les mouvements combinés du flotteur et du bateau de maintenance soumis à des conditions météo-océaniques complexes rendent les opérations dangereuses. La prise de décision opérationnelle nécessite donc des modèles de prévisions météo-océaniques précis et calibrés sur les paramètres environnementaux limitant les opérations (vent et vagues en particulier).

Pour répondre aux besoins de l'industrie sur ces questions spécifiques, France Énergies Marines a lancé en 2021 un projet de recherche collaboratif appelé FLOWTOM (FLoat-

ing Offshore Wind Operations and Maintenance) en collaboration avec des partenaires industriels (EDF, Skyborn renewables, Innosea, Jifmar, SBM offshore, Sofresid Engineering, EOLFI) et académiques (IFREMER, IMT Atlantique). Il a pour but d'explorer des solutions innovantes de levage lourd de flottant à flottant, et de développer des outils de prévisions météo-océaniques basés sur l'apprentissage profond répondant aux contraintes de ces nouvelles opérations. C'est ce dernier point qui fait l'objet de cette thèse qui a été entreprise en collaboration avec l'IMT Atlantique sous la chaire doctorale OCEANIX, sous la direction de Ronan Fablet et Pierre Pinson et la supervision de Pierre Tandeo et Maxime Thiébaut. Elle est soutenue par l'ANR via le projet FLOWTOM et la subvention CIFRE.

Pour une opération donnée, des limites opérationnelles sur certaines variables environnementales comme la hauteur des vagues ou la vitesse du vent sont définies à partir des réponses du systèmes aux conditions. Pour prendre en compte les incertitudes de prévision, ces limites sont dégradées d'un facteur appelé «alpha-factor», estimé de manière semi-empirique dans des tables de normes calibrées sur la mer du Nord [3]. L'application de ce facteur de sécurité rend la prise de décision plus conservatrice. De nombreux travaux récents insiste sur l'importance d'une prise en compte plus complète des incertitudes dans la prise de décision [4–8]. Les travaux présentés s'inscrivent dans la continuité de ces études, avec l'ambition de faire le lien entre prévision probabiliste multivariée, prévision de fenêtre météorologiques, prise de décision opérationnelle, et évaluation opérationnelle des modèles de prévisions.

L'état de l'art des prévisions météorologiques est basé sur des modèles de prévision numériques qui simule l'évolution de l'atmosphère en discrétisant l'espace-temps et en résolvant les équations de circulation à partir d'un état initial estimé. Ces modèles ont permis des avancées spectaculaires dans les dernières décennies [9], mais sont limités par leur formidable coût de calcul, notamment dans leur résolution spatiale et dans estimation des incertitudes. L'intégration de briques de modèles basées sur l'apprentissage profond a explosé ces dernières années, au point que des modèles entièrement basés sur l'apprentissage profond - modèles end-to-end - ont vu le jour ces dernières années, surpassant selon certaines métriques les capacités des meilleurs modèles numériques [10–13]. De nombreuses applications de problèmes inverses [14], de post-traitement pour l'éolien [15] ou encore d'émulation d'incertitudes [16, 17] ont également vues le jour. Dans ces travaux, nous explorons des méthodes d'apprentissage profond pour la prévision multivariées probabiliste de variables météo-océaniques, en nous situant dans le domaine du

post-traitement des modèles numériques et l'émulation d'incertitudes.

Par ailleurs, les modèles sont développés selon les contraintes des opérations de maintenance en mer, et évaluer dans ce cadre. Ainsi, nous relions les travaux de prévision jointe vent-vagues de Wu et al. [18], et la prise de décision probabiliste évaluer dans un cadre opérationnel comme introduit par Catterson et al. [6] et développé par Taylor and Jeon [7].

A la lumière de cette rapide introduction de la littérature scientifique, les questions auxquelles ces travaux essaient de répondre sont les suivantes :

- ***Comment les méthodes d'apprentissage peuvent-elles être utilisées pour améliorer les prévisions météo-océaniques pour l'industrie éolienne en mer ?***
- ***Qu'est-ce l'implémentation de ces méthodes implique pour l'évolution des standards, le déploiement de mesures en mer et l'évaluation des modèles de prévision ?***

Résumé par chapitres

Nous donnons ici un résumé de chacun des chapitres constituant la thèse. Le Chapitre II.4.3 dresse un état de l'art de la prévision en mer et des opérations de maintenance. Il s'adresse au lecteur non-familier avec ces notions. Le Chapitre II introduit l'ensemble des concepts mathématiques utilisés dans cette thèse, et ses fondements méthodologiques.

Les Chapitres III et IV sont données sous la forme d'articles publiés dans des revues scientifiques internationales. Ils peuvent être considérés comme des travaux de recherche en tant que tel et comportent des introductions, description de méthodes, résultats et conclusions.

Le Chapitre III décrit une méthode d'échantillonnage parcimonieux basé sur le clustering non-supervisé de données de modèles numériques. Des recommandations pour le déploiement de réseaux de capteurs sur les zones de développement de l'éolien en mer en France y sont données.

Le Chapitre IV décrit le développement d'un modèle convolutif pour la prévision probabiliste du vent sur l'île de Porquerolles. Il est basé sur l'utilisation du jeu de données open source Météo Net [19] dont il hérite des limitations.

Le Chapitre V fais le lien avec les opérations de maintenance de l'éolien en travaillant sur une cible jointe vent et vagues au large de Marseille. La méthodologie pour la prévision

de fenêtres météorologiques et l'évaluation des modèles dans un contexte opérationnel y sont détaillées.

Chapitre 1 - Prévision météo-océaniques et opérations en mer

Le travail présenté dans cette thèse est pluridisciplinaire. Il fait appel à des notions de prévisions météorologiques, de dynamique de l'atmosphère et des océans, d'opérations en mer pour l'éolien, de mathématiques appliquées et d'apprentissage profond.

Opérations en mer et conditions météo-océaniques

Dans une première partie, nous présentons le contexte général des opérations de maintenance de l'éolien en mer, le contexte particulier de l'éolien flottant, et lien avec les prévisions météo-océaniques. Les opérations de maintenance représentent 30% des coûts totaux des projets éolien en mer [20]. Pour l'éolien flottant, les mouvements combinés du flotteur et du bateau rendent plus compliquées les opérations de transfert de personnel ou de charge lourde (pale, hub, etc.). Différentes stratégies de maintenance peuvent être implémentées. La maintenance prédictive qui se base sur une estimation du vieillissement des composants par du suivi en service pour anticiper leur défaillance, et la maintenance corrective qui régit à la défaillance non prévue d'un composant. Ces deux types de maintenance ne représentent pas le même risque opérationnel. Lors d'une maintenance corrective, la production de l'éolienne est arrêtée jusqu'à ce que l'opération soit menée à bien. Le coût associé est lié à la perte de production pendant le temps d'arrêt de la machine. Les sources de délais sont décrite par [21]: la disponibilité des pièces détachées, d'un bateau et d'un équipage, leurs mobilisations, et le délai lié à l'attente d'une fenêtre météorologique en sont les principales.

Les standards de sécurité pour les opérations sont délivrés par le Det Norske Veritas (DNV) [3]. Ils y différencient deux types d'opérations en fonction de la durée de l'opération. Les opérations durant moins de 72h (Et 96h en contant la marge de sécurité) sont considérées comme «limitées par la météo», et sont sujettes a des limites d'opérabilité censées garantir l'occurrence d'événements dangereux sous un certain seuil de probabilité. Une méthode générale pour définir les limites opérationnelles est décrite dans [22].

Lors d'une phase d'opérations, la décision ou non de tenter l'opération se fait à la lumière de modèles de prévisions météo-océaniques. Ceux-ci étant sujets à de nombreuses incertitudes, une méthodologie est mise en place au niveau industriel pour limiter les risques. Les limites d'opérabilité sont dégradées d'un facteur $0 \leq \alpha \leq 1$ nommé l'alpha-

factor, qui rend les limites plus conservatrices et donc limite le risque de conditions dangereuses. Ce facteur est déterminé à partir de critères semi-empiriques et de tables de valeurs calibrées sur la mer du Nord. Les détails du calcul de l'alpha-factor sont décrits dans [8]. Des travaux récents proposent des méthodologies pour améliorer l'alpha-factor pour qu'il prenne en compte la réponse dynamique du bateau aux conditions [4, 5, 18]. L'amélioration des prévisions météo-océaniques et l'intégration des incertitudes dans la prise de décision sont identifiés comme des facteurs importants de la réduction des coûts d'opération pour l'éolien en mer [23].

Bases de prévisions météo-océaniques

Dans les sections suivantes, nous donnons des éléments de contexte sur les prévision météo-océaniques, les mesures in-situ en mer et la météorologie du Golfe du Lion. Les prévisions numériques rassemblent l'ensemble des méthodes basées sur la résolution des équations dynamiques de l'atmosphère ou de l'océan sur une grille discrétisée. Ils ont connus un développement fascinant depuis leur théorisation par Abbe, Bjerknes et Richardson [24–26], et il est estimé qu'ils gagnent un jour de prévision supplémentaire par décennie [9]. Le formidable développement des moyens de calculs associé à des progrès considérables dans la paramétrisation des phénomènes physico-chimiques et a une quantité exponentielle de données d'observation de l'atmosphère ont permis cette prouesse informatique. L'atmosphère étant un système complexe souvent assimilé à un système chaotique en très grande dimension, les incertitudes liées à l'estimation de l'état initial de l'atmosphère et les incertitudes liées aux approximations numériques amènent les simulations à diverger pour des horizons de prévisions grandissant, limitant dans le futur les capacités de prédiction. Dans ces travaux nous utilisons le modèle AROME qui est un modèle de très haute résolution sur la France avec une maille de 1.3 km. Il fournit des prévision 42 heures en avance toutes les 6 heures [27]. Pour estimer les incertitudes, les modèles numériques font plusieurs simulations en parallèle à partir d'un état initial perturbé et avec des paramétrisations perturbées. Le nombre de membres d'ensemble est un paramètre majeur pour caractériser la complétude de l'estimation de l'incertitude. Le modèle AROME utilise 16 membres d'ensemble, ce qui est une limitation, tandis que le modèle européen IFS de l'ECMWF en produit 50. Les prévisions d'ensemble donnent une idée de la prédictibilité de la situation météorologique en générant des échantillons d'une distribution postérieure en grande dimension non explicitée.

Récemment, un grand nombre d'équipes de recherche ont cherché à implémenter des modèles basés-données ou d'apprentissage profond pour des applications liées à la prévision météorologique. Celles-ci ont comme avantage d'avoir un coût de calcul bien plus faible que les modèles numériques, tout en proposant des propriétés non-linéaires très intéressantes. En revanche, la dynamique de l'atmosphère n'y est pas explicitée et les modèles peuvent être perçus comme des boîtes noires n'offrant que peu de possibilités d'interprétabilité des résultats. Nous donnons dans cette section une revue de la littérature complète en différenciant les modèles relevant de modèles «statistiques», souvent paramétriques et ne faisant pas appel à des réseaux de neurones, les méthodes utilisant l'apprentissage profond qui ont essaimé ces dernières années, et enfin les méthodes hybrides qui s'attachent à intégrer des contraintes physiques dans leur modélisation. L'intérêt de ces méthodes pour générer des prévisions améliorées pour des applications spécifique est évident au regard de cette littérature.

Le lien entre prévision météorologiques et océanique est important. Les oscillations de surface de l'océan sont principalement dues au vent qui génère par friction de surface des vagues de gravité. Les caractéristiques de celles-ci dépendent de la durée et la distance durant laquelle le vent a soufflé sur la surface [28]. Les paramètres considérés dans ces travaux comme influençant les opérations de maintenance sont la hauteur significative des vagues définie comme la moyenne du tiers supérieur des vagues observées, et la période moyenne des vagues. Comme les modèles numériques de vent, les modèles numériques de vagues (Wave Watch III forcé par AROME utilisé dans ces travaux) sont gourmands en temps de calcul. De la même manière, de nombreux travaux fleurissent sur des applications de modèles d'apprentissage pour la prévision des vagues [4, 29–33]. La relation entre le vent et les vagues étant complexe, certain travaux s'attardent sur leur prévision jointe, surtout dans des cadres de réanalyse on de génération de séries temporelles historiques [34–37]. Pour la planification des opérations, nous avons besoin de prédire ces séries temporelles jointe. Ce nouveau besoin n'est traité à notre connaissance que dans les travaux récents de [38], et constitue une partie importante de la motivation de cette thèse.

Pour le développement de méthodes d'apprentissage pour la caractérisation de sites en mer, des données de mesures in-situ en mer sont nécessaires. Celles-ci sont rares, pas toujours accessibles, et sur des durées limitées (campagne HyMeX [39]). La France a besoin d'installation de mesures pérennes pour que la recherche sur le sujet puisse avancer et pour favoriser le développement des éoliennes en mer. L'exemple des mats FINO allemands est frappant, car les données qu'ils produisent sont utilisés dans une majorité des travaux de

recherche européens sur le sujet. Dans cette thèse nous avons du nous contenter d'un jeu de données open-source MétéoNet [19] limité pour notre cas d'étude. C'est un exemple de partage de données qui doit être répliqué. Nous utilisons également des données de mesure sur l'île du Planier au large de Marseille ou un LIDAR a été déployé par France Énergies Marines qui fournit des données de grande qualité et dont la durée d'exploitation doit être prolongée pour bénéficier à la filière.

Enfin, ces travaux sont appliqués au Golfe du Lion, zone propice au développement de l'éolien flottant et enjeu majeur pour la décarbonation des activités industrielles sur le littoral. Il jouit d'une situation très ventée dominée par des vents régionaux marqués que sont le Mistral dans sa partie Est, la Tramontane dans sa partie Ouest, et le Marin qui est un vent de Sud-Est chargé d'humidité et responsable d'événements de précipitation intense sur les reliefs littoraux. Le développement de parcs éoliens commerciaux renforcent les enjeux en mer sur cette zone et doit être l'occasion d'y multiplier les efforts de recherche pour mieux en caractériser les caractéristiques météo-océaniques.

Chapitre 2 - Background mathématique et méthodologique

Ce chapitre donne un cadre mathématique et méthodologique aux travaux qui suivent. En particulier, les méthodes d'apprentissage profond et leurs principes méthodologiques sont présentés. Une introduction à la prévision probabiliste et à son évaluation est ensuite donnée. L'ensemble des équations et du formalisme développé dans la thèse n'est pas décrit ici par souci de synthèse.

Bases de l'apprentissage profond

La théorie de l'apprentissage machine a été décrite par [40] en ces termes : «Un algorithme apprend d'une expérience E par rapport à une tâche T et selon une mesure de performance P si sa performance pour la tâche T mesurée par P s'améliore avec l'expérience E.». Les modèles d'apprentissage profond sont un sous-ensemble des modèles d'apprentissage machine qui utilisent plusieurs couches de neurones. Il a été démontré par [41] qu'ils étaient des approximateurs universaux si leur nombre de neurones est suffisant. Pour une description complète des principes sous-jacents et des modèles existants en apprentissage profond, le lecteur est dirigé vers le livre de Goodfellow et al. qui fait référence dans le milieu [42].

Un modèle d'apprentissage est dit supervisé s'il est entraîné en calculant une erreur

par rapport à une cible. S'il n'y a pas de cible, on parle d'apprentissage non-supervisé. On distingue dans les types de modèles existants les modèles de régression et de classification. La régression traite de sorties continues tandis que la classification prédit des sorties discrètes (assigne les données à des classes). Les deux sortes de modèles sont utilisés dans ces travaux.

Différentes architectures peuvent être utilisées pour construire des modèles d'apprentissage profond. Les réseaux convolutifs utilisent des filtres convolutifs pour diminuer la quantité de coefficients du modèle en partageant les filtres pour toutes les données d'entrées. Ils sont très utilisés pour traiter les images. Les réseaux récurrents possèdent des cellules internes qui retiennent l'information des données précédentes, ce qui les rend performants pour la prévision de séries temporelles. Les perceptrons multicouches consistent en une succession de couches de neurones complètement interconnectées et sont les briques de bases des modèles d'apprentissage profond. Enfin de nouvelles architectures dites génératives bénéficient d'une notoriété grandissante pour leur capacité à approximer des distributions de probabilité complexes. Parmi celles-ci les flots normalisant utilisent une composition de transformation bijectives et inversible pour transformer un espace latent simple en une distribution complexe.

Prévision probabiliste

La prévision probabiliste consiste à prédire l'évolution d'un phénomène incertain dans le futur. Le développement de modèles probabilistes en météorologie constitue un changement de paradigme dans les années 1980 avec le développement de la prévision d'ensemble [43]. Pour l'éolien en mer, l'estimation de l'incertitude est primordiale par exemple pour l'équilibrage du réseau électrique [44], la participation au marché de l'électricité [45], ou bien l'estimation de fenêtres météorologiques [5]. La prévision probabiliste consiste à estimer une distribution postérieure i.e. pour les sorties du modèle. Cette distribution peut être décrite de manière paramétrique, en prédisant les paramètres d'une distribution connue (e.g. Gaussienne), ou bien non-paramétrique, en prédisant des observations indirecte de la distribution : niveaux de quantiles, intervalles, échantillons. Enfin pour les séries temporelles, la prévision peut être désignée pour générer des scénarios comme c'est le cas pour les prévisions numériques, ce qui peut être paramétrique ou pas, mais qui est en très grande dimension. Ces différentes approches sont utilisées dans les travaux.

Évaluation des prévisions

L'évaluation d'une prévision probabiliste n'est pas triviale et nécessite l'introduction de certains concepts. Le travail de Tilmann Gneiting est une référence pour l'évaluation des modèles probabilistes [46–49]. Nous référons le lecteur également à l'excellent travail de Messner et al. [50] qui recense les méthodes pour évaluer les prévisions dans un cadre énergie éolienne.

Selon Gneiting, un bon modèle de prévision probabiliste doit être «sharp subject to calibration». Le terme «sharp»fais référence à l'écart à la moyenne, et le terme «calibration»est lié à la consistance statistique de la distribution prédite. Un modèle correctement calibré observera 10% des observations tombant dans chacun des déciles de sa distribution.

Au-delà de l'aspect qualitatif des modèles, il est primordial d'évaluer leur valeur vis à vis d'une application donnée. Comme montré par [51], la valeur d'une prévision pour la participation au marché de l'électricité diffère de la mesure de qualité. Dans [50], l'importance de considérer des métriques pénalisant de manière proportionnelle aux coûts associés les erreurs de prévisions est appuyée.

Chapitre 3 - Design d'un réseau de capteurs pour l'échantillonnage optimal de la ressource en vent.

Le développement de modèles de prévisions basés sur l'apprentissage profond requiert la construction de jeux de données conséquents. Pour l'éolien en mer cela repose sur des séries temporelles de mesures. Partant du constat que ces mesures sont rares et devraient être développées à grande échelle sur les zones de développement de l'éolien en mer, nous présentons des travaux pour la construction de réseaux de capteurs pour l'échantillonnage parcimonieux optimal de la ressource en vent. Nous faisons référence à la littérature d'échantillonnage parcimonieux, qui propose des méthodes d'analyse de données pour trouver des points saillants dans un jeu de données. En particulier les méthodes présentées par [52], [53], [54] sont pris pour références.

En utilisant des données de prévision d'AROME, nous présentons une méthode de clustering non-supervisé à l'aide de mélange de Gaussiennes (GMM). Un ensemble de Gaussiennes est entraîné sur une projection des données de vent sur un sous-espace obtenu par décomposition en composante principales. Cet entrainement fait appel à l'algorithme d'Expectation Maximization [55], et permet de décrire l'espace en question à partir d'un nombre réduit de distributions multivariées, en maximisant la vraisemblance des obser-

vations. Par cette méthode, nous obtenons les centroïdes des distributions qui sont des points de grilles du modèle portant le maximum d'information. Nous montrons que ces points sont de très bons candidats pour la reconstruction du champ de vent à partir de leur seule mesure.

Avec cette méthode, nous proposons une méthodologie complète pour la définition d'un réseau de capteurs optimal, notamment définissant le nombre optimal de capteurs à déployer. Les résultats obtenus surpassent largement les résultats obtenus par les méthodes de l'état de l'art pour ce cas particulier, et nous permettent de donner des recommandations de réseaux de capteurs pour les 3 principales zones de développement de l'éolien en mer en France. Dans le futur, l'intégration d'autres variables comme les vagues, ou bien des variables biochimiques ou en lien avec la biodiversité pourraient être intégrées dans l'optimisation pour le déploiement de plateformes de mesures multimodales et le suivi en service des impacts des parcs éoliens (biodiversité, effets de sillages, pollution chimique). A l'échelle d'un parc, cette méthode gagnerait à être implémentée sur des modèles haute résolution de type Large Eddy Simulation pour le placement fin des capteurs sur une zone d'exploitation.

Ce travail a été publié dans le journal *Wind Energy Science*, présenté à la conférence internationale *Seanergy 2022*, et a été utilisé par Météo France pour fournir des recommandations à la Direction Générale de l'Energie et du Climat pour le déploiement de capteurs au niveau des grandes façades maritimes françaises.

Chapitre 4 - Prévision très court-terme probabiliste du vent sur le jeu de données MétéoNet

Dans ce chapitre, nous présentons les modèles développés pour la prévision probabiliste multivariée en mer. Nous utilisons un jeu de données open source de Météo France - MétéoNet - et faisons de la station de mesure de l'île de Porquerolles notre cas d'étude. Le but de l'étude est d'explorer l'apport des méthodes d'apprentissage profond pour la prévision probabiliste à un point en mer non observé. L'importance de différentes sources de données est étudiée, avec notamment pour idée de corriger les prévisions de modèles numériques avec des données de mesures récentes des stations côtières voisines de notre point d'intérêt.

A l'aide de réseaux de neurones convolutifs, nous assimilons des données de modèles numériques en grande dimension et des données de mesures multivariées des stations

voisines, en les projetant sur un espace latent d'où sont extraits des séries temporelles de paramètres de distributions Gaussiennes en sortie. Nous le nommons ConvE-STF dans la thèse pour Convolutional Encoder for Short-Term Forecasting. Un schéma du modèle est proposé en Figure IV.5. Nous proposons une hypothèse supplémentaire utilisant des flots normalisant pour relâcher l'hypothèse Gaussienne du postérieur et nous montrons que cela amène des propriétés intéressantes pour un coût de calcul raisonnablement plus élevé pour approximer des phénomènes très non-Gaussiens (ici la discrétisation de la cible). Ces deux modèles sont comparés avec le point de grille du modèle AROME le plus proche de la cible, ainsi que deux méthodes de l'état de l'art, les analogues [56, 57] qui utilisent des situations observées dans un catalogue comme possible futures évolutions de la situation, et une Gradient Boosting Machine qui est une méthode d'apprentissage automatique basée sur l'utilisation d'un ensemble d'arbres de régression, et qui est implémentée dans des travaux de recherche proches [4].

Les résultats obtenus montrent que l'apprentissage profond est capable d'assimiler une grande quantité de données sans se soucier du pré-traitement, ce qui in-fine permet une correction bien plus performante des prévisions que pour les modèles de référence. L'importance des mesures de stations voisines est mis en exergue et ouvre la porte à une correction proche de l'instantané des modèles numériques. Par ailleurs cette correction est plus importante pour les vents soufflant depuis la côte, montrant les difficultés liées aux effets orographiques côtiers pour les modèles numériques et l'apport de données de mesures récentes en amont dans la direction du vent pour sa correction. Cette étude est limitée par la taille du jeu de données à des prévisions 6 heures en avance, ce qui ne permet pas le développement d'un cadre opérations de maintenance, ou des prévisions à au moins 24h sont nécessaires. En revanche elle pose un cadre de développement et d'évaluation pour les prévisions probabilistes multivariées du vent en mer. La cible de l'entrainement est un mat de mesure sur une île dont le relief est important. La très grande différence en performance avec le modèle AROME montre les perturbations dont le site est victime, qui rend très favorable le post-traitement. La généralisation à un site non perturbé en mer est importante. Les conclusions sont spécifiques au site et devraient également être généralisées à d'autres façades avec d'autres régimes de vent.

Chapitre 5 - Prévision jointe vent-vagues et applications pour la prise de décision en opérationnel

Grâce à des séries temporelles de mesures de vent et de vagues sur l'île du Planier au large de Marseille, nous avons construit un jeu de données permettant un cas d'études réaliste pour les opérations de maintenance de l'éolien en mer. Nous adaptons le modèle ConvE-STF et son pendant utilisant les flots normalisant au cas de la prévision jointe entre la vitesse du vent à 100m, la hauteur significative des vagues et la période des vagues, trois paramètres capitaux pour les opérations. En contraste avec les résultats du chapitre précédent, les prévisions AROME au point de grille le plus proche de la cible se sont révélées très pertinentes pour la vitesse du vent, et nos modèles n'améliore pas la prévision de la moyenne (ils sont dans la marge d'erreur). Nous montrons en revanche que l'entrainement de modèles d'apprentissage profond par rapport à un critère de vraisemblance permet de prédire des incertitudes bien calibrées par rapport à ce qui pourrait être fait de manière empirique. Nous montrons par ailleurs que l'usage de flots normalisant dans ce cas est très bénéfique pour la calibration du modèle, ce qui supporte l'hypothèse selon laquelle il est bénéfique de traiter ces trois variables de manière jointe, et que la forme des incertitudes entre ces variables corrélées n'est pas bien décrite par une distribution Gaussienne.

Afin d'appliquer ces prévisions pour la planification des opérations, nous devons obtenir des trajectoires afin de calculer la probabilité d'occurrence de fenêtres météo. Celles-ci sont liées à des probabilités de dépassement de seuil sur une fenêtre glissante et nécessitent donc des échantillons corrélés dans le temps reflétant l'incertitude de la prévision, et un cadre d'évaluation dédié.

Nous proposons différentes approches pour générer des échantillons corrélés temporellement de manière satisfaisante. Celles-ci sont dans la littérature scientifique basées sur l'utilisation de copules [58]. Nous proposons un copule Gaussien obtenu avec une matrice de covariance empirique pour la composante temporelle et une covariance conditionnelle pour la corrélation entre les variables. Pour les flots normalisant, nous introduisons deux méthodes innovantes afin de s'affranchir de l'hypothèse Gaussienne pour la dépendance temporelle. La première, basée sur un échantillonnage similaire pour chacun des pas de temps dans l'espace latent utilisant une astuce algorithmique, se montre d'une grande simplicité et est plus performante que la référence utilisant des copules Gaussiens. La seconde utilise un copule Gaussien dans l'espace latent, mais ne permet pas d'obtenir des résultats satisfaisants. D'autres travaux récents offrent des pistes de recherche intéressantes sur ce

point particulier [44, 59].

Les modèles ainsi obtenus sont évalués grâce à des métriques dédiées aux opérations de maintenance. Avec des hypothèses ad-hoc sur une durée d'opération et des limites d'opérabilité, la qualité des fenêtres météo prédites est évaluée montrant un avantage pour les méthodes introduites par rapport aux modèles numériques. Par ailleurs nous introduis la notion de risque pour l'évaluation de la valeur des prévisions, en partant de l'observation de [60] que l'utilisation de métrique économique pour l'optimisation tend à pousser les équipes à travailler dans des conditions dangereuses. Notre cas d'étude corrobore cette hypothèse, et nous proposons une manière d'augmenter les métrique économique de [6] par cette notion de risque.

Perspectives

Ces travaux démontrent l'intérêt des méthodes d'apprentissage profond pour la prévisions météo-océaniques et proposent un cadre méthodologique et d'évaluation dédié aux opérations de maintenance en mer. Le points suivants ont été soulevés et présentent un intérêt de recherche fort :

- Plus de campagnes de mesures en mer et sur de plus longues périodes sont requises pour la recherche et le développement de parcs éoliens en mer. Nos résultats proposent des cadres de déploiement qui permettraient d'échantillonner la ressource en vent à l'échelle des façades maritimes.
- Les modèles proposés offrent un cadre compétitif pour la prévision probabiliste multivariées des variables météo-océaniques. Ceux-ci doivent être comparés à l'état de l'art des prévisions probabilistes à savoir les prévisions d'ensemble des modèles numériques. Pour ce faire un partage plus large des données de modèles serait bénéfique.
- Des variables additionnelles pourraient être ajoutées aux modèles pour améliorer les prévisions. Par essence, ceux-ci sont capables d'assimiler une large quantité de données hétérogènes. Nous pensons aux données de température de surface de la mer ou de rugosité de surface.
- La génération conditionnelle de scénarios à l'aide des flots normalisant semble très prometteuse et nécessite des travaux de recherche dédié dans une littérature scientifique très récente.
- Les cadre d'évaluation proposé doit être implémenter avec des données réelles

d'opérations de maintenance. La simulation de l'utilisation de ces modèles comme outil de prise de décision opérationnelle sur des séquences de maintenance réelle serait un très beau cas d'application et une démonstration de l'importance des partenariats recherche / industrie.

- L'éolien en mer possède d'autres aspects qui bénéficieraient grandement de modèles de prévision probabilistes. Nous pensons ici à la participation au marché de l'électricité, à l'équilibrage des réseaux électriques, au contrôle actif des turbines, à la prévision de la turbulence ou d'événements de rampes de vent. Des cadres d'évaluation dédiés doivent être développés pour chacune de ces applications.

INTRODUCTION

General context

Humanity faces unprecedented threats due to its impacts on the environment. Amongst them, anthropogenic carbon dioxide emissions since the industrial revolution are causing a global warming and climate change that put ecosystems at an existential risk, and threatens hundreds of millions of people around the globe [1]. The majority of anthropogenic carbon dioxide emissions results from the combustion of fossil fuels for energy applications. In 2022, fossil fuels still represented a staggering 80.1% of the global final energy consumption [61]. The share of electricity in the final energy consumption would need to increase by 4% per year to achieve net zero emissions by 2050, a prerequisite for the limitation of global warming to a maximum of 2°C. The massive development of low-carbon energy sources is then a worldwide necessity for the decarbonisation of energy systems, and a historic industrial challenge.

Renewable electricity generation regroups energy generation from solar irradiance, wind, tides, waves, rivers etc. It globally represented 30% of the global electricity production in 2022, and is expected to cover 40% of it by 2028 [62]. Solar photovoltaic and wind energy are the two main drivers in the current increase of renewable energy sources, with 510GW installed globally in 2023, 75% of which being solar photovoltaic [62]. While offshore wind energy represented only 0.3% of the global energy production in 2022, mostly in Europe, it has been estimated that the global bottom-fixed offshore wind energy potential was 1.5 times the current total electricity consumption, and the floating offshore wind energy potential up to 11 times [63]. In this context, offshore wind energy will become one of the main sources of low-carbon electricity globally, and a key technology for mitigating climate change in the next decades.

The French energy context is a particular case. While the share of fossil fuels in the final energy mix is 60%, its electricity mix is one of the least carbon-intensive in the world thanks to massive historical nuclear energy, and large hydro-power capacities. Its carbon intensity was estimated in 2021 to be $62\,gCO_2/kWh$ in contrast with the mean European carbon intensity of $317\,gCO_2/kWh$ [2]. However, there is a need for electrification of en-

ergy uses. There is an uncertainty in the industrial capabilities for the replacement and increasing of an ageing nuclear fleet. Furthermore, the hydro-power exploitation is reaching saturation. For all these reasons, the development of solar and wind energy will be the cornerstone of the French energy transition in the next decades. Offshore wind energy allows the installation of more powerful turbines in very large areas where the wind resource is greater. France has one of the largest maritime areas of the European Union, with a variety of wind regimes in the North Sea, the Atlantic and the Mediterranean coast. It could then become one of the leaders in the global offshore wind energy market. Despite some delays in the development of the first commercial-scale wind farms, which have been commissioned in 2022, France still has a key role to play in the worldwide development of this technology. In particular, floating offshore wind energy, which is to be deployed in waters that are more than 60m deep requires massive research and development efforts to reach industrial maturity. That is why the French government awarded 3 pilot floating wind farms in the French Mediterranean Sea, the first one of which has been installed in 2023 off the coast of Port-Saint-Louis du Rhône. In 2024, the first call for tender for an industrial-scale 250MW floating offshore wind farm in Southern Brittany was awarded at a staggering 86.45 €/MWh, showing the rapidly increasing maturity of such projects. In France, up to 18.5 GW of floating and 40.5 GW of bottom-fixed offshore wind capacity is envisioned by 2050.

In this context, France Énergies Marines, the French research institute dedicated to marine renewable energies (MRE), conducts collaborative research projects with its industrial and academic partners to increase the maturity level of MRE. Its research topics are raised by industry-relevant issues related to environmental and societal integration, site characterization, farms optimization and systems design. The collaborative construction of the research projects then encourages data and knowledge sharing, improves the industrial transfer of research, and ensures the industrial relevance of the research work.

One of the main difficulties that offshore wind energy faces compared to its land-based counterpart is the added challenges for operations. The installation, maintenance, monitoring, decommissioning of offshore wind turbines require the use of specific vessels, specialized crews in a challenging environment. This brings the operational expenditures up to 30% of the total cost of projects [20], and more than twice the operational expenditures per MW installed of land-based wind projects. For floating offshore wind energy, the very possibility of performing heavy-lift operations in a floating-to-floating configuration is still uncertain. Compared to the knowledge acquired with bottom-fixed wind

turbines, the sensitivity of different maintenance operations to the floater's motions is to be estimated and mitigated.

To answer this research gap, France Énergies Marines started in 2021 the research project FLOWTOM (FLoating Offshore Wind Turbines Operations and Maintenance) in collaboration with industrial partners (EOLFI, skyborn renewables, EDF, Innosea, Jifmar, SBM offshore, Sofresid Engineering) and academic partners (IFREMER, IMT Atlantique). This project aimed at exploring innovative solutions for floating-to-floating heavy lift maintenance through method-statement validation with basin tests and numerical modelling. It furthermore proposed to explore metocean (meteorological and oceanic) forecasts models, and how it impacts maintenance costs. The latter was explored through this Ph.D. thesis, in collaboration with the Lab-STICC at IMT Atlantique under the OCEANIX chair, with the financial support of the ANR through the CIFRE granting framework.

Scientific context

Offshore operations are constrained by environmental conditions for their safe execution. For a given operation, operability limits are estimated based on vessel and floater response to metocean conditions. The decision whether to carry out the operation is then based on the compliance of environmental conditions with the operability limits for the duration of the operation based on metocean forecasts. To account for forecast uncertainties, operability limits are to be degraded by a factor depending on semi-empirical criteria [3]. This alpha-factor is used because of the lack of reliable uncertainty estimate of metocean variables, and several recent studies advocate for a more comprehensive uncertainty consideration during operational decision-making [4–7, 18, 64]. This thesis positions itself in the continuation of these studies, trying to connect multivariate probabilistic forecast emulation [4, 64], weather window prediction, operational decision-making [5, 18] and forecast value evaluation [6, 7] to provide a full forecast-to-decision-making workflow.

State-of-the-art metocean forecasts are issued by numerical models of the atmosphere or the ocean. In the last decades, they have provided results of ever-increasing quality [9], in particular for uncertainty estimation with the development of ensemble forecasting [65]. However, they face a high computational cost that limits their resolution and the number of ensemble members that can be run in parallel. Deep learning integration in weather forecasts applications has recently drastically increased [66], be it for end-to-end

model emulation [10–13], model post-processing [15], inverse problems [14] and uncertainty emulation [16, 17]. In the context of offshore metocean forecast for operations, the low inference computational cost appears appealing for short-term forecast emulations from recent measurements. This thesis explores learning-based methods for the multivariate probabilistic post-processing of numerical predictions of wind and wave conditional to recent neighbouring measurements.

The development of learning-based forecast models for offshore applications requires the gathering of a substantial amount of in-situ data, where environmental conditions can be very complex, and at prohibitive water depths. Such data sources are scarce and expensive, and their deployment is one of the main barriers to offshore wind energy research [67]. As a strategic asset for future offshore wind energy development, offshore wind speed sensing networks should be deployed. Where to place those sensors to optimally sample the wind field then appears as an important optimization question. In this thesis, we use a learning-based algorithm on numerical weather prediction data to define optimal sensors networks for offshore wind speed resource.

The recent literature on wind or wave post-processing using deep learning is flourishing, driven by the rise of offshore marine renewable energies [68, 69]. The joint post-processing of wind and wave parameters however has rarely been studied in a forecast configuration, but rather in a historical time series construction perspective [37]. A recent work from [8] developed a joint wind and wave probabilistic forecast model to be integrated in decision-making in the form of a response-alpha factor, hence eventually considering a point forecast for decision-making. In this thesis we aim at building a probabilistic forecast of limiting parameters that could be used for probabilistic decision-making and be evaluated in the framework described by [6] and [7].

As described in [64], the multivariate probabilistic forecasting problem is an active field of research. Novel generative architecture bring new possibilities in complex distributions estimation in high dimensional space which can be used for multivariate probabilistic forecasting [44], as an alternative to more classical probabilistic tools such as multivariate copulas [58]. For weather window forecasting, multivariate temporal scenarios should be generated. The applicability of generative models for scenario generation is still to be validated, and its quality and value versus classical copula-based scenario generation has not been studied to the best of our knowledge and is investigated on a real case study.

Considering the above literature, the questions that are investigated in this Ph.D. are the following:

- ***How learning-based metocean forecasting methods can improve operations at sea for the offshore wind energy industry ?***
- ***What does the implementation of such methods imply for in-situ measurements, standards evolution and forecast evaluation ?***

Overview

The work presented in this thesis is cross-disciplinary and requires background in research fields as distant as offshore wind turbines maintenance operations, probabilistic forecasting, weather and wave forecasting, applied mathematics and deep learning. Chapters I and II are intended for readers who are unfamiliar with these notions. Literature reviews of the different aspects of the thesis are given in these chapters to help define the scope of the research work.

In Chapter I, general background is given about the research fields this thesis refers to. Firstly, the industrial constraints associated with offshore operations for wind turbines are detailed, especially in connection to environmental conditions, to serve as specifications for the different experiments. An introduction to wind and wave forecast is then given, introducing important concepts about numerical modelling of weather systems and waves. Data-driven methods are also introduced. The complex interactions between wind and waves are highlighted. A focus is then made on in-situ data gathering for offshore applications. The meteorological features of the Gulf of Lion, the main study area, are finally described.

A methodological and mathematical background is then provided in Chapter II, introducing notations and concepts associated with multivariate time series probabilistic forecasting, forecast evaluation and deep learning. All the applied mathematics concepts and model architecture used in this thesis to build data-driven models are introduced in connection to the time series forecasting of environmental variables. The different approaches to multivariate probabilistic forecasting and their evaluation framework are detailed. The economic value of forecasting is introduced, in connection to maintenance operations planning and execution. Eventually a thorough literature review on data-driven models for offshore wind energy is given.

Chapters III and IV are given as published articles in peer-reviewed journals. They can be read as independent research contributions. These two chapters also contain introduction of scientific context, literature review and used methods. Their main conclusions

are detailed in their last section.

Chapter III presents a data-driven strategy using numerical weather prediction data to define an optimal sensors network for offshore wind characterization. It is applied to the main offshore wind energy development areas in France and compared to state-of-the-art sparse sampling methods. The in-situ measurement of wind and wave is a prerequisite to the development of data-driven offshore forecast model, and the lack of measurement data was found to be one of its main limitations. This study is published as a peer-reviewed article in the open source journal *Wind Energy Science*.

In chapter IV, a deep-learning approach for the probabilistic post-processing of wind forecasts at sea is proposed. It uses an open-source dataset to mimic a realistic case study. Several approaches are compared both in terms of deterministic and probabilistic metrics. The probabilistic output of the model is expressed as quantile levels, parametric distribution and generative non-parametric distribution. These approaches and their pros and cons are thoroughly discussed. It has been published as a peer-reviewed article in the journal *Artificial Intelligence for Earth Systems*.

Eventually in Chapter V, the probabilistic forecast model of Chapter IV is adapted to a more realistic dataset based on combined wind and wave measurements at an offshore site. It is applied to the joint probabilistic forecasting of limiting parameters for operations. From this probabilistic forecast, multivariate scenarios are generated to permit the modelling of operational decision-making under uncertainty. Different strategies are compared for scenarios generation in terms of quality and value. The definition of forecast value metrics for offshore operations defines a new evaluation framework and paves the way for a more comprehensive uncertainty consideration in operational decision-making.

The main results from all the chapters are gathered in the conclusion section, with discussions on the study limitations and perspectives for future research. Owing to the fact the manuscript is based on published articles, the mathematical notations can be redundant from one part to another. Though efforts were made to harmonize them at the maximum, the notations should be considered chapter dependent.

Outreach and industrial transfer

This thesis work is by construction meant to be transferred to the industry. The discussions with the FLOWTOM project partners not only helped defining the scope of work but enlightened the results analysis with operational realism. A workshop gathering

different stakeholders within marine operations was organized on the topic of decision-making under forecast uncertainty. Evaluation metrics relevant for operational decision-making were proposed, and the forecast models were integrated in an online forecast visualisation platform for the project partners to compare the different decision-making strategies.

The proposed methodology for offshore wind sparse sampling was transferred to Météo France, the French weather service, for real case application. They produced recommendations for the deployment of floating LIDAR arrays in the future offshore wind energy tender areas in France using the results from the model proposed in Chapter III.

A study on the integration of wave data from the published work of Chapter IV was performed by Oscar Gauvrit, a student from IMT Atlantique during his master thesis project. I fully supervised the internship, and built on his work to integrate wave data in Chapter V.

In addition to the two peer-reviewed articles, this work was presented at two international conferences: the Seanergy conference in 2022 in Nantes, and the Wind Europe conference in 2023 in Copenhagen.

The research work performed in the FLOWTOM project is to be continued at France Énergies Marines. I participated in defining the scope of a research project NEMO dedicated to offshore wind turbulence measurements and modelling. The deep learning approaches for probabilistic time series forecasting will be applied to offshore turbulence and low-level jets predictions for wind turbine active monitoring.

Chapter I

Weather forecasting for offshore wind energy operations

Preamble

The work presented in this thesis is multidisciplinary. First driven by an industrial technological barrier on floating offshore wind energy maintenance operations, it uses mathematical notions from the weather forecasting and applied mathematics communities. This chapter aims at giving background and related literature on the different aspects addressed throughout the thesis.

The development of novel deep-learning-based models for metocean characterization and forecasting is motivated by the challenges faced by the offshore wind energy industry. In Section I.1, we give background on offshore wind energy maintenance operations to introduce the framework of this work. In link with the work carried out in the FLOWTOM project, the specific challenges faced by floating platforms are detailed, thus defining the scope of the study case. The industrial standards for the operational use of metocean forecasting bulletins are described, and their identified limitations are the starting point of the research work. A literature review on operational decision-making under uncertainty for offshore operations is given.

Section I.2 is dedicated to the forecasting of metocean variables. We introduce notions of numerical weather prediction, in link with the forecasting of wind speed and waves. The good understanding of numerical predictions is important to evaluate the added value of data-driven models. The strengths and weaknesses of state-of-the-art numerical models are detailed and explained. Numerical weather predictions serve as the main input data for Chapter III, Chapter IV and Chapter V. It is also the main baseline model for forecasts comparison in Chapter IV and Chapter V. We then describe state-of-the-art statistical, and hybrid forecast models as research topics in weather forecasting.

The challenges associated with in-situ measurement at sea, and the different data

sources used in this thesis are presented in Section I.3, and a brief introduction on the weather patterns and hydrodynamics of the study area, the Gulf of Lion is given in Section I.4.

I.1 Weather limited offshore wind energy operations

I.1.1 Background on maintenance operations for offshore wind energy

Operations and maintenances (O&M) take place throughout the lifetime of an offshore wind farm. It comprises all the offshore operations for the installation, operation and decommissioning of the wind farm. The sum of all costs associated with operations and maintenance is called the operational expenditures (OPEX), which are estimated to represent a third of total costs in offshore wind energy projects [20]. The environmental conditions have an impact on the structures' fatigue and on the site accessibility, making the operational expenditure an additional challenge compared to onshore wind energy.

Throughout the different phases of an offshore wind farm, different types of operations are performed. During the installation phases, several types of vessels are needed for cable laying, foundation installation (which can require piling or drilling), and turbine installation (requiring heavy-lift operations). For maintenance operations, crew transfers are regularly needed and require either a helicopter transfer with landing on the nacelle deck, or crew transfer vessels. Occasional major component repair require heavy-lift means (e.g. blade replacement, drive-train replacement). Specific maintenance vessels have been developed in the last decades for the maintenance of bottom-fixed wind turbines, and most notably heavy-lift vessels capable of major component replacement on large turbines up to 15MW [70]: dynamic positioning vessels and jack-up vessels. The latter position themselves above the waterline to eliminate sea-state-induced movements and can perform heavy-lift operations in water depths of up to 60 meters. Each operation is different and is limited by the environmental conditions, most importantly by waves, wind and current, to be performed safely.

During the wind farm operational phase, different maintenance strategies are implemented [71, 72]. Preventive maintenance relies on the monitoring of the components' ageing to prevent failures and extend their lifetime through planned maintenance campaigns, either conditioned on the state of the system, or based on regular operations.

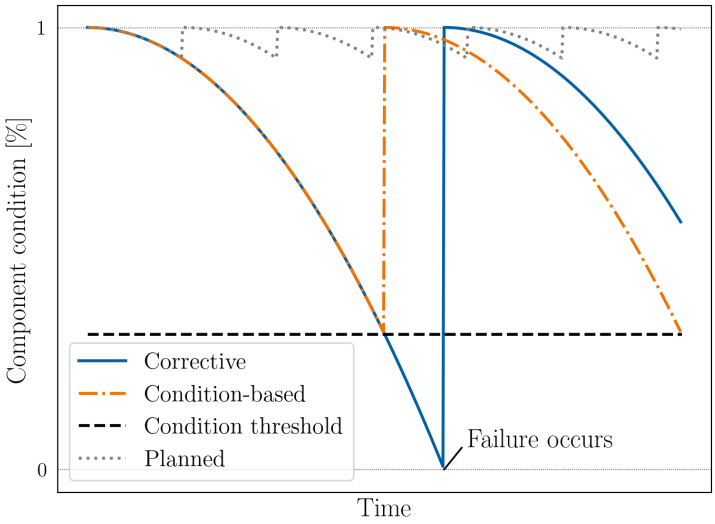


Figure I.1 – Schematic of the different maintenance strategies for offshore wind energy components.

These strategies are illustrates in Figure I.1.

During a planned maintenance operation, the production is halted only during the operation. On the other hand, when unplanned failure occurs, a corrective maintenance is required. In this case, the wind turbine production is interrupted until maintenance is made. This downtime results in an opportunity cost that represents the energy production loss. Spare parts availability, vessel and crew mobilisation and weather downtime are then the main drivers for total downtime [21].

I.1.1.1 Maintenance operations for floating offshore wind energy

Floating offshore wind energy faces significant additional challenges for maintenance operation compared to bottom-fixed offshore turbines because of the combined movements of the floater and the maintenance vessel. Dynamic cables for grid connection, mooring lines and anchors are additional challenges for the installation and are prone to fatigue and bio-fooling, hence requiring additional subsea monitoring. The floater's motion due to metocean conditions might induce supplementary loads on the structure and have an impact on the turbine's components fatigue. It is an additional limiting factor for accessibility compared to bottom-fixed wind turbines. This implies special modelling to assess the wind turbines workability under various conditions [73]. Turbine motions can also become a limitation for technicians to work efficiently due to sea sickness with a

decrease in accessibility of up to 5% estimated in [74].

Floating offshore wind farm are to be deployed in deeper seas, potentially further away from the coast, hence facing harsher environments. Jack-up vessels can't operate in water depth superior to 60 meters. Floating-to-floating operations hence require the development of dedicated maintenance solutions. For heavy maintenance operations, several strategies are studied throughout the industry.

- Tow-to-port maintenance strategies require the turbine's towing to dockside for performing the operation. This strategy applies to floater designs with limited draft and guaranteed stability during towing. It also implies complicated connection/disconnection operations and requires deep water quays. The Kincardine project installed offshore Aberdeen in Scotland had to tow its turbines to Rotterdam for major component replacements in 2023, which resulted in important downtime.
- Heavy-lift semi-submersible vessels have large ballast capacities. They can stabilize on site by lowering their centre of gravity and perform heavy lift operations in deep waters. These vessels have extremely high mobilization and chartering costs, and their low availability make them not scalable for large maintenance campaigns at the moment.
- Self-stabilized cranes are mounted on movement compensation platforms to cancel out the vessel's motions. These technologies are very expensive and relatively immature. They might not scale up for heavy lift operations.
- Self-hoisting and turbine-mounted cranes are dedicated cranes that are installed on the wind turbine or on the foundation. They can perform lift operations in the wind turbine referential hence getting rid of vessel's motions impact. These technologies are relatively immature and require research work to be developed. Their design interfaces with the wind turbine and floater designs, making industry collaboration a key development factor.

Turbine mounted cranes could become a game-changer in floating offshore wind energy maintenance market. They nonetheless face important technological challenges. The crane build-up on the foundation requires transfer from a vessel and will face operability limitation specific to the crane, vessel and foundation. The main case study of this thesis considers the crane build-up of the WindSpider solution on a Deep-C semi-submersible foundation from the refitted tanker Windsor Knutsen vessel. This case study was the subject of numerical and experimental studies in the FLOWTOM project. The operational

sequence and operational limits considered are realistic and issued from the numerical and experimental campaign.

I.1.2 Weather limited operations

Offshore operational standards issued by the Det Norske Veritas (DNV, [75]) distinguish two types of offshore operations, based on their reference duration:

$$T_R = T_{POP} + T_C \tag{I.1}$$

with T_{POP} the planned operation time equal to the duration between the last issued forecast before operation starts, and the end of the operation. T_C is a contingency time that should take into account uncertainties on the operation duration and should cover any contingency situation. A weather-limited operation verifies:

$$\begin{cases} T_{POP} \leq 72h \\ T_R \leq 96h \end{cases} \tag{I.2}$$

For unrestricted operations, the weather forecasts are considered not reliable enough to be used as such, and vessels should be able to sustain extreme weather during operations. In this case, the environmental limitations are based on extreme values. Weather-restricted operations on the other hand are to be executed using weather forecasts. Knowing operability limits and considering the weather forecast uncertainty, one can determine weather windows to carry out the operation. Though an installation campaign of an offshore wind farm lasts several months in total, each individual operation is weather limited, and is depending on environmental conditions for weather window.

I.1.2.1 Definition of operability limits

The basic criteria for carrying out an offshore operation is that the dynamic response of the system should not exceed a pre-defined limit. This is referred to as operability limits. The general methodology to compute operability limits for an offshore operation can be found in [22]. A given operational procedure will have several identified critical events. The numerical modelling of the system's responses to all metocean conditions, either in time domain or frequency domain, should be made to identify the limiting conditions for each of these critical events. These models can be calibrated by performing a basin test of

the system. Eventually, an operability region of allowable metocean conditions is obtained and define the operability limits of the operation. For the installation of a turbine mounted crane onto a floating wind turbine, a typical critical event will be a contact between the load and the turbine with a too high relative velocity.

The operability limits are used in the planning phase for the estimation of the total operational expenditures. All the different operational sequences are compared to historical metocean conditions to compute the mean accessibility of the wind turbine.

During operational phase, the weather forecasts are compared with the operability limits to decide whether to start the operation or not. The operation can be performed if the forecast variables are within the operability region.

I.1.2.2 Operational decision-making under uncertainties

During operations, the weather forecasts should be compared with the pre-defined operability limits. The forecasts used for the operations are agreed upon between all parties (Developer, vessel operator, maritime warranty surveyor). Operational standards recommend the use of minimum two different weather forecasts from different models. The decision is to be taken considering the worst forecast throughout the operation. The weather forecasts used for decision-making have inherent uncertainties that are not necessarily detailed. To account for weather forecast uncertainty, the operability limits are degraded by a factor called the alpha-factor $0 \leq \alpha \leq 1$ that make the decision-making more conservative. The alpha-factor methodology is the state-of-the-art procedure in the industry and is detailed by the Det Norske Veritas [75]. It should ensure that the operational criteria does not exceed the operability limit of more than 50% with a probability of 10^{-4}. More details on the computation procedure of the alpha factor can be found in [8]. Given an operational criteria OP_{LIM} defined from numerical modelling, the forecast operation criterion OP_{WF} is obtained as:

$$OP_{WF} = \alpha \, OP_{LIM}. \tag{I.3}$$

In Figure I.2, an example of weather window estimation from a point forecast is given. The operability limit plotted as a black line is degraded by the alpha-factor, resulting in a more conservative limit in dashed black. The weather window forecast is obtained by looking from time t to $t + T_{POP}$ if the predicted weather is below limits.

In practice, alpha factors are taken in the tabulated values proposed by [75]. These

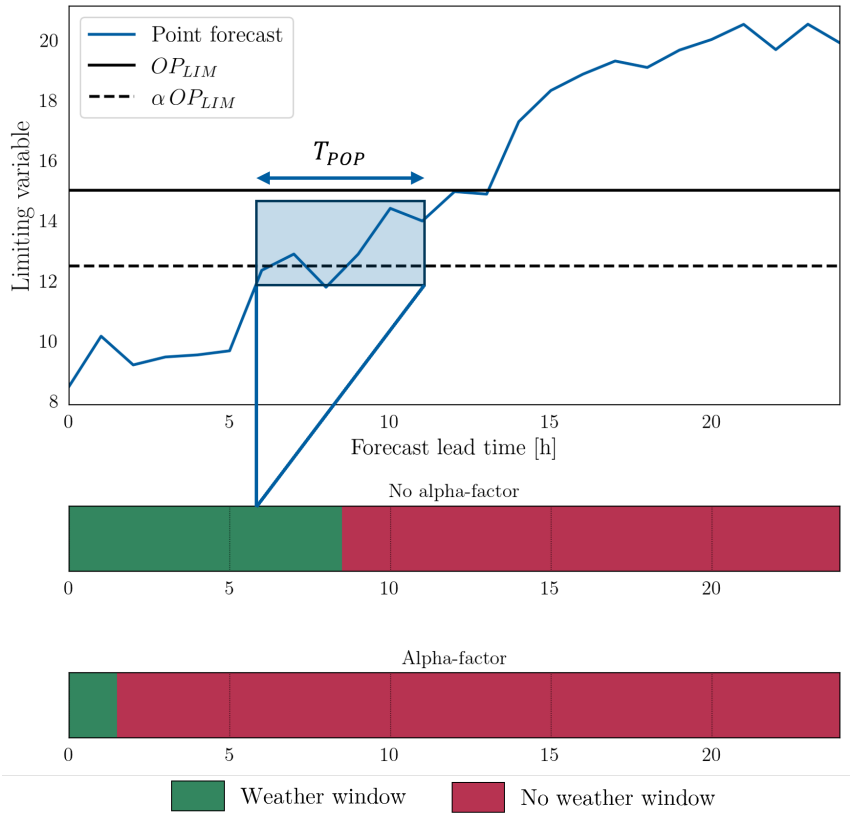


Figure I.2 – Illustration of weather window forecasting from point forecast. The given operability limit is degraded by an alpha-factor, making the decision-making more conservative.

values are calibrated for North Sea conditions and do not take into account the specificities of the weather forecast used, of the operation or of the geographical specificities. [75] describe alpha factors depending on the operation duration, the number of weather forecasts used, the presence of environmental monitoring on site, the presence of a meteorologist onboard and the criticality of the operation. It is important to note that the use of such factors is needed because of the lack of reliable uncertainty estimation in current weather forecasts. It is specified that subject to reliability, ensemble weather forecasts could be used to replace the use of alpha factors.

In [18], a novel methodology for alpha factor computation is proposed to take into account the system's response to metocean forecast uncertainty. Illustrated on a single blade installation on a bottom-fixed wind turbine, this methodology shows the importance of propagating the weather forecast uncertainty onto the system, and finally in the alpha factor to correctly account for uncertainties. It however requires the reliable probabilistic forecast of all limiting environmental variables. The prediction of vessel's response to uncertain metocean forecasts is used in [5] as a direct input for decision-making. The exceedance probability based on probabilistic forecasts is directly used to determine weather windows. It is compared to alpha-factor-based decision-making and found to be less conservative, hence increasing the accessibility. Similar approaches are implemented in [4] with a data-driven surrogate model of the vessel for a crew transfer operation. The vessel's response to a statistical probabilistic forecast of wave peak period and significant wave height is proposed as an alternative decision-making support.

The improvement of metocean forecast models is identified as a key factor in O&M cost reduction in floating offshore wind energy [23]. The development of dedicated maintenance solution, the specificities of each floater technology and the higher sensitivity to metocean conditions advocate for a better uncertainty characterization of weather forecasts for offshore operations. The development of such methods then paves the way for a more comprehensive decision-making under uncertainty for offshore wind energy operations.

I.2 Forecasting of metocean variables

Metocean forecasts ar not only essential for the safe execution of offshore operations as described in Section I.1, but they also impact the development of wind farms, their market participation, the grid management etc. [64]. In this section, we provide a literature review and conceptual introduction to metocean forecast models with a focus on offshore wind

energy applications.

A forecast model Ψ predicts a variable or a set of variables $\hat{\boldsymbol{y}} \in \mathbb{R}^N$ from an initial time $t = 1 \dots T$ and for a period of time in the future. The time t at which the forecast is issued is called the forecast issue time. The forecast model has an internal computation time step for discretization called the model time step (typically 1 minute). The forecast is not expressed for every model time step but has a time step of its own called the forecast time step (typically 1 hour). The maximum predicted time K is called the forecast horizon, and the difference between an instant in the future and the forecast issue time is called the lead time $k = 1 \dots K$. The frequency at which the model outputs a new forecast is called the model frequency (typically 6 hours).

In all that follows, we consider a time series of predictions $\hat{\boldsymbol{y}}_{t:t+K}$ between forecast issue time and forecast horizon, from an ensemble of input data at time t $\boldsymbol{X}_t$. The symbol $\boldsymbol{X}$ describes the input data, and $\boldsymbol{y}$ the output data. The predictions are depicted with a hat superscript, and the observed values without superscript. The predictions can be compared with observations $\boldsymbol{y}_{t:t+K}$. The term model refers to the operator Ψ that maps from the input space to the forecast space. This can refer to any kind of model, numerical, statistical, or neural models.

$$\hat{\boldsymbol{y}}_{t:t+K} = \Psi(\boldsymbol{X}_t). \tag{I.4}$$

I.2.1 Numerical Weather Prediction

Numerical weather prediction describes the principle of predicting the future state of the atmosphere by solving the atmospheric circulation equations on a discretized grid in space and time. This theoretical approach was originally proposed by Abbe and Bjerknes in the early 1900s [24, 25], and first put into practice by Lewis F. Richardson in 1922 [26, 76]. V. Bjerknes described numerical weather prediction as an initial value problem with a two-step process: a diagnostic step at which the initial state of the atmosphere is estimated from measurement data, and a prognostic step during which the atmosphere's changes with time is estimated by solving the governing differential equations. The first results obtained by Lewis F. Richardson were unrealistic and he declared: "*Perhaps some day in the dim future it will be possible to advance the computations faster than the weather advances... But that is a dream.*". He estimated that 64,000 people would be needed to scan the atmosphere, each person taking care of a grid point. The research field had to wait for the development of the first computers to provide evidence of the possibility of

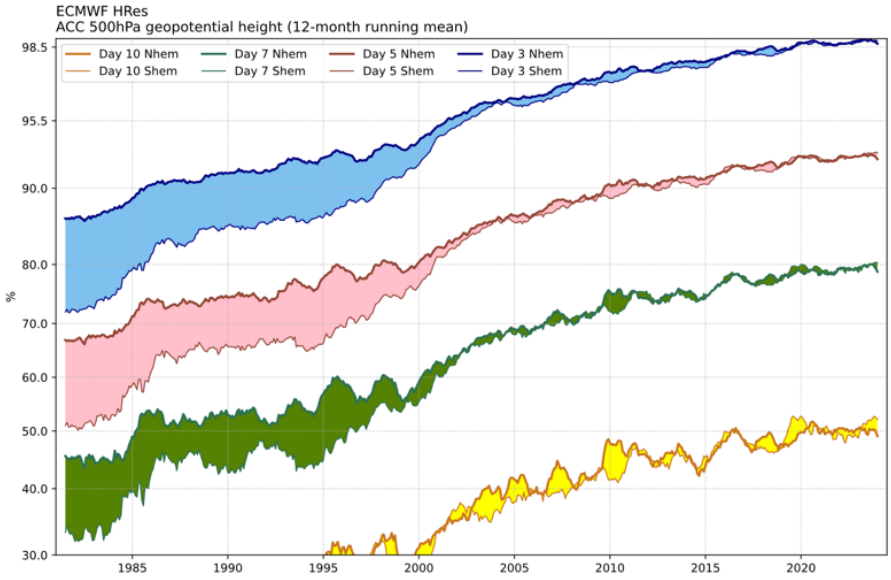


Figure I.3 – Evolution of forecast skill in predicting the 500hPa pressure level for the Northern hemisphere (NH) and Southern hemisphere (SH). The gap reduction after 1999 corresponds to the variational assimilation of satellite data that compensates the lack of measurement data in the Southern hemisphere. Reproduced with permission from [9], data updated by the ECMWF.

creating a skilful forecast using numerical weather prediction.

Since then, enormous advances in computational power and numerical integration schemes have been made. Furthermore, the quantity of observations of the atmosphere has drastically increased, leading to a more precise estimation of the initial state of the atmosphere. The forecast skill of numerical weather prediction is commonly evaluated using the height of the 500 hPa pressure level, and it is estimated that a day of skilful forecast is gained every decade, as illustrated in Figure I.3 [9]. Nowadays, numerical weather predictions routinely solve the circulation equations on grids of kilometric scale totalling $\approx 10^8$ grid points, for weeks ahead.

Despite great reduction in resolution, the most refined numerical weather prediction models have a grid size of around 1km and are often regional models. Global models such as the Integrated Forecasting System (IFS) from the European Centre for Medium-range Weather Forecast (ECMWF) have grid sizes of around 10km. Such a resolution does not permit the full solving of atmospheric circulation, and some phenomena have to be parametrized such as convection, diffusion, or cloud dynamics. An illustration of parametrized phenomena from [9] is given in Figure I.4. These parametrizations are im-

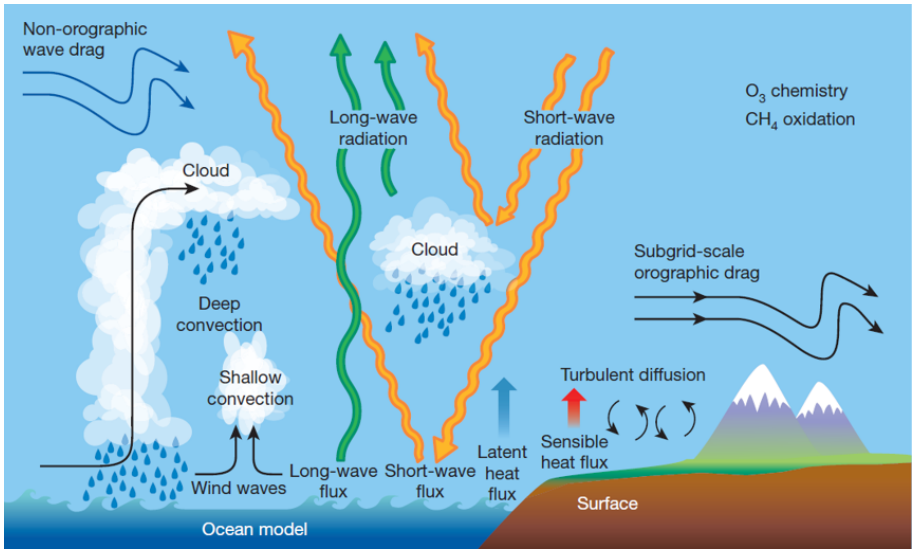


Figure I.4 – Schematic of the main physical and chemical sub-grid processes that are parametrized in numerical weather prediction models. Reproduced with permission from [9]

portant sources of uncertainty and are an active research field [9].

Numerical modelling is a dynamic system framework, where the predictions are obtained by solving a system of differential equations from an estimated initial state of the system $\boldsymbol{X}_t$. The system's evolution can be written as a dynamical system equation

$$\frac{d\boldsymbol{X}}{dt} = \Phi(\boldsymbol{X}) \tag{I.5}$$

where $\boldsymbol{X}$ is a vector in the phase space $\mathcal{P}$ that fully characterizes the state of the system. The operator $\Phi : \mathcal{P} \to \mathcal{P}$ defines the dynamics of the system. When integrating over time, the flow operator $\psi : \mathbb{R} \times \mathcal{P} \to \mathcal{P}$ is obtained such that

$$\psi(k, \boldsymbol{X}_t) = \boldsymbol{X}_{t+k}. \tag{I.6}$$

For numerical weather predictions, the system state at time $t + k$, $\boldsymbol{X}_{t+k}$ contains various variables (wind, pressure, temperature, etc.) at every point of the space-time grid. The system's state is subject to physical laws (Navier-Stokes, thermodynamics etc.) that constitute the model Φ and flow ψ through time integration. It explicitly describes the expected evolution of the system from an initial state. At each forecast issue time t, an initial state of the atmosphere $\boldsymbol{X}_t$ is estimated, and propagated through lead times $k = 1 \dots K$ in the forecast window with the flow operator. The prediction $\hat{\boldsymbol{y}}_{t+k}$, $k = 1 \dots K$ is

extracted from the full state of the system through an observation operator $\mathcal{H} : \mathcal{P} \to \mathbb{R}^N$ such that

$$\hat{\boldsymbol{y}}_{t+k} = \mathcal{H}(\psi(k, \boldsymbol{X}_t)) = \Psi(\boldsymbol{X}_t) \tag{I.7}$$

We refer to these models as numerical models or physics-based models.

The chaotic nature of the global weather system implies that small errors in the initial state or in the model parametrization can grow rapidly and lead to diverging situations [77]. Some weather situations will be stable for a longer time than others, and this is true for both local and synoptic scales. Researchers developed in the 1990s the ensemble method that runs several scenarios in parallel to estimate the forecast spread and hence the predictability of the weather situation [65]. The different scenarios are called members of the ensemble, and are perturbed in initial conditions, model parameters or both. An ensemble prediction from the Météo France model AROME with 16 members is shown in Figure I.5 to illustrate the process. The deterministic forecast is then a weighted combination of the ensemble members. In Figure I.5, it can be observed that the spread between ensemble members grows with the forecast lead-time (i.e. the time ahead of forecast issue time), showing the uncertainty increase due to the chaotic nature of the system. Ensemble prediction has become one of the main research areas in the last decades and is still a major scientific challenge [9].

The very high dimension of the system and the large amount of heterogeneous noisy measurement data makes the system's state estimation an important challenge. Data assimilation refers to the process of correcting a system's state estimation with measurement data. The finding of the best state of the system (the analysis state) given observations, prior information and forecast model is an inversion problem that is to be optimized in space and time [78]. Around a forecast time step, observations within the assimilation window are assimilated to correct the state of the system. Operational 4-dimensional variational assimilation systems (4D-Var) appeared in the late 1990s and led to great improvements in global weather forecasting [79].

The high-resolution model AROME (Applications de la Recherche à l'Opérationnel à MEsoéchelle) is the operational regional model of Météo France [27] and is used throughout this thesis as a reference model. It has a grid size of 1.3km and outputs hourly predictions up to 42 hours ahead. It uses a 3D-Var assimilation scheme on 3-hour assimilation windows and is initialized from the Météo France global forecasting model ARPEGE (Action de Recherche Petite Echelle Grande Echelle) [79] which features a 4D-Var assimilation

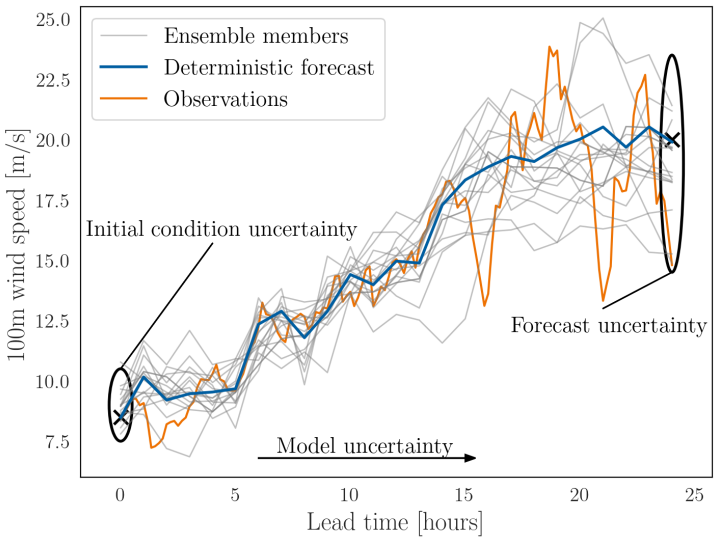


Figure I.5 – Schematic view of an ensemble forecast, illustrated from a 16-members pressure 24h ahead forecast from Météo France AROME model. The initial condition and the model parameters are perturbed to run N parallel ensemble members. Their evolution and divergence show the growing uncertainty in the forecast model.

scheme.

Numerical weather prediction is state of the art for weather forecasting and has shown remarkable progress in the last decades [9]. However, it is still prone to limitations because of its computational cost and because of the complexity of the system. Model frequencies are limited by the computational cost: for example, the Météo France AROME forecast can only be issued every 6 hours for the next 42 hours. Computational cost is an issue and implies a trade-off between model resolution, model frequency, model horizon and the number of ensemble members. Current operational ensemble models use tens of ensemble members (16 for AROME, 50 for EPS) to estimate the forecast spread in the future. Probability distributions can be approximated from these weighted trajectories, but it is still a relatively low amount of samples to estimate the uncertainty from, especially given the dimension of the system.

Eventually for offshore applications, the lack of measurement data impacts both model calibration and evaluation. The development of offshore wind energy brings additional challenges to numerical weather prediction. The need for improved metocean forecasts at different time and spatial scales at sites with low measurement availability is a great challenge [80]. In particular, the marine boundary layer interaction with the sea surface is still an active field of research. In [81], authors found similar RMSE for offshore and

onshore wind measurements but highlights stronger errors in the wind profile. As shown in [82], the offshore wind speed prediction accuracy is dependent on the height level. Lower wind levels tend to be more sensitive, highlighting the difficulty of parametrization of air/sea exchanges. This is coherent with the results from [83], which shows the difficulty of predicting wind profiles in nocturnal stable and low-level jets conditions. The validation of numerical weather prediction against measurements is made through LIDAR measurements. The bias in the horizontal wind speed increase with height and is site-dependent even for close locations as shown by [84]. To overcome these limitations, statistical and especially deep learning methods are more and more being applied to offshore wind speed forecasting, as pre-processing, post-processing or end-to-end methods.

I.2.2 Data-driven methods for metocean forecasting

In contrast with physics-based forecast models, data-driven methods do not explicitly model the weather but they use statistical methods to link a target phenomenon (say the wind speed at hub height) to an set of input explanatory variables. The development of data-driven methods is not new but has drastically accelerated with the increase in collected data and the progresses in computational capacities in the last decades. A data-driven model is parametrized by a set of parameters Ξ that are to be tuned so that the forecast errors are minimized

$$\hat{\boldsymbol{y}}_{t:t+K} = \Psi_{\Xi}(\boldsymbol{X}_t). \tag{I.8}$$

The model parameters Ξ should minimize a cost function that measures how far the predictions are from the observations

$$\min_{\Xi} \mathcal{L}(\hat{\boldsymbol{y}}_{t:t+K} | \boldsymbol{y}_{t:t+K}). \tag{I.9}$$

Several categories of models can be found in the literature.

Firstly, data-driven methods can be used for the post-processing of numerical weather prediction. In this case the model Ψ_{Ξ} is trained to apply a correction to this forecast. The input $\boldsymbol{X}_t$ then contains a set of explanatory variables used for the forecast correction. The output $\hat{\boldsymbol{y}}_{t:t+K}$ can be set as the residuals of the numerical model, i.e. the forecasts errors, or as the observations. The use of residuals can be beneficial for forecasts probabilistic post-processing as discussed in [**xie2015normality**], since it gets rid of the point-forecast.

Parts of the data-driven models can be used for data pre-processing. Using a set of

heterogeneous input variables usually implies pre-processing. Dimension reduction is a widely used tool for large input data, and normalization is an important step for machine learning based models.

Eventually, recent developments in artificial intelligences for geosciences pave the way for end-to-end data-driven models, as described in [66]. This means that the data-driven model fully replaces the numerical model by learning to make forecasts only from observations. Recent quantum leaps from various teams around the world opened up tremendous perspectives in the last years [10–13].

A literature review on data-driven forecast models for wind energy applications is given in Section II.4.

I.2.3 Hybrid methods

One of the main drawbacks of machine learning models for earth system application is their "black box" nature, that complicates the interpretability and the implementation of physical constraints. They can be very performant in interpolation, but offers no guarantee for extrapolation. It implies that the performances of the model on weather situations that were not present in the training dataset can not be controlled. These are natural features of numerical models, and recent research work aim at incorporating the knowledge about the dynamical system from decades of numerical model development into the deep learning architecture. These are known as hybrid methods [15, 85].

[86] advocate for a joint research effort between machine learning models and data assimilation. Indeed, data assimilation methods result from decades of research on how to handle noisy, indirect and sparse data into an inverse model problem and this experience should be used in machine learning models for earth systems, which must deal with the same input data. The author notes that data assimilation and machine learning can be gathered under the same Bayesian framework, highlighting the similarities between 4D-Var data assimilation and recurrent neural networks. Recent work on the development of learnable data assimilation schemes for earth systems inverse problems have been carried out by [87] and applied to satellite altimetry [88], sea surface current and sea surface wind retrieval from passive underwater acoustics [14, 89]. Deep learning can furthermore be used for model identification based on observation. [90] use convolutional neural networks to learn hidden governing laws and hidden variables for sea surface height prediction in complement of a Quasi-Geostrophic advection model.

I.2.4 Joint wind and wave forecasting

I.2.4.1 Sea state forecasting

The deformation of the sea surface forms gravity waves, that can be induced by the wind, seismic activity, or the attraction of the Moon and the Sun. Wind induced waves' periods range from 1s to 30s and extract their energy from the wind blowing over the sea surface. For simplicity, In all that follows, we will note "waves" the wind-induced waves. The energy transferred by the wind to the wave is partially lost to surface currents and dissipated through wave breaking. The wave field then grows in the time scale of hours to days under wind forcing. The length of wave growth under wind forcing is called the fetch, and the longer the fetch, the higher the wave energy [28]. Waves locally generated by the wind are called wind-waves and have relatively low periods and wave height. Waves resulting of a long fetch and dissipating their energy by propagating beyond the wind blowing area are called swell. They can be associated with longer periods and higher wave heights. An illustration of a bi-modal wave spectrum is given in Figure I.6a.

The characteristics of the sea surface at a certain location is called the sea state and is characterized by the power spectrum density of the sea surface elevation. It is mainly composed of a wind-sea generated by local winds and swells from remote winds. The sea state is an energy spectrum $E(f, \theta)$ in frequency f and direction θ, as illustrated in Figure I.6b. The usual variables used to describe the sea state are the significant wave height and the wave peak-period. The significant wave height H_S is defined as the mean wave height of the 1/3 highest waves. In the frequency domain, it can be retrieved as H_{m_0} from the 0^{th} moment of the energy spectrum

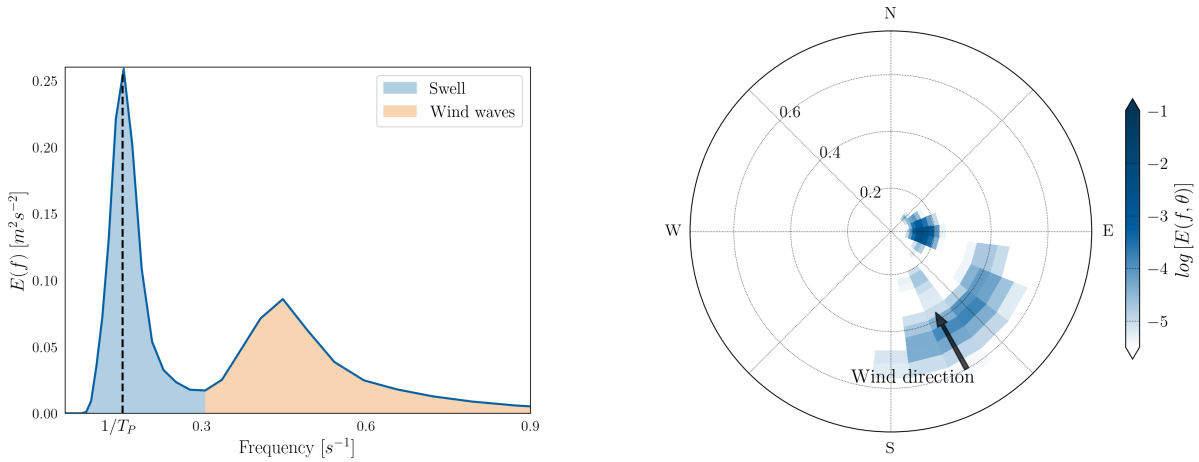
$$H_S = H_{m_0} = 4\sqrt{\int E(f)df} \tag{I.10}$$

with $E(f)$ the univariate power spectrum in frequency obtained from the integral of the power spectrum in directions

$$E(f) = \int_0^{2\pi} E(f, \theta)d\theta. \tag{I.11}$$

It is defined as such because most of the measurement devices used for sea state are not sensitive to wave direction [91], so it is usual to work with the frequency power spectrum $E(f)$. The wave peak period T_P is defined as the period of the most energetic peak in the

power spectrum density, as illustrated in Figure I.6a. The mean period T_m is defined as the mean period of individual waves. Eventually the maximum wave height H_{max} is defined as the maximum observed wave height in a given period of 3 hours. All these variables can be expressed for the total sea, or for individual systems (wind sea, swell) by clustering the total power spectrum. Most of the operational criteria for offshore operations are based on these variables, in link with wave direction and wind direction. In particular, cross-seas with non-aligned wind and swell components can be problematic for vessels' operations.



(a) 1D spectrum

(b) Directional spectrum

Figure I.6 – Illustration of a wave spectrum at the Planier island location off the coast of Marseille. The spectrum show a bi-modal distribution with a low frequency swell component and a high frequency locally induced component. The wind waves are less energetic and more dispersed than the swell, showing their local generation.

The forecasting of sea states dates back to the 1950s, for the need of military operations, and in conjunction with the development of numerical weather prediction. It is based on an evolutionary equation for the 2-dimensional power spectrum density. At a given location $\boldsymbol{x}$ and time t, the energy balance equation is

$$\frac{\partial E(\boldsymbol{x}, t, f, \theta)}{\partial t} = S(\boldsymbol{x}, t, f, \theta) \tag{I.12}$$

where $S(\boldsymbol{x}, t, f, \theta)$ is a source term including the wave generation (wind induced), the wave dissipation (breaking, wave-ice interaction, bottom friction, bathymetry breaking) and the non-linear wave interactions [92, 93]. The parametrization of the source term S and the solving of the energy balance equation in space and time on a computational grid is at the core of numerical wave modelling. The operational sea state modelling in real time

is then referred to as sea state forecasting. Seemingly to numerical weather prediction, it depends on the estimation of an initial state of the system from measurement data. Wave buoys, drifting buoys and satellite measurements are the main sources of in-situ data but are still relatively scarce and inequitably installed globally [91]. The wind forcing for wave generation makes it dependent from numerical weather predictions as input. In this study we use the Wavewatch-III model (WW3) which is a 3rd generation model run on an unstructured grid. In the Gulf of Lion a configuration with AROME wind forcing is used.

The high computational cost associated with numerical wave forecasting motivates research in data-driven models [69]. Deep learning models can be developed and trained as surrogate models for wave forecasting [30], for model post-processing [31] or for statistical downscaling [29, 32, 94]. For offshore wind energy operations, dedicated data-driven forecast models are developed to improve the forecast quality [95], or to emulate uncertainty estimation [4].

I.2.4.2 Topics in joint wind and wave forecasting

The complex relationship between sea state and wind speed time series requires special care. Despite the wind forcing used as input of numerical wave models, the forecast and hindcast of wind and wave time series is not performed jointly. Special care is then needed to construct joint wind and wave times series which are essential for availability analysis for offshore operations [34–36]. The difficulty of estimating the joint distribution of metocean parameters increases with the number of parameters to take into account. Furthermore, the complex relationship between parameters such as the highly non-linear relationship between significant wave height and wave period that is physically limited by the wave breaking phenomenon requires the development of specific statistical tools [96]. A review of statistical methods for the joint wind and wave time series modelling can be found in [37]. Copulas are used to model the dependency structure between wave height and wind speed in [34] to generate long-term time series for a cable installation campaign in an offshore wind farm. The joint generation of wind and wave time series is shown to be essential for the duration estimation of the campaign, as the independent generation leads to unrealistic time series. Copulas are the most common model for joint time series generation [97, 98]. Markov-switching models can alternatively be fitted to data to generate stochastic joint time series [99, 100]. Joint estimation of wind and wave is also essential for offshore wind turbines and foundations design, because of the occurrence of joint extreme events [101,

102]

In addition to the complex correlation structure between wind and wave time series, the uncertainty associated with their forecast is also to be taken into account for offshore operations scheduling. As describe in Section I.1, the forecast uncertainty is still considered independently for the limiting parameters, and applied as in a pseudo-deterministic way using the alpha-factor and leading to conservatism in decision-making. The development of multivariate probabilistic metocean forecast recently gained interest as a path to maintenance operations optimization. A data-driven multivariate probabilistic forecast model for significant wave height and wave period was developed by [4] and expanded in [64], they modelled the dependency between uncertainties during the scenario generation with empirical copulas. They show that the addition of temporal cross-dependencies do not improve multivariate scores. To the best of our knowledge, the only study to deal with the joint probabilistic forecast of wind and wave is presented in [38] and [8]. An adaptative neuro-fuzzy inference system (ANFIS) is applied to the multi-step ahead forecasting of wave height, wave period and surface wind speed. The relevance of multivariate probabilistic metocean forecasting for offshore operations planning and execution and the relative immaturity of forecast models for this application motivated the work presented in this study.

I.3 Offshore in-situ data

The acquisition of offshore in-situ data is identified as one of the main research needs for offshore wind energy development in [103]. Offshore measurements are scarce and challenging but are essential for the development and calibration of offshore forecasting models. An ideal offshore measurement platform would monitor wind speed at different levels through the rotor swiping area, wave parameters, sea surface temperature, surface atmosphere parameters (pressure, humidity, temperature), and current speed. Additionally, biodiversity monitoring equipment should be installed to assess the impact of offshore wind development.

For wind speed, anemometers can be deployed on offshore buoys, offshore meteorological masts or wind turbines structures. Buoys-mounted anemometers have the advantage of measuring wind speed together with sea state, but at a limited height. Turbine-mounted anemometers are essential for wind farm operations but are influenced by the structure and are not available prior to the installation. Offshore meteorological masts permit the

Figure I.7 – The Planier LIDAR operated by France Energies Marines installed on the Planier island for the recording of long-term time series of offshore wind profiles.

long-term monitoring of a variety of parameters but are expensive infrastructure and may not reach rotor heights. The recent development of optical remote sensing instruments made possible the measurement of wind speed up to hundreds of meters. LiDAR (Light Detection And Ranging) are optical devices that measure the wind speed indirectly through the doppler shift of laser light backscattered by aerosols carried by the wind [104]. LiDARs can be installed on offshore platforms, on turbine hubs, or on buoys. Floating LiDARs offer a great potential cost reduction for offshore wind speed measurement compared to meteorological masts [105], though care must be given to compensate the buoy's movement. In a recent study, [106] highlight the need for motion compensation algorithms for error reduction in the turbulence measurements from a floating LiDAR by comparing the measurements from a hexapod-mounted mobile LiDAR and a ground-based LiDAR.

Satellite-based remote sensing measurements can be used for wind speed and wave measurements. The sea surface roughness measurements from satellite-based Synthetic Aperture Radar (SAR) can be used to estimate the surface wind speed with very high spatial resolution but low temporal resolution. In combination with deep learning models, it can be used to downscale numerical weather prediction [107, 108], but care must be given during rain events [109]. Significant wave height can be measured via altimeters, again with low temporal resolution [91]. Overall, satellite data are formidable sources of offshore measurements, but require research work to be assimilated in forecast models. No satellite data will be used in this work.

In the North Sea and Baltic Sea, the FINO platform provide long-term measurements, and are widely used in the literature for weather forecasts development and evaluation [5,

33, 95, 110, 111]. In the French territorial waters, no such installation is in place. The main sources of data come from buoy measurements, island-based meteorological stations and coastal meteorological stations. Large data acquisition campaigns have been sporadically organized, such as the HyMeX campaign across Gulf of Lion in 2013 [39] mainly focusing on sea-atmosphere coupling and heavy precipitation events [112], but they do not provide long-term time series. The Fécamp meteorological mast, previously private-owned for the development of the Fécamp wind farm, has been acquired by France Energies Marines to serve as a research platform dedicated to marine renewable energy in the French North Sea. In this study, a LIDAR deployed and operated by France Energies Marines on the Planier island off the coast of Marseille, France (Figure I.7) is used as a target for offshore wind forecasting. Such infrastructures are key in the development of offshore wind energy.

I.4 Characteristics of the study area

The main study area of this work is the Gulf of Lion, situated in the North-Western Mediterranean Sea. Being relatively shallow for the Mediterranean Sea with depth of less than 500m, and having a remarkable wind resource, it is one of the main development areas for floating offshore wind energy in France with a 4GW to 7.5GW pipeline for 2050. It also features climatic specificities that makes it a challenging area for weather modelling. In combination with rising financial stakes at sea with offshore wind energy development, the development of specific post-processing models is necessary.

The Gulf of Lion and the North-Western Mediterranean Sea are characterized by a complex coastal orography, that drives complex winds. In the Gulf of Lion, there is a strong predominance of offshore blowing winds namely the Mistral in the Eastern Gulf of Lion and the Tramontane in the Western Gulf of Lion, that are associated with a pressure dipole between the Ligurian Sea and North-Western Europe. These winds are channelled through the Rhone valley for the Mistral and the Pyrenees - Massif Central channel for the Tramontane [113]. During the passing of atmospheric lows in the North-Western Mediterranean Sea, sudden wind changes can happen between North-Western winds and South Eastern winds. The latter are less frequent than North-Western winds, and are associated with cyclonic conditions, strong winds and waves in the Gulf of Lion and heavy precipitations.

In a hydrodynamic perspective, the strong wind forcing induce important up and down-welling phenomenon along the coastal bathymetry, that creates sudden sea surface

temperature changes at the local scale [114]. This is shown to have a strong impact on the wind resource in [115], but are hardly modelled in the numerical weather modelling. Strong changing winds and short fetches create complicated cross-seas with low period waves and potentially high significant wave heights. These seas can raise very quickly because of sudden wind ramps and be dangerous for offshore operations. [116] showed that high significant wave heights are largely underestimated by numerical models.

For all these reasons, the Gulf of Lion appears like an interesting study area, with few in-situ measurements and high economical stakes. The complex orography, hydrodynamics and climate make it a challenge for numerical modelling, and data-driven models could be used to post-process forecasts for specific applications, provided that a sufficient amount of in-situ data is gathered.

Chapter II

Mathematical and methodological background

Preamble

This chapter describes the mathematical and methodological background used in the next chapters. The mathematical background for deep-learning-based models is given in Section II.1. A general introduction is given on deep learning models, and a more detailed description of the concepts used in Chapter III, Chapter IV and Chapter V is given.

We give background on probabilistic forecasting, a very active research field in the weather forecast community with important applications for offshore wind energy, and we describe the mathematical tools for parametric and non-parametric methods for probabilistic posterior description. The generation of scenarios from probabilistic forecasts is then described as an important feature for decision-making problems in offshore wind energy. The mathematical functions for evaluating the forecasts' quality is described in Section II.3 for deterministic and probabilistic forecast, and a focus on metocean forecasts' value for offshore operations is given.

II.1 Basics of machine learning and deep learning

In this section, we give a theoretical background of the concepts used in the following chapters. For a thorough description of deep learning architectures and tools, the interested reader is directed to [42]. A general introduction covering the governing concepts of deep learning and the main hyper-parameters and training procedures is given in Section II.1.1. Then, the main architectures used in this study are described in details, namely Gaussian Mixture Models for unsupervised classification, Convolutional Neural Networks for large data processing in weather applications, and Normalizing Flows as a generative architecture for probabilistic forecasting.

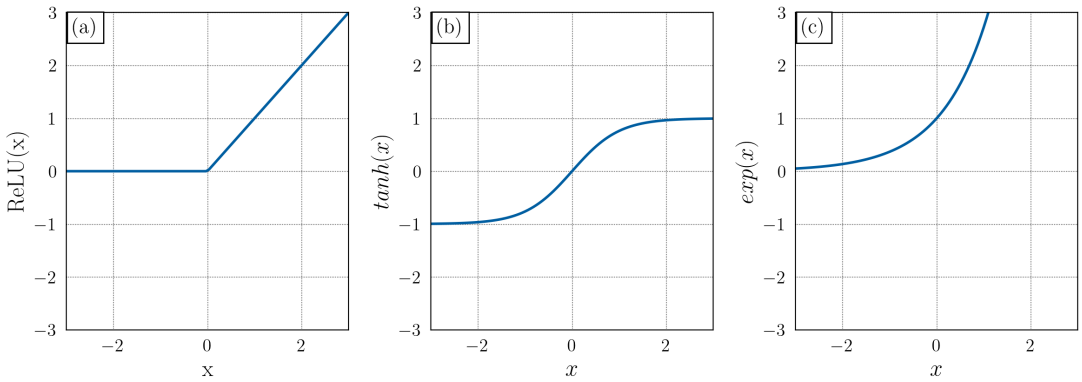


Figure II.1 – The ReLU (a), hyperbolic tangent (b) and exponential (c) activation functions are used to introduce non-linearities in neural networks.

II.1.1 General introduction

Machine learning is a sub-category of the larger data science field. It refers to the development of models f that learn from data to perform a certain task, without being explicitly parametrized. The theory of machine learning is described by [40] in these terms : «A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks T, as measured by P, improves with experience E ». A machine learning model is then associated with a specific task, and a dedicated cost function that is to be minimized by training on data.

Deep learning models are a subset of machine learning models that use networks of artificial neurons to approximate complex relationships. They are shown to be universal approximators given a sufficient number of neurons [41]. An artificial neuron is a function inspired by the functioning of brain cells, that applies a non linear function g to an input x

$$g(x) = \sigma(wx + b) \tag{II.1}$$

where w and b are trainable parameter, and σ is a non-linear function called an activation function. Common activation functions are Rectified Linear Unit (ReLU), hyperbolic tangent, exponential function, sigmoid function etc. Activation functions are used to introduce non-linearities in the neuron function. In certain cases, specific activation functions can be used to force properties on the output: exponential activation for positive output, hyperbolic tangent for output in $[-1, 1]$, sigmoid for binary classification. The three activations used in this study are plotted in Figure II.1.

Interconnecting neurons together yields to the construction of artificial neural net-

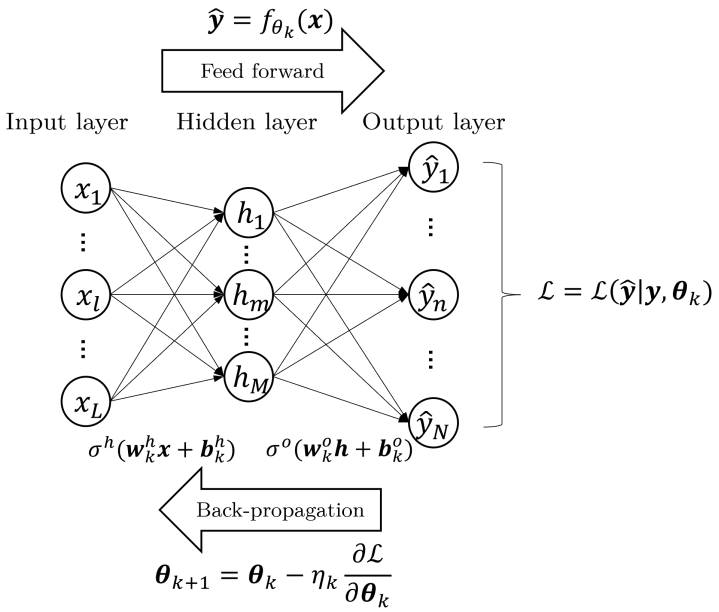


Figure II.2 – Schematic of the functioning of a basic MLP. The input is forwarded through layers of neurons, and the loss function is back-propagated to update the parameters of the model.

works. A layer of neurons connected together is called a fully connected layer. The stacking of fully connected layers is called a Multi-Layer Perceptron, or Feed Forward Neural Network. A schematic of a 1-hidden layer Multi-Layer Perceptron is given in Figure II.2.

In a Multi-Layer Perceptron, the hidden neuron h_m is connected to all the input $x_l,\ l = 1 \dots L$ such that

$$h_m = \sigma_m^h \left(\sum_{l=1}^{L} w_{ml}^h x_l + b_{ml}^h \right) \tag{II.2}$$

and the output $\hat{y}_n,\ n = 1 \dots N$ is obtained from the hidden neurons $h_m,\ m = 1 \dots M$

$$\hat{y}_n = \sigma_n^o \left(\sum_{m=1}^{M} w_{nm}^o h_m + b_{nm}^o \right) = \sigma_n^o \left[\sum_{m=1}^{M} w_{nm}^o \sigma_m^h \left(\sum_{l=1}^{L} v_{ml} x_l + b_{ml}^h \right) + b_{nm}^o \right] \tag{II.3}$$

with σ_n^o the output layer activation function and σ_n^h the hidden layer activation function. The parameters $\boldsymbol{\theta}$ of this simplified model are then the weights and biases $\boldsymbol{\theta} = \left\{ \boldsymbol{w}^h; \boldsymbol{b}^h; \boldsymbol{w}^o; \boldsymbol{b}^o \right\}$ such that $\hat{\boldsymbol{y}} = f_{\boldsymbol{\theta}}(\boldsymbol{x})$. The output $\hat{\boldsymbol{y}}$ is then obtained by forwarding the input $\boldsymbol{x}$ through layers of neurons, with sufficient flexibility to approximate complex relationships.

Now the success of such architecture is due to the rise of back-propagation algorithms.

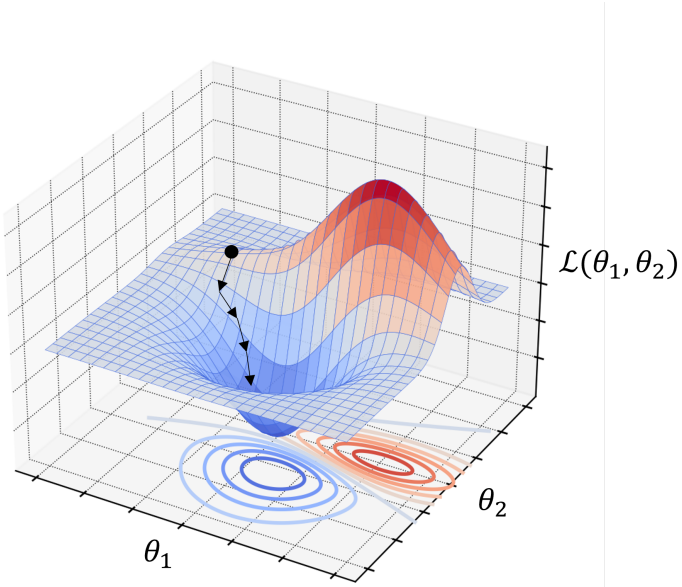


Figure II.3 – Illustration of a gradient descent to find the minimum of a loss function.

Back-propagation refers the to process of computing error gradients with respect to the weights of the model. Given a loss function $\mathcal{L}$, the weights of the model are iteratively updated using stochastic gradient descent such that for every training step k, any individual parameter θ of the model is updated as

$$\theta_{k+1} \leftarrow \theta_k - \eta_k \frac{\partial \mathcal{L}}{\partial \theta_k} \tag{II.4}$$

with η_k an hyper-parameter called the learning rate, which controls the speed at which the gradient descent is computed. High learning rate can induce instabilities while slow learning rates induce low-convergence and potentially falling into a local minimum. A common strategy is to implement a learning rate scheduler that updates the learning rate during the training, to start from a high learning rate and reduce it when getting closer to the global minimum. In this study we use an exponential scheduler. The gradient is computed for a loss function computation on a batch of input samples. The batch size can impact the convergence of the model. The computation of the loss gradient is for each parameter of the model is easily parallelized and can then be sped up by using a Graphical Processing Unit (GPU) which is performant at making several parallelized small computations.

The training of a deep neural network using stochastic gradient descent requires a

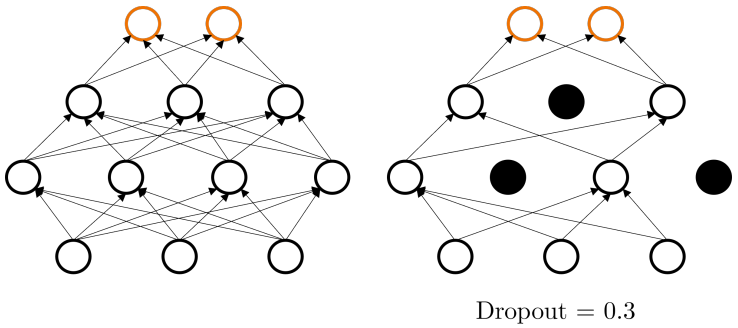


Figure II.4 – Illustration of dropout for a Multi-Layer Perceptron. During inference, with a dropout rate of 0.3, 30% of the neurons are deactivated.

monitoring for stopping the training. Otherwise, the model can over-fit the training data perfectly in theory. The validation data is then used after each computation on the full training dataset (an epoch). The loss function is then evaluated on the validation dataset without updating the model weights. The training is to be stopped when the validation loss reaches a minimum, which indicated the instant when the improvement on the training dataset does not generalized to unseen data and hence the beginning of over-fitting.

Eventually important concepts for gaining robustness for deep learning models are the weight decay and dropout layers. The weight decay adds a regularization to the model parameters so that specific ways through the neural networks do not take to much importance. Dropout layers randomly drop a certain proportion of connections at each computational step. By adding randomness it forces the model to diversify computational paths in the model and improves the performance and generalizability (see Figure II.4). Dropout layers can be used to perform Monte-Carlo simulations and estimate the epistemic model uncertainty.

II.1.2 Clustering with machine learning

When the target is discrete and un-ordered, the machine learning model is said to perform a classification task, where input points should be assigned to discrete classes. When these classes are known in advanced, the machine learning model is said to be supervised. On the other hand, when the target data is unlabelled, the task is unsupervised. In this case, the model is to assign discrete classes to data points, and we refer to such tasks as clustering problems. An illustration of the classification and clustering problems is given in Figure II.5.

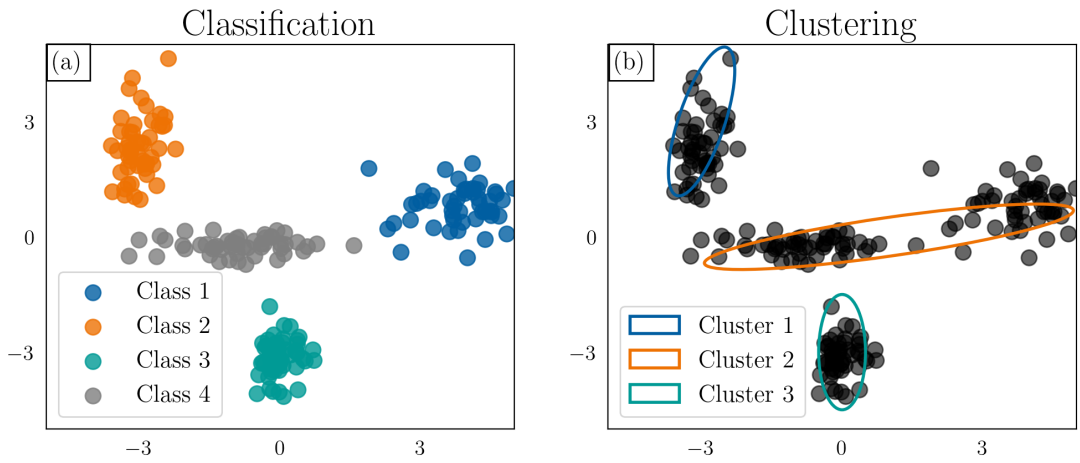


Figure II.5 – Illustration of the classification (a) and non-supervised clustering (b) problems.

The unsupervised data clustering consists in finding patterns in the multivariate data to organize the data points as clusters, according to a certain similarity measure. Partitional methods aims at describing a partition of the input data by iteratively minimizing a criterion of the quality of the partition. A thorough overview of clustering methods can be found in [117]. The most common method for unsupervised clustering is the use of the K-means algorithm that selects K points in the data, assigns all the data points to their closest centroid according to a certain similarity measure, and update the centroids of the cluster from their assign data points. By iteratively updating the K centroids, K clusters minimizing the similarity can be found. Another classic method is to base the clustering on a mixture of probability distributions, usually Gaussian distributions. The fitting of these distributions is usually performed with the Expectation-Maximization algorithm. Iteratively, an expectation step and a maximization step are computed until convergence. The expectation step computes the probability for each data point to belong to each distribution of the mixture, then the parameters of the distributions are updated to maximize the likelihood. An illustration of fitting a 4-component Gaussian Mixture Model on 2D data is given in Figure II.6. The number of cluster to fit on the data is a parameter of the model that needs to be optimized independently.

In Chapter III, we use a Gaussian Mixture Model for the unsupervised clustering of wind data to determine sample points using the clusters centroids. A mathematical description of the Gaussian Mixture Model can be found in Section III.4.5.

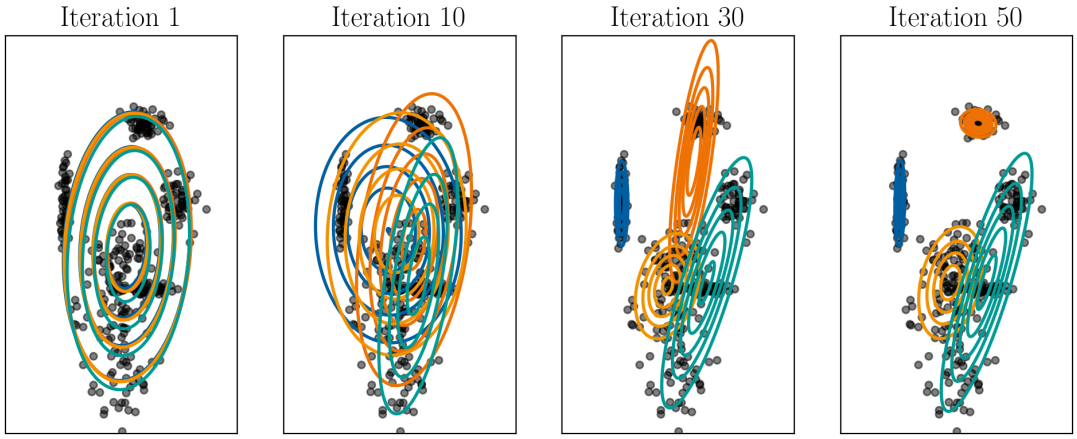


Figure II.6 – An illustration of the fitting of a 4-component Gaussian Mixture Model on 2D data with a random start using the Expectation-Maximization algorithm.

II.1.3 Convolutional neural networks

The MLP architecture presented in Section II.1.1 has a number of coefficient proportional to the dimensions of the input data and the hidden layer. For this reason, the number of coefficient can significantly increase for large input data. The development of convolutional architectures brought significant improvements in dealing with large input data, and especially for computer vision problems. Numerical weather prediction data having very high dimensions in space and time, the use of convolutional architectures is essential [94].

Convolutional architecture, often called Convolutional Neural Networks (CNN) are a specific type of architecture that uses convolutional layers to gradually decrease the dimension of the data by the use of convolution filters. These filters are matrices or kernels that apply the same weights to all the regions of the image as illustrated in Figure II.7. By doing so, the number of coefficients in the model is drastically decreased through weight sharing. A 2D convolution filter has a size (F_1, F_2), and is convolved through the input with a certain stride (S_1, S_2). The input can be padded with P zeros on its outer limits for example if the size needs to be preserved. If the initial input has size (N_1, N_2), then the size of the obtained feature map is (M_1, M_2) with

$$M_i = \frac{N_i - F_i + 2P}{S}. \tag{II.5}$$

The values h_{m_1,m_2}, $m_1 = 1 \dots M_1$, $m_2 = 1 \dots M_2$ of the feature map are obtained by

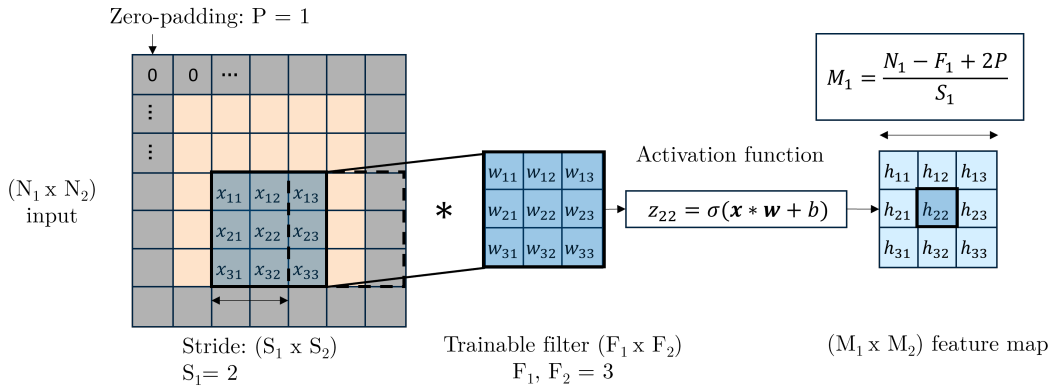


Figure II.7 – Schematic view of a 2D (3x3) convolutional filter with stride of 2 and zero-padding of 1. The trainable filter $\boldsymbol{w}$ is successively applied to subsets $\boldsymbol{x}$ of the input to obtain a feature map through a non-linear activation function σ.

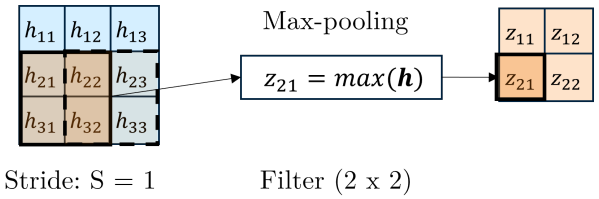


Figure II.8 – Schematic view of a 2D (2x2) max-pooling filter with stride of 1. This non-trainable filter is applied to reduce the dimension if the feature maps while helping capturing multi-scale features.

multiplying the subsets of the input in the filter window $\boldsymbol{x}_{m_1,m_2}$ by the filter weights $\boldsymbol{w}$, adding a bias b and applying an activation function σ

$$h_{m_1,m_2} = \sigma(\boldsymbol{x}_{m_1,m_2} * \boldsymbol{w} + b). \tag{II.6}$$

As illustrated in Figure II.9, convolutional layers are coupled with pooling layers, which apply a simple filter for resizing the data. Typical pooling filters use the max, min, *average* or $L2 - norm$ functions. Pooling layers are useful for reducing the dimension of the feature maps, and to allow the extraction of multi-scale features. An illustration is given in Figure II.8.

A Convolutional Neural Network is then classically composed of a succession of convolutional layers and pooling layers before applying one or several fully-connected layers to the flattened latent data to obtain the output. A simple 2 layers architecture is illustrated in Figure II.9.

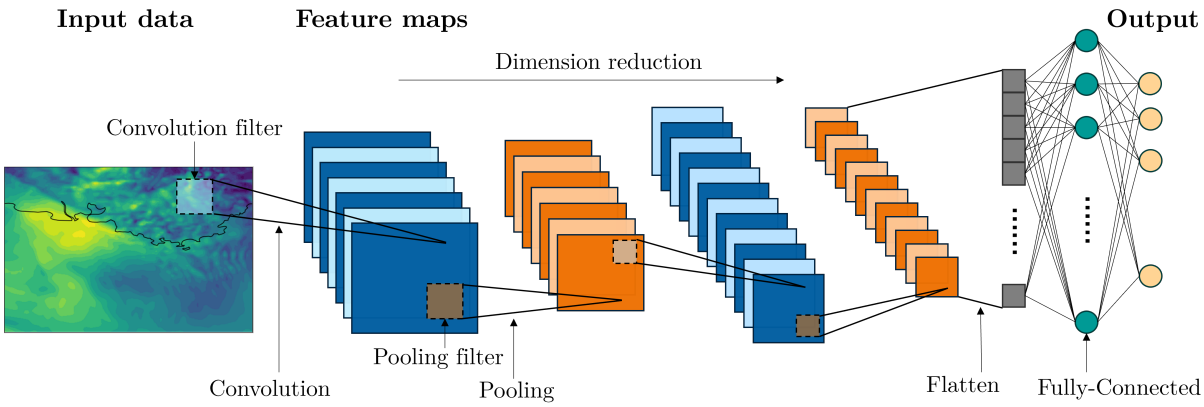


Figure II.9 – Schematic view of a 2D convolutional neural network. Successive convolution and pooling filters reduce the dimensionality of data while limiting the complexity of the model by weight sharing. Final fully-connected layers map to the output.

Convolutional architecture can be applied in 1, 2, 3 or more dimensions with the same principle. They have been used in many different applications, and are a key tool for dealing with spatio-temporal data in metocean forecasting applications as described in Section II.4.3.

II.1.4 A generative architecture: normalizing flows

The concept of generative modelling is based on the idea that the underlying manifold of a complex process can be expressed on a simpler latent space from which samples from the true data distribution can be generated. In other words, generative models learn to approximate an underlying distribution $\mathcal{Y}$ from samples $\boldsymbol{y}_1, \dots \boldsymbol{y}_K \in \mathbb{R}^N$. The use of deep learning for generative modelling recently achieved impressive results for very high dimension problems such as image generation.

A deep generative model usually relies on a generator network $\mathcal{G}_\theta : \mathbb{R}^M \to \mathbb{R}^N$ that transforms a simple latent distribution $\mathcal{Z}$ on $\mathbb{R}^M$ which can be easily sampled (e.g. multivariate Gaussian) into the target distribution $\mathcal{Y}$. Hence there is no explicit description of the underlying distribution $\mathcal{Y}$. Let $\boldsymbol{z} \sim \mathcal{Z}$ be a sample from the latent distribution $\boldsymbol{z} \in \mathbb{R}^M$. A synthetic sample of the target distribution $\hat{\boldsymbol{y}} \in \mathbb{R}^N$ can be obtained through the generator $\mathcal{G}_\theta$

$$\hat{\boldsymbol{y}} = \mathcal{G}_\theta(\boldsymbol{z}). \tag{II.7}$$

The intuition behind the development of deep generative models is that the underlying

distribution p_y can be approximated by the distribution p_θ obtained by transforming the latent distribution p_z such that

$$p_\theta(\mathcal{G}_\theta(z)) \simeq p_y(\boldsymbol{y}). \tag{II.8}$$

This requires the measure of the similarity between the underlying unknown distribution $p_y \sim \mathcal{Y}$ and the predicted distribution $p_\theta \sim \mathcal{G}_\theta(\mathcal{Z})$ which is the main challenge for deep generative models.

Generative Adversarial Networks do not rely on any estimation of the posterior probability density function and rely on the joint training of a generator that transforms samples from a latent distribution with $M << N$ into data samples, and a discriminator that is to separate generated samples from real samples. The joint training of these two components through a min-max optimization game forces the generator to learn to fool the discriminator and then to generate samples that are realistic given the data distribution in the training set. This architecture is assumption-free on the target distribution, but is known to require fine tuning to avoid instabilities during training and mode collapse when the generator $\mathcal{G}_\theta$ collapses on a limited number of modes of the target distribution. Though they have produced remarkable results and advances in generative modelling (e.g. rainfall nowcasting [118]), especially for high dimension problems, their lack of explainability and their instability are limitations.

Variational Auto-Encoders use a Bayesian framework to estimate the data distribution f_y. From the data space, an encoder and a decoder are jointly train to map the data distribution to a simple latent space and back to the real space. The encoder $\mathcal{F}_\phi$ and decoder $\mathcal{G}_\theta$ are two neural networks that generate the parameters of the conditional distributions $p_\phi(\boldsymbol{z}|\boldsymbol{y})$ and $p_\theta(\boldsymbol{x}|\boldsymbol{z})$, often assumed Gaussian. The computation of the likelihood of the marginal distribution

$$p_y(\boldsymbol{y}) = \int p_\theta(\boldsymbol{y}|\boldsymbol{z}) p_\phi(\boldsymbol{z}) d\boldsymbol{z} \tag{II.9}$$

is often intractable even for simple prior distributions. Variational Auto-Encoders hence use the Evidence Lower Bound (ELBO) of the distribution to approximately maximize the likelihood of the reconstruction. The ELBO provides both a regularization term for the latent space and a reconstruction error for the real space. By doing so, both the encoder and the decoder can be jointly trained using an approximation of the posterior likelihood.

In contrast with the two most classic architectures briefly described above, Normalizing Flows (NF) are a relatively new type of architecture that allow for an exact computa-

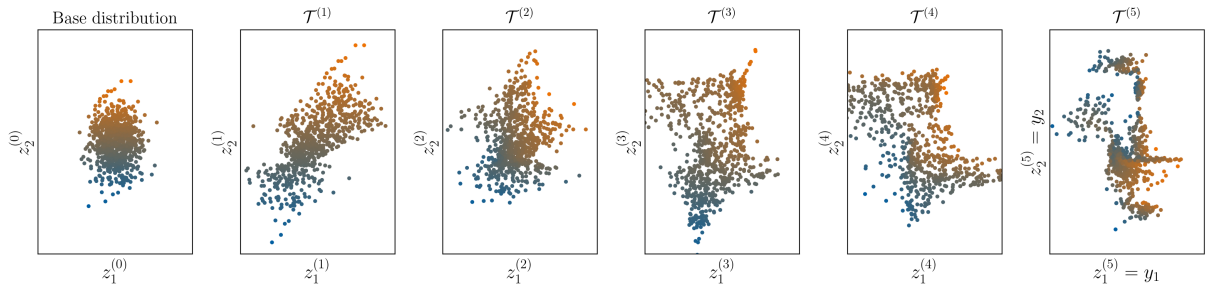


Figure II.10 – Illustration of a normalizing flow that transform a base Gaussian distribution (left) into a complex distribution (right) with 5 affine coupling layers.

tion of the posterior likelihood. It was proposed for variational inference in [119], and generalized to density estimation in [120]. A normalizing flow $\mathcal{G}_\theta$ is constructed as a diffeomorphism between the real space and a latent space of same dimension $M = N$. $\mathcal{G}_\theta$ is to be differentiable, bijective, and has a differentiable inverse $\mathcal{F}_\theta = \mathcal{G}_\theta^{-1}$. By the change of variable formula, and using the properties of the Jacobian of a diffeomorphism, the posterior density p_y can be obtained as

$$p_y(\boldsymbol{y}) = p_z(\boldsymbol{z}) \left| det \left(\frac{\partial \mathcal{F}_\theta(\boldsymbol{y})}{\partial \boldsymbol{y}} \right) \right| = p_z(\boldsymbol{z}) \left| det \left(\frac{\partial \mathcal{G}_\theta(\boldsymbol{z})}{\partial \boldsymbol{z}} \right) \right|^{-1}. \tag{II.10}$$

Using this composition rule, we can construct a transformation $\mathcal{T}_\theta$ as the composition of diffeomorphisms $\mathcal{T}_\theta^{(m)}$, $m = 1 \dots M$ such that

$$\boldsymbol{y} = \mathcal{T}_\theta(\boldsymbol{z}) = \mathcal{T}_\theta^{(1)} \circ \dots \circ \mathcal{T}_\theta^{(M)}(\boldsymbol{z}) \tag{II.11}$$

where the posterior distribution can still be explicitly computed as

$$\log(p_y(\boldsymbol{y})) = \log(p_z(\boldsymbol{z})) - \sum_{m=1}^{M} \left| det \left(\frac{\partial \mathcal{T}_\theta^{(m)}(\boldsymbol{z}^{(m)})}{\boldsymbol{z}^{(m)}} \right) \right|. \tag{II.12}$$

A normalizing flow process is illustrated in Figure II.10 with 5 transforms. If we implement the diffeomorphisms $\mathcal{T}_\theta^{(m)}$, $m = 1 \dots M$ as neural networks, we construct a trainable diffeomorphism between a simple latent distribution $\mathcal{Z}$ and a real distribution $\mathcal{Y}$ while keeping explicit the computation of the posterior likelihood. These neural-network-based diffeomorphism furthermore need to have a tractable determinant and inverse to scale up for high dimension while keeping a reasonable computational cost.

Starting from the fact that the Jacobian of triangular matrices is equal to the product

of diagonal elements, two types of transforms have been proposed for normalizing flows namely coupling layers and autoregressive layers.

Coupling layers

Coupling layers have been proposed by Dinh at al. in [121] and [120]. They natively construct block-diagonal matrices by splitting the input in two, leaving the first part unchanged and transform the second part as a function of the first part. A coupling layer is implemented as follows:

1. the input $\boldsymbol{z}^{(m)} \in \mathbb{R}^N$ is partitioned into two subsets $[\boldsymbol{z}^{(m)}_{1:n},\ \boldsymbol{z}^{(m)}_{n+1:N}]$ with $n \in [\![1, N]\!]$,
2. the first part of the input is left unchanged $\boldsymbol{z}^{(m+1)}_{1:n} = \boldsymbol{z}^{(m)}_{1:n}$,
3. the parameters ϕ_i of a coupling function g_{ϕ_i} are obtained as function of the first part of the input $\phi_i = \Phi(\boldsymbol{z}^{(m)}_{1:n})$ with Φ a trainable function, typically a deep neural network,
4. the second part is computed as $z_i^{(m+1)} = g_{\phi_i}(z_i^{(m)})$ for $i = n+1 \dots N$.

When constructed this way, a coupling transform $\mathcal{T}^{(m)}$ has a lower triangular Jacobian

$$\frac{\partial \mathcal{T}^{(m)}(\boldsymbol{z}^{(m)})}{\partial \boldsymbol{z}^{(m)}} = \begin{bmatrix} \mathbb{I}_n & \mathbf{0}_n \\ \frac{\partial \boldsymbol{z}^{(m)}_{n+1:N}}{\partial \boldsymbol{z}^{(m)}_{1:n}} & \frac{\partial g_\phi(\boldsymbol{z}^{(m)}_{n+1:N})}{\partial \boldsymbol{z}^{(m)}_{n+1:N}} \end{bmatrix} \tag{II.13}$$

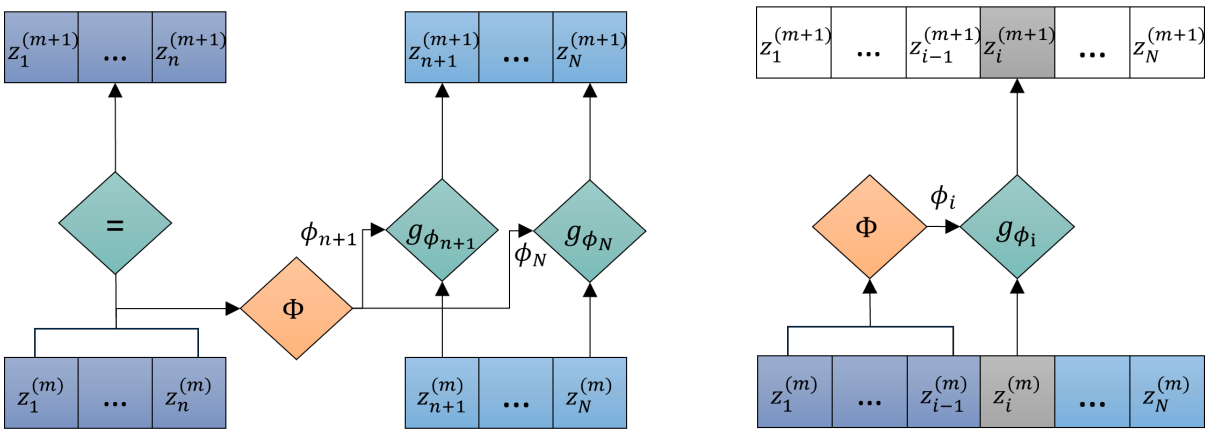
and its determinant is easily computed as the product of the derivatives of the coupling function

$$\det \left| \frac{\partial \mathcal{T}^{(m)}(\boldsymbol{z}^{(m)})}{\partial \boldsymbol{z}^{(m)}} \right| = \prod_{i=n+1}^{N} \left| \frac{\partial g_{\phi_i}(z_i^{(m)})}{\partial z_i^{(m)}} \right|. \tag{II.14}$$

A coupling layer is often combined with a permutation layer so that all input entries are transformed when stacking several layers. An illustration of a coupling layer is give in Figure II.11a.

Autoregressive flows

As developed in [122], coupling layers can be seen as a special case of autoregressive flows. The latter express each entry $z_i^{(m)}$ as a function of the previous inputs $z_i^{(m)} = g_{\phi_i}(\boldsymbol{z}^{(m)}_{1:i-1})$. A triangular Jacobian is thus obtained, and a greater flexibility can be achieved. An illustration of an autoregressive layer is given in Figure II.11b.



(a) Coupling layer (b) Autoregressive layer

Figure II.11 – Illustration of the forward passes for a coupling layer (a) and an autoregressive layer (a) to ensure tractable Jacobian determinant and easy inversion.

Bijective function

Now that layers permitting an easy inversion and Jacobian determinant computation have been constructed, one needs to choose a bijective function g_θ to be implemented as a coupling function in the transform. As described above, this function needs to be a diffeomorphism and should offer a tractable inverse an derivative. In Non-linear Independent Components Estimation (NICE) [121], a simple additive coupling function is proposed, that has a unit determinant and a trivial inverse but does not offer great non-linear capabilities. In RealNVP (standing for Non-Volume Preserving) [120], this concept is generalized as affine coupling function, where $g_{\phi_i}(z_i^{(m)}) = z_i^{(m)} \exp(s(z_{1:n}^{(m)})) + t(z_{1:n}^{(m)})$ with s and t that can be set as deep neural networks.

Polynomial splines functions [123] and cubic splines functions [124] have been proposed to increase the non-linear power of the transform. These functions interpolate between knots while ensuring a monotonic function. In [125], a generalized framework for rational quadratic spline function (RQSF) is proposed that permits easy Jacobian determinant computation, easy inversion and great flexibility. It is shown to overcome numerical difficulties and offer greater flexibility than previous splines functions. An example of a rational quadratic spline function is shown in Figure II.12.

The rational quadratic spline function is defined by K monotonic knots in $[-B, B]$ defined by their widths θ_k^w and heights θ_k^h. The derivative δ_k at each knot k are kept

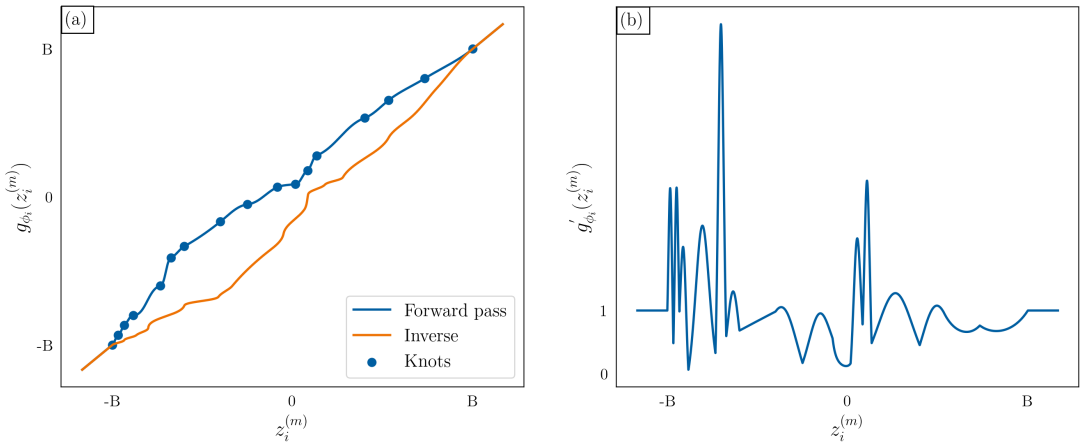


Figure II.12 – An example of a rational quadratic spline function, its inverse (a) and derivative(b).

positive. Outside the boundary domain $[-B, B]$, to be fixed in accordance with the data boundaries, linear tails are implemented so that no samples can be out of bounds. Setting $s_k = (\theta^h_{k+1} - \theta^h_k)/(\theta^w_{k+1} - \theta^w_k)$ and $\xi(z_i^{(m)}) = (z_i^{(m)} - \theta^w_k)/(\theta^w_{k+1} - \theta^w_k)$, the rational quadratic spline g_k in bin k is defined as

$$g_k(\xi) = \theta^h_k + \frac{(\theta^h_{k+1} - \theta^h_k)[s_k\xi^2 + \delta_k\xi(1-\xi)]}{s_k + [\delta_{k+1} + \delta_k - 2s_k]\xi(1-\xi)} \tag{II.15}$$

which is analytically invertible and derivable as described in [124]. The parameters of the transform are then obtained in a coupling or autoregressive layer as

$$\phi_i = \Phi(z^{(m)}_{1:i-1}) = [\boldsymbol{\theta}^w_{1:K}, \boldsymbol{\theta}^h_{1:K}, \boldsymbol{\delta}_{1:K}]. \tag{II.16}$$

The rational quadratic spline function is used as the base bijective function throughout this study for its flexibility.

Autoregressive neural networks: MADE

In an autoregressive setup, the neural network Φ needs to have autoregressive properties to ensure that its output for a certain input $z_i^{(m)}$ only depends on the previous input $\boldsymbol{z}^{(m)}_{1:i-1}$. To achieve this, masked neural networks are used in which connections are skipped to ensure autoregressive properties. In this study we use the Masked Auto-Encoder for Density Estimation (MADE) defined by [126] and adapted to normalizing flows by [122].

MADE is a feed-forward neural network to whom a binary mask is applied to ensure autoregressive properties by masking connections.

II.1.5 Experimental framework for machine learning based forecasts

When training and evaluating forecast models of environmental data, car must be given to time series auto-correlation and data representativity. We detail here strategies for implementing train, validation, test splits for environmental data, and cross-validation to alleviate data reprensentativity issues. A thorough description of deep learning framework for weather forecast is given in [66].

Train, validation, test split for environmental data

A common practice when developing data-driven models is to split the dataset into training, validation and testing split. The model is trained on the training split, the validation split is used for hyper-parameters tuning and over-fitting control, and the model is finally evaluated on the test split. In order for the model to be correctly trained and evaluated, each split should be independent and representative of the actual data distribution of the dataset.

For environmental data, several difficulty arise. The time series have strong auto-correlation, which means that the samples cannot be randomly drawn from the dataset to form the train, validation, test split. Indeed, doing so would yield to spurious correlations between the splits. Furthermore, the time series have strong cyclic dependencies with inter-annual (El Niño phenomenon), seasonal, and diurnal variabilities. Each set should be sufficiently long to be representative of these variabilities. An integer number of year is often taken as a condition for a representative train, validation, test split.

Eventually, the complexity and chaotic nature of weather systems imply that the length of the dataset should be sufficient to capture long-term trends. This requires long-term measurement campaign which is a major limitation for metocean forecasting as describes in Section I.3. Furthermore the assumption of a stationary climate is to be questioned for long-term applications due to climate change. All these points should be discussed and addressed when implementing metocean forecast model.

Cross-validation

To alleviate the impact of data representativity, cross-validation can be used. It consists in alternating the periods used for training, validation and testing, while maintaining independence between the sets. Each simulation with a specific train, validation, test split is called a fold, so the term K-fold cross-validation is used. The error metric obtained by averaging the errors obtained for the K-fold is called the generalized error, and its variability indicates the impact of dataset split in the performances of the model. For weather applications, the number of folds that can be constructed is limited by the constraints described above. The generalized error is then given either as the standard deviation, or as a 50 % inter-quantile range around the mean error. It is important to note that this is not the only source of uncertainty in the scores. For example sampling dependence should be alleviated with bootstrapping. An exhaustive description of uncertainties for data-driven models is given in [127]. For the experiments described in this manuscript however, error variability was found to be largely dominated by splits variability. This is why the scores variability is only described as inter-quantile ranges or standard deviation of cross-validation splits.

Cross-validation is common practice in deep learning applications to evaluate the error variability due to the data used for training and testing. For environmental applications it is also a mean to evaluate the impact of data representativity. When the variability in the measure is high, it means that the underlying data distribution in the training, validation and/or testing set is not representative of the full dataset distribution, and that the dataset used is to short to ensure satisfying representativity. An example of cross-validation keeping the temporal correlation and seasonal variability for a limited weather dataset is shown in Section IV.2 Figure IV.4.

II.2 Probabilistic forecasting

Predicting the evolution of an uncertain phenomenon in the future is the nature of forecasting, which makes it by essence a probabilistic problem. The development of probabilistic forecast models notably in the weather forecasting field marks a paradigm shift since the 1980s with the development of ensemble forecasting [43, 47]. The development of probabilistic forecast models is a theoretical requirement for the optimal use of forecast products for decision-making problems. In the wind energy industry, probabilistic fore-

casts of wind energy production is essential for energy systems management and market participation [44, 45] or for the estimation of weather windows for offshore operations [5, 7].

The uncertainty of a forecast model can be divided into two categories: the aleatory and the epistemic uncertainties. The aleatory uncertainty refers to all the sources of randomness in the modelled phenomenon, including but not limited to chaos-related uncertainties or observational errors. The epistemic uncertainty on the other hands refers to all the sources of uncertainty from the imperfect knowledge of the system such as model parameters error. Epistemic uncertainty can be alleviated in deep learning models by running Monte-Carlo simulations using dropout layers, or by simultaneously running different architectures on the same input data to generate an ensemble of predictions. A thorough study on uncertainty quantification using data-driven models can be found in [127].

A probabilistic forecast model describes the output $\boldsymbol{y}_{t+k}$ as the realisation of a random variable $\hat{\boldsymbol{Y}}_{t+k}$. The probability distribution of this random variable is called the posterior distribution. Its probability density function is noted $\hat{f}_{t+k}$, and cumulative density function $\hat{F}_{t+k}$.

To describe the random variable $\hat{\boldsymbol{Y}}_{t+k}$, several strategies can be adopted. We regroup the strategies used in the litterature under three categoris: parametric and non-parametric methods for the time dependent posterior, and scenario generation.

The term parametric refers to all methods that use parametric assumptions for the posterior. We describe here all the methods that assume this assumption for each predicted lead time. The time-dependant posterior distribution $\hat{f}_{\boldsymbol{\theta}_{t+k}}$ is parametrized with parameters $\boldsymbol{\theta}_{t+k}$ that fully describe the law of the random variables.

On the contrary, non-parametric methods predict summary statistics of the posterior distribution for each time step. Quantiles levels for example can be used to describe the inverse of the cumulative density function.

We eventually introduce scenarios-based methods that are widely implemented in the literature. These can be parametric or non-parametric, but have in common that they describe the posterior distribution as the distribution of the random variable $\hat{\boldsymbol{Y}}_{t:t+K}$, which samples are multivariate trajectories $\hat{\boldsymbol{y}}_{t:t+K} \in \mathbb{R}^N \times \mathbb{R}^K$. Because of the dimension increase it implies, it requires the use of specific methods that justify this classification.

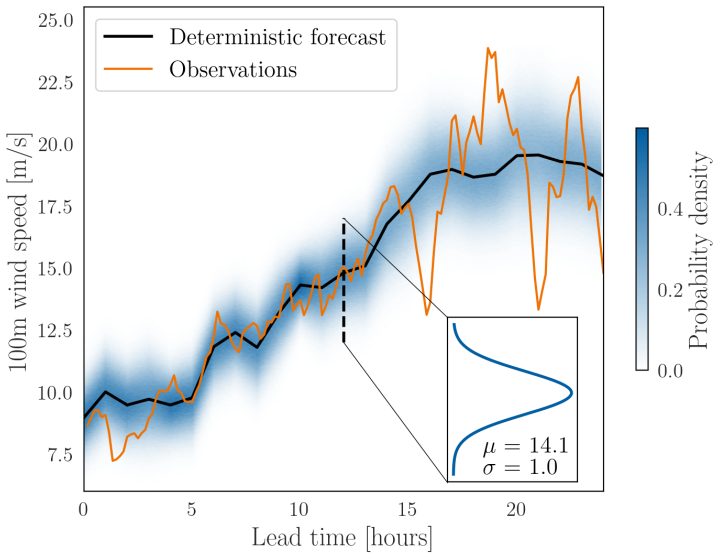


Figure II.13 – An example of a 24 hours ahead probabilistic forecast under Gaussian assumption. The parameters of a Gaussian distribution are predicted for each lead time.

II.2.1 Parametric methods

A simple assumption for approximating a probabilistic posterior is to consider it takes the shape of a parametric distribution $\hat{f}_{\boldsymbol{\theta}_{t+k}}$, which parameters $\boldsymbol{\theta}_{t+k}$ are predicted by the forecast model Ψ

$$\Psi(\boldsymbol{X}_t) = \{\boldsymbol{\theta}_{t+k}\}_{k=1\dots K}. \tag{II.17}$$

The parametric assumption of the posterior often permits closed-form computations of the likelihood $\hat{f}_{\boldsymbol{\theta_{t+k}}}(\boldsymbol{y}_{t+k})$, which permits optimization. Some distribution also permit the closed-form computation of the continuous ranked probability score (CRPS, see Section II.3). It is associated with sampling capabilities. Since the law of the random variable $\hat{\boldsymbol{Y}}_{t+k}$ is perfectly known, samples $\hat{\boldsymbol{y}}^{(l)}_{t+k},\ l = 1 \dots L$ can easily be drawn from it.

Common assumptions for the posterior distributions are Gaussian [16, 56, 128–130] log-normal [131–133], Beta [134], Sinh-Arcsinh [127], Laplace [135], Gaussian Mixture Models (GMM) [136]. An example with Gaussian posterior assumption is shown in Figure II.13.

In Chapter IV and Chapter V, a multivariate Gaussian assumption is made for the posterior allowing for exact computation of the likelihood and easy sampling. The relaxing of this assumption is shown to be marginally improving the probabilistic scores of the forecast. The multivariate Gaussian distribution is described by its mean $\boldsymbol{\mu}$ and covariance

matrix $\mathbf{\Sigma}$, and its density function f reads

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right) \tag{II.18}$$

for $\boldsymbol{y}$ a realization of the random variable $Y \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$. The vector $\boldsymbol{\mu}$ contains the means of the marginal distributions, and $\mathbf{\Sigma}$ the covariances. The matrix $\mathbf{\Sigma}$ has to be definite positive to be a valid covariance matrix i.e. $\boldsymbol{z}^T\mathbf{\Sigma}\boldsymbol{z} > 0 \; \forall \boldsymbol{z} \neq \mathbf{0}$. The achievement of definite positiveness require special care and can become complicated for high dimensional problems. For low dimensions, the covariance matrix can be decomposed into an unique positive-diagonal lower triangular matrix L such that

$$\mathbf{\Sigma} = LL^T \tag{II.19}$$

with L the Cholesky decomposition of Σ. In [137], a Cholesky-based parametrization is used for the modelling of the temporal covariance matrix for a 10-step ahead temperature forecast obtained with regression. However, the number of parameters to estimate for a N-dimensional Gaussian distribution is $\frac{N(N+1)}{2}$ and becomes intractable for high dimensional cases. In this case, low-rank parametrization should be considered for example when modelling spatio-temporal dependencies such as the Matérn covariance matrix which is isotropic and stationary [138]. Recent work in the multivariate energy forecasting advocate for the development of low-rank non-stationary non-isotropic covariance matrices. [139] proposed to generalize covariance functions by allowing their parameters to vary conditioning on input explanatory variables, hence increasing the flexibility and explainability of the matrices, but loosing the guarantee of definite positiveness for non-stationary processes. In this study, the cross-variable dependency per time step is modelled by a multivariate Gaussian distribution under Cholesky assumption since the dimension of the problem is low (2 variables in Chapter IV and 3 variables in Chapter V).

II.2.2 Non-parametric methods

To avoid making assumptions on the shape of the posterior distribution, non-parametric methods can be used. In contrast with parametric methods, they do not rely on parametrized functions. The most used methods in the literature consist in predicting summary statistics of the posterior distribution such as quantile levels or intervals.

The idea behind quantile forecasting is that a cumulative density function of a random

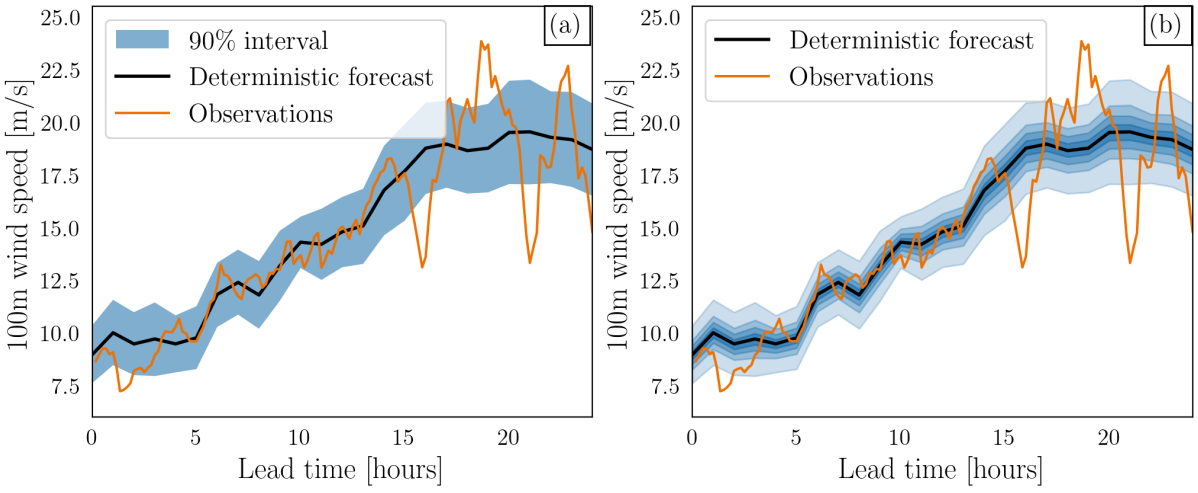


Figure II.14 – An example of a 24 hours ahead non-parametric probabilistic forecast. (a) shows an interval forecast with $\beta = 90\%$. (b) is a quantile forecast showing the quantiles 5%, 20%, 30%, 40%, 60%, 70%, 80% and 95%.

variables can be approximated using its quantiles. A quantile q_α associated with a proportion $\alpha \in [0, 1]$ is defined as the value under which a realization of the random variable has a probability α to be. The quantiles define the quantile function F^{-1} which is the inverse of the cumulative density function F

$$q_\alpha = F^{-1}(\alpha). \tag{II.20}$$

From the predicted quantiles, and interpolated quantile function is obtained using e.g. cubic spline interpolation is then used [140, 141]. Quantile forecasting has the advantage of being assumption free, but it requires the training of one model per quantile, resulting in a substantial computational cost and potential quantile crossing issues. It furthermore has limitations for multivariate forecasting, since it cannot be used to model the dependency between variables. It is however a common assumption for the development of probabilistic forecast models [4, 142, 143].

Interval forecasting is another type of non-parametric model that predict an interval in which a realization of the random variable has a nominal probability $(1-\beta)$, $\beta \in [0, 1]$ to lay in. The interval defined by an upper bound α_u and lower bound α_l which are indeed quantiles. A description and evaluation framework of interval and quantile methods is presented in [144]. Both techniques are illustrated in Figure II.14

Eventually, recent developments in the generative artificial intelligence domain offers

new methods for non-parametric posterior descriptions. Generative models are designed to approximate very complex distributions and can be used for high dimensional probabilistic problems such as image or speech generation. In a sense, the posterior distribution is encoded implicitly in the structure of the model. In Section II.1 we give a detailed description of generated approaches. Especially, we describe normalizing flows and how they can be used to approximate complex posterior distributions for forecasting.

II.2.3 Scenario generation

The probabilistic forecasts presented above are issued per lead time, describing the cross-variable dependency but not the temporal correlation between lead-times. For time-dependent decision-making problems such as the vessel and crew dispatch for offshore operations, the temporal evolution is essential [145]. As described in Section I.1, offshore operations require the forecasting of weather windows during which the environmental conditions are below a certain threshold. Considering the forecast uncertainty at each time step independently yields to an incorrect estimation of the probability of weather window occurring. On the contrary, an ensemble of realistic scenarios is needed that respect the marginal distributions for each time step, and have a realistic temporal evolution.

Ensemble numerical weather prediction is the most intuitive scenario forecasting technique. It has become a new paradigm in numerical weather prediction, and it provides an ensemble of scenarios in a very high dimensional state, modelling boss the cross-variable and temporal correlation natively through its set of differential equations. [146] use ensemble forecasts from which density distributions are estimated with a kernel method to predict the wind power output of an offshore wind farm. The number of members issued by the ensemble being limited to 10 to 50 for current operational systems, there is a need to fit a distribution per lead time to have a full probabilistic description. [147] calibrate the ensemble distribution to fit a Gaussian distribution for the tuple of wind speed. [148] fitted a Gaussian distribution on the temporal evolution of a raw numerical weather prediction ensemble to increase the number of scenarios and to output a reliable forecast of power ramp events. Eventually [17] used GANs to generate additional members of an ensemble forecasts showing promising results. Given the limitations of ensemble numerical weather forecasts in terms of computational cost, it is appealing to implement statistical methods for scenarios forecasting.

Another approach is to explicitly model the temporal dependency with a probability density, but this imply describing a density function in a very high dimension space. A

common assumption for approximating complex multivariate dependency structures are copulas. Copulas were first introduced by [149] to describe multivariate distributions with uniform marginals. The Sklar's theorem states that any K-dimensional distribution G with marginal distributions $F_1, \ldots, F_K$, there exists a copula function C such that

$$G(y_1, \ldots, y_K) = C(F_1(y_1), \ldots, F_K(y_K)). \tag{II.21}$$

Furthermore, if the marginal distributions are continuous, this function is unique. The copula function then describes the joint distribution between the random variables $u_1, \ldots, u_K$ with $u_k = F_k(y_k)$ uniformly distributed on $[0, 1]$. Several copula functions can be implemented to model the joint distribution: Gaussian, Clayton, Gumbel, Frank, Vine. For Gaussian copulas, the only requirement is the construction of a valid $(K \times K)$ covariance matrix.

If we assume a time series of predicted cumulative density functions $\hat{F}_{t+k}$, and a covariance matrix $\mathbf{\Sigma}_t$ of dimension $(K \times K)$ describing the temporal dependency, scenarios can be generated using the following process under Gaussian copulas assumption.

1. L samples $\{\hat{\boldsymbol{s}}^{(l)}_{t:t+K}\}_{l=1\ldots L}$ are drawn from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t)$
2. For each lead time $k = 1 \ldots K$, the realizations $\{\hat{z}^{(l)}_{t+k}\}_{l=1\ldots L}$ of the uniform variables $\hat{Z}_{t+k}$ are obtained using the inverse probit function Φ:

$$\hat{z}^{(l)}_{t+k} = \Phi(\hat{s}^{(l)}_{t+k}),\ l = 1 \ldots L,\ k = 1 \ldots K. \tag{II.22}$$

3. For each lead time k, the samples $\{\hat{z}^{(l)}_{t+k}\}_{l=1\ldots L}$ are then transformed to respect the marginal distribution $\hat{f}_{t+k}$ using the inverse cumulative density function $\hat{F}^{-1}_{t+k}$ to generate predicted samples $\{\hat{y}^{(l)}_{t+k}\}_{l=1\ldots L}$

$$\hat{y}^{(l)}_{t+k} = \hat{F}^{-1}_{t+k}(\hat{z}^{(l)}_{t+k}),\ l = 1 \ldots L,\ k = 1 \ldots K. \tag{II.23}$$

4. The obtained samples $\{\hat{\boldsymbol{y}}^{(l)}_{t:t+K}\}_{l=1\ldots L}$ are then scenarios which marginal distributions are $\hat{F}_{t+k},\ k = 1 \ldots K$, and which temporal dependency is described by the Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t)$.

For high dimensional cases, the estimation of a valid $(K \times K)$ covariance matrix can cause difficulties, and [150] proposed a low-rank covariance matrix for minimizing the number of coefficients to be predicted. For scenario forecasting, one can model the temporal dependency between time steps that the copulas should describe the temporal

dependency between time steps, and that the marginal distributions per time step are predicted independently. This was proposed in [151] for the generation of power production scenarios from probabilistic forecasts output. They furthermore proposed a recursive estimation of the temporal covariance structure. [145] showed that the Gaussian copulas is appropriate for modelling the temporal dependency for wind power scenario forecasting. Scenario forecasting can also be useful for extreme events prediction, such as ramp forecasting for power market operations [152].

The above copula structure does not accommodate multivariate temporal scenarios, in the sense that it can only approximate joint density distribution with uniform marginals on $[0, 1]$. For multivariate scenarios forecasting, both the temporal covariance and cross-variable covariance need to be modelled. For offshore operations, metocean variables and their associated uncertainties are correlated in time. The wave height resulting from the time integral of the wind speed, and the wave numerical models being forced by wind numerical models, both variables' evolution and uncertainty should be treated jointly. Conditional copulas can be developed for modelling multivariate scenarios [153] and temporal dependent copulas [154] extend the definition of conditional copulas to pseudo-copulas with non-uniform marginal distributions. To the best of our knowledge, such copulas methods are not applied to multivariate scenarios generations for environmental forecasting. The wealth of the economic literature on this topic should be considered for future applications. A thorough review of copulas methods is given in [155].

Alternatively, copulas can be coupled with other methods for handling both the cross-variable and temporal dependencies. [150] developed a method based on low-rank Gaussian copulas conditioned by the state of a recurrent neural network. The cross-variable dependency is modelled by the copulas, and the temporal dependency is latent in the state of the deep learning model.

Recently, the rise of generative models in the deep learning community led to new possibilities in scenario forecasting. The capabilities of generative architecture in estimating very high dimension latent distributions could allow for the direct generation of multivariate scenarios given sufficient data and model complexity. [44] compared GANs, VAE and normalizing flows for the multivariate probabilistic forecasting of renewable energy generation. Though they conclude that normalizing flows are easier to implement, they do not explicit the temporal dependency, while GANs and VAE are trained to output trajectories, normalizing flows are trained on a per time lead basis. [59] on the other hand, coupled a normalizing flows with a recurrent neural network to implicitly integrate

a temporal dependency in the flows' conditioning. These novel approaches can be designed to seamlessly generate multivariate scenarios, and are an interesting research topic for multivariate probabilistic forecasting. We present two innovative methods implying normalizing flows in Chapter V and compare them to Gaussian copulas based methods.

II.3 Forecast evaluation frameworks for probabilistic forecasting

The development of various forecasting techniques requires the formalisation of an evaluation framework. The choice of the metric used for model evaluation is crucial, and should reflect the actual cost associated with the forecast error. The best model according to a certain metric might not be for another metric, and several metrics can be combined to evaluate different characteristics of the forecast.

The models' performances should be evaluated on out-of-sample data. The datasets are to be split in training, validation and testing sets, which should all be independent, and representative of the full distribution of the data. When dealing with meteorological data, special care should be given to the data auto-correlation, and to cyclic correlation within the data. Meteorological phenomenon are know to be influence by annual and seasonal patterns, can have diurnal variations, and are auto-correlated in the range of several days [66].

The framework for evaluating probabilistic forecasts requires specific metrics, because a full distribution should be scored with a single observation. Furthermore, the predicted uncertainty should be representative of the underlying process' uncertainty. This property of statistical consistency is called the forecast calibration. The elements of probabilistic forecasts evaluation can be found in [46].

Eventually, a difference should be made between the forecast quality in the sense of its performance regarding evaluation metrics, and its value, which is linked to the impact of the forecast use for a specific application [6].

In this section we describe the statistical tools for computing metrics for the evaluation of deterministic forecasts and probabilistic forecasts. A thorough description of evaluation tools for wind energy applications can be found in [50]. We then present the advancement in forecast value for offshore wind operations.

II.3.1 Deterministic forecast evaluation metrics

We first describe the metrics used to evaluate single-point forecasts. We consider a forecast issued at time $t = 1 \dots T$ that issues predictions for lead times $k = 1 \dots K$ of the target N-dimensional observation $\boldsymbol{y}_{t+k} \in \mathbb{R}^N$. The forecast issued at time t for a lead time k is noted $\hat{\boldsymbol{y}}_{t+k} \in \mathbb{R}^N$. For probabilistic scenarios, the associated single-point forecast is the mean or the median of the predicted distribution. It is important to note that if the observation vectors contains different variables with different physical units (e.g. wind speed and significant wave height) they should be normalized before the metrics are computed and averaged between variables. Alternatively, the metrics can be computed and expressed per variable.

The metrics are computed per lead time k, and noted accordingly with a subscript k. If noted without subscript, they refer to the mean value of the metric across lead times. For a given metric $\mathcal{M}$:

$$\mathcal{M} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{M}_k \tag{II.24}$$

The average difference between the observation and the forecast is called the bias.

$$bias_k = \frac{1}{T} \sum_{t=1}^{T} (\hat{\boldsymbol{y}}_{t+k} - \boldsymbol{y}_{t+k}) \tag{II.25}$$

The bias only measure the ability of a forecast to predict the same mean value as the observed mean.

The mean absolute error (MAE) is defined as:

$$MAE_k = \frac{1}{T} \sum_{t=1}^{T} |\hat{\boldsymbol{y}}_{t+k} - \boldsymbol{y}_{t+k}| \tag{II.26}$$

For probabilistic forecasts, the median value of the distribution should be considered for the mean absolute value. It measures the mean absolute difference between forecasts and observations. Since errors are penalized proportionally it is suited for applications where the cost of bad predictions is proportional to the error.

The Mean Square Error (MSE) penalizes outliers stronger by averaging the squared error between observations and forecasts. The mean of probabilistic forecasts should be

considered for mean squared error.

$$MSE_k = \frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol{y}}_{t+k} - \boldsymbol{y}_{t+k})^2 \tag{II.27}$$

The Root Mean Squared Error (RMSE) is a widely used metric and is simply the square root of the MSE. It has the same unit as the observations space which makes it more explainable than the MSE.

$$RMSE_k = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol{y}}_{t+k} - \boldsymbol{y}_{t+k})^2} \tag{II.28}$$

The Quantile Score is analogous to the mean absolute error, but it weights differently positive and negative error.

$$QS_k(p) = \frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol{y}}_{t+k} - \boldsymbol{y}_{t+k})(\mathbb{1}(\hat{\boldsymbol{y}}_{t+k} > \boldsymbol{y}_{t+k}) - p) \tag{II.29}$$

It is optimized when $\hat{\boldsymbol{y}}_{t+k}$ is the p-th quantile of the distribution and is useful for training quantile forecasting models.

II.3.2 Probabilistic forecast evaluation metrics

Probabilistic forecasts can be issued as quantiles or intervals, ensemble, or full distributions. We denote $\hat{f}_{t+k}$ the predicted probability density function at time t for lead time k, and $\hat{F}_{t+k}$ the associated cumulative density function. We give the literal expressions of the different metrics as function of the predicted distribution for parametric outputs, and as function of samples of the predicted distribution for non-parametric and ensemble predictions.

According to Gneiting et al. [46], a good probabilistic forecast should « *maximize sharpness subject to calibration* ». The forecast sharpness refers to the concentration of the predicted distribution around the mean value, and the calibration is the statistical consistency between the forecast and the observations. For a calibrated forecast the distribution of the forecast errors is on average equal to the predicted distribution. In other words, in the long run 10% of the observations should fall on the first decile of the predicted distributions and so on.

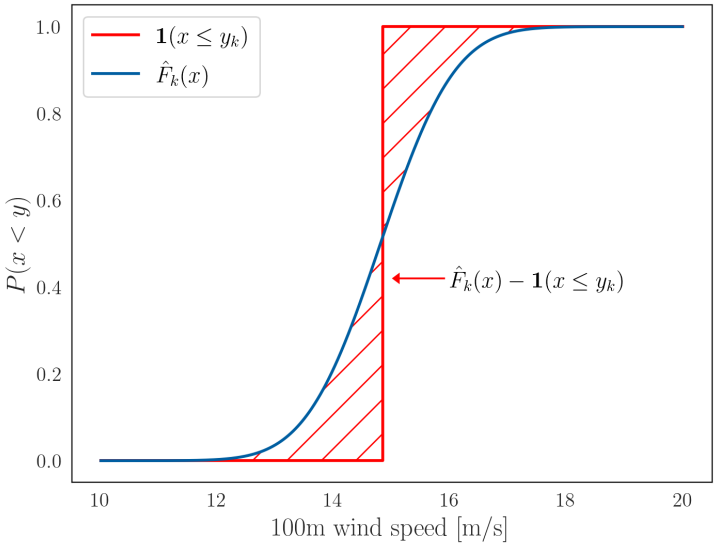


Figure II.15 – The CRPS is the integral of the squared area between the predicted cumulative density function (blue) and a step function at the observation (red). A sharp distribution centred on the observation will have the lowest CPRS.

II.3.2.1 Evaluation metrics for continuous variables

The evaluation of probabilistic forecasts requires the evaluation of both the sharpness and the calibration of the forecast. A common way of evaluating the quality of a probabilistic forecast is to use single-value scores that should be strictly proper scoring rules [49].

Univariate forecasts

For univariate forecasts, the most common evaluation metric is the Continuous Rank Probabilistic Score (CRPS). It evaluates the quality of the cumulative density function by integrating the squared difference between the cumulative density function and a step function at the observed value as illustrated in Figure II.15.

$$CRPS_k = \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{+\infty}\left[\hat{F}_{t+k}(y) - \mathbb{1}(y \leq y_{t+k})\right]^2 dy \tag{II.30}$$

As shown in [49], the CRPS can be expressed as function of the expected values of $\hat{Y}_k$ and $\hat{Y}'_k$, two independent univariate random variables of cumulative density function $\hat{F}_{t+k}$, which can be used for ensemble forecasts and forecasts which cumulative density

function cannot be analytically integrated.

$$CRPS_k = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{2}E|\hat{Y}_{t+k} - \hat{Y}'_{t+k}| - E|\hat{Y}_{t+k} - y_{t+k}|\right] \tag{II.31}$$

The CRPS can be decomposed into reliability (REL), resolution (RES) and uncertainty (UNC) [156]. The reliability of a forecast is equivalent to its calibration. Usually, probabilistic forecasts are evaluated given a single observation point, making the evaluation of the reliability difficult. The resolution of a forecast is the capacity of the model to discriminate between different situations. A forecast model that issues the same mean forecast every day has a low resolution. The uncertainty defines the remaining uncertainty of the underlying process and is independent from the forecast model. The decomposition reads

$$CRPS = REL - RES + UNC \tag{II.32}$$

The ignorance score (IS) also called logarithmic score is analogous to the likelihood defines in Section II.1. The logarithm function makes it a more discriminatory score than the CRPS, but can induce instabilities and strong influence of outliers.

$$IS = \frac{1}{T}\sum_{t=1}^{T}\log(\hat{f}_{t+k}(y_{t+k})) \tag{II.33}$$

The IS can only be computed for continuous distribution forecasts since the probability density function cannot be easily computed from quantiles or samples. The IS can naturally be computed for multivariate forecasts when the probability density function is analytically tractable.

Multivariate forecasts

In a multivariate setup, assuming that all predicted variables are normalized or have the same scale and units, univariate scores can be used. The CRPS can be generalized for multivariate forecasts as the Energy Score (ES) [49]

$$ES_k = \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{+\infty}\left[\hat{F}_{t+k}(\boldsymbol{y}) - \mathbb{1}(\boldsymbol{y} \leq \boldsymbol{y}_{t+k})\right]^2 d\boldsymbol{y} \tag{II.34}$$

And it can seemingly be computed from samples using the Euclidean norm $||\cdot||$

$$ES_k = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{2}E||\hat{\boldsymbol{Y}}_{t+k} - \hat{\boldsymbol{Y}}'_{t+k}|| - E||\hat{\boldsymbol{Y}}_{t+k} - \boldsymbol{y}_{t+k}||\right] \tag{II.35}$$

As shown in [58], the ES is mostly sensible to 1st moment errors, and is not very discriminant for correlation structures between variables. The Variogram Score (VS) proposed by [157] is used to score the correlation structure of the predictive distribution, but is invariant to first moment errors :

$$VS_k(p) = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(|y_{t+k,i} - y_{t+k,j}|^p - E|\hat{Y}_{t+k,i} - \hat{Y}_{t+k,j}|^p)^2\right] \tag{II.36}$$

where $y_{t+k,i},\ i = 1 \dots N$ are the components of the multivariate observation $\boldsymbol{y_{t+k}}$ at lead time k, and $\hat{Y}_{t+k,i},\ i = 1 \dots N$ are the components of a random vecotr $\hat{\boldsymbol{Y}}_{t+k}$ distributed on the predicted multivariate distribution $\hat{F}_{t+k}(\boldsymbol{y}_{t+k})$. w_{ij} are positive weights that can be assigned if desired and p is the order of the variogram, shown to be optimal at $p = 0.5$. In all that follows, the $VS(p = 0.5)$ is considered and noted VS.

In Chapter V, the ES and VS are computed in the scenarios space of dimension $(N \times K)$. The above equations can still be applied by replacing the random variable $\hat{\boldsymbol{Y}}_{t+k}$ in the variable space by the random variable $\hat{\boldsymbol{Y}}_{t:t+K}$ in the scenarios space.

Evaluate the statistical consistency of forecasts

The evaluation of the forecasts calibration is classically made by graphically analysing the verification rank histogram or the Probability Integral Transform (PIT) of the forecast. The verification rank histogram is the histogram of the ranks of the observations in an ensemble forecasting, and is also called the Talagrand diagram. The closely related reliability diagram is a plot of PIT of the forecast i.e. the observed frequencies as function of the predicted frequencies. Both graphs are show in Figure II.16. A perfect rank histogram should be flat, meaning than 10% of the observations fall in the 10% quantile and so on. Seemingly, a perfect reliability diagram is a $[1:1]$ line showing a perfect prediction of the observed frequencies. A U-shaped rank histogram shows an under-dispersive model since too many observations fall in the extreme quantiles of the distribution. In this case the model underestimates the uncertainty. On the contrary a bump-shaped histograms is obtained for over-dispersive modes.

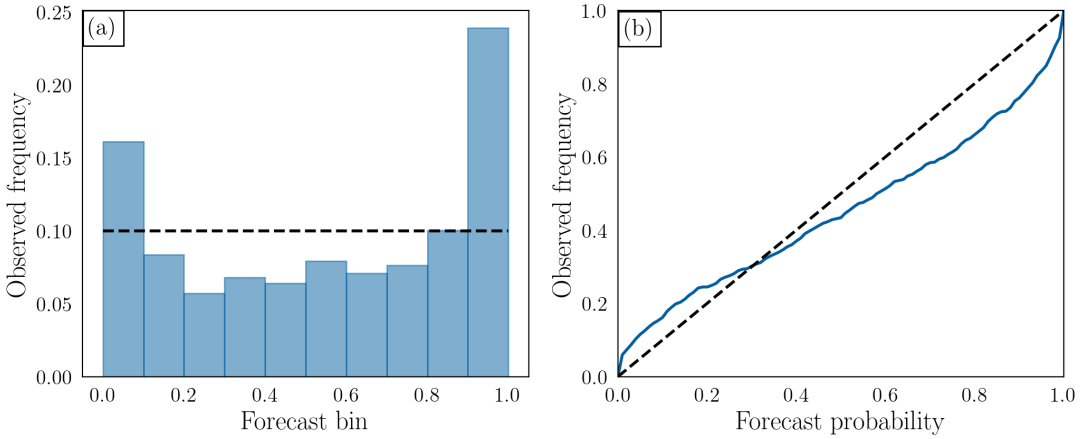


Figure II.16 – (a) an example of an under-dispersive rank histogram from the AROME ensemble forecast. A perfectly reliable forecast is shown as a dotted line. (b) the same information presented as a reliability diagram showing the PIT of the forecast.

It is possible to derive a scoring rule from the rank histogram called the reliability index or discrepancy index [48]

$$REL = \frac{1}{Q}\sum_{j=1}^{Q} |\hat{b}_j - \frac{1}{Q}| \tag{II.37}$$

where $\hat{b}_j$ is the observed frequency of rank j.

A multivariate generalization of the verification rank histogram is proposed in [48]. To compute the rank, we note

$$\boldsymbol{x} \preceq \boldsymbol{y} \text{ if and only if } x_j \leq y_j,\ j = 1 \dots N. \tag{II.38}$$

Given $\hat{\boldsymbol{y}}_{t+k,l} \in \mathbb{R}^N,\ l = 1 \dots L$ L samples of the random variable $\hat{\boldsymbol{Y}}_{t+k}$ following the predicted distribution $\hat{F}_{t+k}$, we define the pre-rank $\rho_{t+k,l}$ of the sample $\hat{\boldsymbol{y}}_{t+k,l}$ as the number of samples $\hat{\boldsymbol{y}}_{t+k,m}$ such that $\hat{\boldsymbol{y}}_{t+k,m} \preceq \hat{\boldsymbol{y}}_{t+k,l}$

$$\rho_{t+k,l} = \sum_{m=1}^{L} \mathbb{1}(\hat{\boldsymbol{y}}_{t+k,m} \preceq \hat{\boldsymbol{y}}_{t+k,l}) \tag{II.39}$$

The multivariate rank of the sample $\hat{\boldsymbol{y}}_{t+k,l}$ is then drawn randomly between $s^< + 1$ and

$s^{=} + s^{<}$ defined as

$$s^{<} = \sum_{m=1}^{L} \mathbb{1}(\rho_{t+k,m} < \rho_{t+k,l}) \text{ and } s^{=} = \sum_{m=1}^{L} \mathbb{1}(\rho_{t+k,m} = \rho_{t+k,l}). \tag{II.40}$$

In this way, it is possible to represent the calibration of multivariate probabilistic forecasts as a rank histogram.

II.3.2.2 Event-based performance metrics

When forecasting binary events such as the existence of a weather window for offshore operations, the observation b_{t+k} is an integer value $c_{t+k} \in \{0, 1\}$, where $b_{t+k} = 1$ if the event occurs, and $b_{t+k} = 0$ if not. The forecast $\hat{b}_{t+k}$ can be an integer $\hat{b}_{t+k} \in \{0, 1\}$ or a real value $\hat{b}_{t+k} \in [0, 1]$ and then correspond to the probability for the event to occur.

The Brier Score (BS) is equivalent to the mean square error and is a common single-value score for evaluating binary forecasts

$$BS_k = \frac{1}{T} \sum_{t=1}^{T} (\hat{b}_{t+k} - b_{t+k})^2. \tag{II.41}$$

The forecast results can be summarized by counting the true positives (TP) when the event was predicted and occurred; the true negatives (TN) when the event was not predicted and did not occur; the false positives (FP) when the event was predicted but did not occur; and the false negatives (FN) when the event was not predicted but occurred. The counting of these four categories yields a contingency table (Figure II.17 (a)) from which single-value scores can be defined. The True Positive Rate (TPR) is equal to the proportion of predicted positive outcomes among the positive outcomes

$$TPR = \frac{TP}{TP + FN}. \tag{II.42}$$

Seemingly, the False Positive Rate (FPR) is the proportion of predicted positive outcomes among the negative outcomes

$$FPR = \frac{FP}{FP + TN}. \tag{II.43}$$

The Receiver Operation Characteristic (ROC) is the plotting of TPR as function of FPR for different probabilistic thresholds. The area under the ROC curve is called the AUC score, and is equal to 1 for a perfectly discriminant forecast with a step ROC curve.

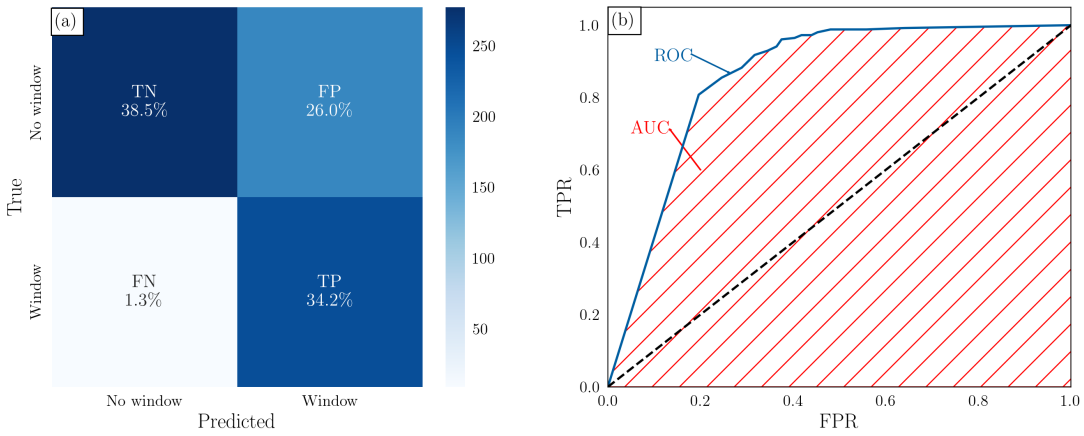


Figure II.17 – (a) shows a graphical visualisation of the confusion matrix, gathering the informations about TN, FN, TP and FP for the forecast of the binary event, here a 3 hours weather window with 100m wind speed < 12 m/s from the AROME ensemble forecast of January 2023 at the Planier island. The AROME ensemble forecast has 16 ensemble members, form which the ROC curve is computed and plotted in (b). The hatched area is the AUC score. A perfect forecast will have a step ROC curve and an AUC of 1.

A random forecast of 0 and 1 would yield to a diagonal ROC curve. An example is given in Figure II.17 (b).

II.3.3 Value of forecast models for offshore operations

In addition to the metric described above, the application-specific value of the forecast should betaken into account. [50] advocate for a metric selection as function of the application to correctly reflect the forecast value in its optimization and training. [51] showed that the forecast value can differ from the forecast quality for specific applications, here a wind power market participation. While for cost/loss situations, an optimal model can be obtained by training on a specific quantile [158], for some complex applications, no proper scoring rules can be applied to faithfully reflect forecast value.

For wind power applications, the value of probabilistic forecasts of wind power for market participation has been an active research field in the last decades [44, 159, 160], including the human decision-making based on probabilistic forecasts and their different illustrations [161].

For offshore operations, recent research papers explored the different factors impacting the cost of offshore operations in link with metocean forecasts and decision-making. [162] compare probabilistic forecast models of wind and wave conditions using the mean wind

farm availability, the total energy produced and the total revenue including power losses, spare parts costs and vessels logistics costs using a full O&M model. Seemingly, [163] developed a mixed integer linear problem for considering crew dispatch, energy production and accessibility. Such complex O&M models are difficult to implement operationally and lack explainability. [99] take the example of a specific operations to compute the mean operability after weather window estimation to evaluate their Markov-witching model wind speed and wave height. [5] proposed a response-based alpha factor for weather window estimation, and evaluated scenarios by computing the wind farm accessibility. All these metrics are interesting for comparing models together, but they can't be derived analytically for operational decision-making. [6] proposed and Economic Forecasting Metric (EFM) based on the cost of predicted a weather window when the conditions are not workable (Cost of False Below C_{FB}) and the cost of missing a weather window (Cost of False Above C_{FA}). From this the EFM can be computed as

$$EFM = C_{FA} + C_{FB} \tag{II.44}$$

where C_{FA} reflects the opportunity cost associated with production loss when missing a weather window, and C_{FB} the chartering cost of dispatching a vessel. When considering the problem this way, it becomes a cost-loss decision-making problem, shown to be optimal in [7] for a probabilistic threshold of weather window existence

$$p = \frac{C_{FB}}{C_{FA} + C_{FB}} \tag{II.45}$$

that is then dependent on the weather window duration, operability limits, vessel chartering cost and wind power production. These papers pave the way for optimal decision-making under uncertainty, provided that the metocean forecasts are properly reliable. However, as noted recently in [164], the race towards optimal decision-making for offshore operations can cause an increase in the number of operations carried out in marginal conditions, and hence an increased risk on material and staff. In Chapter V, we propose a theoretical framework to penalize risk in the decision-making process under uncertainty.

II.4 Literature review on data-driven forecast models for wind energy applications

In this section we give an overview of data-driven models for metocean forecasting, with a focus on wind energy applications. We distinguish statistical methods, machine learning methods and deep learning based methods.

II.4.1 Statistical methods

A fundamental class of models are the autoregressive moving average (ARMA) models and their variants introduced in [165]. They are widely explored for wind speed and wind energy forecasting, and often serve as a statistical baseline [166, 167], often for ultra-short-term forecasting (< 12h), for which the time series dynamic is the main source of variability. An ARMA model of orders (p, q) is composed of an auto-regressive model of order p and a moving average of order q

$$\hat{\boldsymbol{y}}_t = \sum_{i=1}^{p} \phi_i \hat{\boldsymbol{y}}_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{II.46}$$

with $\epsilon_t, \epsilon_{t-1}, \ldots$ independent identically distributed random variables, and $\phi_1, \ldots, \phi_t$; $\theta_1, \ldots, \theta_t$ the parameters of the model. The ARMA models and its variants are stat-of-the-art benchmarking models for time series modelling. Evolutions of the ARMA model with generalized autoregressive conditional heteroscedasticity (GARCH) can also be used to model the autocorrelation of the covariance matrix for probabilistic forecasting [7, 110]. Despite their simple implementation, they have limitations in the non-linearities they can approximate.

Markov chains are another group of statistical methods that can be used for time series forecasting. They use a transition matrix to model the probability for the system to move from a certain state to the next, hence discretizing the system's state. They have been used for wind speed forecasting and remain a competitive baseline in recent research papers [33, 99, 162, 168].

Furthermore, statistical methods are widely used for the data pre-processing in forecast models [169], for data decomposition, dimension reduction or data correction. Wavelet transform is used to decompose non-stationary signals into a sum of orthogonal wavelet functions. It is widely used as a pre-processing step for wind speed time series which

are not stationary [170–173], and wave time series [174, 175]. Another method for non-stationary data decomposition is the empirical mode decomposition (EMD) developed by [176], that uses a finite number of intrinsic mode functions applied in the time domain to decompose the signal. The modes are creating by iteratively linking up the time series extrema with cubic interpolation, hence creating a series of nearly orthogonal functions. This method is widely applied in data pre-processing for wind speed time series to extract wind dynamics at different temporal scales [177–179]. Dimension reduction methods are useful for high dimensional problems such as weather forecasting. In particular, the statistical methods presented in this section have limitations for large dimensional problems. Feature selection is a common practice to estimate the input variables carrying the most information. Features can otherwise be selected based on their correlation with the output or using cross-validation. Feature selection strategies look for variables carrying the most information, while feature extraction methods construct latent variables from the input variables to carry the most information possible. One of the main methods for feature extraction in high dimension dataset is the Principal Component Analysis (PCA), in which the feature space is expressed as a sum of orthogonal vectors sorted by explained variance. These vectors and associated weights are obtained through singular value decomposition. It is a common tool to reduce data dimension for metocean forecasting [180, 181] and is used throughout this study.

Eventually analogs forecasting, a simple yet efficient class of models for dynamical systems forecasting has gained recent interest in the weather forecast community. Given a certain initial state, one can look into a catalogue of past weather situations for analogs, i.e. closest situations to the initial state given a certain distance metric. The evolution of these analogs in time can then be considered as possible scenarios from this initial state, and a probability distribution can be estimated [56, 57, 182–185]. Analogs can be implemented without relying on the output of physics-based models and can then be considered as end-to-end data-driven models. The analogs method is computationally cheap and is natively probabilistic. However, it requires long-enough catalogues and relies on the assumption that past weather situations are sufficient to accurately forecast a phenomenon from its initial state. Analogs are implemented as statistical probabilistic baseline in Chapter IV.

II.4.2 Machine learning methods

ARMA approaches are linear and parametric which limits their regression capabilities in most cases. Learning-based regression models have gained interest as a compromise between simple time series model and deep learning methods. They rely on the training of an algorithm to minimize a certain cost function based on training data.

Gradient boosting machines are regression models that use an ensemble of regression trained with a boosting algorithm [186]. It is described thoroughly in Section IV.4.3, and is a common statistical model for metocean forecasting. It is applied for the multivariate forecasting of wave conditions in a deterministic framework in [31], and for probabilistic forecasting in [4]. It is often combined with multi-quantile forecasting, as applied in [187] for wind speed forecasting in the GEFCom2014 competition [188].

Many other regression frameworks are implemented throughout the literature such as support vector machines [172, 189] in which hyperplanes are fitted to high dimensional data to separate it with a max-margin criterion, or generalised additive models that use a sum of smooth functions of the input to describe the output. In [4], they are applied with a boosting algorithm and a maximum likelihood framework to predict the parameters of a distribution for the probabilistic forecast of waves.

II.4.3 Advancement in deep learning forecasting models

Deep neural networks are a subset of machine learning algorithms that use several layers of interconnected neurons. Deep-learning-based forecast models have been surging in the recent years, due to the development and generalization of deep learning tools such as open-source coding libraries (PyTorch [190], Tensorflow [191], etc.), the use of Graphical Processing Unit for parallelization and the exploitation of large datasets. It triggered a wide interest in the forecasting community. We give here a literature overview on the applications of deep learning for metocean forecasting and offshore wind energy applications, and a more theoretical background is given in Section II.1.

Deep learning models use deep neural networks optimized by back-propagation to approximate highly non-linear functions. Provided that they have a sufficient number of hidden neurons and non-linear activation functions, deep neural networks are shown to be universal approximators [41]. Once trained, they are relatively fast to run, which makes them an attractive alternative to physics-based models such as numerical weather prediction. For all these reasons, they have been widely applied in metocean forecasting

[68].

Deep learning methods can be used at different stages of a forecast model, as a pre-processing, assimilation, prediction or post-processing step. Several types of architectures can be used, depending on the nature of the problem, convolutional architectures for high dimension data, recurrent architecture for time series, generative models, auto-encoders, residual networks, etc. All these methods have in common that they are trained to approximate a non-linear relationship between an ensemble of heterogeneous data and a specific end-user's product. In the following paragraph, we provide references of recent development on deep learning models applied to metocean forecasting and energy-related applications, giving examples for different modelling stages and architectures. This literature review is not exhaustive and the interested reader is referred to [68, 192] for a systematic overview of deep learning forecast models.

Metocean forecasting use a variety of data with spatio-temporal correlation at different scales. Convolutional neural networks (CNN) are deep learning architecture that use weight-sharing to deal with high dimension data with high correlation. They are widely used to deal with the spatialized data from numerical weather prediction. In [193], a 3D CNN is used to extract saptio-temporal features from numerical weather prediction for wind farm production forecast, showing improvement compared to principal component analysis and 2D-CNN. In [194], a specific architecture of CNN called U-Net is used to post-process cloud cover forecasts from the ARPEGE model, outperforming operational regression models using Random Forest at Météo France. The relevance of CNN for dealing with spatialized data make them a regularly used tool in metocean forecasting [32, 130, 143, 195–198]. CNN can be adapted to deal with time series by using causal convolution layers that ensure that only past information is used for forecasting. These models are called temporal convolutional networks (TCN) and have recently been applied to metocean forecasting [199–201].

More traditional approaches to time series are recurrent neural networks (RNN) which have a memory hidden state that is updated at each time step to take into account the temporal evolution. Long-Short-Term-Memory networks (LSTM) are variants of RNN that are designed to improve the gradient computation and avoid issues of gradient vanishing or exploding in RNN. They feature a cell state that stores long-term information. Recurrent architecture are common tools for metocean forecasting [171, 180, 202, 203]. [95] use LSTM, bi-directional LSTM and Gated Recurrent Unit (GRU) for forecasting wind speed and wave height at an offshore location for accessibility improvement for offshore

wind farms, showing the capabilities of recurrent networks for extracting information from input time series. [204] use the power of LSTM to forecast the probabilistic multivariate renewable energy generation for the next 24h based on forecast weather conditions and past observations. They show that LSTM can be used to capture both the temporal and cross-variables correlations. Recurrent architecture can be combined with convolutional layers to extract both the spatial and the temporal correlation [205]. The impact of wind speed on wave height is complex and involve different saptio-temporal scales. [29] used a combination of CNN and LSTM to learn the spatio-temporal dependencies between wind speed and wave height to improve wave height forecasting at a specific location.

Attention based models have recently gained a lot of interest, in particular for large language models with the development of the Transformer architecture. Attention mechanism use dynamic weighting with past time steps, allowing the network to select relevant past time steps [206]. [207] show that Transformers can bring interpretability to forecast models by highlighting the temporal relationships used for making a prediction. [143] use a combination of CNN for extraction spatial information, LSTM for temporal information, and attention mechanism to improve the selection of relevant input features for hourly probabilistic wind speed forecasting.

Generative methods have drawn attention due to their performances for images generation. They use deep neural networks to learn the distribution of the observations from which samples can be generated stochastically. Such methods are useful for forecasting as they can learn the underlying multivariate distribution of the system by directly maximizing the likelihood and be used for generating scenarios and probabilistic forecasts. The main generative architectures to date are Generative Adversarial Networks (GAN), Variational Auto-Encoders (VAE), Normalizing Flows (NF) and diffusion models. GAN jointly train a generator that is to generate samples from stochastic noise, and a discriminator that should differentiate true samples to generated samples [208]. They have been successfully adapted for rainfall nowcasting in [118] or ensemble forecast emulation in [17], two very high dimension problems. However, their adversarial construction makes them sensitive to mode collapse during training, hence requiring careful hyper-parameters tuning. VAE are deep latent models that jointly train an encoder and a decoder by minimizing the variational lower bound of the posterior distribution [209]. [210] implemented a recurrent VAE (VRAE) for probabilistic wind speed forecasting, with the ability of generating full wind speed sequences. Diffusion models are latent variable models that learn an iterative de-noising process starting from white noise. They have recently been applied to multivari-

ate time series forecasting in [211] using a LSTM conditioner for the de-noising process, showing great performance on state-of-the-art forecasting datasets. Finally, normalizing flows are a class of methods that use a composition of invertible non-linear functions to conditionally transform a tractable latent distribution into an arbitrarily complex distribution [119]. They have the advantage of offering an exact computation of the likelihood of the posterior compared to GAN and VAE. [44] compared VAE, GAN and normalizing for a renewable energy load day-ahead forecasting from the Global Energy Forecasting Competition 2014 open-source dataset. The authors showed that NF are more accurate in forecast quality and value, while being easier to implement than VAE and GAN. A thorough introduction of normalizing flows is given in Section II.1.4.

Recent advancement in end-to-end deep learning models for weather forecasting have opened-up completely new paradigms, where numerical models are completely replaced by deep learning models. For specific applications such as rainfall nowcasting, deep learning has proved in [118] to increase forecast quality and value up to 90 minutes ahead against state-of-the-art advection methods. The use of Graph Neural Networks recently brought a significant breakthrough in the field of end-to-end weather forecasting. Training on a reanalysis database to predict 227 variables per grid point on a 0.25° grid for 10 days, the DeepMind's 37 million coefficients GraphCast [11] significantly improves forecast skill compared to operational forecast system, while requiring only 1min of single TPU time to create a 10 days ahead prediction. Recent work at the European Centre for Medium-Range Weather Forecast describe the first operational end-to-end deep learning based global weather forecast [12]. Their computational cost is brought down significantly by a team at the INRIA [10]. This is a strong incentive for continued research in deep learning models for weather forecasting. [66] identify specificities of weather forecasting problems that require dedicated machine learning developments: the coupling in large spatio-temporal scales, the need for uncertainty quantification in a non-Gaussian environment, the auto-correlation of time series and dynamic correlations in the system, the imbalance of data, the importance of extreme events and the amount of noisy, imperfect data to assimilate.

A great amount of deep-learning-based forecast papers are dealing with specific datasets and do not permit a fair comparison between methods [212]. In this context, it is difficult to conclude which model is definitely the best, as to some extent several architectures can yield the same range of performances given sufficient tuning effort. That's why the existence of competitions and open-source datasets is also paramount for supporting the development of deep learning architecture for weather forecasting. Weather datasets such

as WeatherBench [213] developed by Google DeepMind or MétéoNet [19] proposed by Météo France are representatives of the complexity of weather data. For energy forecasting applications, competitions such as the Global Energy Forecasting Competition [188] are great steps in the development of forecast models for specific applications. However, compared to other machine learning communities, the availability of benchmark datasets, established comparison metrics and thorough specifications is still largely insufficient.

Chapter III

Design of optimal observation networks of offshore wind speed: a data-driven sparse sampling approach

Introducing the article

Offshore wind project need metocean time series for design, construction and operations. Wind speed measurements in particular are crucial for power production estimate, wind turbine layout optimisation, system design, wind turbine control and accessibility forecasting for maintenance operations. In link with the development of data-driven methods, long-term in-situ time series should be gathered. Weather patterns have an important daily, seasonal and multi-annual variability so several years of data are required to train and evaluate the models. This should anticipated by the installation of long-term measurement facilities.

At the regional level, the development of offshore wind energy requires the deployment of large scale observation networks. In France, the wind energy roadmap paves the way for the installation of 40 GW of offshore wind energy projects across its maritime facades. Strategical facade documents are issued at the national level to schedule the installation of successive wind farms. To support the choice of development areas, and to provide long-term time series to call for tender participants, regional-scale wind speed measurement networks should be deployed. Offshore in-situ measurements being scarce and expensive, the optimization of the observation network should be considered. In this chapter, we investigate the following questions:

* Given a geographical area, how many sensors are required to reconstruct the wind field?

* Where to place those sensors to optimally reconstruct the wind speed?

We propose a data-driven approach based on the training of a Gaussian mixture model on numerical weather prediction data. We define an evaluation framework for this sparse sampling problem that is adapted to the offshore wind speed reconstruction application. By using unsupervised clustering on model data, we can find salient points on the model grid, and we show that the obtained reconstruction error is lower than when using state-of-the-art semi-empirical methods. The optimal number of sensors to be deployed is discussed in terms of statistical and empirical criteria. Eventually, we propose optimal sampling strategies for three of the main offshore wind development areas in France.

This work has been used by the French weather institute Météo France to provide guidelines to the general direction of energy and climate (DGEC) for their measurement campaigns. It is published as such in *Wind Energy Science* as *Robin Marcille, Maxime Thiébaut, Pierre Tandéo, Jean-François Filipot, Gaussian mixture models for the optimal sparse sampling of offshore wind resource, 2022*

III.1 Introduction

Offshore wind energy is key in the decarbonation of the global energy production and the reaching of net-zero targets as developed in [214]. With 11 million km² of territorial waters under French jurisdiction and 20,000 km of coastline, France has an extensive and windy seafront. It benefits from the second largest offshore wind potential in Europe, after the United Kingdom with up to 80 GW of foundation-based offshore wind and 140 GW of floating offshore wind that could be exploited according to [63]. Offshore wind can then be a leading sector for the development of renewable energies in France. The French roadmap currently plans 1 GW of tender per year from 2024 onwards for fixed and floating wind farms. This was confirmed and reinforced in early 2022, with 40 GW of installed capacity envisioned by 2050.

During the development phase of a wind project, the wind resource assessment is a key step to determine its financial feasibility. It can be carried out with numerical weather prediction (NWP) hindcast data such as WRF (Weather Research and Forecasting model) data. However, field observations are necessary to estimate the uncertainties of the models and to assess higher resolution wind dynamics [215].

LiDARs, standing for Light Detection And Ranging, are remote sensing devices that measure wind speed using lasers. Floating LiDARs are certified devices for offshore wind

resource assessment, they are LiDAR units integrated onto a standalone floating structure. These wind sensors offer the potential for reduced costs compared to meteorological masts [216], however, they can be expensive to install and require regular maintenance. Their number and siting thus need to be optimized in order to compose an optimal network of sensors in an offshore wind development area. Such networks are expected to capture most of the dominant wind dynamics from a minimum number of sensors.

Numerous efforts have been undertaken in different scientific fields to optimize sparse sensor siting, a combinatorial problem not solvable by standard approaches such as convex optimization. Sparse sampling is about selecting salient points in a highly dimensional system. It then requires a dimension reduction of the data, such as the use of Empirical Orthogonal functions (EOF). EOF analysis projects the original data onto an orthogonal basis derived by computing the eigenvectors of a spatially weighted anomaly covariance matrix. Therefore, EOF of a space-time physical process can represent mutually orthogonal space patterns where the data variance is concentrated, with the first pattern being responsible for the largest part of the variance, the second for the largest part of the remaining variance, and so on. EOF are then very useful for the data reduction of any complex data set such as climate data. By projecting the original data onto a limited subset of relevant orthogonal vectors, it reduces the dimensionality of the system and helps explain the variance of the data. In the past few decades, EOF analyses were used to study spatio-temporal patterns of climate variability, such as the North Atlantic oscillation, the Antarctic Oscillation or the variability of the Atlantic thermohaline circulation (e.g., [217–220]).

EOF are often at the origin of methods employed to determine the optimal sensors' locations for signal reconstruction. In the field of geoscience, [52] employed simulation results from different regional ocean models to define an efficient sensor placement. The authors used the EOF technique to determine the spatial modes of different simulated ocean dynamics systems. The extrema of the EOF spatial modes were found to be good locations for sensors' placement and accurate field reconstruction. [221] added to the Empirical Orthogonal Functions' extrema (EOF extrema) method a constraint on the cross products of EOF to select the sensors' locations, and applied it to Pacific sea surface temperature reconstruction. Using the same kind of constrained EOF analysis, [222] proposed a data-driven framework based on a Proper Orthogonal Decomposition (POD) to determine the optimal locations for power system oscillation monitoring and state reconstruction. In this study they selected iteratively the locations with highest POD amplitude and lowest

cross coupling between the modes. In [223], the EOF extrema are used for ocean dynamics reconstruction, introducing an exclusion volume to avoid redundancy, account for gappy data and for uncertainty.

[53] proposed a data-driven method based on a QR pivoting greedy algorithm on a reduced basis to determine optimal sensors' placements for face recognition, global sea surface temperature and flow reconstruction around a cylinder. The QR pivoting method decomposes a matrix into an orthogonal matrix and an upper triangular matrix using columns pivot. By iteratively selecting the column with the highest two-norm as pivot, this algorithm for QR factorization is suited for the selection of salient points. In [224], the QR-pivot decomposition is modified to include cost constraints and is applied on the three same data sets. The QR greedy algorithm described in these studies is a often used in recent studies due to its easy implementation.

Recent studies proposed innovative methods to improve the capabilities of sparse sampling. To improve the performance of the reconstruction, [225] proposed a method for grid augmentation to allow for continuous sensor placement, off-grid sensor selection and convex optimization problem formulation. In [226], the authors took inspiration from insects' neural activation during flights to derive a sparse sampling method in complex flows to create an encoder for flight mode classification including both the spatial and temporal dependencies of the data. In [227], the use of Voronoi tessellation, a method to optimally partition the space into n cells given n input points using a distance measure d, helps creating a viable input for super-resolution from sparse sensors using a Convolutional Neural Network (CNN). This reconstruction technique is then tested on sea surface temperature reconstruction globally, showing the possibility to use sparse sampling on very high dimension problems.

The optimal sensors' placement problem has also been investigated for wind energy measurements applications. [228] uses the QR greedy algorithm described in [53] to determine the optimal locations of sensors to improve the overall estimation precision of the flow field within a wind farm. In this study, the number of sensors is directly computed using a user-defined threshold with regards to reconstruction error. A similar strategy is implemented in this article as presented below. The obtained results show good performance compared to randomly selected grid points, with an improvement of 8% in flow field reconstruction, and shows the interest in applying sparse sampling methods to the wind energy sector. At even finer scales, [229] uses low-dimensional classifiers applied to the Proper Orthogonal Decomposition of a LES wake simulation to obtain sensors' locations

for the reconstruction of wind turbine wakes. Using the method of sparse sensor placement optimization for classification described in [54], it shows the interest of sparse sampling for the control of wind turbines, using a Deep Learning algorithm to predict the wake fluctuations from sensors' measurements. Results show that most sensors are placed in the transition region, and the reconstruction yields to more than 92% correlation between predicted and real values.

However, to the best of our knowledge, such methods were never applied at the regional scale for wind energy resource assessment. In our opinion this is due to site selection procedures at the political level that do not necessarily rely on wind resource assessment at the regional level, and to smaller required spatial scales at the wind farm developer level, where only one or two sensors are deployed at the extremities of the area, assuming spatial representativity. The application of sparse sampling methodologies to offshore wind reconstruction is an addition of this work. Using NWP spatial wind data as input, the study proposes an unsupervised clustering framework for the identification of salient points in the spatial grid, similar to what can be obtain through EOF extrema analysis in [52] or QR pivoting in [53]. In the application-driven experimental set-up of this study, the two state-of-the-art methods fail to capture wind dynamics at the regional level. Unsupervised clustering automatically discriminates points that are too similar, making it a good candidate for sparse sampling in this case, while keeping the whole method simple and easily implementable.

The objective of the present study is twofold. For conducting offshore wind resource assessment of any targeted area:

1. What is the optimal number of offshore wind sensors to be deployed to best characterize the wind resource?
2. What is the optimal location of each wind sensor?

The optimal number of sensors refers to a trade-off between wind field reconstruction accuracy, and overall cost and computational cost. The optimal locations given a certain number of sensors is the configuration giving the lowest reconstruction error. The two aspects are presented in this work, though realistic cost considerations are not covered.

To do so, this paper presents a data-driven method based on NWP data unsupervised clustering to estimate optimal sensors' locations for offshore wind field reconstruction using a Gaussian Mixture Model. It is compared to state-of-the-art methods used in the above literature (EOF extrema, QR pivoting, randomly selected sensors). The method is then implemented on three areas identified for offshore wind energy development in

France. An optimal wind sensors network is proposed for each area, to help for the development of offshore wind energy in France.

III.2 Study data set

III.2.1 Study areas

The three areas investigated in this study are located off the coast Normandy, off the coast of Southern Brittany and in the Mediterranean sea, three major development areas for offshore wind in France with numerous planned offshore projects, listed in Table III.1, with future tender processes for respectively 1.5GW of fixed offshore wind, 250MW of floating offshore wind and 2 x 250MW of floating (expected date of commissioning in 2030).

With water depth not exceeding 50 m (Figure III.1), the area located off the coast of Normandy area is particularly suitable for the deployment of fixed offshore wind farms. Current projects will be installed off the coast of Fécamp, Courseulles-sur-Mer and Dieppe - Le Tréport (Figure III.1). The total capacity of each wind farms will be 450-500 MW with a starting date of commissioning expected in 2023-24 (Table III.1). In addition, the French Government has recently announced a new project of a wind farm located 32 km off the coast of Normandy (Figure III.1). This future wind farm will generate 1 GW. The starting date of commissioning is expected by 2028.

The area off the coast of Brittany is endowed with water depth up to 100 m which make it a very favourable area for the development of floating wind farms. The French Government aims at developing 250MW of floating wind energy in the area (Figure III.1).

Because of its very favorable and regular wind regimes and deep bathymetry, the Mediterranean Sea has significant wind potential for floating wind energy. This led to the development of three pilot floating wind farm projects (Leucate, Gruissan and Provence Grand Large) in the gulf of Lion. These projects will rely on 3 full scale 8-10 MW floating turbines, whose generated power will be injected in the French power grid by 2022-2023 (Table III.1). In addition, two commercial wind farms with power capacity over 250 MW each will be in operation by 2029.

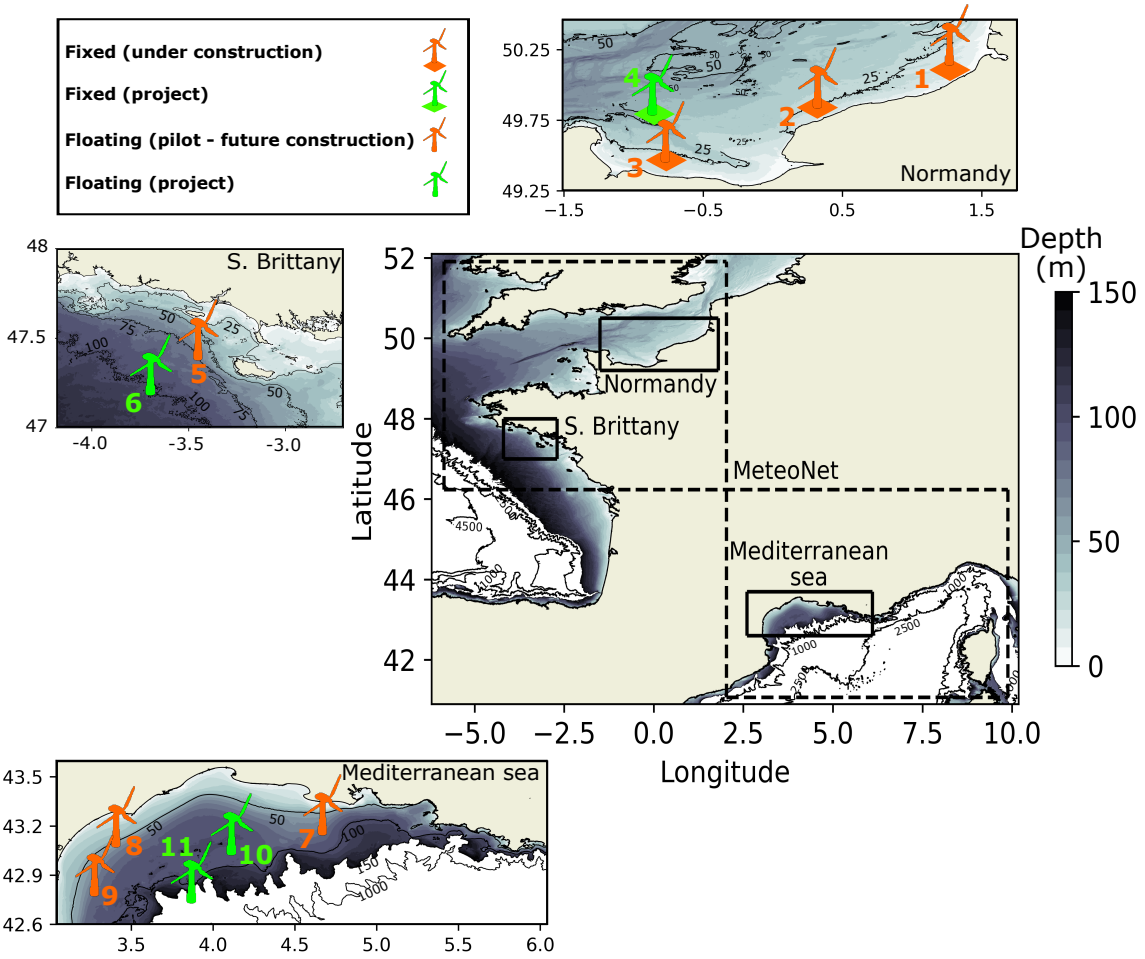


Figure III.1 – Overview of the French coasts with the bathymetry. The color shading shows the water depth until 150 m. Areas with water depth exceeding 150 m are shown in white. Black contours are used to identified depth of 1000m, 2500 m and 4500 m. The three study areas are shown by black rectangles. Each area is presented on different panels where the locations of the future foundation-based and floating wind farms are shown with their different stages of development. On the main panel, the dashed black lines delimit the areas covered by the MeteoNet data set.

III.2.2 The MeteoNet data set

MeteoNet is a meteorological data set developed and made available by Meteo-France [19], the French national meteorological service. The data set contains full time series of satellite and radar images, NWP models and ground observations. The data covers two geographic areas of 550km x 550km on the Mediterranean and Brittany coasts (Figure III.1), and spans from 2016 to 2018. Hourly 10-meter wind output of the high resolution NWP

p

Table III.1 – Characteristics of the foundation-based (Normandy) and floating (Southern Brittany, Mediterranean sea) future wind farms planned for the next decade on the study areas [230].

Areas	Wind Farm	Index in Figure III.1	Capacity (MW)	Number of wind turbines	Expected date of commissioning
Normandy	Dieppe- Le Tréport	1	496	62	2026
	Fécamp	2	497	71	2023
	Courseulles-sur-mer	3	448	64	2025
	AO4 call for tender	4	1 000	N/A	2030
Southern Brittany	AO5 call for tender	5	250	N/A	2025-2030
Med. sea	Faraman - Port-Saint-Louis-du-Rhône	6	24	3	2023
	Gruissan	7	30	3	2024
	Leucate - Le Barcarès	8	30	3	2024
	AO6 call for tender	9 and 10	2 x 250	N/A	2028-2029

model AROME are available. AROME is operational at Meteo-France since December 2008 [231]. It was designed to improve short range forecasts of severe events such as intense Mediterranean precipitations, severe storms, fog or urban heat during heat waves. The physical parametrisations of the model come mostly from the Méso-NH research model whereas the dynamic core comes from the Non-Hydrostratic model ALADIN [232]. The resolution of the AROME grid is 1.3 km. The model is initialized from data assimilation derived from the ARPEGE-IFS variational assimilation system [233] and adapted to the AROME finer resolution.

For each area of interest, the 10-meter zonal (u) and meridional (v) wind speed are extracted from AROME. The open-source MeteoNet data set only contains surface parameters of temperature, humidity, pressure and precipitation, and 10-meters wind speed (u10, v10), which are considered in this study. The assumption is then made that relevant measurement points at 10 meters are equally relevant for hub height estimation, though this assumption should be tested with a suitable data set. Since the focus is on offshore wind, grid points at land were excluded from the analysis. The characteristics of each area are then the following:

— Normandy: 4272 grid points ($\sim$ 7 000 km^2).

— Southern Brittany: 1837 grid points ($\sim$ 3 000 km^2).

— Mediterranean sea: 3571 grid points ($\sim$ 5 800 km^2).

A total of 65 days ($\sim$ 6%) of the 3-year data set are unusable due to largely missing data. The missing data days are similar for each area and were removed from the analysis.

III.3 Preliminaries

III.3.1 Problem statement

The problem tackled in the presented work is to find D measurement points out of K NWP grid points to minimize the reconstruction error of the offshore wind field. A formalism for this sparse sampling problem is proposed in this section.

In all that follows, the full state matrix $\boldsymbol{X}$ refers to the concatenation of zonal and meridional wind speeds on the K grid points of the model for all time steps. The list of sensors' locations $\boldsymbol{\gamma}$, which is the output of the methods described in this paper, contains the locations of the D sensors to sample the offshore wind field. The associated sparse measurement matrix $\boldsymbol{Y}_\gamma$ corresponds to the measured zonal and meridional wind speed at the $\boldsymbol{\gamma}$ locations for every time step.

The formalism developed in this section is applied to the MeteoNet data set presented in Section III.2.2. The data set is split into a training and testing period. The training is performed on two thirds of the data set, composed of years 2016 and 2017, while the methods are scored on year 2018. By taking an integer number of year, the seasonality bias of weather data is limited.

III.3.2 Reduced order model

The reduced order model used to decrease the dimension of the input data is the Empirical Orthogonal Function analysis (EOF). Also known as Principal Component Analysis (PCA), it decomposes the data set onto an orthogonal basis. Practically, it is linked to the singular value decomposition of a matrix $\boldsymbol{X}$ such that:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{III.1}$$

With $\boldsymbol{\Sigma}$ a diagonal matrix of positive σ_k singular values, $\boldsymbol{U}$ a matrix whose columns are the vectors of the orthogonal basis, and $\boldsymbol{V}$ the weights of the associated vectors.

The singular vectors are orthogonal vectors on which the variance of the projected

data is maximized. The diagonal elements of $\mathbf{\Sigma}$ are sorted per value, and are equal to the percentage of variance of the data set explained by each principal components. The variance explained by the first r EOF is then:

$$\frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_{i}^{K} \sigma_i^2} \tag{III.2}$$

The number of EOF to use in the reduced order model can be set so that the variance explained by the reduced basis is above a certain threshold.

For the study dataset, EOF of zonal and meridional wind speed are computed. In the NWP model, the grid points are strongly correlated spatially, hence, only a low number of EOF is needed to describe the vast majority of the data set variance. The number of EOF was set to 10, both for the zonal and meridional components of wind so that the reduced basis explains more than 95% of the total variance for the 3 areas. The $\mathbf{\Phi}^r$ reduced basis is then the concatenation of the $\mathbf{\Phi}_u^{r/2}$ and $\mathbf{\Phi}_v^{r/2}$ EOF for zonal and meridional wind speed, with $r = 20$.

III.3.3 Sparse sampling formalism

III.3.4 State description

Let us consider a system described by its time-varying state $\boldsymbol{X}(t)$ that evolves according to unknown non-linear dynamics. It can be described on an orthogonal basis $\{\phi_i\}$ (e.g. the EOF) as:

$$\boldsymbol{X}(t) = \sum_{i=1}^{K} a_i(t)\boldsymbol{\phi}_i \tag{III.3}$$

To reduce the complexity of the model, the state of the system can be approximated using the first r modes:

$$\boldsymbol{X}(t) \approx \sum_{i=1}^{r} a_i(t)\boldsymbol{\phi}_i = \boldsymbol{X}^r(t) = \mathbf{\Phi}^r \boldsymbol{a}(t) \tag{III.4}$$

Where $\mathbf{\Phi}^r$ is the reduced basis matrix containing the first r modes, and $a_i(t)$ are the time varying coefficients of the system's state on the reduced basis.

The given system is then sampled according to a set of index $\boldsymbol{\gamma} = [\gamma_1, ..\gamma_D] \in [1, K]^D$, $\gamma_i \neq \gamma_j$ which represents the sensors' locations. From this, a sampling matrix is constructed $\boldsymbol{C}_{\boldsymbol{\gamma}} \in \mathbb{R}^{D \times K}$ that extracts the D measured locations out of the K grid points of the full state. The sampling matrix is composed of lines of zero with ones at the sensors' locations.

With the canonical basis vectors $\boldsymbol{e}_{\gamma_j} = (\boldsymbol{\delta}_{\gamma_j,k}) \in \mathbb{R}^K$, the sampling matrix is:

$$\boldsymbol{C}_{\boldsymbol{\gamma}} = \begin{bmatrix} \boldsymbol{e}_{\gamma_1} & \boldsymbol{e}_{\gamma_2} & \cdots & \boldsymbol{e}_{\gamma_D} \end{bmatrix}^T \tag{III.5}$$

The sparse measurement matrix $\boldsymbol{Y}_{\boldsymbol{\gamma}}$ is then obtained by multiplying the full state $\boldsymbol{X}$ by the sampling matrix $\boldsymbol{C}_{\boldsymbol{\gamma}}$:

$$\boldsymbol{Y}_{\boldsymbol{\gamma}}(t) = \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{X}(t) \tag{III.6}$$

III.3.5 Full state reconstruction from sparse measurements

From the sparse measurement matrix, the full state is reconstructed using the coefficients of the reduced basis. A linear model is constructed to link the matrix of EOF coefficients, $\boldsymbol{a}$, to the sparse measurement matrix $\boldsymbol{Y}_{\boldsymbol{\gamma}}$:

$$\boldsymbol{a} = \boldsymbol{\beta} \boldsymbol{Y}_{\boldsymbol{\gamma}} + \epsilon \tag{III.7}$$

With ϵ an additive Gaussian error.

The model fitting is performed on the training split of the data set. Let $\boldsymbol{Y}_{\text{train}}$ be the sparse measurement matrix on the training split, and $\boldsymbol{a}_{\text{train}}$ the true coefficients of the full state on the reduced basis for the training split. Using the Ordinary Least Squares formula, the $\boldsymbol{\beta}$ matrix can be estimated as:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Y}^T_{\boldsymbol{\gamma},\text{train}} \boldsymbol{Y}_{\boldsymbol{\gamma},\text{train}})^{-1} \boldsymbol{Y}^T_{\boldsymbol{\gamma},\text{train}} \boldsymbol{a}_{\text{train}} \tag{III.8}$$

On the test data set, only the wind speed measurements at the $\boldsymbol{\gamma}$ locations are available. The coefficients of the reduced basis are computed using the least squares matrix estimated on the training data set:

$$\hat{\boldsymbol{a}}_{\boldsymbol{\gamma},\text{recons}} = \hat{\boldsymbol{\beta}} \boldsymbol{Y}_{\boldsymbol{\gamma},\text{test}} \tag{III.9}$$

And the full state is reconstructed using the reduced basis:

$$\hat{\boldsymbol{X}}_{\boldsymbol{\gamma},\text{recons}} = \boldsymbol{\Phi}^r \hat{\boldsymbol{a}}_{\boldsymbol{\gamma},\text{recons}} \tag{III.10}$$

III.3.6 Reconstruction error

The reconstruction $\hat{\boldsymbol{X}}_{\boldsymbol{\gamma},\text{recons}}$ is then compared to the reconstruction with perfect knowledge on the reduced basis coefficients $\boldsymbol{X}_{\text{real}} = \boldsymbol{\Phi}^r \boldsymbol{a}_{\text{real}}$, assuming that the actual coefficients of the reduced basis are perfectly known.

The reconstruction error associated with the sensors' locations $\boldsymbol{\gamma}$ is then the root mean squared error of the reconstructed state:

$$E_{\boldsymbol{\gamma},\text{recons}} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\hat{\boldsymbol{X}}^k_{\boldsymbol{\gamma},\text{recons}}(t) - \boldsymbol{X}^k_{\text{real}}(t)\right)^2} \tag{III.11}$$

The optimisation problem that needs to be solved can then be stated as the minimization of the reconstruction error over all locations' combinations $\boldsymbol{\gamma}$ and number of sensors D:

$$\arg\min_{\boldsymbol{\gamma},D} E_{\boldsymbol{\gamma},\text{recons}} \tag{III.12}$$

III.4 Sparse sampling methods used in this study

In this section, the methods applied for the sparse sampling are described in detail. The novel data-driven method based on Gaussian Mixture Model is presented together with the three baselines emerging from the literature review. These are the random selection of locations (Monte Carlo), the dominant spatial modes' extrema (EOF extrema), and the QR greedy algorithm (QR pivots). All the methods described in this section should output a list of sensors' locations $\boldsymbol{\gamma}$ given a number of sensors D.

III.4.1 Baseline methods

The selected baseline methods are emerging from the literature as simple yet efficient methods for sparse sampling in different situations. They are implemented to compare their performances with the Gaussian Mixture Model for this specific application.

III.4.2 Monte Carlo simulations

The first baseline consists in picking random sensors' locations. For each area, and for a number D of sensors ranging from 1 to 10, a hundred random combinations of locations $\boldsymbol{\gamma} \in P_D([1, K])$ are considered. For each $\boldsymbol{\gamma}$ combination of sensors' locations,

the reconstruction error is computed. From this ensemble of simulations, statistics on the reconstruction error are computed.

The median Monte-Carlo scenario for each area and number of sensors is then considered a benchmark for the study. It also gives information about the spread in reconstruction error resulting from all possible combinations.

III.4.3 Dominant spatial modes' extrema

In [52], the extrema of the spatial dominant modes are found to be relevant locations if not optimal for the reconstruction of the flow field. Those points can be seen as salient points, that best characterize the spatial modes. It is then intuitive to select those to reconstruct the full state from the reduced basis. How many extrema are chosen from each variable and mode is studied specifically in [234], it is empirical and thus case specific.

In the case study of [234], the EOF decomposition gives modes that are highly spatially correlated. Moreover, in this study, points nearby the coast are influenced by the orography and show strong variability. Hence, sorting the points per coefficient and selecting the N first ones will lead to the selection of neighboring points, and/or irrelevant coastal points for our performance metric.

The extrema are then chosen manually, as performed in [52], from the visualization of the first EOF for both zonal and meridional wind speed. For each parameter and EOF rank, the extrema are selected, and discarded if redundant (manual process). Then, they are sorted per absolute value and per EOF number for the two parameters.

The selected input points from EOF extrema for the Mediterranean Sea are shown in Figure III.2. From the first to the fourth EOF for zonal and meridional wind, the extrema are selected if they are not too redundant or close to the coast / border. This unsatisfactory workflow is a way to ensure minimum relevance for the obtained sensors array.

III.4.4 QR pivots

The QR decomposition of the reduced basis matrix $\boldsymbol{\Phi}^r$ is the finding of two matrix $\boldsymbol{Q}$ orthogonal and $\boldsymbol{R}$ triangular superior such as:

$$\boldsymbol{\Phi}^r = \boldsymbol{Q}\boldsymbol{R} \tag{III.13}$$

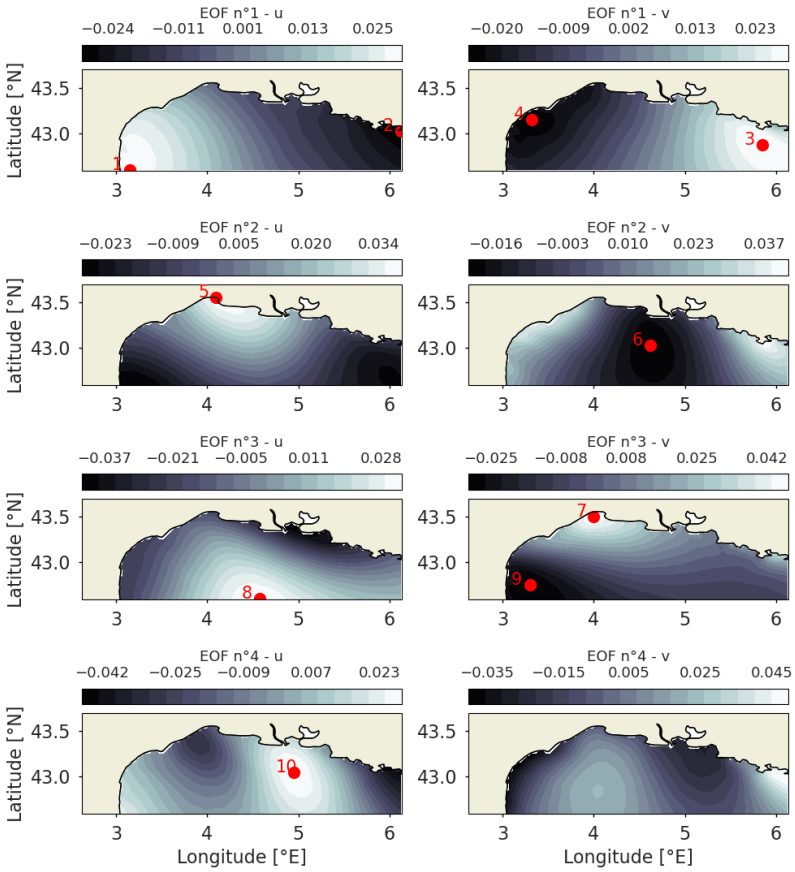


Figure III.2 – Selected sensors for the EOF extrema baseline in the Mediterranean Sea. The EOF coefficients are displayed as background, and the selected salient points, and their associated ranks are displayed as red dots. The two columns are zonal (u) and meridional (v) wind speed, and the rows correspond to the EOF rank.

The $\boldsymbol{Q}$ and $\boldsymbol{R}$ matrix are obtained using Gram-Schmidt process [235], which consists in iteratively removing each column's orthogonal projection onto a pivot column. The QR algorithm can be performed using column pivoting, i.e., at each iteration, the matrix $\boldsymbol{\Phi}^r$ is multiplied by a permutation matrix $\boldsymbol{P}$ such that the column taken for pivoting has the maximum two-norm. The decomposition is then:

$$\boldsymbol{\Phi}^r \boldsymbol{P}^T = \boldsymbol{Q}\boldsymbol{R} \tag{III.14}$$

The permutation matrix $\boldsymbol{P}$ is constructed so that the diagonal elements of $\boldsymbol{R}$ are decreasing. It is applied to the matrix of the reduced basis to identify pivot locations. It then

contains the ranked index of sensors' locations to build the sensors' locations list:

$$\gamma_j = \boldsymbol{P}_{jj} \quad \forall j \in [1, D] \tag{III.15}$$

The QR pivots method is described in [53] as a simple yet efficient method for sparse sensors' placement. It is used to determine model data driven sensors networks to reconstruct flow fields for the flow past a cylinder and the sea surface temperature retrieval, two situations that are analogous to the case study. It is even tuned to include costs constraints for the search of Pareto optimal sensors' placement in [224]. For wind fields estimation, it was applied to Computational Fluid Dynamics data in [228] to best reconstruct the flow in a wind farm. All in all, it represents a simple yet competitive baseline method for spare sensor placement.

III.4.5 Gaussian Mixture Model clustering

The proposed method in this study uses unsupervised clustering of the data to define sensors' locations. Gaussian Mixture Models use machine learning to fit multivariate normal distributions on the data.

III.4.6 Gaussian mixture

A Gaussian mixture model (GMM) is a probabilistic model for representing normally distributed sub-populations within an overall population [236]. Each Gaussian distribution represents a group of points, i.e., cluster. The model is a mixture, i.e., superposition, of multivariate Gaussian components which define a probability distribution $p(x)$ on the data:

$$p(\boldsymbol{x}) = \sum_{j=1}^{D} \pi_j \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \tag{III.16}$$

π_j being the mass of the Gaussian component j, with $0 \leq \pi_j \leq 1$ for all $j = 1, ..., D$ and $\sum_{j=1}^{D} \pi_j = 1$. $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ being the Gaussian density distribution such that:

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^{\mathrm{r}} det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right) \tag{III.17}$$

with $\boldsymbol{x}$ being the r-dimensional input vector, $\boldsymbol{\mu}$ the r-dimensional mean vector, and $\boldsymbol{\Sigma}$ $(r \times r)$ the covariance matrix.

III.4.7 Expectation — Maximization algorithm

The core of GMM lies within Expectation Maximization (EM) algorithm, developed by [55]. It iteratively modifies the model's parameters to maximize the log-likelihood of the data.

The log-likelihood, $\log(\mathcal{L})$, of the observations is given by:

$$\log(\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})) = \sum_{k=0}^{K} \log \left(\sum_{j=1}^{D} \pi_k \mathcal{N}(\boldsymbol{x}_k | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right) \tag{III.18}$$

Then the empirical means, $\boldsymbol{\mu}_j$, covariances, $\boldsymbol{\Sigma}_j$ and weights, π_j of the different clusters are computed. The weights (mixing coefficients) represent the mass of the different clusters. The mass of the cluster is the proportion of data points assigned to this cluster. For the first iteration, the mean and covariance matrices are initialized randomly, and the weights matrix is equal for each cluster.

The second step of the algorithm is the expectation step, E-step. The model parameters are updated to increase the log likelihood of the data. For each data point, $\boldsymbol{x}_k$, the probability that this point belongs to the cluster, c, is computed such that:

$$r_{kc} = \frac{\pi_c \mathcal{N}(\boldsymbol{x}_k | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)}{\sum_{j=1}^{D} \pi_j \mathcal{N}(\boldsymbol{x}_k | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{III.19}$$

The E-step computes those probabilities using the current estimates of the model's parameters. In this step, "responsibilities" of the Gaussian distributions are computed. They are represented by the variables r_{kc}. The responsibility measures how much the c-th Gaussian distribution is responsible for generating the k-th data point using conditional probability.

The third step is the maximization step, M-step. In this step, the algorithm uses the responsibilities of the Gaussian distributions (computed in the E-step) to update the estimates of the model's parameters. π_c, $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are updated using the following equations:

$$\pi_c = \frac{\sum_{k=1}^{K} r_{kc}}{K} \tag{III.20}$$

$$\boldsymbol{\mu}_c = \frac{\sum_{k=1}^{K} r_{kc} \boldsymbol{x}_k}{\sum_{k=1}^{K} r_{kc}} \tag{III.21}$$

$$\boldsymbol{\Sigma}_c = \frac{\sum_{k=1}^{K} r_{kc} \left(\boldsymbol{x}_k - \boldsymbol{\mu}_c\right)^2}{\sum_{k=1}^{K} r_{kc}} \tag{III.22}$$

These updated estimates are used in the next E-step to compute new responsibilities for the data points. This algorithm is applied iteratively until algorithm convergence, when the log likelihood of the data is maximized. The strict monotony of the likelihood in the E-M algorithm is demonstrated in [237].

III.4.8 Optimal number of clusters

GMM requires to impose as input the number of clusters in the model. The optimum number of clusters can be defined through the calculation of the Bayesian Information Criterion (BIC) score [238]:

$$\mathrm{BIC} = -2\ln(\mathcal{L}) + G\ln(K) \tag{III.23}$$

with $\mathcal{L}$, the maximized value of the likelihood function of the model, G the number of parameters in the mean vectors and covariance matrices of the Gaussian components, and K, the number of data points. The BIC score penalizes too complex models to avoid the over-fitting of the data set. In this way, it limits the number of components of the GMM with the $G\ln(K)$ term.

The lower is the BIC, the better is the model. However, the curve of the BIC score can be monotone, and the identification of a minimum, i.e., the optimal number of clusters can be difficult. An alternative is the calculation of the gradient of the BIC score. The identification of the optimal number of clusters is hence done by the identification of an elbow in the curve of the gradient of the BIC score. The elbow is often not directly associated with one single specific number of clusters but rather encompassed two or three possible solutions. Thus, one can say that the gradient of the BIC score gives an indication on the range of optimal number of clusters. An extra step is required to determine the optimal number of clusters. In this study, this step is done through the determination of an error reconstruction threshold of the wind field. The number of clusters associated with the minimum error is considered as optimal for the GMM.

III.4.9 Implementation for the study case

Figure III.3 shows the workflow in this study. A two-dimensional data set composed of $K = 3571$ grid points with 20 EOF features is used to feed the GMM. The 20 features are composed of the 10 first EOF of zonal and meridional velocities. The clustering is

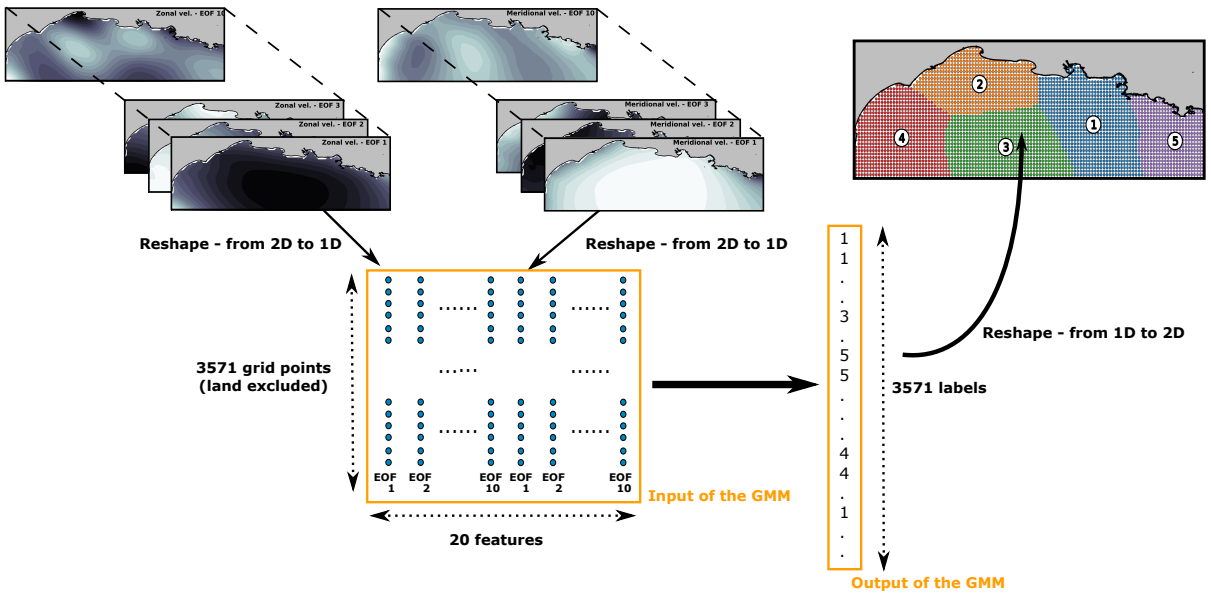


Figure III.3 – Schematic of the clustering procedure of wind data using Gaussian Mixture Models. It is illustrated for the unsupervised clustering of the Mediterranean Sea wind field.

then optimized spatially, so the entries (the grid points) are assigned to clusters, based on their features (their coefficients on the first 20 EOF). The output of the model is then a list of labels for each grid points, creating spatial clusters in the study areas.

The GMM procedure will find clusters of grid points that are correlated in the reduced basis. The centroids of the clusters, i.e., the point of maximum likelihood for a given cluster, are then chosen to be sensors' locations, as they are the most representative points of the clusters:

$$\gamma_j = \arg\max_{\boldsymbol{x}} \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \forall j \in [1, D]. \tag{III.24}$$

III.5 Results

In this section, the methods presented in Section III.4 are implemented on the three identified areas (Mediterranean Sea, Normandy and Southern Brittany) and compared with respect to the wind field reconstruction error. A method for the selection of an optimal number of sensors is described, and the suggested sensors' locations for the three areas are given.

III.5.1 Optimal number of sensors

The number of sensors to place on the grid is an input of the GMM. The BIC score described in Section III.4.8 computes a trade-off between the likelihood of the obtained distribution, and the complexity of the model. Being sensible to the likelihood of the model and to its complexity, it is usually used to determine the number of clusters for the GMM by finding its minimum. However, there is no guarantee that there will be a minimum BIC score corresponding to an optimal number of clusters, and there is no guarantee that this number of clusters is actually optimal for the considered metric. Indeed, this metric is a heuristic criterion to hint the trade-off between accuracy and complexity, to avoid over-fitting. If there is no minimum to the BIC score, one can look for an elbow in the BIC score's gradient, showing a number of clusters after which the marginal gain of BIC score is no longer significant. In this study, the BIC score showed no minimum up to 50 clusters, so its gradient was studied. However this technique is not very accurate, and the results should be interpreted carefully. For example, the knee identification is very dependent on the cut-in and cut-out of the curve for the definitions of the asymptotes. Furthermore, the GMM results are dependent on the initialization of the algorithm. As shown in Figure III.4, the obtained optimal number of sensors can range between 4 and 7, though it shows clear convergence for a number of clusters above 10. The gradient of BIC score was computed for 20 random GMM initializations for the three areas, and the mean gradient plotted as dashed line, with its 95% confidence interval as envelope. The BIC score was normalized to compare the three areas together. Similar trends can be observed, with stronger gradients in the Mediterranean Sea for the first clusters showing a bigger underlying complexity. For Southern Brittany, the associated uncertainty is bigger, showing weaker global minimum for the Expectation Maximization.

While the BIC score gives an indication on the range of optimal number of clusters, it does not necessarily translate into equivalent reconstruction for the wind fields. Although the clustering itself might find an optimum of 5 clusters for the Mediterranean Sea, this can lead to much higher reconstruction error than for the other areas as illustrated in Figure III.5(a). In particular for the Mediterranean Sea, the considered region is wider with several different wind regimes, which implies a higher variability. It then seems natural that more sensors than other areas would be needed to reach the same error level. Furthermore, the uncertainty on the optimal number of sensors shows an underlying property of this spatio-temporal data which has strong correlations between points, and for which clusters are not well separated. All in all, there is a need to cross-validate

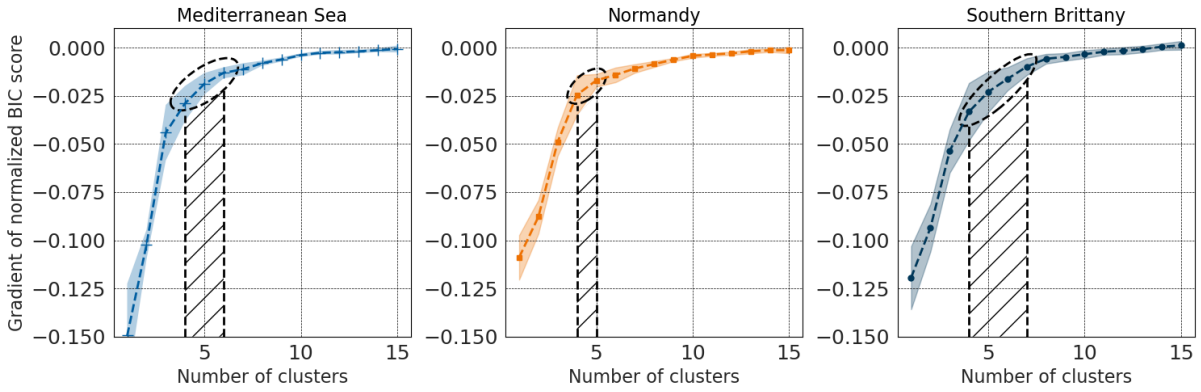


Figure III.4 – The gradient of normalized BIC score is shown for the three areas, for a number of clusters ranging from 1 to 15. The curves' envelopes are the 95% confidence interval obtained from 25 different initializations of the GMM training. The determination of an optimal number of sensors from these curves is uncertain, ranging from 4 to 6 for the Mediterranean Sea (shaded area), 4 to 5 for Normandy and 4 to 7 for Southern Brittany.

the computation of the optimal number of sensors. It is then proposed to validate the number of sensors from the computation of the reconstruction error. Exploring the range of number of clusters obtained through the BIC score gradient, the final number of sensors is chosen using a reconstruction error threshold.

To compare the three areas which have different wind regimes, the error threshold is defined as the reconstruction error of the normalized wind (Normalized Root Mean Squared Error or N-RMSE). The optimal number of clusters is then computed as the minimal number of clusters required to reconstruct 75% of the map with an error lower than the threshold.

It is then up to the final user to define an empirical error threshold to derive the optimal scenario. As shown in Figure III.5 (a), while the BIC score gradient curves are similar for the three areas, the normalized reconstruction error is significantly higher for the Mediterranean Sea for the same number of input points, thus necessitating a higher number of clusters to reach 75% of the map under threshold. The threshold of 0.2 normalized reconstruction error is shown in Figure III.5 (b). It yields to coherent results with regards to the BIC score analysis. The final numbers of clusters are then 4 for Normandy and Southern Brittany and 7 for the Mediterranean Sea. This workflow for the definition of the optimal number of sensors ensures similar performance between the three areas.

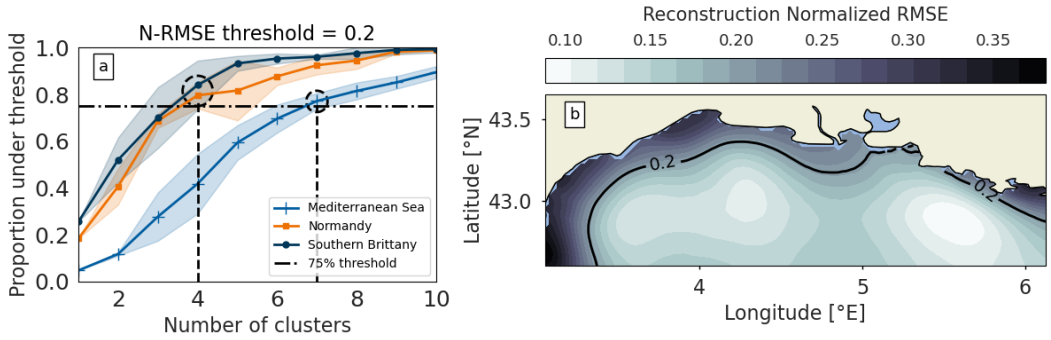


Figure III.5 – The computation of the proportion of the map under a certain error threshold for the three areas and 20 different GMM initializations allows for optimal number of sensors selection as the minimum number of sensors required to reach 75% of the map under threshold (a). The obtained reconstruction error map with the threshold contour shown illustrates the selection on the Mediterranean Sea (b).

III.5.2 Clustering-derived sensors performance

The sensors' locations for the base case scenario with optimal number of sensors of 4 for Normandy and Southern Brittany and 7 for the Mediterranean Sea are then computed on the three areas for the four methods: Monte Carlo, QR pivoting, EOF extrema, and GMM.

The obtained sensors' locations are displayed in Figure III.6 as red dots. It can be visually noted that the sensors array derived from the GMM method (second row) is more evenly distributed than the benchmark sensors arrays. QR pivots' locations (third row) are concentrated near the coast or at the maps' limits, and so are EOF extrema (fourth row). It shows how the GMM method allows for homogeneous sampling of the area, while benchmark methods tend to give too much weight to coastal and bordering points. This can be either artificial, due to spatial discontinuity at the limits of the maps, or because of the orographic impact of the coast. Indeed, the wind near the coast shows more variability, and while those points are contained in wider spatial structures in the GMM, they can be considered as salient points in the QR pivoting method or in the EOF extrema.

The resulting reconstruction at different time steps is illustrated as background for the three areas and four methods in Figure III.6. The first row is the reference case, reconstructed with perfect knowledge on the 20 EOF coefficients (EOF reference). For the Mediterranean Sea, this specific time step shows a combined Mistral and Tramontane winds blowing in the Mediterranean Sea. It is a complex and standard situation with

different wind regimes, strong offshore blowing winds in the North and West of the Gulf, and South-Eastern winds on the Eastern extremity. It can be noted that the GMM method correctly reproduces the intensity of those three phenomenons, while other techniques tend to overestimate or underestimate their effects.

For Normandy, the benchmark sensors array are largely off target on this specific case, predicting little to no wind offshore due to their exclusive coastal sampling, while the GMM method better captures both the coastal low winds and offshore wind cell.

For Southern Brittany, the effects of the sensors array is less clear, possibly explained by the smaller area, or by a simpler wind regime. However, the GMM method still performs largely better in terms of reconstruction error and wind patterns than benchmark methods.

Three different metrics are computed for the optimal scenarios on the three areas, and displayed in Table III.2. Along with the reconstruction error described in Sect. III.3.6, the error in the reconstructed mean and maximum wind speed are displayed. For the three areas, the GMM method clearly leads to good reconstruction error and mean wind speed estimation. However, the EOF extrema method yields to better estimation of the maximum wind speed for Normandy and Southern Brittany. It illustrates the fact that the GMM method is good at reconstructing the synoptic situation, while discarding high variability points that can be relevant for extreme events. Indeed, coastal points that can have a high variability due to the coastal orographic effects, are selected as salient points by the EOF extrema and QR pivot, and discarded by the GMM that assign them to a wider cluster. This is efficient to reconstruct the mean situation in the whole map but can lead to higher errors on high variability areas.

The proposed GMM method is scored against the three baselines methods on the Mediterranean sea area, for a number of sensors ranging from 1 to 10. The results are displayed in Figure III.7, showing the great interest of the clustering derived method compared to benchmark methods for the offshore wind reconstruction from sparse sampling. QR pivoting sensors and EOF extrema sensors fail to surpass the Monte Carlo simulation for low number of sensors. The GMM method yield to reconstruction errors systematically below the minimum of the boxplots (i.e., first quartile minus 1.5 times the inter-quartile range which is equal to 99.65 % of the data in the Gaussian case.), showing the near-optimal reconstruction. The benchmark methods' errors eventually decrease for a high number of sensors and surpass the Monte Carlo median scenario for 10 sensors. However it is expected that the different methods should converge for high number of sensors, as

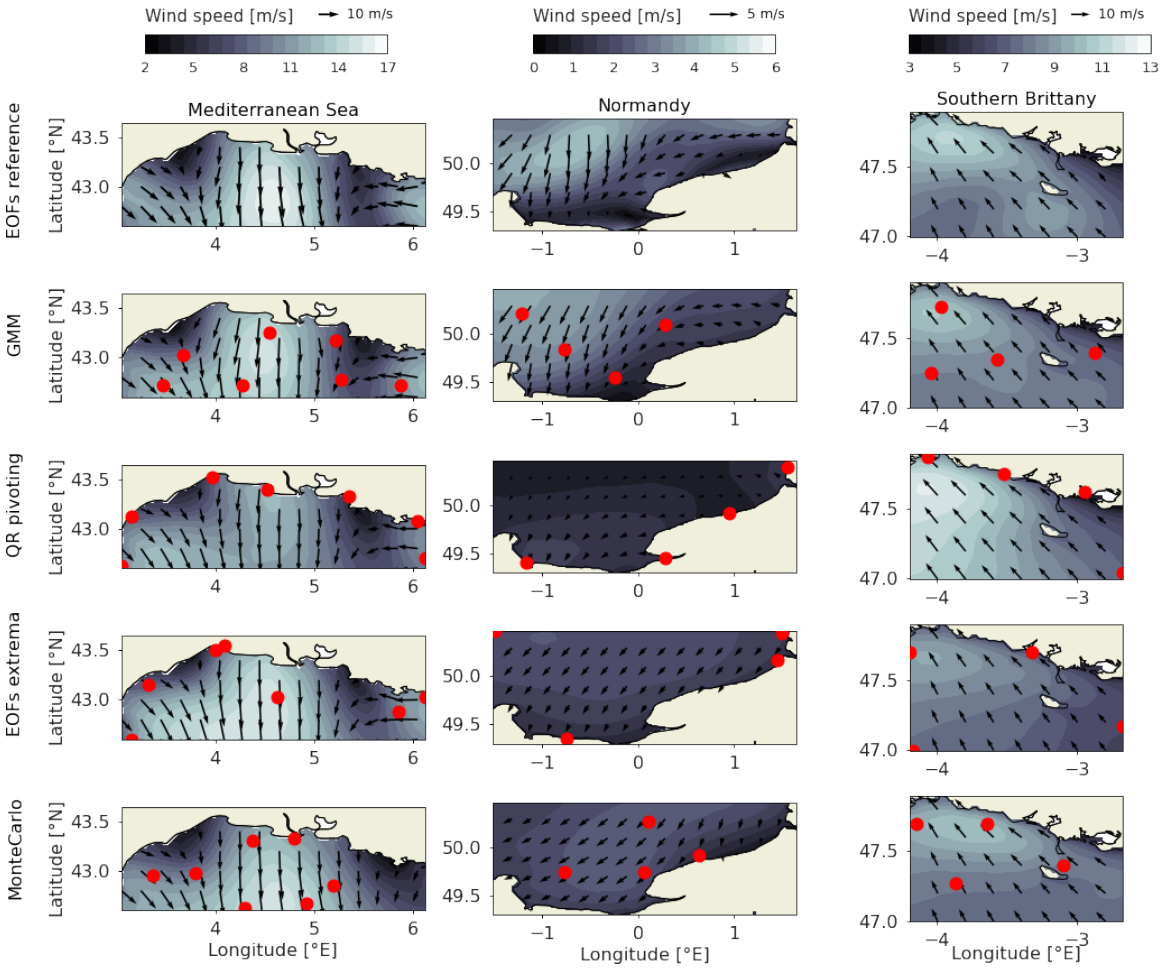


Figure III.6 – Reconstruction example for the optimal scenario on the three areas, from the reduced basis (EOF reference), from GMM clustering (GMM), and from the 3 baselines, QR pivoting, EOF extrema and Monte Carlo. The color grading shows the wind speed, the black arrows the wind direction, with length proportional to the wind speed, and the red dots are the locations of the sensors on the optimal scenario, with 7 sensors on the Mediterranean Sea, and 4 for Normandy and Southern Brittany.

the system is more and more constrained. It is illustrated by the decreasing spread within the Monte Carlo simulation.

For the GMM curve and the Monte Carlo boxplots, the reconstruction error seems to inflect for a number of sensors around 7, cross-validating the obtained optimal number of sensors for the base scenario. It can be noted that the reconstruction error for the EOF extrema method drastically decreases with the addition of the sixth sensor. As shown in Figure III.2, the sixth sensor is a central offshore point, while the 5 first locations are near

Table III.2 – Reconstruction errors computed for the 3 areas and the 4 sampling methods, including the random scenario displayed in Figure III.6. The best performing method is displayed in bold for each area. The reconstruction error (RMSE) and the errors on the max and mean wind speed at each time step are computed.

Area	Method	Max wind speed RMSE [m/s]	Mean wind speed RMSE [m/s]	RMSE [m/s]
Meditteranean Sea	GMM	**0.94**	**0.17**	**0.9**
	QR pivoting	1.42	0.42	1.77
	EOF extrema	1.28	0.2	1.37
	Monte Carlo	1.28	0.35	1.41
Normandy	GMM	2.0	**0.1**	**0.85**
	QR pivoting	1.84	0.56	1.83
	EOF extrema	**0.89**	0.42	1.44
	Monte Carlo	1.4	0.23	1.08
Southern Brittany	GMM	1.32	**0.09**	**0.7**
	QR pivoting	2.04	0.29	1.33
	EOF extrema	**0.89**	0.16	0.96
	Monte Carlo	1.87	0.17	0.93

the coast. For number of sensors above 6, the EOF extrema method then compares to the Monte Carlo median scenario.

For low and optimal number of sensors, compared to state-of-the-art techniques, the GMM method allows for the efficient sensors' placement for offshore wind reconstruction. The obtained reconstruction errors are displayed in Table III.3, along with the RMSE gain relative to the Monte Carlo median score. In the three areas, the GMM method improves significantly the reconstruction error on the base case scenario by 13% for Southern Brittany, and more than 20% for Normandy and the Mediterranean Sea. The QR pivoting method proves irrelevant for this application with a 50% increase in reconstruction error in Normandy, and around 30% in Southern Brittany and the Mediterranean Sea. The extrema method is closer to the Monte Carlo median case, though above, probably thanks to the manual removal of irrelevant extrema.

To visualize the effect of the sensors' locations, and the origin of the reconstruction error, the reconstruction error is computed as the root mean square error for each grid point, and displayed for the main scenario on the three areas in Figure III.8.

The coastal sensors arrays from the QR pivoting method displayed in the second row do not allow for offshore wind reconstruction, as those points are strongly influenced

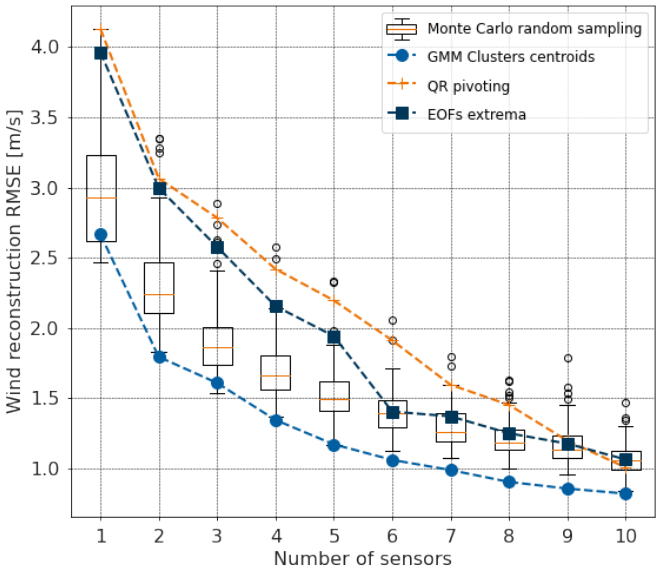


Figure III.7 – QR pivoting method (orange plus), EOF extrema method (deep blue squares) and GMM method (blue dots) are compared to Monte Carlo simulations, displayed as boxplots in terms of wind reconstruction error.

Table III.3 – RMSE and RMSE percentage gain versus Monte Carlo median value for the base case scenario on the three areas, for number of clusters of 7 for the Mediterranean Sea and 4 for Normandy and Southern Brittany. The bold numbers show the best performances.

	Mediterranean Sea	Normandy	Southern Brittany
Score	RMSE [m/s] | % gain	RMSE [m/s] | % gain	RMSE [m/s] | % gain
GMM	**0.99** | **-22%**	**0.90** | **-24%**	**0.82** | **-13%**
QR pivoting	1.60 | +27%	1.83 | + 55%	1.33 | + 39%
EOF extrema	1.37 | + 9%	1.44 | +22%	0.956 | +0.3%
Monte Carlo	1.26 | -	1.18 | -	0.952 | -

by coastal effects. Strong reconstruction errors of more than 2 m/s are then obtained far offshore for Normandy and the Mediterranean Sea. For the EOF extrema method displayed in the third row, where some offshore sensors' locations are present in addition to coastal ones, the synoptic wind regime seems better captured with more homogeneous reconstruction error. The reconstruction error patterns show the strong spatial correlation of the input wind data with lower reconstruction errors around the sensors' locations. However, the radius of lowered reconstruction error depends on the location. It can be noted that coastal points in QR pivoting for the Mediterranean Sea have a small radius

Table III.4 – Final locations selected for the deployment of floating LIDARs in French offshore wind development areas

sensor #	Mediterranean Sea (Latitude [°N], Longitude [°E])	Normandy (Latitude [°N], Longitude [°E])	Southern Brittany (Latitude [°N], Longitude [°E])
1	(42.775, 5.275)	(49.546, -0.242)	(47.721, -3.967)
2	(43.25, 4.55)	(50.096, 0.283)	(47.396, -2.867)
3	(42.725, 4.275)	(50.221, -1.217)	(47.346, -3.567)
4	(42.725, 3.475)	(49.846, -0.767)	(47.246, -4.042)
5	(43.175, 5.225)		
6	(43.025, 3.675)		
7	(42.725, 5.875)		

of influence, as opposed to some offshore points in the EOF extrema method. Coastal areas, where the wind is influenced by the coastal orography and thermodynamic effects, have lower spatial correlations or higher variability. As such, they are considered by the QR pivoting method and the extrema method as salient points. But in the end, they barely help to reconstruct the whole area's dynamics. It illustrates the importance of smart sparse sampling for the reconstruction, and the non-adequacy of QR pivots and EOF extrema as locations for the sparse sampling in this case.

For the three areas, the GMM obtained sensors' locations are homogeneously spatially distributed, allowing for good reconstruction on the whole map, though somewhat neglecting coastal locations. The locations, as centroids of clusters, are the most representative points of maximum likelihood clusters for a given number of sensors. As such, every point of the map belonging to a certain cluster, it allows for satisfactory reconstruction on most of the map. On the other hand, the QR pivoting method and EOF extrema method select salient points that do not necessarily correlate well with neighboring points, hence lowering the performance for reconstruction. The inadequacy of the considered benchmark methods for the proposed problem is then highlighted.

The final suggested sensors' locations for the three offshore wind development areas are given in Table III.4. These locations should be considered preferred locations for the deployment of floating LIDARs for wind resource assessment in the French offshore wind development areas.

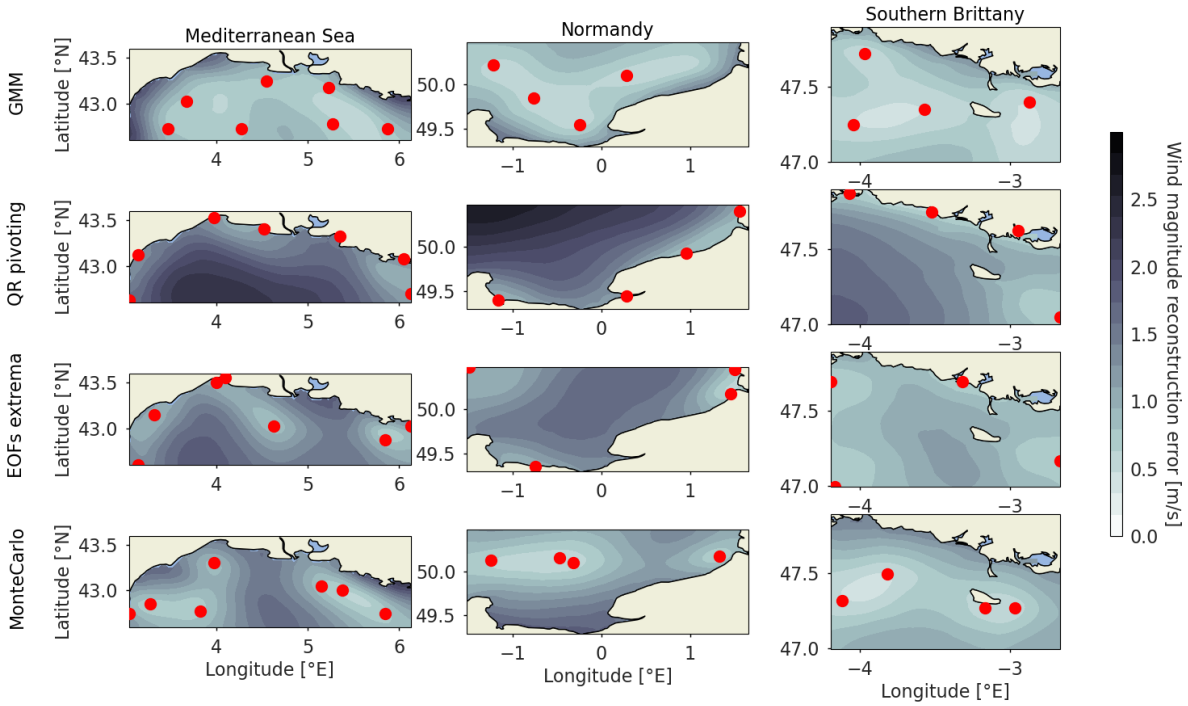


Figure III.8 – Wind magnitude reconstruction error temporally averaged per grid point on the Mediterranean Sea (left), Normandy (Center), Southern Brittany (right). The reconstruction error is computed for the optimal number of sensors determined in Sect. III.5.1 using the 3 baselines and the proposed clustering method. The red dots are the grid points used as input for the least squares reconstruction.

III.6 Discussion

In this study, an optimal sparse sampling is proposed using a Gaussian Mixture Model on high-resolution NWP data from Meteo France's MeteoNet data set (AROME model). The method used is simple yet efficient for the optimal sparse sampling of offshore wind field. Applied to offshore wind resource assessment, it can be a useful tool for the design of observation networks. It is compared to state-of-the-art solutions that fail to efficiently sample this specific problem, and a method is proposed to estimate the optimal number of sensors to deploy. The authors nonetheless raise attention on the following points to interpret and discuss the obtained results.

The metric that is used to measure the performance of the sparse sampling in this paper advantages the GMM method, because its homogeneous sampling allows for a

correct reconstruction of the synoptic situation. Coastal points are not well reconstructed using the GMM method, but this does not reflect on the scoring. Since the metric averages the reconstructed wind field's error over the grid points, a method that performs fairly good over the entire area is preferred.

As a consequence, both the obtained sensors' location and its performance are depending on the selected area. In this study, the selected sites are simple rectangles over the future development areas. But it could be interesting to reconstruct the wind field and score the performance on specific sites defined by operational limits for example bathymetry on the Mediterranean Sea, or coastal exclusion area for the three cases. This could lead to different results, and the sensitivity of the proposed method would then be studied.

Given the high variability of the wind near the coast and the possible impact on the obtained results, the 20 first kilometers from the coast were excluded to test the sensitivity of the methods. It roughly corresponds to the distance to the coast for future offshore wind parks, and ensures that the impact of the coastal orography is limited. It turns out that it does not make the state-of-the-art methods more relevant for this application as they still tend to select bordering points as input points.

Now, in the context of the development of marine energies in French waters, not only the wind should be considered but other variables such as wave variables, physicochemical parameters (turbidity, sea surface temperature, salinity) which are important for environmental impact monitoring. The installation of a network of sensors would therefore gain traction if optimized with regards to multiple variables. A follow-up of this work would be to include model data for each of the variables of interest, and perform the clustering on the stacked 10 first EOF of each variable, for the design of a multi-parameters observation network.

Getting even closer to the industrial reality of the sensors' network, it would be of great interest to include a cost function dependent on the location (depth, distance from shore, other constraints). This method could be declined to find the Pareto optimal sensors network. The optimal number of sensors would therefore become the number at which the sensors marginal cost exceeds the reconstruction gain.

The data in this study is derived from the NWP model AROME data, with a 1.3km grid size and hourly definition. The parametrization of the model offshore is not perfect, in particular for the sea/atmosphere coupling that can lead to discrepancies in surface parameters, as shown during the Mediterranean HyMex campaign [239]. The learnt dy-

namics might then be a coarse description of the reality, and the derived sensors' locations be limited for real time wind reconstruction, since only trained on low spatio-temporal resolution patterns. The obtained locations are in this case optimal only for reconstructing the dynamics of the NWP model, and the representativeness of the used data compared to the reality needs to be questioned. It could then be of interest to run such study on higher resolution data, either from Synthetic Aperture Radar measurements or Large Eddy Simulations, though on shorter periods, or comparing the reconstruction on a set of measurements offshore.

Publicly available through the MeteoNet data set, only 10 meters wind speed are used in this paper. For offshore wind application, hub height wind speed are to be considered (heavy maintenance, loading, energy production). The described method is agnostic to the input data, though it would be of interest to validate the obtained sensors network with 100m wind speed data. It is not direct that the obtained sensor network will be the same with above 100m wind speed data, since the extrapolation will be depending on the grid point and the wind speed and direction (changing the sea surface roughness). The non-linear transformation of data can then change the weight given by the clustering model at each timestep and grid point.

Seemingly, performing the clustering with the power-curve transformed data can potentially lead to different results. The study focuses on wind speed, as this can apply to wind energy production but also maintenance operations planning or wind turbine loading. A specific study could focus on wind power, applying vertical extrapolation and wind power curve. The proposed method can be easily implemented to different data inputs.

The used data set compiles 3 years of data. The model is trained on 2 years and tested and scored on the third year. It could be of interest to carry the same study on a longer data set from global reanalysis models such as ERA5. The high-resolution regional AROME model from Météo France with its 3 years of open-source data from the MeteoNet data set offers a higher number of grid point, making it more relevant for the sensor siting on small areas as the ones in this study. The features that need to be captured by the reconstruction are smaller scale than global models' grid size. Comparing the results obtained from the two sources could be of interest.

The used benchmark methods from the sparse-sampling literature i.e., the QR pivoting method and the EOF extrema method do not prove efficient for the stated problem. For generalization purposes, this method would need to be compared to state-of-the-art method on benchmark data set such as the simulated flow past a cylinder used in [53].

This paper does not aim at generalizing a method, but develops an efficient solution to an identified problem, for which state-of-the-art methods seem to fail. The sensitivity of the methods to the size and the choice of the domain could also be considered. It is possible that the benchmark models would perform better on wider areas, or for a bigger amount of sensors.

Eventually, the use of Gaussian Mixture Model seems appropriate for the sparse sampling of offshore wind resource. It is an easy method to implement with relatively low computational cost. It is flexible and can in principle be applied to higher dimensional systems. This could be of interest for offshore wind energy, allowing the inclusion of environmental parameters in the siting optimization. The method also shows good consistency on the three development areas tested with very different wind regimes. It is however important to stress the difficulty associated with the optimal number of sensors. As proposed in this paper, the number of sensors is derived indirectly from an error threshold. In this context it seems difficult to include cost or environment constraints as such in the sensors siting.

III.7 Conclusions

A method for the finding of an optimal sensors network for offshore wind reconstruction is presented in this paper, and applied to three of the main offshore wind energy development area in France. The sparse sensors' placement problem is stated on a reduced basis of the 3 years AROME prediction of wind from the MeteoNet data set. State-of-the-art techniques of sparse sensors' placement for reconstruction (QR pivoting and extrema methods) are compared to the proposed method, based on the Gaussian Mixture Model clustering of Empirical Orthogonal Functions of zonal and meridional wind offshore. By selecting the clusters' centroids as proposed locations for sensors, the GMM method homogeneously partitions the domain into spatially correlated clusters. In this way, the reconstruction error on the whole domain is minimized, leading to a 20% decrease in wind reconstruction error compared to the median Monte Carlo case. On the other hand, state-of-the-art methods fail to reconstruct the whole wind field because they are attracted by salient points with high variability (bordering points). However, these points are not very spatially correlated to neighboring points, yielding to a reconstruction error higher than the median Monte Carlo case. The GMM clustering method gives indications on the optimal number of sensors to deploy, though this estimation should be

refined either by the integration of cost or environmental constraints, or by the definition of a reconstruction error threshold.

GMM clustering method seems to be a simple yet efficient solution for sparse sensors' placement. Applied to offshore wind reconstruction, it allows for the optimal placement of sensors, and paves the way for smart marine monitoring in the era of offshore wind energy development. Further work should focus on the technique's generalization to benchmark problems, and question the representativeness of the used data set. For wind energy applications, the multivariate case should be studied for multi-instrumental sensors' placement, and the economic constraints should be implemented for the definition of the Pareto optimal number of sensors.

In the light of this study, the authors suggest the deployment of 7 sensors in the Mediterranean Sea, 4 sensors in Normandy and 4 sensors in Southern Brittany at optimal locations to reconstruct the offshore wind field and to help with the wind resource assessment on these areas.

Meteorological data used in this study are available online through the MeteoNet data set. The code developed for offshore wind resource sparse sampling using Gaussian Mixture Models can be accessed through `https://github.com/rmarcille/gmm_sparse_sampling.git`

This work was supported by France Energies Marines and the French government, managed by the Agence Nationale de la Recherche under the Investissements d'Avenir program, with the reference ANR-10-IEED-0006-34. This work was carried out in the framework of the FOWRCE_SEA and POWSEIDOM projects.

Chapter IV

Very short-term probabilistic offshore wind speed forecasting using deep learning

Introducing the article

As described in Chap.III, the deployment of offshore measurement network is expensive and of high importance for wind energy projects. The quality of numerical weather prediction at sea is limited by the lack of in-situ measurements and the limited coupling with the oceanographic models. Provided that we have access to an offshore wind speed time series - e.g. from a future offshore wind farm development site - it is then of interest to explore the post processing of numerical weather prediction at this specific site. In this study, we focus on the added value of neighbouring measurement data, and more specifically coastal measurements. These are already installed, and much cheaper to operate than their offshore counterpart. In the French Mediterranean Sea, offshore blowing winds are dominant. This makes coastal measurements upwind from the target site and potentially more informative for the post-processing.

As described in Chap.I, numerical weather prediction have a high computational cost and are usually limited to a duration of 6 hours between consecutive forecasts. Moreover, the uncertainty estimation from numerical weather prediction is obtained with ensemble modelling, running several perturbed members in parallel to estimate the spread of predicted variables. For high-resolution models such as the AROME model from Météo France, it consists in 16 members, which gives important information about the spread of the predictions, but is limited for the estimation of the posterior distribution, especially in high dimensional problems. Despite being limited by current model capabilities, these two aspects are key for offshore wind farms operations:

- Short-term decisions (1 to 12 hours ahead) are to be made when operating a wind

farm. Maintenance operations require close to real-time metocean monitoring to ensure safety and to optimize weather windows use. Wind power prediction for market participation also requires short-term to ultra-short-term forecasts that can't be issued by numerical weather prediction.

- Decision-making at sea is subject to uncertainty. When performing offshore operations, the risk associated with forecasts errors needs to accounted for and mitigated. Numerical weather prediction uncertainty is often reduced to a reliability index. Using such a restricted uncertainty information leads to conservatism and is sub-optimal.

Time series models and learning-based models can offer alternatives for these two aspects. Their relatively low computational cost permits close to real-time forecast and recent measurement assimilation. They can be designed to produce probabilistic forecast, hence providing a full uncertainty estimation without post-processing. Deep learning problems in particular can easily accommodate a large amount of input data which can be helpful to integrate explanatory variables for the forecast uncertainty.

In this article we propose a deep learning framework for the probabilistic post-processing of numerical weather prediction at sea. We use an open-source dataset in the French Mediterranean Sea to simulate the situation of future offshore wind projects. A relatively short offshore measurement time series is to be predicted from coastal measurements and numerical weather predictions. Several probabilistic frameworks are proposed and compared in terms of forecast quality, reliability and computational cost:

* We implement analogs forecasting [56] and gradient boosting machine [4] as state-of-the-art statistical approaches for probabilistic forecasting.
* We then implement a convolutional neural network approach that can accommodate both spatialized numerical weather prediction data and coastal in-situ data to predict the bi-dimensional wind speed under Gaussian posterior assumption. We show that the implicit pre-processing of large input data is a key advantage of convolutional approaches.
* We relax the posterior parametric assumption using normalizing flows, and we show that it can be a powerful yet simple tool to emulate probabilistic forecasts without parametric assumptions while keeping sampling and scoring capabilities.

This work is published in the journal *Artificial Intelligence for Earth Systems* as *Robin Marcille, Pierre Tandéo, Maxime Thiébaut, Ronan Fablet, Pierre Pinson, Convolutional encoding and normalizing flows: a deep learning approach for offshore wind speed proba-*

bilistic forecasting in the Mediterranean Sea, 2024 [240]. ©American Meteorological Society. Used with permission.

IV.1 Introduction

Weather forecasting in offshore environments is challenging due to the scarcity and sparsity of offshore observations, both in space and time [103, 241]. These limitations affect data assimilation systems, especially initial state estimation, and validation processes. Moreover, wind profiles are challenging [81] and influenced by various factors, such as air/sea exchanges [82], diurnal variations [83], and site-dependent effects [84], which are difficult to model accurately. Offshore weather forecasts are essential for marine operations, especially at locations where in-situ data is scarce. These forecasts inform decision-making at sea for weather-limited operations. Weather operability limits are computed from simulation to avoid operation failure, and weather windows with critical parameters under the operability limits have to be forecast. Forecast errors imply risks of operation failure, and forecast uncertainty ought to be considered for operations planning and execution. To deal with uncertainty in the offshore wind energy industry, a factor ranging from 0 to 1 (the alpha factor) is assigned to each weather operability limit [3]. According to [4], most existing methods rely on deterministic forecasts and the use of the alpha factor to account for weather forecast uncertainty. This may result in conservative decision making and sub-optimal planning. As illustrated in [5], probabilistic forecasts can address these shortcomings. Under the assumption of reliable weather forecast of the limiting parameters, the uncertainty can directly be transferred to the probability of operation failure. When doing so, one can obtain a large improvement in operational hours compared to the alpha-factor methodology. This requires the reliable joint probabilistic forecasting of limiting wind and wave parameters that impact vessel motions (e.g. Significant wave height, 10-meter wind speed, wave peak-period [34]). The decision-making using probabilistic forecasts is then cost-optimal compared to deterministic forecasts [6, 7], motivating the development of probabilistic post-processing of deterministic forecasts.

State-of-the-art weather forecasting systems generally rely on ensemble methods to assess and describe forecast uncertainty [242]. They generate different scenarios by varying both the initial state of the system and model parameters to estimate the spread of the forecast state. The very high computational cost associated with this forecast process limits the number of members in the ensemble, typically up to a few tens of members. Such

ensembles cannot fully inform the forecasting uncertainties, especially for local processes such as strong convective events in South-Eastern French maritime facade (Gulf of Lion) which is the main study area. The post-processing of Numerical Weather Predictions (NWP) using statistical and machine learning methods then appears appealing to better emulate these forecast uncertainties [243].

A large variety of models can be used for the probabilistic post-processing of deterministic forecasts [244]. We can distinguish models based on the description of the probabilistic output. Non-parametric methods such as interval or quantile forecasting [143], kernel density and ensemble methods, make fewer assumptions about the shape of the target distribution. For instance, gradient boosting machine (GBM) can be used for the quantile forecasting of wave parameters [4]. Parametric approaches assume a certain parametric distribution for the output (e.g. Gaussian, beta, log-normal) [130] which allows for analytical computations. Within parametric descriptions, the Gaussian assumption might be simple, but can characterize satisfyingly the uncertainty of 2-dimensional wind prediction [147]. One can estimate the parameters of Gaussian distribution using analogs of the observed weather situation [56, 182]. Alternatively, regression and deep learning models can emulate a Gaussian covariance matrix from a deterministic forecast as developed in [129] considering a diagonal covariance matrix.

Novel generative deep learning techniques offer innovative methods for the approximation of complex posterior distributions. Variational Recurrent Auto Encoders (VRAE) can be used to generate scenarios at a relatively low computational cost [210], but the output distribution can only be accessed via sampling. VRAE are compared in [44] to Generative Adversarial Networks (GAN) and normalizing flows for wind power forecasting. Normalizing flows are deep learning models based on the composition of parameterized bijective functions, that transform a simple parametric distribution into an arbitrarily shaped distribution. It was proposed for variational inference in [119], and generalized to density estimation in [120]. Compared to analogs methods, it needs no parametric assumption for the posterior distribution. In addition to sampling capabilities, they allow for exact likelihood computation. These two features are advantages compared to quantile forecasting. In contrast with VRAE and GAN, they are relatively easy to implement and train. In [59], conditional normalizing flows are shown to be well suited for multivariate time series forecasting. A fair assessment of their advantages and disadvantages for a real application in probabilistic forecasting is lacking from the literature.

In the light of the work cited above, this paper addresses the post-processing of numer-

ical weather prediction and in-situ measurements using deep learning schemes to improve the probabilistic forecasting of wind speed at sea. Numerical weather prediction act as a physical prior of the future state of the weather system at the considered offshore location, while recent neighboring measurements may better inform the actual state of the system. In this study a parametric Gaussian model and a generative model using normalizing flows are compared with baseline models (analogs, gradient boosting machine, numerical weather prediction) to analyze their performances in terms of probabilistic and deterministic metrics. Models are also compared as function of the weather situation, to highlight the advantages and disadvantages of the method for marine operations. Eventually, the importance of various input data is discussed, to give indications on the required input data for offshore wind speed probabilistic forecasting.

The dataset used for the experiment is described in Section IV.2. The proposed approach and its mathematical formalism are thoroughly presented in Section IV.3, before the baseline methods and metrics used for comparison are detailed in Section IV.4. The obtained results are shared and analyzed with deterministic and probabilistic metrics, and as function of weather situations in Section IV.5. A discussion on the limitations of the experiment is done in Section IV.6 to provide recommendations and perspectives for future work.

IV.2 Dataset

IV.2.1 Case-study area

To develop the methodology, we consider the MeteoNet dataset [245]. It is an open-source dataset developed and shared by Météo France, the french national weather service. It contains time series of weather ground stations data and numerical weather prediction model over a 550km x 550km region in South-East France. It spans between 2016 and 2018 with 65 days of missing data. Hourly forecasts of weather variables (10-meters wind speed, 2-meters relative humidity, 2-meters air temperature, pressure at sea level) from the high-resolution model AROME (Applications de la Recherche à l'Opérationnel à MEsoéchelle) are available. AROME is the operational high resolution model on France operated by Météo France. It has a grid size of 1.3km and outputs hourly predictions. The ground station network covers 484 stations scattered over the South-East of France, as shown in Figure IV.1 (a), with 6-min measured time series of 10-meters wind speed, 2-meters air

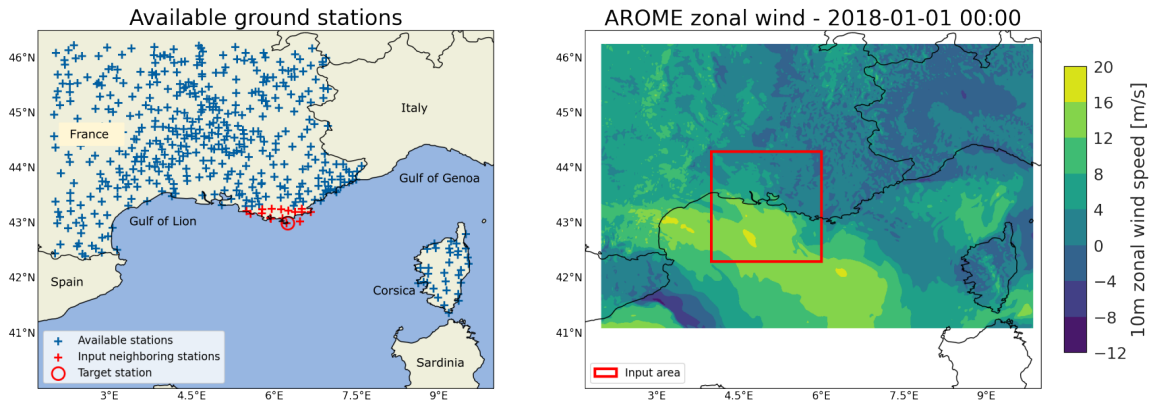


Figure IV.1 – Subset of the MeteoNet dataset selected for methodology development. Coastal stations around Porquerolles target station are selected (a). A geographical subset containing local information from numerical weather prediction is selected to reduce dimensionality of the input (b).

temperature, station pressure, 2-meters dew point temperature, 2-meters relative humidity and precipitation.

The study focuses on the Gulf of Lion, which is situated in the North-East Mediterranean Sea, between the cities of Toulon and Perpignan in South-East France. It is considered one of the main floating offshore wind development areas in France [246]. The study area is characterized by a strong dominance of offshore blowing winds in the northern (Mistral) and western (Tramontane) Gulf of Lion. Those phenomena are due to an orographic channeling in the Rhone and Garona valleys with the pressure difference between the North-East Atlantic (high pressures) and the North-West Mediterranean Sea (Gulf of Genoa, low pressure). When the high pressures are rather localized over central Europe, the region experiences strong South-East winds charged with humidity that can cause heavy precipitation on the coastal areas. These two phenomena are largely driving the wind patterns in the area and are sensitive to continental forcing and local orography. They also apply a strong forcing on the hydrodynamics of the region, with large up-welling and down-welling phenomenon [114].

The target station is the Porquerolles island weather station encircled in Figure IV.1 (a). It is the only offshore station available in the dataset. It is located on the Porquerolles island's semaphore, at 135m of elevation on the top of the island. The 14 closest coastal weather stations in Figure IV.1 (a) are selected to serve as input. The numerical weather prediction input is reduced to a subset of 2 degrees of latitude and longitude around the target station to reduce its dimensionality, see Figure IV.1 (b). The correlation between

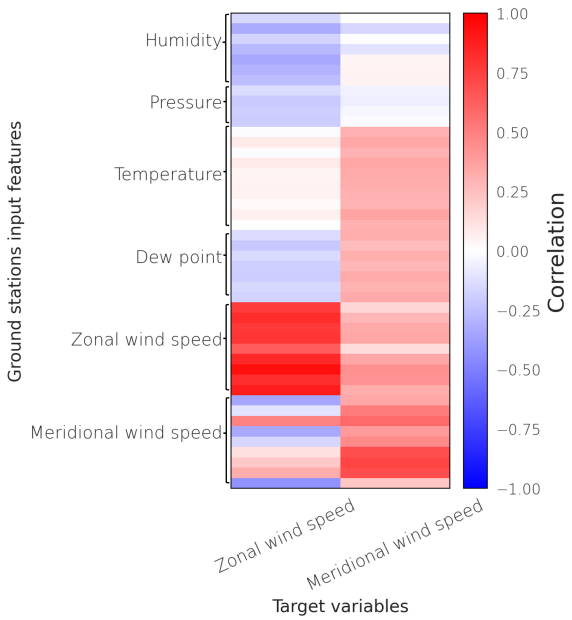


Figure IV.2 – Correlation between measured variables at ground stations and wind speed at the target station.

the measured parameters at the input ground stations and the wind speed measured at the target station is shown in FigureIV.2. Wind speed at coastal ground stations is highly correlated to the target station. Zonal wind speed at target station is negatively correlated to humidity at coastal stations, showing the predominance of eastern wind during rain events. Temperature is correlated to meridional wind, in link with thermal breezes.

IV.2.2 Numerical weather prediction data

The numerical weather prediction input tensor at forecast issue time t, $\boldsymbol{X}_t^{NWP}$, is a 4-dimensional tensor in latitude (80 points), longitude (80 points), weather variables (5 variables) and lead times (6 time steps). The input variables available in the MeteoNet dataset are the 2-dimensional 10-meter wind speed (u, v), percentage of humidity, mean pressure at sea level and 2-meter temperature. The time step of the model data is 1 hour, and the last forecast time step is $\tau_{NWP} = 5$ h ahead. For each forecast issue time t, the AROME input has then $K_{NWP} = 6$ lead times between t and $t + \tau_{NWP}$. The variable and lead time dimensions are merged into a 30-dimensional axis, so the final tensor has dimensions (80, 80, 30). This data correspond to the deterministic forecast of AROME, with no information on the forecast uncertainty. In practice, the AROME forecast is issued

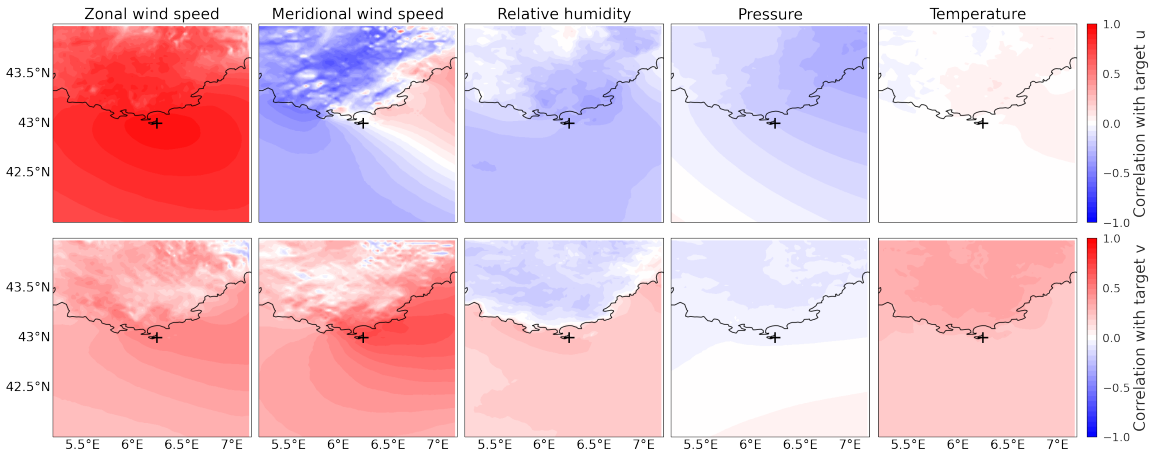


Figure IV.3 – Correlation between AROME forecasts used as input and wind speed at the target station. The correlation is computed for each grid point. The top row shows the correlation with the target zonal wind speed (target u), and the bottom row with the target meridional wind speed (target v).

2 to 6 hours after the initial forecast lead time, so this should be taken into account in an operational setup. The numerical weather prediction data at the forecast issue time should then correspond to the latest numerical forecast available at that time.

Correlation between AROME forecasts and wind speed at the target station is shown in FigureIV.3. Lower pressures on the eastern part of the study area (gulf of Genoa) are negatively correlated to zonal wind speed at the target station, showing the weather systems that channel Mistral northwestern winds. Higher correlations are observed for the zonal wind speed which is more representative of dominant wind systems. Meridional wind speed is more uncertain and is correlated to humidity and temperature.

IV.2.3 In-situ data

The input data from ground stations contain recent observations from the neighboring coastal stations. The ground stations input tensor for the forecast issue time t, $\boldsymbol{X}_t^{GS}$, is a 3-dimensional tensor in stations (14 stations), weather variables (maximum 6 variables, depending on the station), and time steps (60 time steps). The input variables available at each station are the 2-dimensional 10-meter wind speed (u, v), humidity rate, temperature, pressure at sea level and dew point. It has a time step of 6 minutes, and the last $\tau_{GS} = 6$ h of observations are used as input. The ground stations input is then a concatenation of time series of $K_{GS} = 60$ time steps. The stations and weather variables dimensions are

merged so the final tensor has dimensions (80, 60). In practice, a certain latency is to be expected to gather these data. Similarly to numerical weather prediction, it should be considered that the in-situ data is the latest available measurement data.

The output of the dataset is the measured wind speed at the target station. At forecast issue time t, the target vector $\boldsymbol{y}_t$ is a tensor of zonal and meridional wind speed at 10 meters for different lead times. It has a time step of 6 minutes and is to be predicted for the next $\tau_{pred} = 6$ h. The target tensor consists in $N = 2$ time series of $K = 60$ lead times and has dimensions (2, 60).

To deal with missing data, measured variables from the ground stations exceeding 4% of missing data are removed. It corresponds to 3 weather stations and 34 measured weather variables in total. The resulting entries exceeding 4% of missing data are also removed (395 entries). Eventually, the remaining gaps in the data are forward filled. They are then considered as valid data with the assumption that the gaps are small enough to be neglected.

IV.2.4 Training, test and validation datasets

The dataset is split in three parts for training, validation, and testing phases. These three datasets need to be independent but representative of the same statistical distribution [42]. For weather data, auto-correlation at different time scales requires special care [66]. To limit seasonal effects, 2 years of data (two thirds of the dataset) are used for training. The remaining third is split for validating and testing (half a year). 5 days are removed in between the splits to avoid short-term temporal correlation between the datasets. To mitigate data representativity issues, cross-validation on the train-validation-test split is performed. The train, validation, and test sets are shuffled into 6 different splits as shown in Figure IV.4. Results are then computed across those 6 splits. After cleaning and splitting, the final dataset contains 2372 entries in the training split, 779 in the validation split and 798 in the test split. All data sources are standardized with regards to the training dataset to ensure that all features have similar scales.

IV.2.5 Baseline reduced dataset

The full dataset has a very high number of dimensions. To implement statistical baselines that can only accommodate a limited number of features, a baseline reduced dataset is constructed.

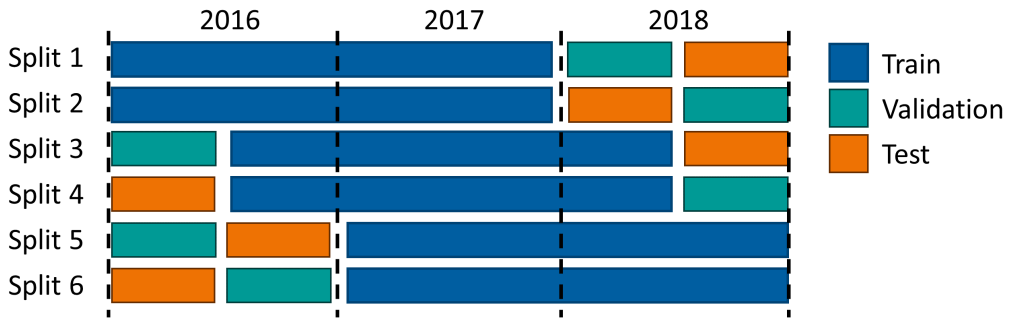


Figure IV.4 – Train validation test splits used for cross-validation.

The reduced dataset contains:

— The 3 first principle components obtained through Principal Component Analysis of the zonal and meridional wind speed of AROME inputs, considered separately and computed on the training dataset,

— The 7 first principle components obtained through Principal Component Analysis of the measured wind speed at the three closest ground stations for the last 6 hours,

— The last wind measurements at the 3 closest ground stations,

— The wind speed forecast from AROME closest grid point.

A sample from the reduced dataset $\boldsymbol{X}_t^r$ is then a tensor of 15 features and 60 lead times. The main dataset has dimensions (80, 80, 30) for AROME input and (46, 60) for ground stations input. The reduced dataset corresponds to 0.5% of the total input data. Principal Component Analysis is used to extract the most relevant features.

The reduced dataset serves as input for the baseline methods presented below. It then allows for fair comparison between different approaches. The selected features of this reduced dataset were optimized to optimize the validation loss of the gradient boosting machine model in Section IV.4.3.

IV.3 Proposed architecture

This section presents the proposed convolutional architecture to emulate a probabilistic multivariate forecast from the input data described in Section IV.2. We first introduce the problem formulation in Section IV.3.1. The convolutional encoding of numerical weather prediction and in-situ data is described in Section IV.3.2. We then detail the Gaussian (Section IV.3.3) and normalizing flows (Section IV.3.4) output probabilistic descriptions.

Eventually Section IV.3.5 gives an overview of the final architecture.

IV.3.1 Short-term wind forecasting at an unobserved location

The goal of the forecast model is to make a wind speed prediction at a target location using numerical weather prediction and ground stations measurements. For a forecast issue time $t \in [1, T]$ and a forecast lead-time $k \in [1, K]$, the model Ψ parameterized by $\boldsymbol{\Xi}$ outputs a vector $\hat{\boldsymbol{\theta}}_{t+k}$ from the input vector $\boldsymbol{X}_t$ such that

$$\hat{\boldsymbol{\theta}}_{t+k}(\boldsymbol{\Xi}) = \Psi_{\boldsymbol{\Xi}}(\boldsymbol{X}_t). \tag{IV.1}$$

The output vector $\hat{\boldsymbol{\theta}}_{t+k}$ is a parameterization of a probability density function $\hat{f}_{t+k}$ of a random variable S_{t+k} from which we can draw samples $\boldsymbol{s}_{t+k}$. The distribution $\hat{f}_{t+k}$ is transformed into a target distribution $\hat{g}_{t+k}$ through a transformation $\mathcal{T}$. Therefore we map a sample $\boldsymbol{s}_{t+k}$ from the initial distribution into a sample $\boldsymbol{z}_{t+k}$ of the target distribution

$$\boldsymbol{z}_{t+k} = \mathcal{T}(\boldsymbol{s}_{t+k}), \tag{IV.2}$$

with $\boldsymbol{z}_{t+k}$ a sample from the random variable Z_{t+k} with probability density function $\hat{g}_{t+k}$. We explore an identity parameterization for transformation $\mathcal{T}$ as well as normalizing flows to account for more complex target distributions. In all that follows, the subscript k refers to $t + k$.

IV.3.2 Convolutional encoding of AROME and ground stations data

The proposed method uses a deep learning architecture to accommodate the large amount of heterogeneous input data. A Convolutional Neural Network (CNN) is a type of deep neural network that uses convolutional layers and pooling layers to efficiently reduce the dimension of input data. Convolutional layers apply convolution filters to the input data, capturing multi-scale features. The convolution filter applies the same weights to the whole input, so the number of model coefficient is reduced. Pooling layers reduce the dimension of the data by applying sub-sampling functions to groups of neighboring points [42]. CNN are extensively used in the forecasting literature when dealing with large numerical model data in 2 dimensions [29] or 3 dimensions [193]. 1-dimensional CNN can

also be used to deal with time series data [143].

For the offshore wind forecasting problem presented in this work, a large amount of data is used as input. Numerical weather prediction data are 80 x 80 images for each time step and each variable. Meteorological variables exhibit features at various scales that need to be extracted. A 2-dimensional CNN is used to encode the numerical weather prediction input into an ensemble of latent time series containing useful information for forecasting. The convolutions are made through space to capture the spatial features, while the weather variables and lead times are taken as channels.

Seemingly, a 1-dimensional CNN is used to encode the ground stations time series onto a latent space. The convolution is performed on the time component, so that the temporal correlations of the time series can be captured. 1D convolutional layers are used, and the different weather variables and stations are taken as channels.

We apply the CNN to numerical weather prediction and ground stations time series to obtain 9 latent time series of 60 time steps. Two additional latent time series are added containing the predicted wind speed at the closest AROME grid point. The final dimension of the latent space is $(11,\ 60)$.

IV.3.3 Gaussian posterior assumption

The basic assumption for the proposed architecture describes the target as a 2-dimensional Gaussian distribution. For a Gaussian posterior the output vector $\hat{\boldsymbol{\theta}}_k$ contains the parameters

$$\hat{\boldsymbol{\theta}}_k = [\hat{\mu}_u(k), \hat{\mu}_v(k), \hat{\sigma}_u^2(k), \hat{\sigma}_v^2(k), \hat{\rho}_{u,v}(k)] \tag{IV.3}$$

such that

$$Z_k \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k\right) \tag{IV.4}$$

with $\hat{\boldsymbol{\mu}}_k$ the mean matrix and $\hat{\boldsymbol{\Sigma}}_k$ the covariance matrix, constructed from the two predicted variances $\hat{\sigma}_u^2(k), \hat{\sigma}_v^2(k)$ and the Pearson coefficient $\hat{\rho}_{u,v}(k)$

$$\hat{\boldsymbol{\Sigma}}_k = \begin{bmatrix} \hat{\sigma}_u^2(k) & \hat{\rho}_{u,v}(k)\hat{\sigma}_u(k)\hat{\sigma}_v(k) \\ \hat{\rho}_{u,v}(k)\hat{\sigma}_u(k)\hat{\sigma}_v(k) & \hat{\sigma}_v^2(k) \end{bmatrix} \tag{IV.5}$$

$$\hat{\boldsymbol{\mu}}_k = \begin{bmatrix} \hat{\mu}_u(k) \\ \hat{\mu}_v(k) \end{bmatrix} \tag{IV.6}$$

A 2-layers Multi-Layer Perceptron (MLP) is used to output Gaussian parameterization

from the latent space. To ensure the positive semi-definiteness of the predicted covariance matrix, the variances should be positive $\sigma_u(k), \sigma_v(k) > 0$, and the Pearson coefficient should satisfy $-1 \leq \rho_{u,v}(k) \leq 1$. A final activation function is applied to the output of the MLP to satisfy these inequalities. The variances are obtained with the use of an exponential activation function, and the Pearson coefficient is obtained through a hyperbolic tangent activation function. The mean values $\mu_u(k), \mu_v(k) \in \mathbb{R}$ need no final activation function.

The loss function $\mathcal{L}_t(\boldsymbol{\Xi})$ used for the optimization is the negative log-likelihood [42]

$$\mathcal{L}_t(\boldsymbol{\Xi}) = \frac{1}{K} \sum_{k=1}^{K} - \log \left[\hat{g}_{t+k}(\boldsymbol{y}_{t+k} | \boldsymbol{\Xi}) \right] \tag{IV.7}$$

with $\hat{g}_{t+k}$ the predicted probability density function of the posterior at lead time k. The negative log-likelihood is a proper scoring rule that has two main advantages. It accounts for the reliability of the prediction defined through the covariance matrix, and it strongly penalizes outliers due to the log function.

Using a Gaussian distribution for the posterior provides an analytical expression for the likelihood which can then be directly computed. For an observation $\boldsymbol{y}_k$ and a predicted 2-dimensional Gaussian distribution with parameters $\hat{\boldsymbol{\Sigma}}_k$ and $\hat{\boldsymbol{\mu}}_k$, the likelihood is equal to [42]

$$\hat{g}_k(\boldsymbol{y}_k | \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k) = \frac{1}{(2\pi)|\hat{\boldsymbol{\Sigma}}_k|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{y}_k - \hat{\boldsymbol{\mu}}_k)^T \hat{\boldsymbol{\Sigma}}_k^{-1} (\boldsymbol{y}_k - \hat{\boldsymbol{\mu}}_k) \right). \tag{IV.8}$$

It is widely used for scoring forecasts versus observations under uncertainty for data assimilation schemes [183] and as a parametric method for multivariate regression [148].

IV.3.4 Normalizing flows

A generative approach is proposed to account for non-Gaussian distributions while keeping the computation of the likelihood tractable, and the sampling capabilities. Normalizing flows are generative deep learning models that use a composition of invertible functions to learn a "flow" from a simple base distribution (here a multivariate Gaussian) to an arbitrarily shaped distribution.

Given a base distribution $h^{(0)}$, and a series of invertible functions $\mathcal{T}_0, ..., \mathcal{T}_M$, the posterior likelihood can be computed using a change of variables from the base to the target distribution. The likelihood of the obtained distribution $h^{(M)}$ can then be obtained through

a change of variable [120]

$$\log\left(h^{(M)}(\boldsymbol{z}_M)\right) = \log\left(h^{(0)}(\boldsymbol{z}_0)\right) - \sum_{m=0}^{M} \log\left(\det\left|\frac{\partial h^{(m)}}{\partial \boldsymbol{z}_m}\right|\right). \tag{IV.9}$$

A sample from the base distribution is transformed to a sample from the target distribution using the following composition of transforms

$$\boldsymbol{z}_M = \mathcal{T}_0 \circ ... \circ \mathcal{T}_M(\boldsymbol{z}_0). \tag{IV.10}$$

A bijective function needs to be selected to compose the layers of the flow. In this work, a rational quadratic spline function is used. As described in [125], it has the advantage of being highly flexible while staying analytically invertible. Compared to more classical affine transformations, it can approximate complicated distributions with fewer transforms. The parameters of the transforms are the knots positions and the derivatives at each knot. These parameters are obtained through a 2-layers Multi-Layer Perceptron from the vector $\hat{\boldsymbol{\theta}}_k$.

Normalizing flows are implemented as an add-on block to the previously described architecture, so it transforms the predicted Gaussian distribution $\hat{f}_k = h^{(0)}$ into an arbitrarily shaped distribution $\hat{g}_k = h^{(M)}$ using $M = 5$ transforms. The number of transform was optimized as an hyper-parameter through Bayesian optimization. The transform applied to the Gaussian distribution $\hat{f}_k$ described in Section IV.3.3 is then $\mathcal{T} = \mathcal{T}_0 \circ ... \circ \mathcal{T}_M$, and the set of parameters $\boldsymbol{\Xi}$ used for optimization contains the parameters of both the encoder and the normalizing flows block.

IV.3.5 Final architecture

The final proposed architecture is shown in Figure IV.5. It uses two convolutional encoder for numerical weather prediction data (2D convolutions in the spatial dimension, 3 layers) and ground stations data (1D convolutions in the temporal dimension, 2 layers) to project the large amount of input data onto a latent space of dimension (13, 60). A Multi-Layer Perceptron of two fully connected layers is added with ReLU activation to obtain a time series of multivariate Gaussian distribution. Final care is given to ensure positive semi-definiteness for the covariance matrix with exponential and hyperbolic tangent activation functions for the correlation matrix.

To avoid over-fitting, dropout layers are added to each of the two encoded blocks. The

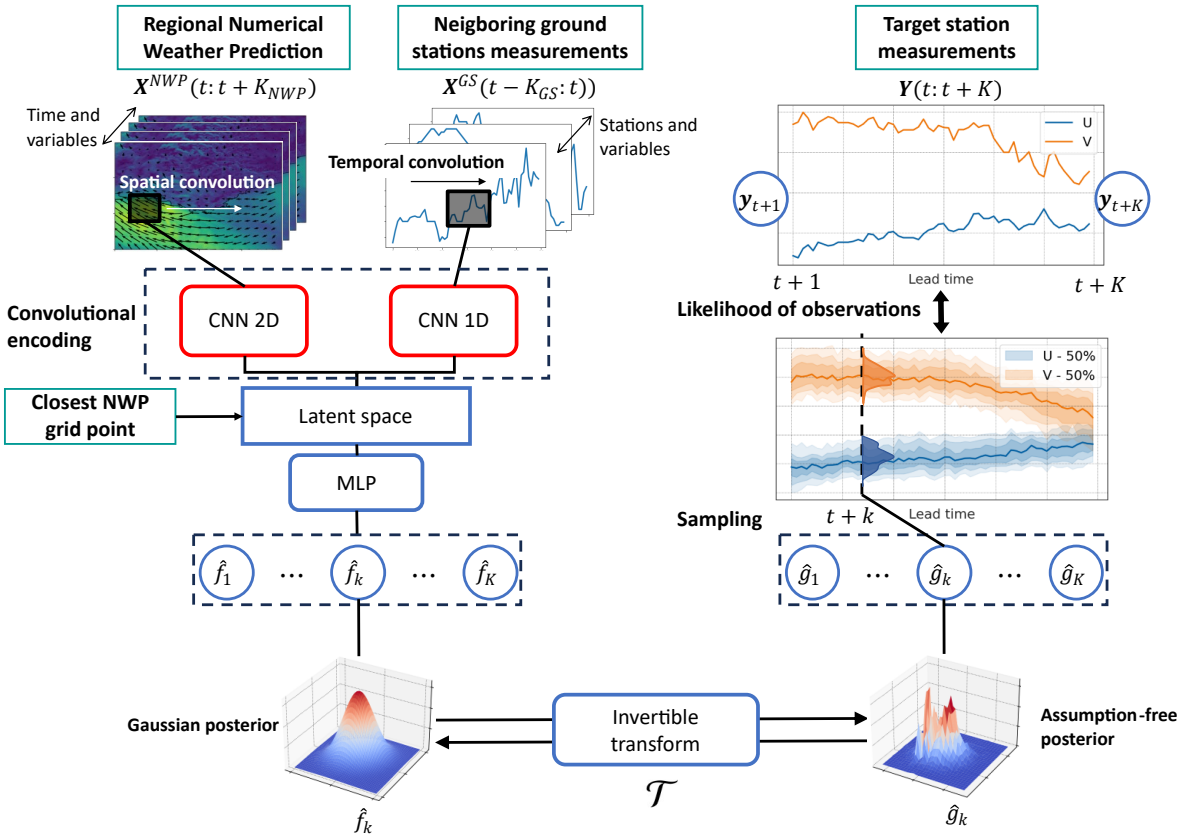


Figure IV.5 – Architecture of the ConvE-STF model illustrated for a forecast issue time t. Probabilistic forecast at the target station for lead times $t+1:t+k$ are obtained from numerical weather prediction $\boldsymbol{X}_t^{NWP} = \boldsymbol{X}^{NWP}(t:t+K_{NWP})$ and recent neighboring ground stations measurements $\boldsymbol{X}_t^{GS} = \boldsymbol{X}^{GS}(t-K_{GS}:t)$. A convolutional encoder outputs a time series of multivariate Gaussian distributions $\{\hat{f}_k\}_{k\in[1,K]}$ that are passed through an invertible transform $\mathcal{T}$ to output the predicted posterior distributions $\{\hat{g}_k\}_{k\in[1,K]}$.

final model with Gaussian outputs has 2.6 million coefficients. Note that under Gaussian posterior assumption, the predicted distribution $\hat{g}_k$ is equal to the Gaussian distribution $\hat{f}_k$.

The normalizing flows add-on block is trained together with the main architecture, transforming the predicted Gaussian multivariate distribution into an arbitrarily shaped distribution. The transformation is made for each time step and is composed of 10 layers parameterized with 1 fully connected layer of 128 hidden features. It adds 0.8 million parameters to the initial model.

The proposed architecture is named thereafter ConvE-STF for Convolutional Encoder for Short-Term Forecasting. When considering a normalizing flows transformation, it is

named ConvE-STF-NF.

All hyper-parameters of the ConvE-STF and ConvE-STF-NF models were obtained using Bayesian optimization presented in Section IV.4.5 to minimize validation loss.

IV.4 Baselines and metrics

We describe below the state-of-the-art methods used as baselines to benchmark the proposed schemes. Considered performance metrics are detailed in Section IV.4.6.

IV.4.1 Closest AROME grid point

The most straightforward baseline consists in considering the output of the AROME numerical weather prediction model at the closest grid point $(i_c,\ j_c)$ to the target station. A linear regression computed on the training split is applied to the prediction

$$\Psi^{AROME}(\boldsymbol{X}_t) = \frac{\boldsymbol{X}^{NWP}_{t,(i_c,j_c)} - \hat{\boldsymbol{\beta}}_0}{\hat{\boldsymbol{\beta}}_1}, \tag{IV.11}$$

with $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ computed using ordinary least squares. $\boldsymbol{X}^{NWP}_{t,(i_c,j_c)}$ is the numerical weather prediction wind speed at the closest grid point from the target station. It is a deterministic output and is noted AROME in all that follows.

IV.4.2 Analogs forecasting

Analogs forecasting is a simple yet efficient statistical method for the forecasting of dynamical systems with unknown dynamics [56]. From a catalog of past trajectories, analog situations are looked for according to a certain distance metric. The D nearest analogs of the current situation are selected, and their trajectories are considered as possible future scenarios. The analogs are weighted according to their distances to the target situation, then mean and covariance matrices are estimated from the ensemble of trajectories under Gaussian assumption.

In this work, the distance metric in the catalog is the Minkowski norm on the variables of the reduced dataset. The weighting of the trajectories and the estimation of the Gaussian distribution is done under locally-constant assumption using $D = 12$ analogs (see e.g., [56] and [182]). Hyper-parameters of the analogs model were tuned with Bayesian optimization to minimize the validation loss.

IV.4.3 Gradient Boosting Machine

Gradient boosting machines are tree-based regressions methods that train an ensemble of weak-learner regression trees to perform a multiple nonlinear regression between output and input. Such methods are implemented in [4] to create probabilistic significant wave height forecasts for offshore wind turbine access forecasting.

The gradient boosting algorithm uses the steepest descent algorithm to optimize the ensemble of regression trees according to a given loss function [186]. Hyper-parameters are the number of regression trees, the number of splits for each tree and a shrinkage parameter that controls the weight of each tree in the ensemble. These parameters were tuned with Bayesian optimization to minimize the validation loss. In this work, a gradient boosting machine is trained with the quantile loss for each predicted quantile $\alpha \in Q = \{0.05, 0.15, ..., 0.45, 0.5, 0.55, ..., 0.85, 0.95\}$, variable $n \in [1, N]$ and lead time $k \in [1, K]$. For a 2-dimensional output, the full model then consists in 1320 individual models. The predicted quantiles form a marginal quantile function for the 2 output parameters for each lead time. The obtained model is noted Ψ^{GBM} and referred to as GBM.

$$\Psi^{GBM}(\boldsymbol{X}_t^r) = \{\Psi^{GBM}_{\alpha,\, n,\, k}(\boldsymbol{X}_t^r)\}_{\alpha \in Q,\, n \in [1,N],\, k \in [1,K]}. \tag{IV.12}$$

Overall the output of each individual gradient boosting machine model contains the quantile prediction $\Psi^{GBM}_{\alpha,\, n,\, k}(\boldsymbol{X}_t^r) = \hat{q}_{\alpha,\, n,\, k}$ for a specific quantile α, variable n and lead time k.

For each time step and variable we approximate the quantile function from the quantiles of the distribution. In addition to second order derivative continuity at the predicted knots, the monotony of the quantile function needs to be preserved. Cubic spline interpolation is then used [140, 141] to obtain the quantile function from the predicted knots. It is a commonly used assumption for quantile function smoothing [4, 142]. Samples can then be drawn from this approximate quantile function to compute scores and generate scenarios. The quantile probabilistic description has the advantage of being assumption-free on the shape of the posterior distribution. However, there is no explicit formulation for the likelihood of the distribution and quantile crossing can appear. It also has a substantial computational cost by requiring one model per quantile, variable, and lead time. It has no explicit control for over-fitting, as it is only controlled by the hyper-parameters of the fitting of regression trees.

IV.4.4 ConvE-STF-reduced

To compare the statistical baselines with the proposed architecture, an additional baseline model is added. It consists of a similar convolutional architecture as the one of the proposed model in Section IV.3, but running with the reduced dataset described in Section IV.2.5 as input. This reduced baseline is noted ConvE-STF-reduced. Its hyper-parameters are tuned using Bayesian optimization to minimize the validation loss.

IV.4.5 Hyper-parameters tuning

We tuned the hyper-parameters of the different models presented in the following sections, and those of the reduced dataset in Section IV.2.5, using a Bayesian optimization framework [247] with the loss metric on the validation dataset as optimization metric. Using the Python package optuna [247], it relies on tree structured Parzen estimators [248] to retrieve optimal hyper-parameters within a pre-defined search space. This Bayesian optimization applies to the following hyper-parameters for the ConvE-STF model: kernel size, pool size, number of convolutional layers, dropout rates, latent space dimensions, number of fully connected layers, number of neurons in the fully connected layers, learning rate, weight decay, learning rate decay rate, batch size. For the gradient boosting machine, it is applied to reduced dataset features, learning rate, number of trees, maximum depth, minimum leaf samples and minimum split samples. For the analogs, it is applied to number of analogs, distance metric and regression mode. Eventually for the normalizing flows, the number of layers, number of hidden features, number of spline function bins and dropout rate are optimized.

IV.4.6 Evaluation metrics

Forecast quality is evaluated using an ensemble of deterministic and probabilistic metrics [50]. Deterministic metrics compare the mean or median of the predicted distribution with observations. The mean value of the predicted distribution $\hat{f}_{t+k}$ is $\bar{\hat{\boldsymbol{y}}}_{t+k}$ and the median value is $\tilde{\hat{\boldsymbol{y}}}_{t+k}$. The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are used in this work. Both metrics do not penalize outliers as strongly. The metrics are computed for each lead time k and noted with a subscript k when given as such. Global metrics across the dataset are averaged over all lead times and are noted without

subscripts.

$$RMSE_k = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{y}_{t+k} - \bar{\hat{\boldsymbol{y}}}_{t+k})^2} \tag{IV.13}$$

$$MAE_k = \frac{1}{T}\sum_{t=1}^{T}|\boldsymbol{y}_{t+k} - \tilde{\hat{\boldsymbol{y}}}_{t+k}| \tag{IV.14}$$

For probabilistic forecasts, the full predicted distribution should be scored against the observations. The Continuous Ranked Probability Score (CRPS) is a proper scoring rule for evaluating the performance of a distribution versus observations [49]. It is a univariate score that is computed for each variable $n \in [1, N]$ and noted with a subscript n for the variables. The global score is averaged across variables and noted without subscript. The CRPS integrates the difference between the predicted cumulative density function and the indicator function at the observation value.

$$CRPS_n = \frac{1}{T}\frac{1}{K}\sum_{t=1}^{T}\sum_{k=1}^{K}\int_{-\infty}^{+\infty}\left(\hat{F}_{t+k}(y) - \mathbb{1}(y \leq y^n_{t+k})\right)^2 dy \tag{IV.15}$$

When the cumulative density function is not tractable, the CRPS can be computed from samples drawn from the distribution. Authors in [49] show that the CRPS can be computed from an ensemble of L samples as

$$CRPS_n = \frac{1}{T}\frac{1}{K}\sum_{t=1}^{T}\sum_{k=1}^{K}\left[\frac{1}{L}\sum_{l=1}^{L}\left(\left|y^n_{t+k} - \hat{y}^{n,(l)}_{t+k}\right|\right) - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{m=1}^{L}\left(\left|\hat{y}^{n,(l)}_{t+k} - \hat{y}^{n,(m)}_{t+k}\right|\right)\right], \tag{IV.16}$$

with $\hat{y}^{n,(l)}_{t+k}$ a sample $l \in [1, L]$ from the predicted distribution of variable n and y^n_{t+k} the corresponding observation. The CRPS is equivalent to the MAE for deterministic forecasts [50].

The Energy Score (ES) is the multivariate generalization of the CRPS and can be computed from samples seemingly to Equation (IV.16) such that

$$ES = \frac{1}{T}\frac{1}{K}\sum_{t=1}^{T}\sum_{k=1}^{K}\left[\frac{1}{L}\sum_{l=1}^{L}\left(\left|\left|\boldsymbol{y}_{t+k} - \hat{\boldsymbol{y}}^{(l)}_{t+k}\right|\right|\right) - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{m=1}^{L}\left(\left|\left|\hat{\boldsymbol{y}}^{(l)}_{t+k} - \hat{\boldsymbol{y}}^{(m)}_{t+k}\right|\right|\right)\right], \tag{IV.17}$$

with $||\cdot||$ the Euclidian norm. CRPS and ES are mostly sensitive to the first moments of the distributions [58] so the Variogram Score (VS) is introduced. It only scores the correlation

structure between the predicted variables and ignores the bias. It can be computed from samples as

$$VS_p = \frac{1}{T}\frac{1}{K}\sum_{t=1}^{T}\sum_{k=1}^{K}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\left|y_{t+k}^{i} - y_{t+k}^{j}\right|^p - \frac{1}{L}\sum_{l=1}^{L}\left|\hat{y}_{t+k}^{i,(l)} - \hat{y}_{t+k}^{j,(l)}\right|^p\right)^2\right], \tag{IV.18}$$

with p the order of the variogram. It is set to 0.5 as recommended by [50], and the scores $VS_{0.5}$ is noted VS for simplicity.

Eventually, rank histograms are used to assess the reliability of the forecasts [249]. A probabilistic forecast is reliable if it predicts probabilities that fits with the observed relative frequencies. In the rank histogram, the quantiles in which fall the observations are counted. For an infinite number of observations, $\frac{1}{Q}$ of it should fall in the $\alpha \in Q$ quantile. The frequency of observed observations are displayed as bar plots and a perfectly reliable forecast should display a flat rank histogram (i.e. uniform distribution). The multivariate generalization of the univariate rank histogram can be found in [48].

The Rank histogram is quantitatively evaluated thanks to the reliability index that measures the mean deviation of the bins to the perfect reliable model. With $\hat{b}_j$ the frequency of observation falling below the j-th predicted quantile $\hat{\alpha}_j$, the reliability index is defined as:

$$REL = \frac{1}{Q}\sum_{j=1}^{Q}|\hat{b}_j - \frac{1}{Q}| \tag{IV.19}$$

IV.5 Results

IV.5.1 Forecast evaluation

Table IV.1 show the generalized scores obtained through cross-validation, with the best values shown in bold. The results proved to be similar for all training-validation-test splits. All implemented methods improve the RMSE compared to the AROME forecast, showing the necessity to post-process the output of numerical weather prediction models for a specific target station.

Baseline models using the reduced dataset as input are all skillfull at post-processing the numerical weather prediction with for instance a 26% decrease in RMSE for the gradient boosting machine forecast. The analogs forecast also improves by 14% the RMSE,

Table IV.1 – Probabilistic and deterministic metrics of implemented forecast models. The best obtained scores are show in bold. The bracket scores show the MAE which is equivalent to the CRPS for deterministic forecasts. The scores are given as mean and standard deviation over the 6 splits.

Model	RMSE [m.s^{-1}]	CRPS [$m.s-1$] (MAE [m.s^{-1}])	ES [m.s^{-1}]	VS
AROME	2.60 ± 0.04	(1.98 ± 0.02)	-	-
Analogs	2.23 ± 0.09	1.20 ± 0.05	1.89 ± 0.07	0.61 ± 0.01
GBM	1.93 ± 0.04	1.05 ± 0.02	1.65 ± 0.03	0.52 ± 0.01
ConvE-STF-reduced	1.93 ± 0.04	1.04 ± 0.02	1.65 ± 0.03	0.53 ± 0.02
ConvE-STF	1.57 ± 0.04	0.84 ± 0.02	1.31 ± 0.04	**0.39 ± 0.01**
ConvE-STF - NF	**1.56 ± 0.07**	**0.82 ± 0.04**	**1.29 ± 0.06**	**0.39 ± 0.02**

with a higher variability. The proposed ConvE-STF architecture largely outperforms the gradient boosting machine by 0.36m.s^{-1} in RMSE and 0.21m.s^{-1} in CRPS, achieving a 40% reduction in RMSE compared to AROME. The ConvE-STF-reduced forecast is just as good as the gradient boosting machine model but is largely surpassed by the ConvE-STF model using the full input. It highlights the presence of explanatory variables in the input dataset and illustrates the capabilities of deep learning architecture to process a large amount of heterogeneous input. The ConvE-STF is 25% better that the gradient boosting machine at predicting the correlation structure between the outputs as shown by the VS, showing that the Gaussian description is competitive for the 2-dimensional wind probabilistic forecast. Eventually, the ConvE-STF with normalizing flows block slightly improves the scores of the Gaussian output, with a higher variability between splits.

The evolution of the generalized RMSE as function of lead time is shown in Figure IV.6. Whereas the error clearly increases with the lead time for the AROME baseline, it is not exactly the case for the other models, for which the error stagnates or even decreases for the first 4 hours of forecast. The trend is not visible in the AROME baseline, and is equally captured by analogs, gradient boosting machine, and ConvE-STF methods. The same error trend is obtained whatever input data is taken as input. It means that the forecast error is more explained 2 hours ahead than 6 minutes ahead, which can either be due to diurnal effects coupled with fixed forecast issue times (6am, 12am, 6pm, 12pm), or lags between closest ground stations and target station. All in all, the proposed approach largely outperforms all baselines for all lead times.

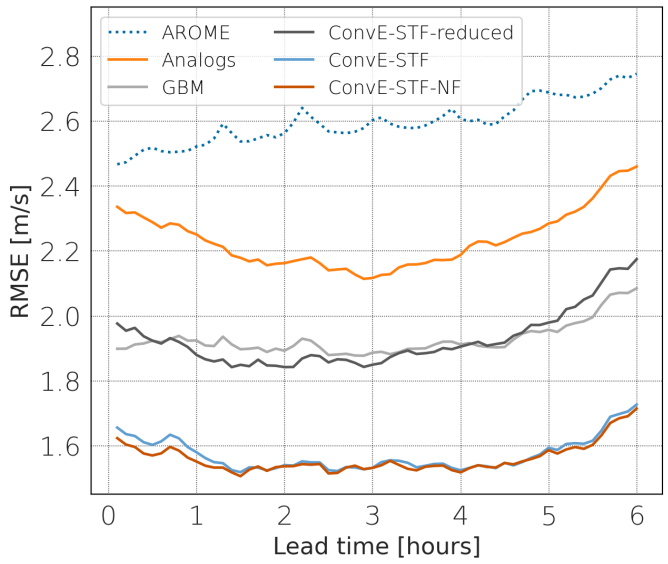


Figure IV.6 – Evolution of the RMSE for all models as function of lead time. The top dashed line is the output of the numerical weather prediction AROME corrected.

The spatio-temporal correlation between the neighboring stations and the target station helps correcting the numerical weather prediction in the very short-term. The ConvE-STF model, with its ability of ingesting a large amount of input data, shows a significant improvement throughout the forecast window.

IV.5.2 Reliability

In Figure IV.7, the observed quantiles are plotted versus the predicted quantiles as a rank histogram for all the forecast models. The dashed line represents a perfectly reliable forecast. The rank histogram is computed for each train-validation-test split and the 50% inter-quantile range between splits is shown as error bars, and found to be the main source of variability in the forecast scoring as described in Section II.1.5. The gradient boosting machine and ConvE-STF-reduced models show clear U-shaped rank histogram, which shows underdispersion (i.e. an under-estimation of the uncertainty). The analogs model, while showing poor deterministic and probabilistic quality metrics, is reliable though slightly overdispersive. Indeed, the analogs estimate a Gaussian distribution from existing trajectories, which guarantee a certain stability in the uncertainty estimation. However, the limited size of the catalog used can explain the overdispersion. The ConvE-STF and ConvE-STF-NF reliability is even more acceptable, with a slight

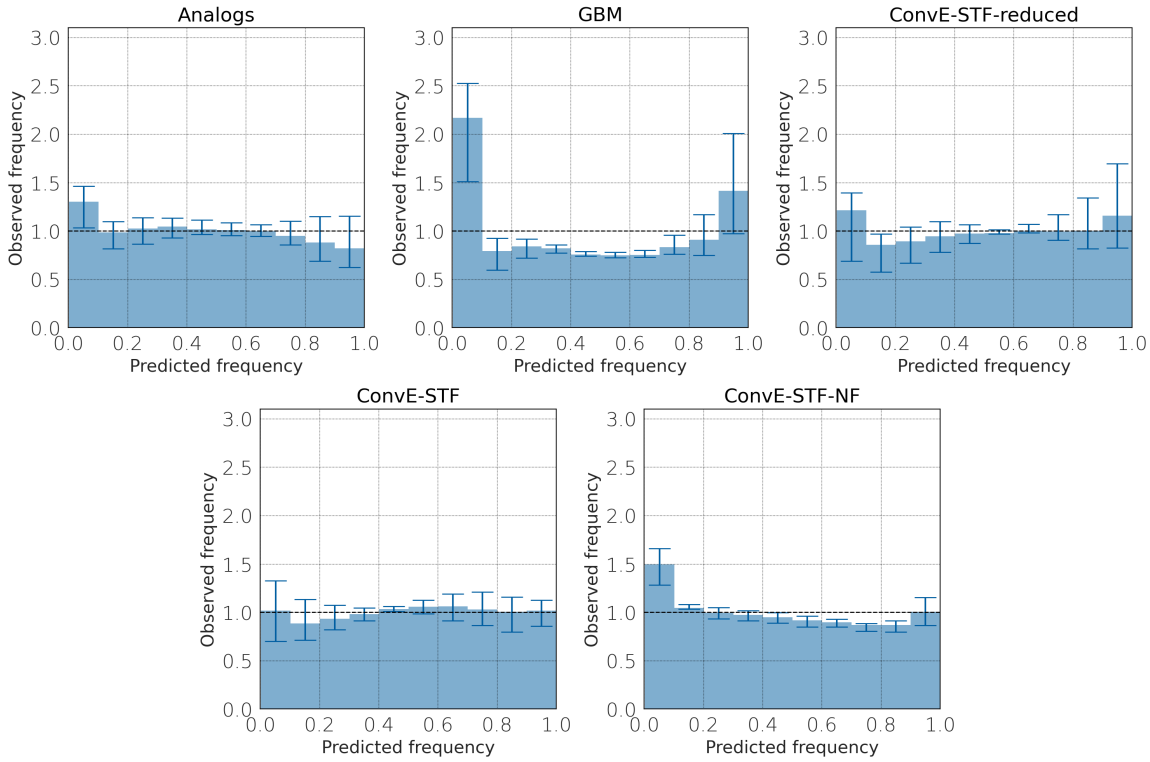


Figure IV.7 – Generalized 2D rank histograms obtained on the test set. Perfect model calibration is showed as a dashed black line.

difference for extreme quantiles. The difference in reliability between ConvE-STF and ConvE-STF-reduced shows that the choice of input data is of greater importance for forecast reliability than the choice of the posterior distribution. The ConvE-STF-NF and ConvE-STF achieve relatively similar reliability patterns with different posterior assumptions but same input data and similar architectures.

The generalized reliability index is given in Table IV.2 to quantitatively assess the models' reliability. The very high variability with cross-validation shows the sensitivity of models' reliability to the training dataset. It highlights the limitations of the obtained models due to dataset length. The ConvE-STF-NF model is the most reliable model with a reliability index of 1.4, and the lowest variability. The use of normalizing flows improves the model's reliability, showing the interest of relaxing the posterior parametric assumption for multivariate probabilistic forecast.

Table IV.2 – Generalized reliability index for all models

Model	Reliability index
Analogs	1.6 ± 0.5
Gradient boosting machine	3.2 ± 0.5
ConvE-STF-reduced	2.1 ± 1.2
ConvE-STF	1.5 ± 0.6
ConvE-STF - NF	$\mathbf{1.4 \pm 0.4}$

IV.5.3 Data representativity

This study relies on a 33 months-long dataset to develop and benchmark deep-learning-based post-processing models. Following similar previous studies [198, 210, 250], we aim to assess the potential impact of the length of training dataset on the generalization performance of the trained models. We then train and assess the proposed ConvE-STF models using training datasets of different lengths from 2 months to 2 years as illustrated in FigureIV.8. Overall, we observe the expected trend, the longer the training dataset, the better the model performance. In FigureIV.8(a), the RMSE skill score shows that from 60-days-long training datasets, we train ConvE-STF models which are more skillful than the AROME forecast. We also note a slower improvement of the forecasting skills from one-year-long datasets, as well as a lower variability between cross-validation splits. Similar results are observed for the model reliability in FigureIV.8(b). These results support the relevance of training datasets covering at least one or two years to retrieve a robust average improvement through the ConvE-STF models of the AROME forecasts.

IV.5.4 Computational cost

The computational cost of the different models was evaluated for training, inference and sampling. The deep learning models (ConvE-STF, ConvE-STF-NF, ConvE-STF-reduced) are trained on a single 32Go NCIDA RTX A6000 GPU. The gradient boosting model is trained on multiple (60) AMD EPYC 7763 CPU. The obtained computational costs are given in Table IV.3.

The training of a gradient boosting machine for quantile forecasting requires the training of a single model for each variable, lead time and quantile. In this study this results in 1320 individual models, and a heavy model file (348 Mb), requiring multi-CPU train-

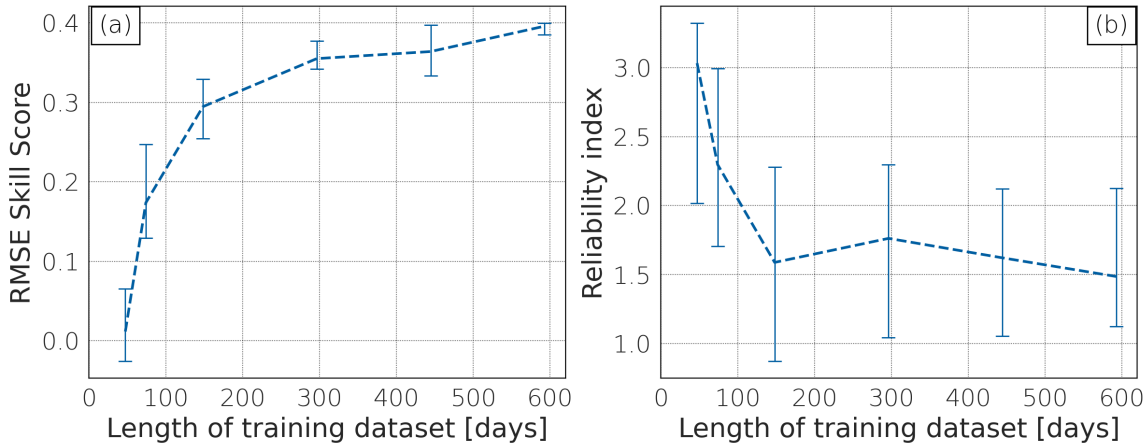


Figure IV.8 – Model improvement with training dataset length obtain with cross-validation. RMSE skill score versus AROME forecast (a) and reliability index (b).

Table IV.3 – Computational cost comparison

Model	Machine	Model size [Mb]	Training CPU/GPU time [s]	Inference time [s]	Sampling time [s]
Analogs	CPU	0	0	0.076	0.076
GBM	60 CPU	348	3600	0.0005	1.05
ConvE-STF-reduced	GPU	1	45 (0.4 s/epoch)	0.008	0.003
ConvE-STF	GPU	10	500 (1.4 s/epoch)	0.015	0.006
ConvE-STF - NF	GPU	16	3500 (4.3 s/epoch)	0.018	0.18

ing. The training time is then $O(NTQ)$, with N the number of samples, T the number of predicted lead times and Q the number of quantiles. Deep learning models are easily parallelized using GPU, resulting in a training time of $\approx 500s$ for ConvE-STF on a single GPU. The addition of normalizing flows implies transformations inversion that adds computational cost for error gradient back propagation, making it 3 times slower to train than ConvE-STF. Analog methods need no training time, making it a very simple to implement probabilistic forecast framework. The sampling from the predicted distributions is more efficient under Gaussian assumption. Normalizing flows transformation makes it 300 times slower than with a simple Gaussian posterior, and the sampling using the empirical quantile function for gradient boosting machine is 2000 times slower.

IV.5.5 Probabilistic wind speed forecasts

The quantile description output by the gradient boosting machine is flexible as it makes no assumption on the underlying distribution. It can in theory capture heavy tail or multi modal distribution. However, it is limited to the prediction marginal distributions, and the correlation structure is not explicitly described. This is observed with the VS in Section IV.5.1. The lack of correlation structure in the gradient boosting machine output is a drawback for the joint probabilistic forecasting of correlated variables. For the 2-dimensional wind speed, it can result in unrealistic sampled wind direction. Though it is hard to measure the impact of the correlation structure with standard statistical metrics, it is expected to strongly impact the generation of multivariate scenarios for offshore wind operations weather window forecasting.

In the ConvE-STF and analogs methods, a multivariate Gaussian assumption is made for the output with $(2, 2)$ covariance matrices. The Gaussian assumption can be relaxed using normalizing flows in the ConvE-STF-NF model but no clear quantitative effects are observed in terms of model performance. However, the normalizing flows approach adds little computational cost to the previous Gaussian assumption. By construction, the likelihood can be easily calculated, and samples can be directly generated. It can in theory adapt to complicated posterior distributions with a limited added model complexity. A sample from the latent Gaussian distribution is passed through several layers of neural splines [125] to be transformed into a sample in real space. The non-linearities within the neural spline flows can approximate very complex distributions and are conditioned by the input data. By doing so, we lift any assumption on the posterior data, compared to the quantile approach or Gaussian assumption.

The shapes of the predicted distributions from the different methods are illustrated in Figure IV.9 for two entries in the test dataset. For the first entry (figures (a), (c), (e)), the gradient boosting machine distribution has heavy tails, showing the flexibility of the quantiles. For the second entry (figures (b), (d), (f)), it has a very low spread, probably due to over-fitting. Figures (a) and (b) show multi-modal distributions obtained with normalizing flows. The distributions shapes are not very different from the Gaussian distributions in figures (c), (d), but show a discretization in wind direction. This is an artefact of the dataset, knowing that the wind direction at the target station is measured with a resolution of 5 degrees. Normalizing flows can partially capture this complicated relationship between the predicted variables in a non-supervised way. It shows the great flexibility of normalizing flows for probabilistic forecasting.

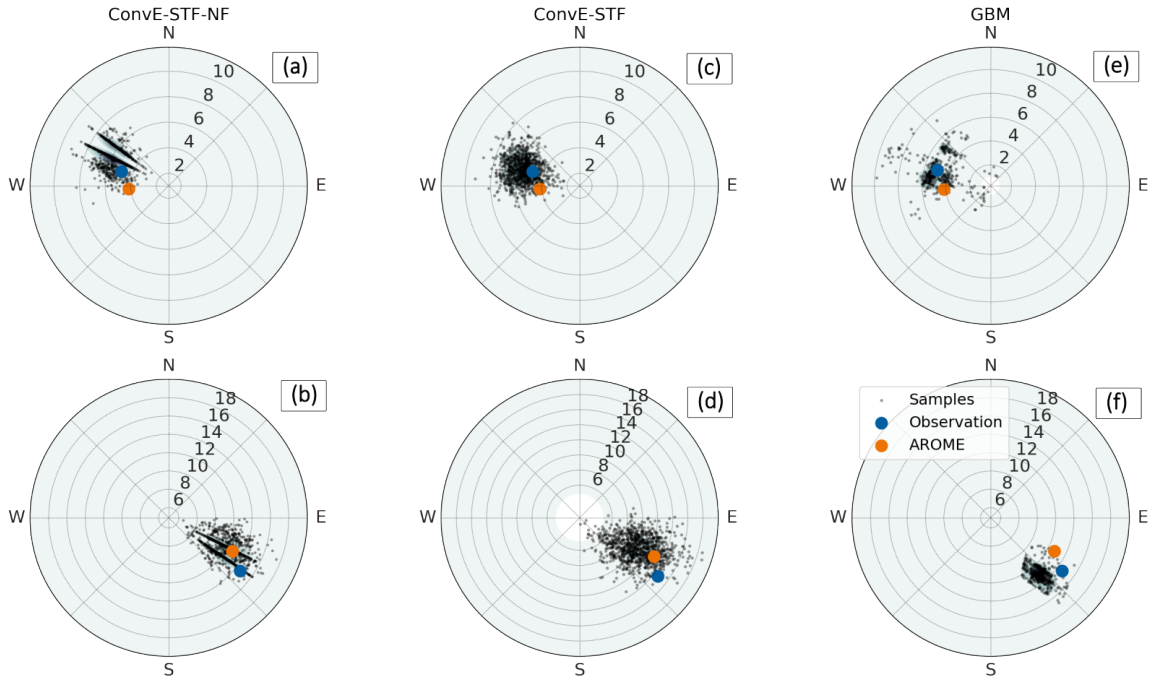


Figure IV.9 – The three different probabilistic approaches are illustrated on two entries of the dataset. Samples generated from the predicted distribution in the generative case (ConvE-STF-NF), Gaussian case (ConvE-STF) and quantile case (Gradient boosting machine, GBM) are scattered on polar plots of wind speed and wind direction. The observation is shown as a blue circle and the AROME prediction as an orange circle.

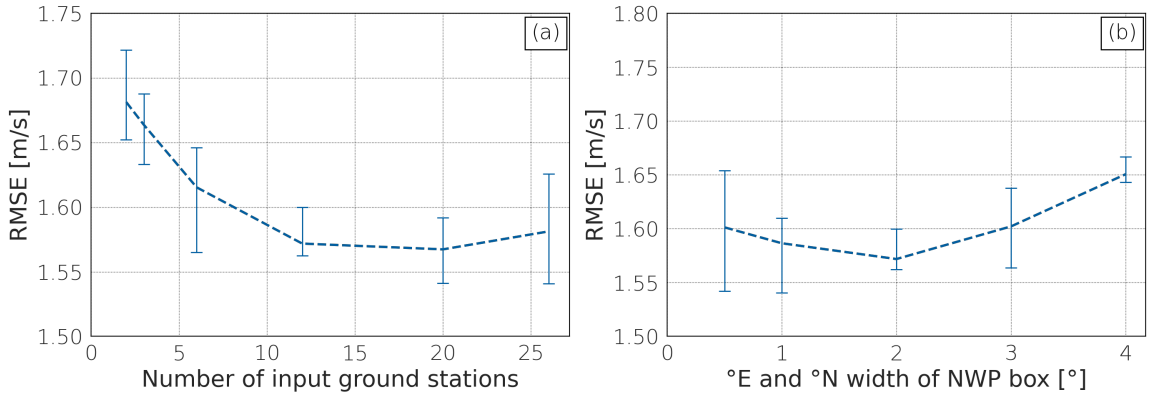


Figure IV.10 – Sensitivity of the RMSE to the number of ground stations taken as input (a) and to the size of the numerical weather prediction input (b). The dashed line is the generalized RMSE, and blue error bars show the 50% inter-quantile range over the 6 splits.

IV.5.6 Input sensitivity

The ConvE-STF is trained with different input sets to compare the value of each data source. The size of the numerical weather prediction domain and the number of neighboring stations are the two main parameters considered for sensitivity. They are crucial parameters for the method generalization, and they can give indications on explanatory variables importance.

In Figure IV.10 (a), the sensitivity of RMSE to the number of ground stations used as input is plotted. A clear trend is identified, with a decreasing RMSE for the 12 closest stations, and a stabilization for an increased number of stations. This validates the choice of 12 closest stations as input for the main model. This optimal number of input stations however is strongly depending on the experimental setup. Firstly, it is site-dependant, and represents the limit of spatio-temporal correlation between measured parameters and target parameters. This is function of the distance and position of the neighboring ground stations, which will be specific for every site. Secondly, it depends on the length of the time series considered as input. In this study, we limited the length of the neighboring measurements time series to 6 hours. Longer time series might then exhibit larger areas of spatio-temporal correlation. Eventually, it is depending on the length of the forecast window, which is for this experiment limited to 6 hours.

In Figure IV.10 (b), the sensitivity to the size of the input numerical weather prediction mask is shown. The change in input size (from 20x20 images to 120x159 images) implies a

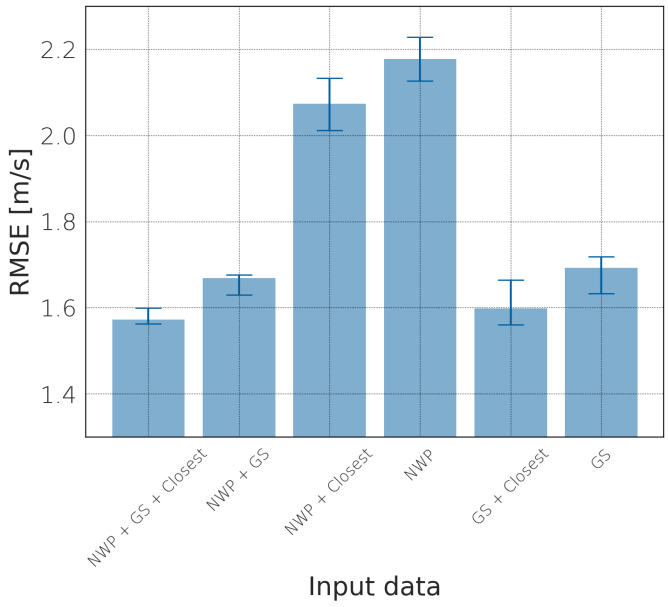


Figure IV.11 – Sensitivity of the RMSE to the input data. Blue bars show the generalized RMSE, and error bars show the 50% inter-quantile range over the 6 splits. Input data is the combination of neighboring ground stations (GS), Numerical Weather Prediction (NWP) and closest numerical weather prediction grid point (Closest).

change in the convolutional architecture (2 to 3 layers). A hyper-parameter tuning for the numerical weather prediction data encoder was made for each input size using Bayesian optimization as described in Section IV.4.5. The link between forecast error and numerical weather prediction input size is not as straightforward and can only be discussed for this specific site. The best performances are obtained with a mask of 2 degrees in latitude and longitude. It is possible that the larger input area in this specific region does not carry more information than the smaller input mask. But there is no guarantee that even larger masks would not bring additional information. In particular, the atmospheric circulation in the eastern Gulf of Lion is notably influenced by the situation in the Gulf of Genoa and Ligurian Sea which would require a wider input mask.

In Figure IV.11, we report the performance of ConvE-STF models using different combinations of wind data as inputs. We consider three wind data sources: namely, the wind measurements from ground stations input (GS), the wind prediction from the operational numerical weather prediction (NWP) for the considered domain, and the wind prediction from the operational numerical weather prediction (NWP) for the grid point the closest to the targeted offshore location (closest) (see FigureIV.5). These results illustrate the rel-

ative importance of the different data sources in the prediction of the ConvE-STF model. The addition of ground stations input greatly improves the RMSE compared to the two central bars. It highlights the importance of neighboring measurements as explanatory variables. From the GS only case, it can be noted that both the addition of numerical weather prediction input and closest grid point input improve the forecast post-processing. It shows that information can be extracted from regional forecasts to improve the forecast at a target station, but that it is hard to capture the forecast at the closest grid point using convolution neural network.

IV.5.7 Qualitative improvements

The forecast quality of ConvE-STF is analyzed as function of the weather situations. The RMSE improvement of ConvE-STF and gradient boosting machine models compared to AROME closest grid point is shown in Figure IV.12. The ConvE-STF model shows general improvement in RMSE compared to AROME, with a RMSE reduction for most wind speed and direction. It shows the model's skills at post-processing numerical weather prediction in most weather situations.

The patterns are relatively similar for both models, but ConvE-STF is notably more efficient than the gradient boosting machine for South-West-blowing winds. This can be due to the processing of a larger amount of coastal in-situ measurements situated upwind from the target station.

However, both models fail to improve the RMSE for North-East and South-West winds with an increased error compared to AROME closest grid point. It is important to note that such winds are relatively rare in the eastern gulf of Lion. Thus, this likely illustrates a shortcoming of the considered training configurations with two-year-long datasets. When such wind situations are not present in the training dataset, deep learning models cannot extrapolate during the test phase for so-called out-of-distribution samples.

IV.6 Conclusion and discussion

This paper proposes a deep learning architecture for the probabilistic wind speed forecast at sea. It uses convolutional neural network to process a large amount of input data and is compared to state-of-the-art statistical methods. Several probabilistic assumptions are proposed for multivariate probabilistic forecasting. A Gaussian posterior is compared

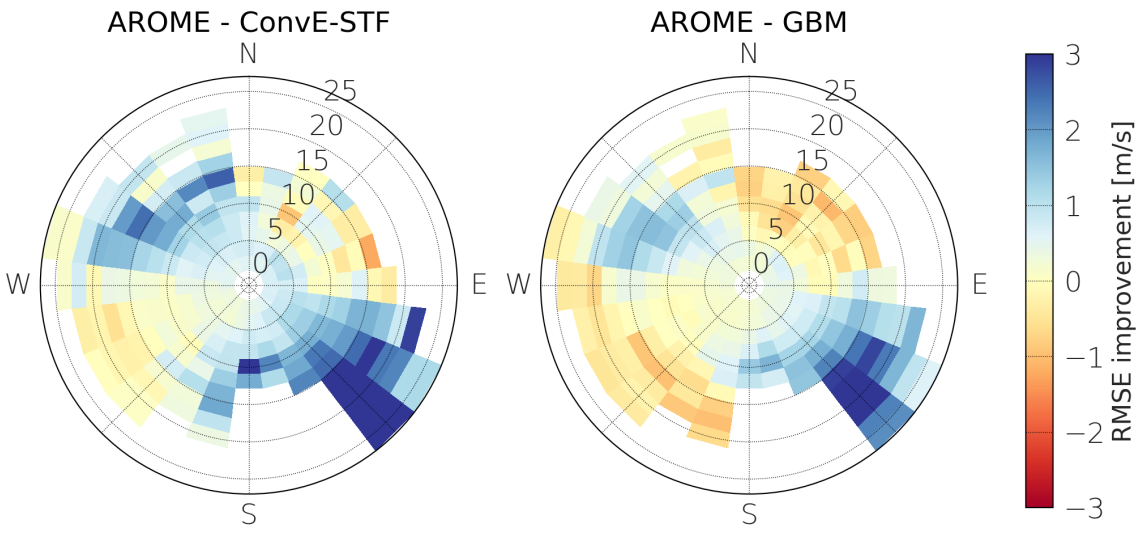


Figure IV.12 – RMSE improvement between ConvE-STF, gradient boosting machine (GBM) and AROME. The RMSE improvement (RMSE(AROME) - RMSE(model)) is shown in color, with blue sectors indicating a RMSE decrease compared to AROME, and red sectors a RMSE increase. The RMSE improvement is plotted as function of wind direction and wind speed.

to a normalizing flows and quantile approaches. The proposed method proves skillful at improving the short-term wind forecast (1 to 6 hours ahead) at a target offshore location, with a 40% reduction in RMSE compared to numerical weather prediction forecast. Other baseline methods improve the forecasts by 14% for analogs to 26% for the gradient boosting machine. It stresses the importance of numerical weather prediction post-processing for offshore applications. Furthermore, the proposed architecture can emulate probabilistic forecasts with a satisfying reliability.

The proposed ConvE-STF architecture shows the best performance in terms of deterministic and probabilistic metrics. It shows an acceptable forecast reliability, with a marginal gain for a Gaussian assumption compared to normalizing flows. Normalizing flows addition can reproduce highly non-Gaussian behaviors for a relatively low computational cost. This can be of great use for multivariate probabilistic metocean forecasting. Other generative models such as GAN, VAE or diffusion models could probably achieve similar results, and were not explored in this study. Normalizing flows however provide a simple yet efficient method to relax the parametric assumption on the posterior distribution.

The use of deep learning methods allows the integration of various sources of data.

It permits the use of recent neighboring measurements that have a great impact on the forecast correction. In the context of offshore operations, it shows the opportunity of post-processing numerical weather prediction using coastal measurements. Moreover, once trained, deep learning models run fast and could enable short-term operational decision making based on high frequency forecasts.

Normalizing flows are used as an add-on block to the ConvE-STF architecture with Gaussian assumption. The normalizing flows transformations conditioning can be constructed in different ways. It is applied in this paper for each lead time independently, and the sampling is to be done for each lead time. The temporal correlation between lead times is not explicit. Whether normalizing flows can be used to jointly model the temporal correlation and variable correlation is still an open question [44].

The considered dataset has inherent limitations. It would be beneficial to complement the study with an extended dataset. The forecast horizon is here limited to 6 hours after forecast issue time. In real operational contexts, offshore operations planning and execution [5], would likely require the extension to 24-hour forecasts. Operational NWP forecasts fulfill this requirement [9]. Furthermore, latency of numerical predictions and in-situ measurements should be included in the experimental framework. Our experiments also assess how the length of the training dataset impacts the forecasting performance of the proposed deep learning scheme. While we retrieve significant average improvement compared with the operational NWP forecast using a 2-year-long training dataset, we also point out limitations for rare events, especially South-West and North-East winds in our case-study. This is likely a limiting factor for a complete forecast evaluation [66], however it shows that a skillful data-driven model can be obtained using 2 years of training data. Related studies applied to wind speed forecasting often use shorter or similar datasets to train post-processing models [198, 210, 250]. Extending the considered dataset to longer times series strongly depends on the availability of longer time series of offshore measurements and requires the deployment of dedicated in-situ observatories [246]. The availability of ensemble NWP forecasts also seems appealing both as a complementary benchmarking baseline as well as to explore how deep learning schemes could benefit from ensemble forecasts as input data [251]. Furthermore, it would be very beneficial to compare the forecasts' reliability with the ensemble prediction of AROME to assess the impact of data representativity on forecast calibration.

Other sources of data could be used to improve the post-processing of numerical wind forecast. For offshore surface winds, sea surface roughness data through satellite Synthetic

Aperture Radar (SAR) images provides high-resolution information [252]. To date, SAR images have a to low temporal availability (2 to 3 days) to be integrated into an operational post-processing model. Further studies on the impact of marine exogenous variables for offshore wind forecasting could be considered.

This study could be extended to jointly forecast wind and wave parameters [253]. Potential non-Gaussian distributions are expected between wind and wave parameters forecast uncertainty. From there, the value of the forecast could be evaluated with regards to probabilistic operational decision making by modelling a realistic maintenance operation [5, 6]. The model reliability is then a crucial parameter to justify the operational use of probabilistic forecasts.

This research has been supported by France Energies Marines and the French Government, managed by the Agence Nationale de la Recherche under the Investissements d'Avenir program, with the reference ANR-10-IEED-0006-34. This work was carried out in the framework of the FLOWTOM project. It is supported by the ANR project OceaniX.

Meteorological data used in this study are available online through the MeteoNet data set. The code developed for the probabilistic short-term forecasting is accessible via `https://github.com/rmarcille/conve_stf_meteonet.git` [254] and uses code from Lguensat et al. [56] for analogs, and Durkan et al. [255] for normalizing flows. The preprocessed dataset is accessible through [256].

Chapter V

Day-ahead probabilistic forecasting of operational weather windows

Preamble

Using the models developed in Chapter IV, this chapter presents results obtained on a realistic dataset for the joint probabilistic forecasting of wind and wave. A dataset based on LIDAR and wave buoy campaigns in the Gulf of Lion off the coast of Marseille is built and presented in Section V.2. This dataset gathers numerical weather and wave predictions, weather in-situ data from coastal stations and neighbouring buoy measurements in the Gulf of Lion. Because of its location next to future development areas for floating offshore wind energy in the French Mediterranean Sea and its unperturbed wind and wave measurements, this site is a great study case for forecast models development.

The models presented in Chapter IV are trained and scored on this dataset for a 24 hours ahead forecast of wind and wave parameters. The models adaptations are presented in Section V.3, with the considered benchmark models. The interest of jointly forecasting wind and wave is discussed in Section V.4, and the validity of Gaussian posterior assumption is evaluated for the parameters of interest.

A description of scenarios generation strategies is given in Section V.3.2. In particular, the Gaussian copula assumption is presented and evaluated for this study case. The possibility to relax the posterior assumptions on the temporal correlation thanks to normalizing flows is discussed. Several frameworks are proposed and experimented. Eventually, the quality of the different scenario generation methods is evaluated. From the scenario forecasts, value metrics for offshore wind energy maintenance operations are presented and proposed in Section V.3.3.3. In connection with the recent literature, a research direction towards an operational risk consideration in the value metric is proposed.

V.1 Introduction

Offshore operations are sensitive to different environmental variables. In Chapter IV, the probabilistic forecasting of wind speed is studied in a multivariate setup. However, most studies applied to weather window forecasting for offshore operations include wave parameters [4, 7, 8, 33]. To create a forecast model that could be used for the weather window forecasting, these wave parameters should be included with the wind speed. As described in Section I.2.4.2, the relationship between wind and wave parameters is complex, and should be modelled carefully. Published work studying jointly wind and wave parameters are applied to the generation of hindcast time series [34–37]. To the best of our knowledge, [38] is the only published work to consider jointly wind and wave parameters in a probabilistic forecast framework. It however focuses on vessel's response to uncertain environmental conditions. In our opinion, the link with weather window forecasting and decision-making should be further studied. To that end, a realistic case study is constructed based on two years of co-located wind and wave measurements off the coast of Marseille. The study site, situated on the Planier island, is particularly representative of offshore wind energy sites in the Gulf of Lion. It appears as a prime location for research studies dedicated to offshore wind energy site characterization in the French Mediterranean Sea.

The integration of probabilistic forecasts for time dependent decision-making requires the sampling of trajectories [144, 145]. For optimal operational decision-making, the probability of weather window should be computed [4, 5]. It corresponds to a probability of exceedance over a sliding time window, and then requires the use of weather scenarios. Indeed, the model presented in Chapter IV is trained to output probability densities for the individual lead times independently. Being able to sample coherently trajectories requires additional model engineering. In Section II.2.3, an introduction to the generation of weather scenarios is given. It describes the importance of methods based on the use of copulas [149] in the literature for generating stochastic scenarios [151, 152]. Copulas are mathematical object that model the high-dimensional dependency between marginal distributions. A review of copulas-based methods for times series can be found in [155]. However, these methods are limited for multivariate conditional scenario generation for several reasons. The construction of the copulas does not necessarily allow for likelihood optimization. For Gaussian copulas, which could in theory be taken conditional to input data, the construction of the covariance matrix is prone to numerical instabilities. More-

over, for the weather window forecasting, multivariate dependent time series are required. Since copulas need univariate marginals, they can not be easily applied in a multivariate framework. Research work on conditional copulas [153] and temporal dependent copulas [154] could be promising solutions. To the best of our knowledge, these methods have never been applied to environmental time series.

Alternatively, the capabilities of normalizing flows could be used to generate multivariate scenarios. The recent work published by [59] is an example of how the temporal relationship between predicted timesteps can be modelled in the latent space. The authors propose to train a RNN to condition the normalizing flow, which would take as input both covariates and the latest observations. By recursively conditioning the normalizing flow in the forecast window, one can train to predict correlated samples through timesteps, by starting from a random sample for lead time $k = 1$. However, the training of the RNN conditioner implies that the target time series is used as input during the forecasting, making it a one-step ahead forecast framework, which is not the case of this study. The RNN conditioner should be realistically trained to be able to capture temporal variations during inference, so a realistic time series should be input.

A more straightforward approach would be to directly train the normalizing flow in the trajectories space, so that the latent distribution encompasses both the temporal and variables dependencies. In this case, the normalizing flow allows for the exact likelihood computation in the trajectories space, and direct sampling of trajectories. However, this leads in this study case to a latent space of dimension NK, which creates great difficulties for training. If a multivariate Gaussian distribution is used in the latent space, it implies the estimation of a $(NK \times NK)$ covariance matrix and NK mean vector for each time step, with a final dimension $D = \frac{NK(NK+1)}{2}$ which is prohibitive for a full matrix prediction, and requires the use of covariance parametrizations. Additional assumptions can be made to simplify the latent space, for example a centred distribution, or a separable covariance matrix with a tractable temporal covariance matrix (exponential, empirical, etc.). However, it then assumes that the normalizing flows has sufficient flexibility to correctly transform samples in dimension NK, and that the conditioning input is sufficiently large and expressive to do so. It becomes likely that the training fails to converge due to the high dimension of the system. It is the case in the recent work from [44] in which a comparison of generative approaches for scenarios forecasting is made. The normalizing flow is constructed as a bijection between a standard normal distribution and the predicted variable (wind power, solar power and power load) which is conditioned by latent

variables extracted from weather forecasts. As opposed to GAN and VAE outputs, the samples generated from normalizing flows have no auto-correlation. The size of the latent space and complexity of the system prevents the model to be expressive enough to generate samples with a realistic temporal evolution. However, the obtained scores are better than with VAE or GAN, this is due to the fact that the temporal correlation does not appear in the scores computations. While the authors advocate for the easiness of use and flexibility of normalizing flows and the quality of the obtained forecasts, we consider that the generated samples should be temporally sound to be used in temporal-dependant decision-making applications such as weather window forecasting.

The application of probabilistic forecast models to the planning of offshore operations requires the development of dedicated metrics [50]. These should be representative of the real costs associated with the use of forecast models. For weather window estimation, the forecast errors result in added costs if a vessel is dispatched with non-operable conditions, or if a weather window is missed. From this observation, [6] proposed an economic metric for forecast evaluation in link with offshore operations. It is shown in [7] that this paves the way to optimal decision-making based on the probability of weather window. In [164], other key indicators for offshore operations are described. In particular, they advocate for the development of indicators in link with security during operations. As described in [60], the pressure for economically optimized decision-making can lead to increased risk during operations. This is confirmed by the discussions that took place within the FLOWTOM project.

This chapter describes on-going research work that bridges all the aspects described above. A real case-study is constructed from joint wind and wave measurements. The probabilistic model ConvE-STF is applied to the forecasting of wind speed at hub height, significant wave height and wave period. Different methods for generating scenarios from its predictions are explored. In particular, two innovative methods based on the use of normalizing flows are proposed. The obtained scenarios predictions are evaluated in an operational framework. A risk forecasting metric including notions of security is introduced.

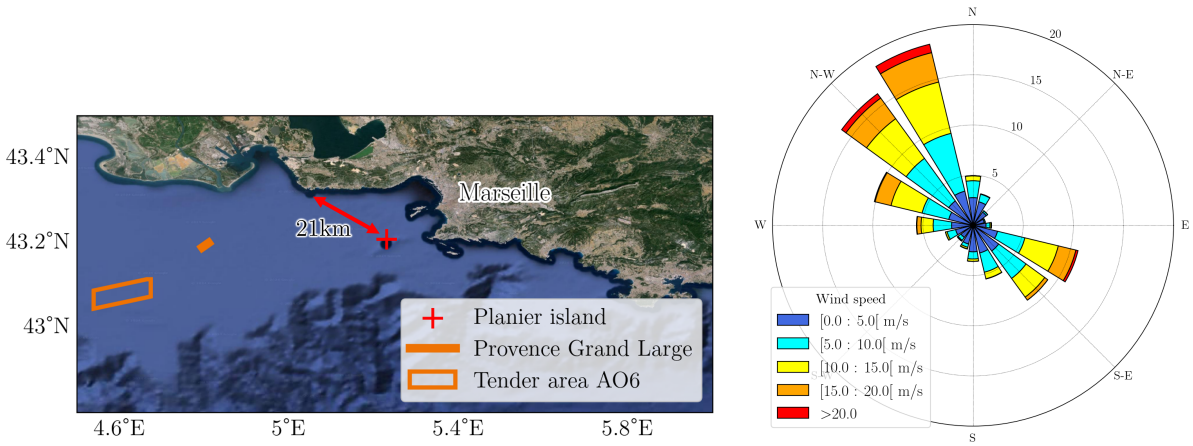
V.2 The Planier island as a realistic study site for offshore wind energy

Similarly to the dataset used in Chapter IV, a dataset composed of in-situ measurements and numerical forecasts is constructed. It is based on in-situ campaigns off the coast of Marseille, on the Planier island (introduced in Figure I.7). This dataset is referred to as the «Planier dataset»hereafter. The Planier situation is illustrated in Figure V.1a, and the measured wind rose for 2023 is illustrated in Figure V.1b. In the directions of the dominant winds (North-West and South-East), the planier island is at more than 20 km of any obstacle, making it a fair study site for offshore wind energy. The pilot floating wind farm Provence Grand Large and its 3 turbines installed in 2023 is in the vicinity of the Planier, and is dominated by the same wind regimes. The future development area for call for tender AO6 is illustrated as an orange box and is located further offshore. This area is supposed to support the development of 750 MW of floating offshore wind capacity in the next decade. This makes the Planier development site a crucial study case for the development of floating offshore wind energy in the Gulf of Lion.

V.2.1 Characteristics of the target measurements

Two LIDAR measurement campaigns at the Planier island are used to construct the wind speed time series. The first one was conducted by the French offshore wind energy developer EOLFI-SHELL spanning two years in 2016 and 2017 using a floating LIDAR next to the island, validated with simultaneous island-based LIDAR measurements. The second one is financed by France Énergies Marines led project POWSEIDOM and has been gathering data with a fixed LIDAR on the island since December 2022 onwards. The wind speed measurements are completed with waves time series measured by the CANDHIS buoy deployed by the CEREMA and Grand Port Maritime de Marseille since 2011 next to the Planier island under the campaign ID 01305. The combination of these three data sources allowed for the creation of a 650 days long offshore in-situ time series.

The LIDAR campaign result in time series of the 10-minutes averaged wind speed at different height levels from 40m to 250m above the LIDAR. Since the limiting parameters for offshore operations are often taken as the 10-meters above surface wind speed and lifting height wind speed, we consider the wind speed at 100m above the surface for target $U_{100}[m/s]$. It might be relatively low for hub-height lifting for future generation



(a) Map of the Planier development site and the neighbouring offshore wind energy sites.

(b) Measured wind rose in 2023 at the Planier development site.

Figure V.1 – The Planier development site is a representative offshore site in the context of offshore wind energy development in the Gulf of Lion. As illustrated in (b), the area is dominated by North-West (Mistral) and South-East winds. The Planier island is not perturbed by coast or obstacles in these dominant directions and is situated 21 km from the coast in the Mistral direction. The pilot site Provence Grand Large and the tender area AO6 for 250MW + 750MW extension floating wind farms are shown in orange in (a).

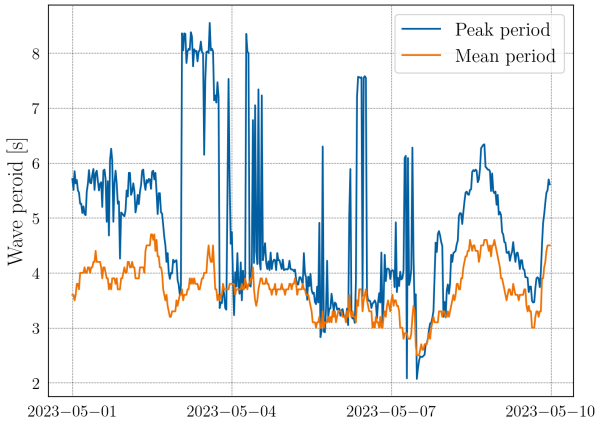
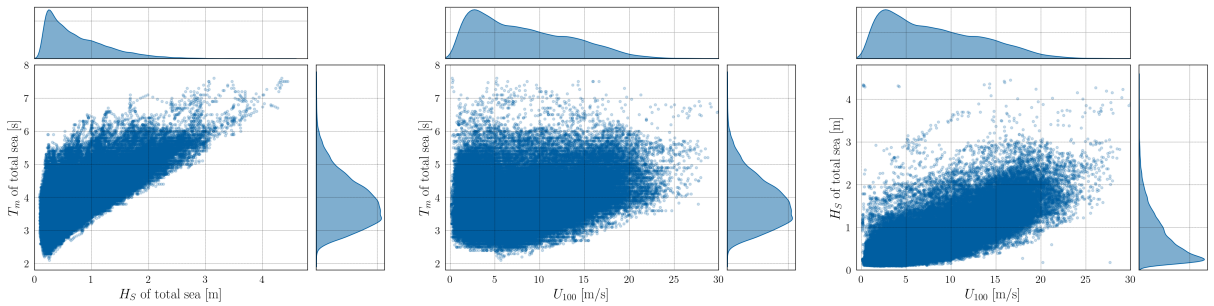


Figure V.2 – Time series of measured wave peak period (blue) and wave mean period (orange) at the Planier CANDHIS buoy. The peak period parameter shows discontinuities, while the mean period parameter is attenuated by lower energy wave systems with lower periods.

turbines, but it corresponds to a height at which the wind speed can be extracted from the AROME numerical weather predictions. The wind direction is only available for the France Énergies Marines campaign. The EOLFI-SHELL campaign is based on a floating LIDAR and the buoy's movements have not been recorded sufficiently well to permit for wind direction correction.

The CANDHIS buoy's measurements time series contain different spectral parameters computed from the wave measurements every 30 minutes. The main limiting parameters for offshore operations are the significant wave height and the wave period. The significant wave height of the total sea is considered as target, which can be the result of different wave systems $H_S[m]$. The wave period parameter is a bit trickier. Usually, the peak wave period is considered. It is defined as the period of the most energetic peak in the wave spectrum. However, its definition brings discontinuities in the time series, corresponding to instants when ne most energetic peak switch from a wave system to another. Though these discontinuities are physically meaningful, and naturally emerge from numerical modelling, they are hard to capture by a data-driven model, and make the peak-period a challenging variable for post-processing. To alleviate these effects, the mean period is considered as the training target and used as operational limitation hereafter $T_m[s]$. The discontinuities in the waves peak period is illustrated in Figure V.2.

The data distributions at the target station are shown in Figure V.3 for 2 years of joint wind and wave measurements. The relationship between the variables is complex

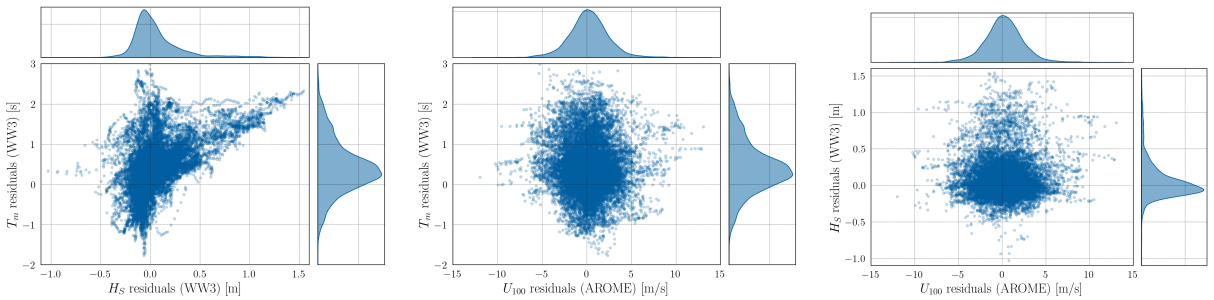


(a) T_m as function of H_S. (b) T_m as function of U_{100}. (c) H_S as function of U_{100}.

Figure V.3 – Scatter plot of the target variables in the test dataset. The individual distributions are shown at the top and right of the scatter plots. The relationship between wind speed and wave parameters is complex which advocates for their joint forecasting.

and illustrate the fact that the variables are influenced by external factors (fetch for wave time series, sea surface temperature for wind speed etc.). The chosen variables do not make a difference between wind waves and swell, which makes their interpretation complicated. For a data-driven model to capture such complex dependencies, appropriate explanatory variables should be used as input. By using a combination of spatialized numerical predictions and environmental in-situ time series from neighbouring sites, we aim at providing a sufficient amount of explanatory variables to accurately reconstruct these relationships.

Not only the designed forecast model should jointly predict the wind and wave parameters, but it should also predict their joint forecast uncertainty. The interest of predicting the joint uncertainty between variables is to be discussed further on. Figure V.4 shows the scatter plot and distribution of residuals for the target variables computed from the forecasts of the closest numerical grid point from the numerical wave and weather predictions (WW3 and AROME). It shows that the forecast error is positively correlated between the wave period and significant height, while it is slightly negatively correlated for wind speed and wave period. The relationship between wind speed residuals and significant wave height residuals is less significant. These graphs depict the error resulting from the deterministic forecast from numerical predictions. The uncertainty associated with each of these forecast points might also show specific correlations. For example, if the wind speed forecast is heavy tailed, i.e. with a risk of high winds, chances are that the significant height of wind waves forecast should also be heavy tailed, and that a sample drawn from the joint wind speed and wave height distribution should show a high wave height only if the wind speed is high. This advocates for a multivariate forecasting of wind and



(a) T_m residuals for WW3 as function of H_S residuals fro WW3.

(b) T_m residuals for WW3 as function of U_{100} residuals for AROME.

(c) H_S residuals for WW3 as function of U_{100} residuals for AROME.

Figure V.4 – Scatter plot of the target variables' residuals for the numerical predictions in the test dataset. The individual distributions are shown at the top and right of the scatter plots. These graphs show the interest in jointly predicting the forecast uncertainty for these variables.

wave parameters. For offshore operations, joint trajectories of wind speed, wave height and wave period should be generated to compute the probability of weather windows, and these trajectories should respect the dependencies between variables.

V.2.2 Input data

The input data used for post-processing the numerical forecasts is similar to the one used in Chapter IV, centred on the new target on Planier island, and mixing wave and weather inputs. This dataset is constructed thanks to the data shared by Météo France for the FLOWTOM project. The different sources of data and their positions relative to the forecast window are illustrated in figure V.5.

The archives of forecast data from AROME model [27] at 0.025° resolution were shared for 2016, 2017 and 2023. The output of the AROME model on the 0.025° structured grid results in (241×81) maps of variables for lead times $k = 0, \ldots, K$ with $K = 24$, with an hourly time step. The used variables are the pressure at the mean sea level, the humidity and temperature at 2 meters above the surface, the wind speed at 10, 50, 100 and 150 meters, the wind gust at 10m, and the wind direction at 10 and 100 meters. This results in an input vector for forecast issue time t_0 of dimensions $(1, 81, 241, 10, 25)$ in (forecast issue time, latitude, longitude, variables, lead time).

In-situ measurements from the Météo France RADOME observation network were also shared to help post-processing numerical predictions with recent neighbouring observa-

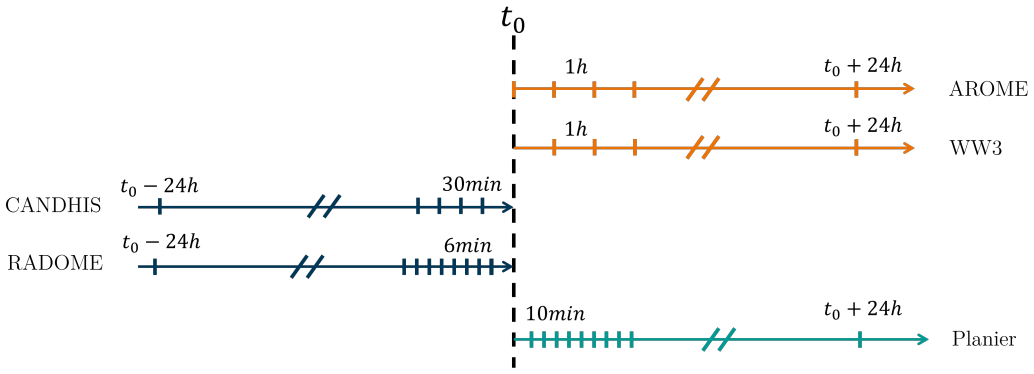


Figure V.5 – Illustration of the input and target data in the Planier dataset. The numerical model data is shown in blue and is available at a time step of 1h in the forecast window of 24 hours. Past measurements of metocean parameters from neigbouring stations for the last 24 hours are used (orange). The target variables at the Planier island are forecast for the next 24 hours at a time step of 10 minutes.

tions. Measurement data from the 18 closest coastal stations is used as input, considering when available the pressure, temperature, wind speed, wind direction and humidity rate for the last 24 hours of observations at 6 minutes resolution. This results after removing non-measured parameters (The pressure data is not available for most stations) to an input vector for forecast issue time t_0 of dimensions $(1, 63, 240)$ in (forecast issue time, variables $\times$ stations, past time step).

For wave data, the numerical model used is an implementation of the Wave Watch 3 model from the NOAA [257], using an unstructured grid on the French Mediterranean Sea (Gulf of Lion and Corsica). It is forced using AROME surface winds which showed to significantly improve the wave propagation due to a better representation of convective events [258]. The data is available on a 0.01° structured grid, which was downsampled to 0.04° to reduce the computational burden. The direction of primary swell, significant wave height of primary swell, wind waves and total sea, the mean period of wind waves and total sea are taken as input variables. An input vector at initial forecast time t_0 from WW3 then consists in the wave forecast for the newt 24 hours (from $t_0 + 1h$) of these 6 variables on a (76×51) structured grid. The input vector has dimensions $(1, 51, 76, 6, 24)$ in (forecast issue time, latitude, longitude, variables, lead time).

In-situ data relies on deployed wave buoys in the Gulf of Lion. The CEREMA network (`https://candhis.cerema.fr`) is used to provide in-situ informations, though the used buoys are far more distant to the target than the coastal stations used for wind

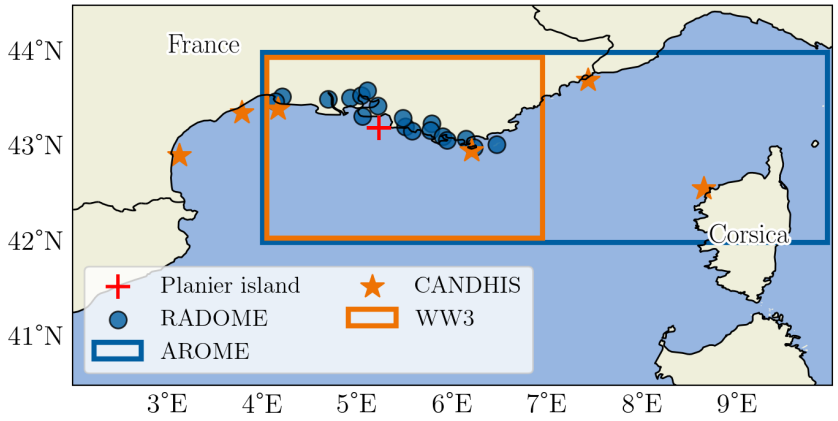


Figure V.6 – Planier dataset construction, with numerical predictions of wave (WW3) and weather (AROME), and in-situ measurements of waves (CANDHIS) and weather (RADOME).

and weather. The used buoys are Leucate, Espiguette, Sète (CEREMA, DREAL Occitanie), Revellata (CEREMA, SHOM), Porquerolles (CEREMA), and Monaco (CEREMA, Monaco, Service des Travaux Publics) buoys. The considered variables are the significant wave height of total sea, maximum wave height, temporal mean period, peak period, and spectral mean period, for the last 24 hours of measurements with a time step of 30 minutes. The input CANDHIS vector at forecast issue time t_0 is then of dimensions $(1, 19, 48)$ in (forecast issue time, variables $\times$ buoys, past time step).

The spatial extend of the input data is illustrated in Figure V.6. The numerical weather prediction data covers a wider area around the target than selected in Chapter IV, to integrate in the training information about the atmospheric circulation in the Gulf of Genoa that is a driver of the local weather and wind.

Similarly to the dataset used in Chapter IV, the length of available data is an issue for data representativity. The dataset is split into train, validation and test splits. The training split is taken as 60% of the dataset, which represents a full year. The validation and test splits are then only 130 days long. Furthermore, the 2 years of available data are not continuous, and the seasons are not evenly represented as illustrated in Figure V.7. The under-representation of autumn and winter months can result in data representativity issues with an under-representation of strong winter events for example. This is alleviated by using cross validation to alternatively train and test the model on different parts of the dataset, but surely advocates for longer in-situ times series. To ensure that all weather regimes are represented in each of the data splits, the data is grouped per quarter, and

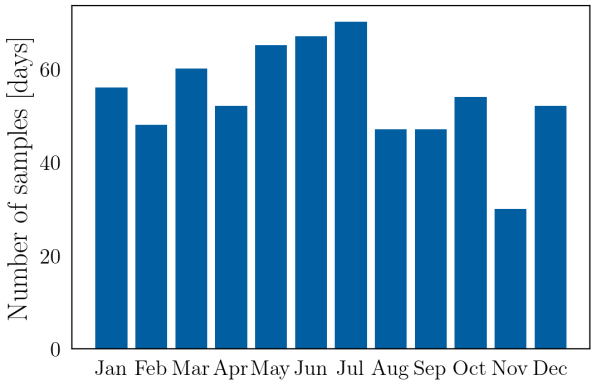


Figure V.7 – Number of samples per month in the Planier dataset after cleaning. The months of October, November and December are somewhat under-represented which can induce data representativity issues in model training and evaluation. Missing data are primarily due to maintenance periods during the campaigns.

each quarter is split in train, validation and test. More strategies for data splits in weather applications are detailed in [9].

V.3 Methods

V.3.1 Convolutional encoding for the joint probabilistic forecasting of wind and waves

With the same notations as Chapter IV, a deep learning model Ψ with parameters Ξ is built to predict a probability distribution $\hat{g}_{t+k}$ for a lead time k ahead of a forecast issue time t, given an input $\mathbf{X}_t$. The posterior distribution $\hat{g}_{t+k}$ is taken either as a Gaussian distribution parametrized by the output $\hat{\boldsymbol{\theta}}_{t+k}$ of the model Ψ, or obtained through a normalizing flows that transforms a simple latent space into an assumption posterior distribution.

We adapt the ConvE-STF architecture of Chapter IV to the Planier dataset by adding the same encoding block for the wave input as for the weather input. Input vectors are encoded in a latent space by individual CNN, 2-dimensional CNN for spatialized numerical predictions of waves and weather, and 1-dimensional CNN for in-situ time series from neighbouring coastal stations and buoys. The direct numerical models output at the closest grid point is stacked to the latent space before a MLP is used to output the parameters of distribution. A schematic view of the ConvE-STF model for the Planier

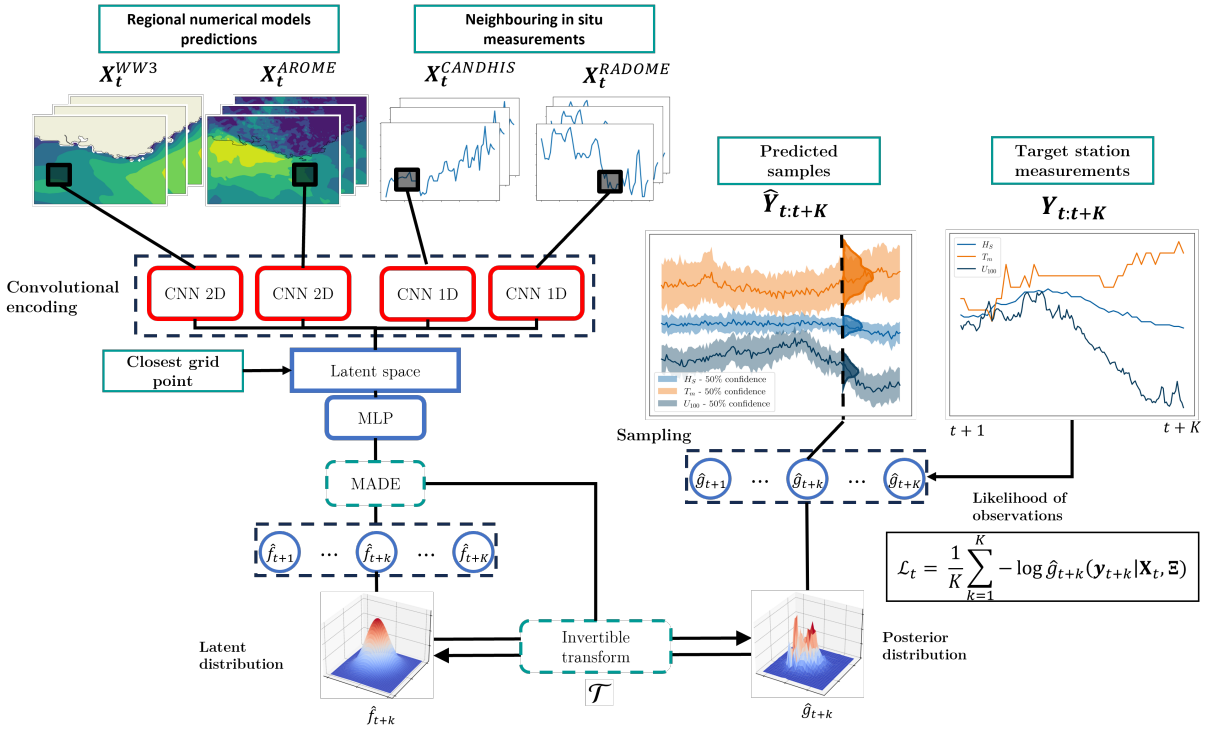


Figure V.8 – Schematic view of the ConvE-STF and ConvE-STF-NF architectures for the joint wind and wave forecasting of the Planier dataset. Convolutional blocks are added to handle new data sources, and the posterior distribution is 3-dimensional. The optional NF blocks are shown in dashed green.

dataset is show in Figure V.8. The optional blocks associated with the normalizing flows (NF) configuration are shown in green.

The NF configuration lifts any assumption on the posterior distribution. As discussed in Section V.2, this is important for the joint probabilistic forecasting of wind and wave variables which have complex relationships, and which forecast uncertainties can be correlated.

V.3.1.1 Gaussian posterior distribution

In contrast with Chapter IV, the target distribution should be 3-dimensional since it describes the forecast joint uncertainty of wind speed, wave height and wave period. To construct a valid definite positive covariance matrix, it is then required to predict its Cholesky decomposition. The Cholesky decomposition states that given a Hermitian positive-definite matrix $\mathbf{A}$, there exists a unique lower triangular matrix with positive

diagonal elements $\mathbf{L}$ such that

$$\mathbf{A} = \mathbf{L}\mathbf{L}^* \tag{V.1}$$

with $\mathbf{L}^*$ the conjugate transpose of $\mathbf{L}$. We then construct the posterior distribution from a time series of parameters $\hat{\boldsymbol{\theta}}_{t+k}$ describing the mean and Cholesky decomposition of covariance matrix

$$\hat{\boldsymbol{\theta}}_{t+k} = \left[\hat{\boldsymbol{\mu}}(t+k),\ \hat{\boldsymbol{\lambda}}(t+k),\ \hat{\boldsymbol{\alpha}}(t+k)\right] \tag{V.2}$$

with $\hat{\boldsymbol{\mu}}(t+k) \in \mathbb{R}^N$ the predicted mean vector, $\hat{\boldsymbol{\lambda}}(t+k) \in \mathbb{R}+^N$ the predicted diagonal elements of the lower-triangular matrix $\mathbf{L}_{t+k}$, ensured to be positive by an exponential activation function, and $\hat{\boldsymbol{\alpha}}(t+k) \in \mathbb{R}^{N(N-1)/2}$ the non-diagonal elements of $\mathbf{L}_{t+k}$ such that for the 3-dimensional case

$$\mathbf{L}_{t+k} = \begin{bmatrix} \lambda_1(t+k) & 0 & 0 \\ \alpha_1(t+k) & \lambda_2(t+k) & 0 \\ \alpha_2(t+k) & \alpha_3(t+k) & \lambda_3(t+k) \end{bmatrix}. \tag{V.3}$$

The covariance matrix of the Gaussian distribution at lead time k can then be computed as

$$\boldsymbol{\Sigma}_{t+k} = \mathbf{L}_{t+k}\mathbf{L}^*_{t+k} \tag{V.4}$$

and be a valid positive-definite covariance matrix. The length of the output vector needed to parametrize the Gaussian distribution is then $\frac{N(N+3)}{2}$, which would become intractable for high dimensional cases.

The model can now be trained to minimize the negative log-likelihood of the observations given the parametrized Gaussian posterior distribution.

V.3.1.2 Normalizing flows for assumption-free posteriors

As described in Section II.1.4 and adapted to ConvE-STF in Section IV.3.4, the use of normalizing flows relaxes assumptions on the posterior distribution. The following assumptions are made for normalizing flows in this experiment:

- The latent distribution is assumed a multivariate Gaussian distribution parametrized as described in Section V.3.1.1. This assumption is still tractable for a 3-dimensional latent space, but would become intractable for higher dimensions.
- The flow conditioning is made through a MADE network with 1 hidden layer and 256 hidden features.

- The bijection transforms are taken as rational quadratic spline functions, and conditioned by the MADE network.
- 6 autoregressive transforms are used with reverse permutation operators between variables.

Constructed as such, the normalizing flow is trained with the negative log-likelihood. It is computed with the change of variable formula through the flow as described in Section II.1.4. In this configuration, each lead time is to be sampled independently. Though the predicted distributions are implicitly temporally correlated, there is no explicit formulation nor any training on it. In Section V.3.2, alternative ways of embedding a temporal relationship during sampling are explored.

V.3.2 Weather scenarios generation

In Chapter IV, a method for the probabilistic forecasting of wind speed was proposed, with a probability density function per lead time, both for parametric (Gaussian) and non-parametric (normalizing flows) frameworks. The obtained uncertainty prediction is in the variable space. For temporal decision-making problems such as weather window estimation under uncertainty, both the cross-variables and temporal correlations are to be considered. An optimal decision-making under uncertainty for maintenance operations execution requires the computation of an exceedance probability on a sliding time window. It is then important to generate scenarios from the probabilistic forecast to allow for this weather window probability estimation.

A scenario is a tensor $\mathbf{S} \in \mathbb{R}^{N \times K}$ containing predicted trajectories of the N variables in the forecast window $k = 1, \dots, K$. It can be seen as a sample from a multivariate distribution $G(\hat{y}_{1,1} \dots \hat{y}_{N,K})$, spanning through the variable and temporal spaces. In the previous chapters, we only considered the marginal distributions in the variable space. In this section, tools are explored to describe the scenarios distribution G to permit the sampling of trajectories. For all that follows, the variable-temporal space of dimension $(N \times K)$ is called the scenarios space.

V.3.2.1 Copulas-based scenarios generation

The modelling of the temporal dependency can be made through the use of copulas, as introduced in Section II.2.3. A forecast multivariate temporal process $(\hat{\boldsymbol{Y}}_{t+1}, \dots, \hat{\boldsymbol{Y}}_{t+K})$ is considered, with $\hat{\boldsymbol{y}}_{t+k} \in \mathbb{R}^N$ a realization of the random variable $\hat{\boldsymbol{Y}}_{t+k}$. It describes

the evolution of the N predicted variables as function of lead time $k = 1, \dots K$. We can flatten the variable space to apply a copula to the forecast $(\hat{Y}_{1,1}, \hat{Y}_{1,2}, \dots, \hat{Y}_{N,K}) = (\hat{\boldsymbol{Y}}_{t+1}, \dots, \hat{\boldsymbol{Y}}_{t+K})$. The random variables $\hat{Y}_{n,k},\ n = 1 \dots N,\ k = 1 \dots K$ have marginal distributions $F_{n,k}$. Then according to Sklar's theorem, there exists a unique copula C such that the cumulative density function G of $(\hat{\boldsymbol{Y}}_{t+1}, \dots, \hat{\boldsymbol{Y}}_{t+K})$ can be written as

$$G(\hat{y}_{1,1} \dots \hat{y}_{N,K}) = C(F_{1,1}(\hat{y}_{1,1}), \dots, F_{N,K}(\hat{y}_{N,K})). \tag{V.5}$$

In this work, we consider Gaussian copula, that uses a NK-dimensional Gaussian distribution $\Phi_{\boldsymbol{\Sigma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with zero mean and $\boldsymbol{\Sigma}$ covariance matrix. An illustration of the sampling process described in Section II.2.3 is shown in Figure V.10.

The $(NK \times NK)$ covariance matrix describes the correlations between variables in the scenarios space. The full estimation of Σ scales as $O(NK)^2$ which is a limit for high dimensional problems. The covariance matrix should furthermore be positive-definite, which constrains the types of matrices that can be considered. In the following paragraphs, several assumptions for covariance matrix construction are presented.

Fixed covariance matrix

The covariance matrix can be set to a fixed covariance matrix using the empirical covariance from training data, provided that the amount of data is sufficient. Given past observations $\boldsymbol{Y}_{i:i+K},\ i = 1 \dots t$ up to time t, the empirical covariance matrix $\boldsymbol{\Sigma}_{emp}$ can be obtained as

$$\boldsymbol{\Sigma}_{emp}(t) = \frac{1}{t} \sum_{i=1}^{t} \boldsymbol{Y}_{i:i+K} \boldsymbol{Y}_{i:i+K}^{\top}. \tag{V.6}$$

An example of empirical covariance is shown in Figure V.9. Each visible block in the matrix is of dimension (K, K) and represents the temporal correlation between two variables. Only the variables axis is labelled for clarity. It can be noted that the covariances for each individual variables, as shown in the block diagonal have relatively similar patterns. The significant wave height tend to have higher temporal correlation and is close to stationarity, i.e. temporal invariance, while the wind speed covariance shows diurnal variations and is less stationary. Low temporal correlation is observed between wind speed and wave period, while the other correlations are not neglectable and justify the joint forecasting of wind and wave.

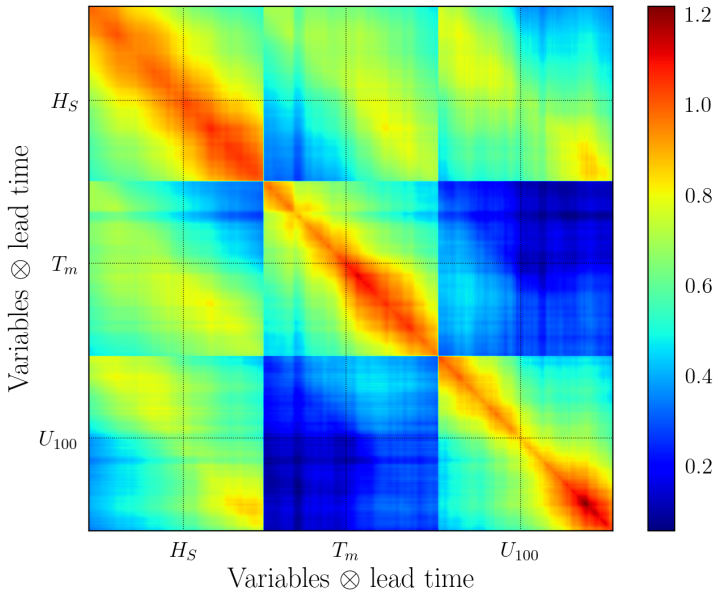


Figure V.9 – Visualisation of the empirical covariance matrix computed on the training dataset. The diagonal blocks show the variables H_S, T_m, U_{100} temporal covariances for lead times $k = 1, \ldots, K$

The assumption of a fixed empirical covariance matrix computed on the flattened training data between lead time and variables is named «empirical Gaussian copula »hereafter, and is noted $\Phi_{\mathbf{\Sigma}_{emp}}$. When doing so, the predicted covariance matrix per time step is not used totally, since only the individual marginal distributions are recovered. The cross-variables correlations and their time dependencies are fully modelled by the empirical covariance matrix obtained from training data.

Since this formulation is by nature recursive, [151] proposed a recursive update of the covariance matrix through a forgetting factor $\lambda \in [0, 1]$, corresponding to exponential forgetting

$$\mathbf{\Sigma}_{emp}(t+1) = \lambda \mathbf{\Sigma}_{emp}(t) + (1-\lambda)\boldsymbol{Y}_{t:t+K}\boldsymbol{Y}_{t:t+K}^{\top}. \tag{V.7}$$

To overcome the limitations of the empirical covariance, [64, 139] proposed parametric forms of covariance matrix for Gaussian copulas use. In particular for high-dimensional problems in energy applications, non-isotropic or non-stationary covariance matrices are required. Such advanced covariance models are not studied in this thesis and would require further research work.

Conditional covariance matrix

The empirical covariance matrix aggregates lots of different weather situations, for which the cross-variables and temporal dependencies can vary. It appears appealing to parametrize the covariance matrix conditionally to the input.

Recent works in the wind energy literature try to model the spatio-temporal relationships between wind turbines or wind farms [259, 260]. To do so, the covariance matrix can be split in a spatial matrix and a temporal matrix under separation assumption. For this study, the covariance matrix can be split into a temporal matrix for different lead times in the forecast window $k = 1 \dots K$ and a cross-variables matrix for the environmental variables $n = 1 \dots N$. A covariance matrix $\mathbf{\Sigma}(n, k)$ is said to be separable if for every $n_1, n_2 = 1 \dots N$ and $k_1, k_2 = 1 \dots K$, the correlation $\Sigma(n_1, n_2, k_1, k_2)$ can be expressed as

$$\Sigma(n_1, n_2, k_1, k_2) = \mathbf{\Sigma}_{var}(n_1, n_2)\mathbf{\Sigma}_{temp}(k_1, k_2) \tag{V.8}$$

given covariance matrices $\mathbf{\Sigma}_{temp}$ for the temporal process and $\mathbf{\Sigma}_{var}$ for the cross-variable dependency. The total covariance $\mathbf{\Sigma}$ is then obtained as the Kronecker product of the two matrices

$$\mathbf{\Sigma}_{sep} = \mathbf{\Sigma}_{var} \otimes \mathbf{\Sigma}_{temp}. \tag{V.9}$$

This assumption means that the correlation between the variables does not depend on the lead time and that the temporal correlation is the same for every variable. In this work, the cross-variable covariance is predicted per time-step, so it can not be used under a separable covariance assumption for copula construction as such. We propose to add the assumption that the cross-variable correlation is constant through the forecast window, which appears reasonable for environmental variables in a 24 hours window. However this might be an issue for wind shifts and wind ramps that imply complex responses of the wave field. In all that follows, we name «separate Gaussian copula»the assumption of a Gaussian copula with separate covariance matrix. The fixed temporal covariance is obtained as the mean of the 3 empirical covariance matrices of U, H_S, T_P. The variable covariance matrix is equal to the mean of the predicted covariances with unit diagonal.

We note $\hat{\mathbf{\Sigma}}_{var}^{(t+k)}$ the (3×3) predicted covariance matrix at lead time $t + k$. We note $\boldsymbol{y}_{t:t+K}^{n}$ the 24 hours of observation of variable n from time t in the training dataset. The predicted cross-variable matrix $\hat{\mathbf{\Sigma}}_{var}^{(t)}$ is obtained as

$$\hat{\mathbf{\Sigma}}_{var}^{(t)} = \left[\frac{1}{K}\sum_{k=1}^{K}\hat{\mathbf{\Sigma}}_{var}^{(t+k)}\right](\mathbf{1}_N - \mathbb{I}_N) + \mathbb{I}_N \tag{V.10}$$

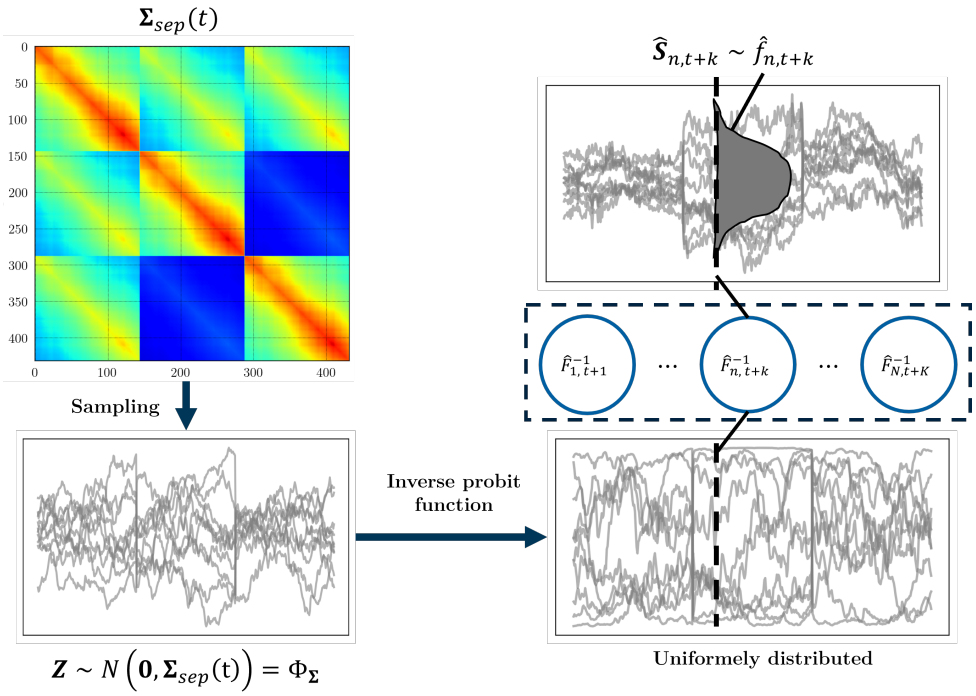


Figure V.10 – Illustration of the sampling process under Gaussian copula assumption. The covariance matrix in the scenarios space is used to sample temporally correlated scenarios, which are transformed to follow the predicted marginal distributions.

and is assumed constant for the forecast window from time t. The temporal covariance matrix is obtained from the empirical covariance matrices of the observations

$$\boldsymbol{\Sigma}_{temp} = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{T_{train}}\sum_{t=1}^{T_{train}} \boldsymbol{y}^{n}_{t:t+K}\boldsymbol{y}^{n\top}_{t:t+K}\right] \tag{V.11}$$

Then the separate Gaussian copula $\Phi^{(t)}_{sep}$ is characterized by the separable covariance matrix

$$\boldsymbol{\Sigma}^{(t)}_{sep} = \hat{\boldsymbol{\Sigma}}^{(t)}_{var} \otimes \boldsymbol{\Sigma}_{temp} \tag{V.12}$$

which cross-variable dependency is conditional to the input data and predicted through the ConvE-STF network. An illustration of a separable covariance matrix obtained with a fixed cross-variable covariance matrix and a fixed temporal covariance matrix equal to the mean of the diagonal blocks of the empirical covariance matrix is shown in Figure V.11. The total process for weather scenarios generation from separable covariance matrix is illustrated in Figure V.10.

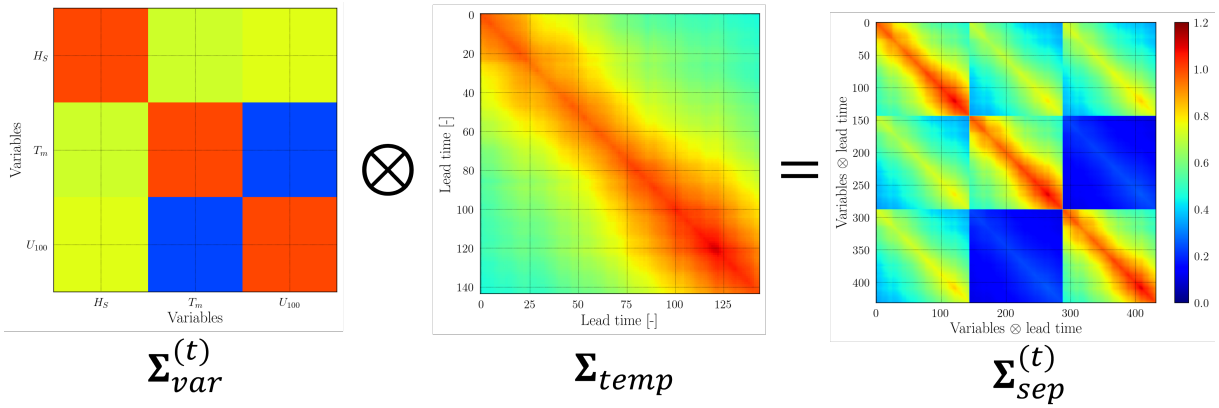


Figure V.11 – Illustration of a variable-temporal matrix obtained under separable assumption.

In this case, the temporal correlation is still assumed constant. It is harder to parametrize due to its high dimension $(144, 144)$, and would require additional assumptions to be conditional to the input. Exponential covariance functions, Matérn matrix or more complex non-stationary functions could be used, such as the ones introduced in [139]. However, this adds serious difficulties and instabilities for negative log-likelihood training. The computation of the Gaussian likelihood requires covariance matrix inversion and log-determinant computation which can be unstable.

Copulas can be applied to ConvE-STF and ConvE-STF-NF outputs. Copulas ensure a multivariate dependency in variables and lead times, while maintaining marginal distributions. This means that the inter-variable dependencies predicted with ConvE-STF are ignored, which is a limitation of the method.

V.3.2.2 Scenario generation with normalizing flows

In Chapter IV, we show that normalizing flows can be used to approximate complex posterior distributions from a simple latent space. It is natural to think that they can also be used to generate multivariate time series scenarios. By embedding a temporal relationship in the latent space, it could in theory be possible to generate multivariate scenarios while alleviating the assumptions on the shape of the marginal distributions, the cross-variable dependency and the temporal evolution.

Seemingly to Chapter IV, the normalizing flow can be trained on the variables dimension, with a latent distribution $\mathcal{Z}$ in $\mathbb{R}^N$. In this case, the temporal dependency should be embedded outside of the latent space formulation itself, and hence can not be used for

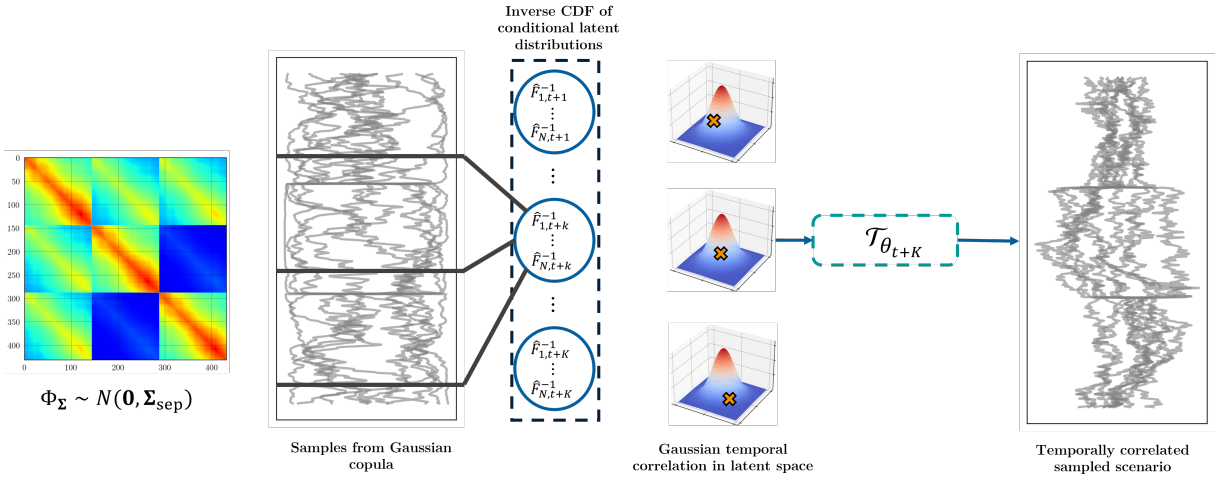


Figure V.12 – Illustration of a Gaussian copulas based sampling in the latent space of a normalizing flows. The generated samples in the latent space are temporally correlated and transformed through the flow to output scenarios in real space.

training. In other terms, the optimization can not include the temporal evolution. In this work, a separable Gaussian copula is used in the latent space for sampling correlated samples from the K lead times. Latent scenarios with a temporal correlation are generated. They are transformed through the flow to output scenarios in the real space. However, when using a multivariate Gaussian latent space in the variables dimension, the copula assumption does not perfectly matches the initial latent distributions. This might result in unrealistic generated scenarios. An illustration of this proposed method is shown in Figure V.12. It is named «NF-latent»hereafter.

Alternatively, we can artificially force the model to sample in the same manner for each time step by forcing the random number generator of our algorithm before each sampling step. This is done by the «seeding»process in PyTorch. We can generate trajectories by fixing the seed to a random number for each generated sample, and repeating the sampling to obtain the right number of trajectories. By doing so, the sampled vector $\mathbf{z}_k$ from the latent distribution $\mathcal{Z}_k$ will be similarly situated in its distribution as the sample $\mathbf{z}_0$ from the latent distribution $\mathcal{Z}_0$. In fact, if $\mathcal{Z}_{k_1} = \mathcal{Z}_{k_2}$, then $\mathbf{z}_{k_1} = \mathbf{z}_{k_2}$. This artificial temporal correlation has no reasons to be relevant in the real space, since it is not present in the loss computation. The loss only depends of the likelihood of the individual samples. An additional loss could be computed from sampling, for example the ES in the scenarios space, but this dramatically increases the computational cost of the optimization. It is illustrated in Figure V.13 and is called «NF-seed»in all that follows.

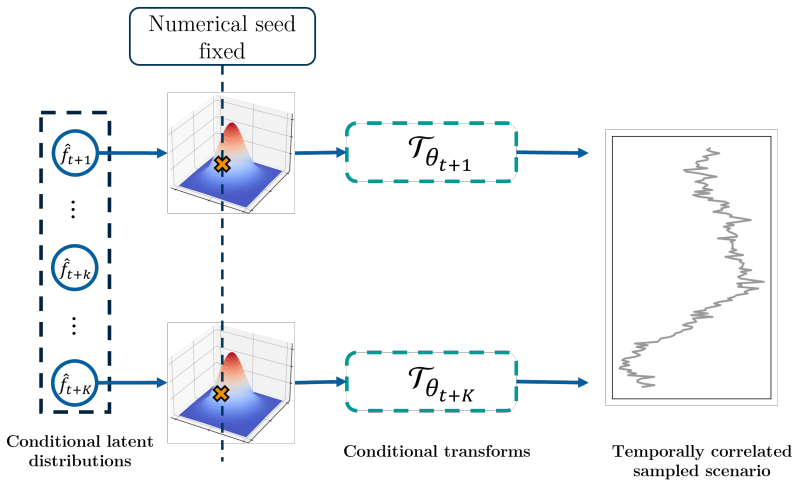


Figure V.13 – Illustration of a seeding-based sampling in the latent space of a normalizing flows. The generated samples in the latent space are numerically correlated by the setting of a seed for random number generation. The output of this inexpensive sampling strategy are scenarios in the real space.

These two innovative methodologies are simple ways to embed a temporal relationship in normalizing flows. However, this temporal relationship is not trainable, which is a limitation. Further studies about temporal embedding in normalizing flows for multivariate scenarios generation should be done. In particular, recurrent conditioning in the latent space [59], or full scenarios latent space [44] should be further explored.

V.3.3 Model evaluation

V.3.3.1 Benchmark model

The proposed models are compared with the numerical predictions from AROME and WW3, extracted at the closest grid point from the target location. This represents a deterministic benchmark model, and is representative to what is currently used operationally in offshore wind energy based on the industrial inputs gathered through the FLOWTOM project. We introduce the observation operators $\mathcal{H}^{AROME}$ and $\mathcal{H}^{WW3}$ which extract the wind and wave forecasts from the closest grid point from the target.

$$\mathbf{X}^{AROME}_{t+k,closest} = \mathcal{H}^{AROME}(\mathbf{X}^{AROME}_{t+k}) \tag{V.13}$$

$$\mathbf{X}^{WW3}_{t+k,closest} = \mathcal{H}^{WW3}(\mathbf{X}^{WW3}_{t+k}). \tag{V.14}$$

We construct the closest grid point vector as the concatenation of the closest grid point from AROME and WW3 inputs

$$\mathbf{X}^{num}_{t+k,closest} = \left[\mathbf{X}^{AROME}_{t+k,closest},\ \mathbf{X}^{WW3}_{t+k,closest}\right]. \tag{V.15}$$

The numerical model prediction is then corrected using linear regression on the training dataset

$$\hat{\mathbf{Y}}^{num}_{t+k} = \Psi^{num}(\mathbf{X}^{num}_{t+k,closest}) = \frac{\mathbf{X}^{num}_{t+k,closest} - \hat{\beta}_0}{\hat{\beta}_1}, \tag{V.16}$$

with $\hat{\beta}_0$ and $\hat{\beta}_1$ computed using ordinary least squares.

The deterministic forecast from the closest grid point is augmented to a probabilistic forecast considering forecast errors in the training dataset. We assume independent Gaussian distributions for the forecast variables with constant variance equal to the empirical variance of the residuals on the training dataset. The observed target from the training dataset $Y_{t+k,n}$ for forecast issue time $t = 1, \dots, T_{train}$, lead time $k = 1, \dots, K$ and variable $n = 1, \dots, N$ is compared the forecast vector from numerical predictions $\hat{Y}^{num}_{t+k,n}$ to compute the forecast errors $\mathbf{e}^{num}_{t+k,n}$ as

$$e^{num}_{t+k,n} = \hat{Y}^{num}_{t+k,n} - Y_{t+k,n} \tag{V.17}$$

and the empirical variance $\sigma^2_{num,n}$ is computed as the variance of the forecasts errors on the training dataset:

$$\sigma^2_{num,n} = \frac{1}{T_{train}}\frac{1}{K}\sum_{t=1}^{T_{train}}\sum_{t=1}^{K}(e^{num}_{t+k,n} - \overline{e^{num}_{n}})^2. \tag{V.18}$$

The predicted random variable Z^{num}_{t+k} is then assumed to follow a multivariate Gaussian distribution $Z^{num}_{t+k} \sim \mathcal{N}\left(\hat{\mathbf{Y}}^{num}_{t+k}, \mathbf{\Sigma}^{num}\right)$ with

$$\mathbf{\Sigma}^{num} = \begin{bmatrix} \sigma^2_{num,1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2_{num,N} \end{bmatrix}. \tag{V.19}$$

From this probabilistic model, predicted scenarios can be generated provided that a an additional assumption on the temporal correlation is made. The empirical Gaussian

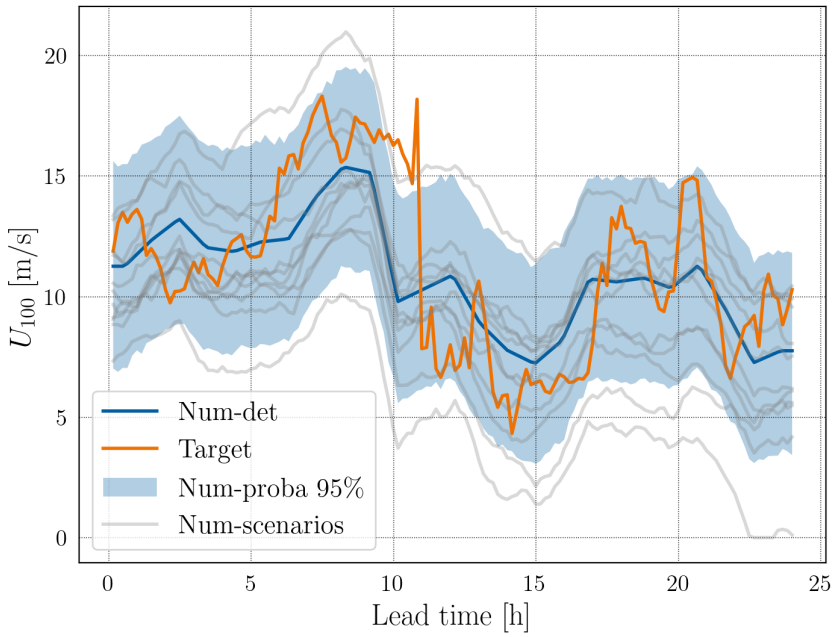


Figure V.14 – Example of prediction from the numerical benchmark model, with deterministic output, probabilistic output in the form of a multivariate Gaussian per time step with constant covariance, and scenarios generated with a separable assumption.

copula is used as described in Section V.3.2.1. An illustration of the obtained prediction from the numerical model (Num) predictions is shown in Figure V.14.

V.3.3.2 Evaluating the quality of scenarios

In Chapter IV, the forecast quality is only evaluated by scoring the predicted distributions per lead time. In addition to all the metrics described in Section IV.4 that are still used in this chapter to evaluate the quality of the posterior distributions, the quality of the generated scenarios should also be computed.

Multivariate metrics in the scenarios space

To do so, the observed trajectory in scenario space can be compared with generated scenarios samples and scored using multivariate probabilistic metrics. Considering a set of L generated scenarios $\{\hat{\boldsymbol{s}}^{(l)}_{t:t+K}\}_{l=1\dots L}$, and the observations expressed in the scenarios space $\boldsymbol{y}_{t:t+K} \in \mathbb{R}^{NK}$, the ES and VS can be expressed in the scenarios space as

$$ES_{scen} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{L}\sum_{l=1}^{L}\left(\left|\left|\boldsymbol{y}_{t:t+K} - \hat{\boldsymbol{s}}^{(l)}_{t:t+K}\right|\right|\right) - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{m=1}^{L}\left(\left|\left|\hat{\boldsymbol{s}}^{(l)}_{t:t+K} - \hat{\boldsymbol{s}}^{(m)}_{t:t+K}\right|\right|\right)\right], \quad \text{(V.20)}$$

$$VS_{scen} = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=1}^{NK}\sum_{j=1}^{NK}\left(\left|y^{i}_{t:t+K} - y^{j}_{t:t+K}\right|^{0.5} - \frac{1}{L}\sum_{l=1}^{L}\left|\hat{s}^{i,(l)}_{t:t+K} - \hat{s}^{j,(l)}_{t:t+K}\right|^{0.5}\right)^{2}\right]. \quad \text{(V.21)}$$

Multivariate rank histograms [48] in the scenarios space can also be used to evaluate the multivariate calibration. Similarly a reliability index can be computed.

Event-based evaluation of scenarios

Eventually, the scenarios can be evaluated based on event detections in connection to the target application. In this thesis, the forecast should accurately predict the existence of a weather window. We consider a certain operation O with duration T_O and operational limits $\boldsymbol{\tau}_O \in \mathbb{R}^N_+$. The predicted probability of a weather window from lead time k onwards $\hat{y}^{ww}_{t+k}$ is equal to the probability of each lead time $i = k, \ldots, k+T_O$ being under the limits of the operation. Writing $\boldsymbol{x} \prec \boldsymbol{y}$ if and only if $x_i < y_i \; \forall i = 1 \ldots N$, we define the predicted probability of weather window as

$$\hat{y}^{ww}_{t+k} = P\left(\hat{\boldsymbol{y}}_{t+k} \prec \boldsymbol{\tau}_O \bigcap \ldots \bigcap \hat{\boldsymbol{y}}_{t+k+T_O} \prec \boldsymbol{\tau}_O\right). \quad \text{(V.22)}$$

For a deterministic forecast, this is equal to the product of indicator functions

$$\hat{y}^{ww,det}_{t+k} = \prod_{i=k}^{k+T_O} \mathbb{1}\left(\hat{\boldsymbol{y}}_{t+i} \prec \boldsymbol{\tau}_O\right). \quad \text{(V.23)}$$

These predicted weather window probability are to be compared to the observed weather window computed on the target observations in the scenario space

$$y^{ww}_{t+k} = \prod_{i=k}^{k+T_O} \mathbb{1}\left(\boldsymbol{y}_{t+i} \prec \boldsymbol{\tau}_O\right). \quad \text{(V.24)}$$

By formulating the problem this way, the role of the temporal correlation becomes clear. If the lead times are considered independent, this probability would artificially tend

to zero.

The probability of weather window can be approximated from a set of L scenarios $\left\{\boldsymbol{s}_{t:t+K}^{(l)}\right\}_{l\in[1\dots L]}$ by counting the number of scenarios that show a deterministic weather window starting at lead time k

$$\hat{y}_{t+k}^{ww} = \frac{1}{L}\sum_{l=1}^{L}\left[\prod_{i=k}^{k+T_O}\mathbb{1}(\boldsymbol{s}_{t+i}^{(l)} \prec \boldsymbol{\tau_O})\right]. \tag{V.25}$$

In the weather window perspective, the forecast output is simply binary, and the quality of the generated scenarios in that prospect can be evaluated with binary forecast metrics presented in Section II.3.

V.3.3.3 Forecast value for offshore wind energy operations

In addition to the forecast quality, its value for operational decision-making is to be evaluated. This section defines metrics based on the probability of weather window to mimic realistic decision-making. A corrective maintenance case is considered, i.e. the turbine is stopped until repair, and a fictive maintenance operation is modelled, with a certain duration T_O and operational limitations $\boldsymbol{\tau}_O$. The forecast is considered to be used for decision-making at the time $t \in [1\dots T]$ when it is issued, for a final go / no-go decision.

Mean downtime

From time t, the duration before the operation starts is considered as «downtime». If this downtime is due to metocean conditions forecast above the operational limitations, it is considered as «weather downtime». It is considered the only source of downtime for scoring (i.e. crew and vessel ready to go without delay). The proposed methodology plays a realistic decision-making scenario for each sample.

A critical threshold $p_{critical}$ is defined - taken as $p_{critical} = 0.5$ for «P50 decision-making». It is the probability threshold above which a weather window is predicted. The predicted existence of a weather window is a binary variable noted $\hat{w}_{t+k}$ for a weather window starting from lead time k

$$\hat{w}_{t+k} = \begin{cases} 1, & \text{if } \hat{y}_{t+k}^{ww} > p_{critical} \\ 0, & \text{otherwise.} \end{cases} \tag{V.26}$$

We define k_O as the first lead time in the forecast window at which $\hat{w}_{t+k_O} = 1$. For a given sample t, the operation is considered a success if in the T_O long time window ahead of k_O, the metocean conditions were indeed observed under the operational limitations. Otherwise, the operation is considered failed for this sample. There are four situations associated with this binary decision-making problem for each sample t.

1. A weather window is predicted from lead time k_O and is indeed observed. This is a True Positive (TP) output. The downtime associated with this decision-making is the time before the first predicted weather window lead time.
2. The predicted weather window at k_O is a False Positive (FP), and the observed metocean conditions are above limits. The operation is a failure, the downtime associated with this decision is equal to length of the forecast window if there was an observed weather window, and zero otherwise.
3. There is no predicted weather window but there was an observed weather window, this is a False Negative (FN). The downtime associated is equal to the forecast weather window.
4. There is no predicted weather window nor observed weather window. No downtime is associated with this True Negative (TN).

The occurrence of a TP event at lead time $t + k$ is represented by the random variable $TP(t + k)$ which is equal to 1 if a TP is observed at lead time $t + k$ and 0 otherwise. The same applies to all other events. Seemingly, the occurrence of a FN throughout the forecast window is represented by the random variable $FN(t : t + K)$ equal to 1 if a weather window was observed but no weather window was predicted through the whole forecast window, and 0 otherwise.

The downtime is noted d_t and is defined from these four cases computed on the set of samples:

$$d_t^{TP} = k_O \tag{V.27}$$

$$d_t^{TN} = 0 \tag{V.28}$$

$$d_t^{FP} = \begin{cases} 0, & \text{if } \sum_{k=1}^{K} y_{t+k}^{ww} = 0 \\ K, & \text{otherwise} \end{cases} \tag{V.29}$$

$$d_t^{FN} = K \tag{V.30}$$

which defines a downtime for sample t as

$$d_t = d_t^{TP}\mathbb{1}(TP(t+k_O)) + d_t^{FP}\mathbb{1}(FP(t+k_O)) + d_t^{FN}\mathbb{1}(FN(t:t+K)) \tag{V.31}$$

and the mean downtime DT as

$$DT = \frac{1}{T}\sum_{t=1}^{T} d_t. \tag{V.32}$$

This proposed metric is realistic to the actual observed downtime due to the imperfect knowledge of the situation and the forecast errors. It represents the generated cost associated with the decision-making on this specific operation.

Economic forecasting metric

Recent work from Catterson et al. [6] advocate for an economic metric, and they associate costs to the situations of FP and FN. The FP cost is associated with the cost of mobilization and operation of a vessel and crew without succeeding in the operation. This cost is linked with the contracted vessel cost $C_{vessel} = 1000\,€/h$ and fuel cost $C_{fuel} = 100\,€/h$ of the vessel in question, times the duration of the operation

$$C_{FP} = C_{vessel}T_O + C_{fuel}T_O. \tag{V.33}$$

These costs were set considering a crew transfer vessel for a floating offshore wind farm. These values are arbitrary and do not intend to be fully realistic due to a lack of industrial information.

The FN cost is related to the cost of missing an operable weather window, i.e. the opportunity cost of not selling electricity until the next repair. In [6], it is considered that the opportunity cost is equal to the energy not sold during the duration of the operation. We claim that this is not a realistic duration to consider, and that the opportunity cost should be computed using the mean duration between two consecutive weather windows. We compute this duration Δ_{ww} as the mean duration of non-weather window periods on the training dataset.

The cost of FN, C_{FN}, is linked with the turbine capacity $P_{max} = 10MW$, mean

capacity factor $C_p = 0.4$ and power price $P_{elec} = 140$ €$/MWh$ as

$$C_{FN} = P_{max} C_p P_{elec} \Delta_{ww} \tag{V.34}$$

Those two costs associated with the frequencies of FN and FP give the Economic Forecasting Metric (EFM)

$$EFM = f_{FN} C_{FN} + f_{FP} C_{FP} \tag{V.35}$$

and can be linked with the expectation of the over-cost associated with the use of a forecast model for a specific operation.

In all that follows, an operation of $T_O = 6h$ is considered, with operational limits set as

$$\boldsymbol{\tau}_O = [\tau_{H_S},\, \tau_{T_m}, \tau_{U_{100}}] = [1,\, 8,\, 15]. \tag{V.36}$$

V.3.3.4 A proposed risk forecasting metrics for offshore wind energy operations

In a recent review, [164] examine the health and security issues related to offshore wind energy. They advocate for the development of specific key risk indicators dedicated to offshore wind energy operations, such as the risk of having a personnel stuck on a wind turbine after a crew transfer. They give a specific analysis on the risks associated with weather uncertain decision-making. Based on the work from [60], they argue that the economic pressure for weather window optimization might increase the risk of performing operations in marginal environmental conditions. They furthermore mention the increased difficulties associated with floating wind operations.

Based on these findings, and following discussions with industrial partners in the FLOWTOM project, developers, operators, maritime warranty surveyors and vessel captains, we propose a framework to integrate the risk taken in addition to economic considerations in the value evaluation metric.

Let's define a dangerous event E taking place during an operation. For a crew transfer vessel, it could correspond to the event of the personnel being stuck on the turbine until next weather window due to bad conditions. This cost is associated with a certain cost C_E, that could correspond to workers compensation for this event. We set this cost as an overhead pay for the stuck personnel for 24 hours at 1000 €. Now the probability of being stuck on the turbine is depending on observed metocean conditions for transfer in

the weather window. In theory, from numerical modelling, the success probability of the operation is computed as function of metocean conditions, so a probability of dangerous event could be computed as function of observed conditions $P(E|\boldsymbol{y})$. The risk is then computed as function of the environmental conditions as

$$R_E(\boldsymbol{y}) = C_E P(E|\boldsymbol{y}). \tag{V.37}$$

We then propose to complement the above defined EFM as a Risk Forecasting Metric (RFM) integrating the risk associated with one or several dangerous output. For FP weather window, the risk is computed from the actually observed metocean condition, hence penalizing decision-making that led to operations taking place in dangerous conditions. In our perception of the situation, it reflects more the asymmetry of the decision-making due to risk aversion. We note $\mathcal{E}$ the ensemble of dangerous event considered. The risk is computed based on the maximum observed metocean condition in the False Positive predicted weather window for all the falsely predicted weather windows $t, k_O \in \{FP\}$. The RFM then reads

$$RFM = f_{FN}C_{FN} + f_{FP}C_{FP} + f_{FP}\sum_{E\in\mathcal{E}}\left[\sum_{t,k_O\in\{FP\}} C_E P(E|\boldsymbol{y}_{t+k_O:t+k_O+T_O}))\right]. \tag{V.38}$$

In the absence of realistic simulation data for characterizing the probability of a dangerous event, a simple exponential function is used between $P(E|y^n_{t+k} = \tau^n_O) = 10^{-4}$ and $P(E|y^n_{t+k} = 2\tau^n_O) = 1$, independently for each metocean variable. The probability of dangerous event is taken as the max obtained probability through the weather window

$$P(E|\boldsymbol{y}_{t+k_O:t+k_O+T_O}) = \max_{i,n} P(E|y^n_{t+k_O+i}). \tag{V.39}$$

V.4 Results

V.4.1 Forecast quality

The quality of metocean forecasts is evaluated with deterministic and probabilistic metrics. The ConvE-STF model is compared to the corrected output from numerical modelling (Num) with empirical uncertainty estimation on the training dataset. The posterior assumption relaxing is assessed with ConvE-STF-NF. The normalized MSE, ES and VS are presented in Table V.4.

Table V.1 – Probabilistic and deterministic metrics of implemented forecast models. The best obtained scores are show in bold. The scores are given as mean and standard deviation over the 6 splits.

Model	MSE	ES	VS	Rel
Num	0.273 ± 0.03	0.584 ± 0.02	0.660 ± 0.01	1.59 ± 0.4
ConvE-STF	$\mathbf{0.240 \pm 0.02}$	0.527 ± 0.02	0.493 ± 0.02	1.31 ± 0.5
ConvE-STF - NF	0.243 ± 0.03	$\mathbf{0.526 \pm 0.03}$	$\mathbf{0.485 \pm 0.02}$	$\mathbf{1.12 \pm 0.5}$

The first observation is that the numerical model is of very high quality, and the gain of post-processing the predictions with ConvE-STF for this dataset is lower than for the dataset constructed in Chapter IV. The ConvE-STF forecast shows a MSE 12% lower than the output of the numerical models. This gain is much lower than the one obtained in Chapter IV, showing the difficulty to generalize the results for different test sites with different characteristics. It appears that the normalizing flows addition slightly deteriorated the post-processing of the model in MSE, because of added complexity and convergence issues, similar to results observed in the previous chapter.

The log-likelihood training of a posterior distribution prediction on the other hand greatly improves the probabilistic metrics. Compared to the empirical uncertainty for the numerical model predictions, the ConvE-STF and ConvE-STF-NF models bring improvements in ES and most notably in VS. The 26% decrease in VS for the ConvE-STF with Gaussian posterior proves that the correlation structure is important to take into consideration for the joint probabilistic forecast of wind and waves. The slight improvement with the posterior assumption relaxing with NF shows the non-Gaussian characteristics of the posterior distribution. The interest for relaxing the posterior assumption is particularly striking when looking at the reliability index that is greatly improved with ConvE-STF-NF. The generalized rank histograms of the three models are shown in Figure V.15, and show that the ConvE-STF model tend to underestimate strong events, while the ConvE-STF-NF is more centred between high and low values. Since the predicted variables are positive outcomes, the extreme values are only observed for high values, and the non-Gaussian characteristics are likely to concern more the extreme values. Relaxing the Gaussian posterior assumption with NF tend to improve the model calibration for extreme events. The Num model has a poor calibration, with strong over-dispersivity and extreme events under-estimation.

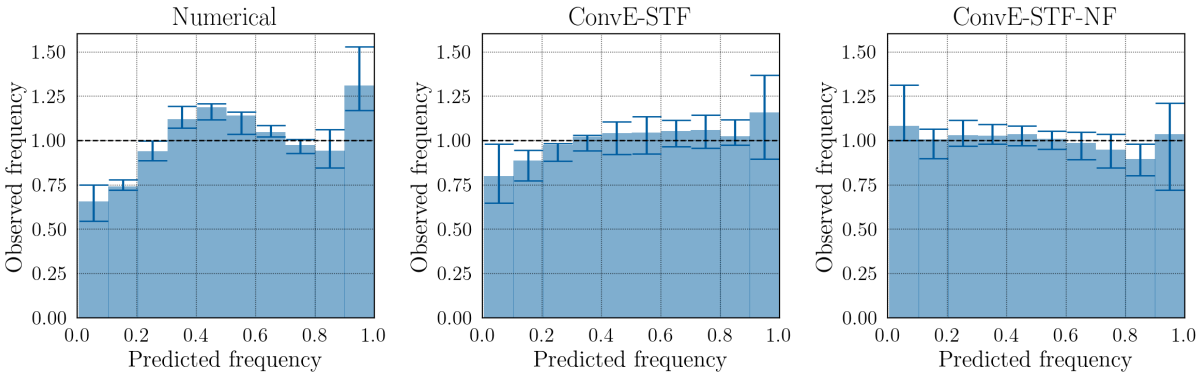


Figure V.15 – Generalized rank histograms for the different models. The ConvE-STF shows underestimation of strong events. The use of NF improves the calibration for extreme events.

The RMSE and CRPS are computed for each variables marginal distributions to understand the added value of the post-processing, and are shown in Table V.2 and Table V.3. The numerical model shows slightly better performances for the wind speed than the ConvE-STF models. With a generalized RMSE of $1.581 \mathrm{ms}^{-1}$, 2% lower than the ConvE-STF, it is within uncertainty range in the cross-validation results. This is an important results that shows the site-dependence of the models performances in terms of post-processing. In particular, the RMSE obtained on the wind speed from AROME model in Chapter IV for a different site was $2.6 m/s$, due to site specificities. Indeed, the wind measurements in Chapter IV were made on top of a small hilly island, with potentially strong orographic disturbances not well captured by the spatial discretization of numerical models. It shows by contrast the quality of the measurement site at the Planier island, its representativeness for offshore site, and the quality of day-ahead AROME wind forecasts at sea in the Gulf of Lion.

The wave parameters are on the other hand well-improved by the post-processing, especially the significant wave height with a 14% decrease compared to numerical models output, but with an important variability between splits in the cross-validation. The improvement is more consistent when looking at the CRPS for the individual variables, and is slightly improved even for wind speed.

The evolutions of the MSE and VS across the forecast window are shown in Figure V.16 for the 3 different models. The improvement in the correlation between variables as illustrated with the VS is rather homogeneous through the forecast window. The use of normalizing flows consistently improves the VS, showing that the correlation structure

Table V.2 – Forecast quality for each individual predicted variable in real values.

Metric	RMSE		
Variable	H_S [m]	T_m [s]	$U100$ [m/s]
Num	0.177 ± 0.02	0.368 ± 0.02	**1.581 ± 0.04**
ConvE-STF	**0.152 ± 0.01**	**0.340 ± 0.02**	1.616 ± 0.06
ConvE-STF - NF	0.154 ± 0.01	0.340 ± 0.02	1.626 ± 0.05

Table V.3 – Forecast quality for each individual predicted variable in real values.

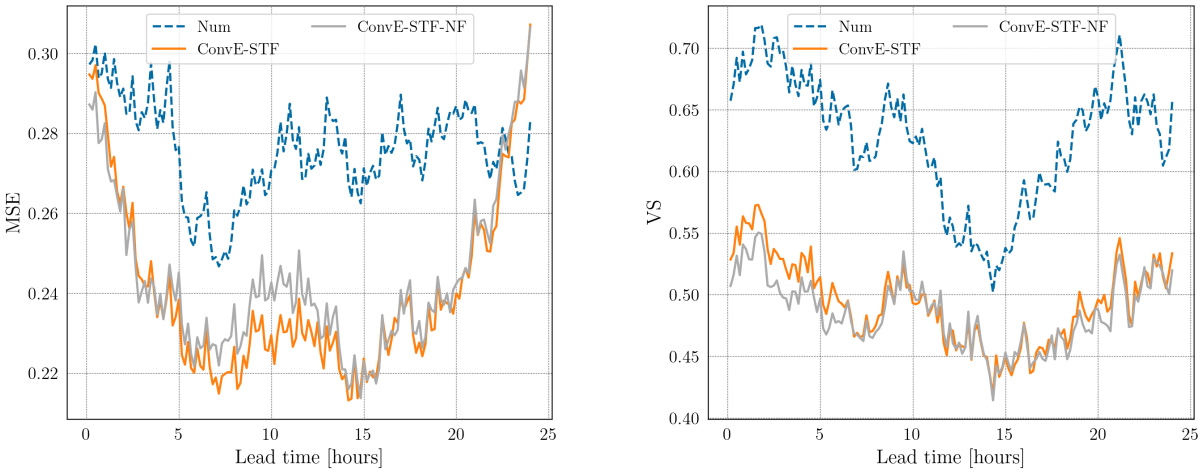
Metric	CRPS		
Variable	H_S [m]	T_m [s]	$U100$ [m/s]
Num	0.138 ± 0.01	0.264 ± 0.02	1.175 ± 0.03
ConvE-STF	**0.110 ± 0.01**	**0.243 ± 0.01**	**1.171 ± 0.05**
ConvE-STF - NF	0.111 ± 0.01	0.244 ± 0.02	1.172 ± 0.03

between wind and wave forecast uncertainty can be complex. It shows how a likelihood-based model training helps capturing important probabilistic features of the joint wind and wave forecast uncertainty.

The MSE evolution in Figure V.16a shows that the post-processing consistently improves the forecast through the forecast window, up to 22 hours ahead, before it abruptly increase to overcome the numerical model error. The ConvE-STF capabilities seems to be limited by symmetrical edge effects at the start and end of the forecast window. These edge effects are not present in the input dataset, as shown by the dashed blue line of the numerical models. It is likely that this feature is due to the model architecture that fails to capture explanatory variables for post-processing the edges of the forecast window. Further work should be done to assess how this can be alleviated.

V.4.2 Scenario generation of wind and waves

From the probabilistic forecast per lead time analysed in the previous section, probabilistic scenarios are generated to link with the operational decision-making process. The scenarios are generated using different assumptions. The Num and ConvE-STF models are combined with a Gaussian copula in the scenario space to generate scenarios. The covariance of the Gaussian copulas is constructed using a separable assumption described



(a) MSE (b) VS

Figure V.16 – Evolution of deterministic (a) and probabilistic (b) metrics with lead time. Edge effects are observed for ConvE-STF models in (a), probably due to model architecture effects. The correlation between variables is highly improved with likelihood training of ConvE-STF in (b).

in Section V.3.2.1. Samples with uniform marginal distributions are drawn from this copula, and transformed to Gaussian marginals using the predicted variances and means per variable and lead time. By doing so, some of the information in the correlation structure between variables is lost, but the marginal uncertainties are maintained. These two models are called Num-copula and ConvE-STF-copula.

We compare this state-of-the-art strategy to two newly proposed strategies for normalizing flows. These are introduced in Section V.3.2.2, and try to take advantage of the non-linear capabilities of normalizing flows to relax assumption on the posterior in the scenario space. The first strategy uses seeding in the latent space to generate correlated samples It is noted «ConvE-STF-NF-seed»in all that follows. The second assumption is to sample in the latent space using the same Gaussian copula assumption as above. This method is described as «ConvE-STF-NF-latent»in all that follows.

The generated scenarios are evaluated in terms of ES_{scen} and for the weather window forecasting with the limits described in Section V.3.3.2. The scenarios quality metrics are shown in Table V.4. With a relatively high variability between splits, the ConvE-STF-NF-seed outperforms both the Num-copula and the ConvE-STF-copula. It can be noted that the scenarios generated by seeding the latent distributions show a higher quality than the ones obtained by latent copula assumption. This is probably due to the fact that the

Table V.4 – Quality metrics of implemented forecast models in the scenarios space. The best obtained scores are show in bold. The scores are given as mean and standard deviation over the 6 splits. The BS is computed considering probability of weather window. The metrics obtained with the direct output of numerical models without uncertainty considerations is given as Num-det.

Model	ES_{scen}	BS
Num-det	-	0.14 ± 0.02
Num-copula	7.44 ± 0.4	0.084 ± 0.01
ConvE-STF-copula	7.12 ± 0.3	$\mathbf{0.075 \pm 0.01}$
ConvE-STF-NF-seed	$\mathbf{7.06 \pm 0.3}$	$\mathbf{0.075 \pm 0.01}$
ConvE-STF-NF-latent	7.50 ± 0.3	0.080 ± 0.01

latent distributions are fully respected with seeding, which is not the case for the latent copula. This can lead to the sampling of irrelevant samples in the latent space.

The BS, which is equal to the MSE of the weather window probability, is largely better for probabilistic outputs that softens the penalization of FP and FN compared to a deterministic output. The ConvE-STF-copula and NF-seed show a similar BS for weather window forecasting, while the NF-latent is outperformed by all probabilistic models.

An illustration of the decision-making process based on scenarios forecasting is shown in Figure V.17. The output of the ConvE-STF is transformed into a set of scenarios, over which the probability of a weather window starting at a certain lead time is computed. It is compared to the observed weather window. The last 6 hours of forecast can not show weather window since the operation duration is 6 hours. They are not considered for metric computation. The operational limit of wind speed is showed in this figure. The same illustration is shown but for the ConvE-STF output without temporal embedding with copulas in Figure V.18. It illustrates the importance of emulating the temporal dependencies between variables. Though the parameters of the marginal distributions in the variables space are implicitly correlated through the conditioning, it is not sufficient to compute weather window probabilities.

The probabilistic forecast of weather window occurrence is evaluated by estimating its calibration, i.e. its statistical consistency. It is illustrated in Figure V.19. The frequency of weather window observed for a given weather window probability forecast is plotted for all models. The weather window under-estimation of the Num-copula and NF-latent, and over-estimation of the ConvE-STF-copula and NF-seed are illustrated by their distances

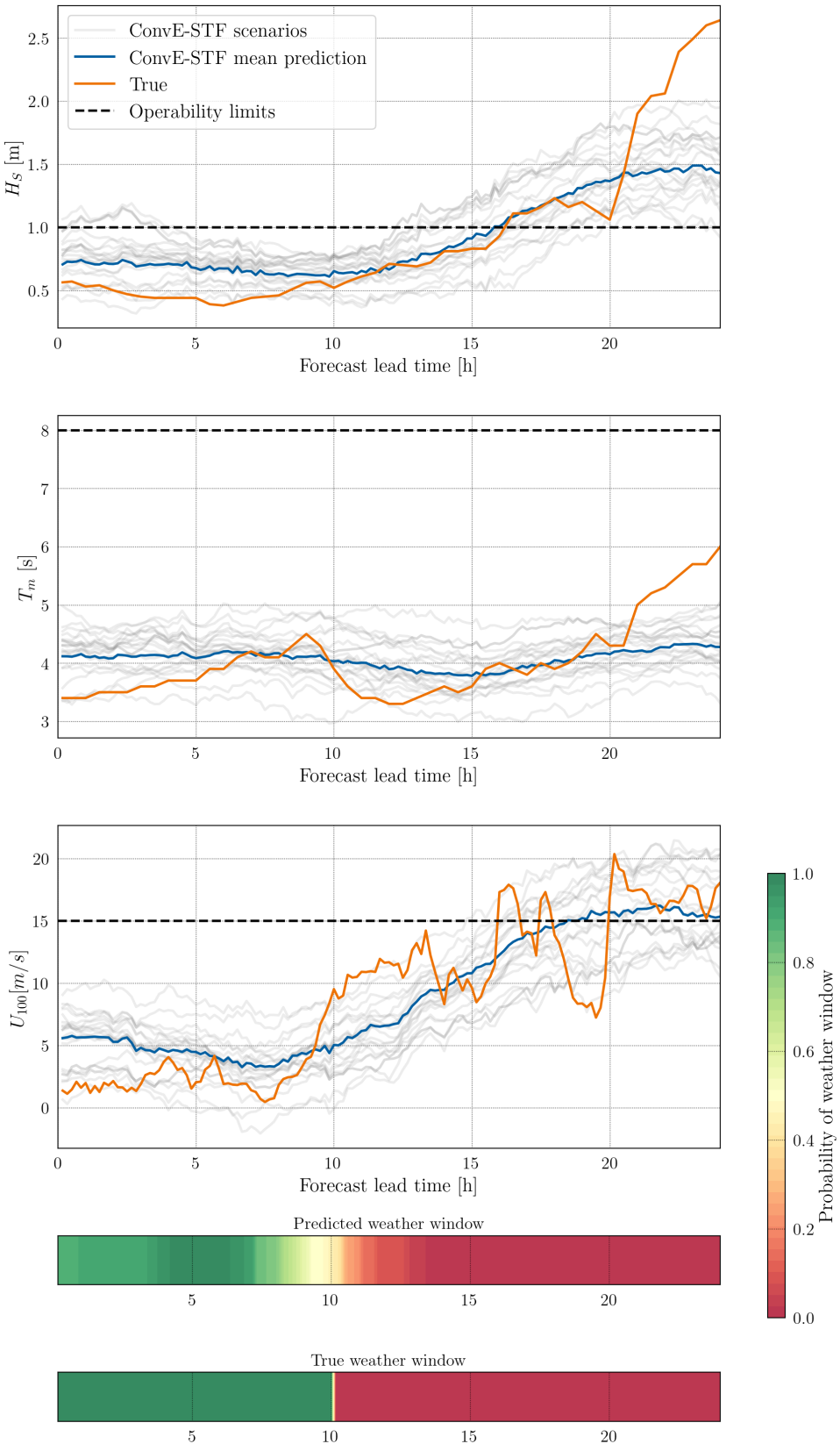


Figure V.17 – Illustration of a weather window probability prediction from the ConvE-STF-copula. The operability is computed from the generated set of multivariate scenarios.

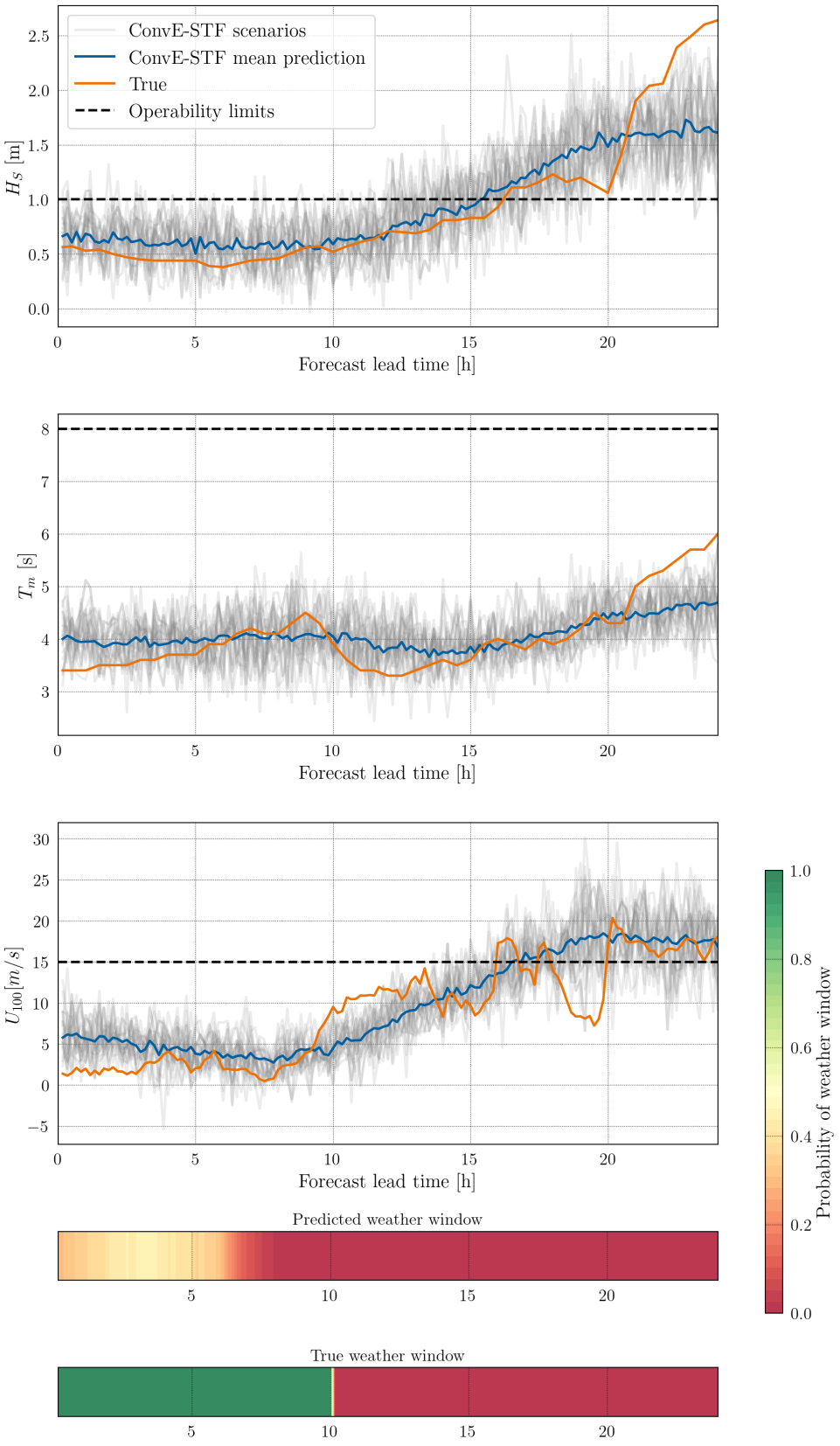


Figure V.18 – Weather window probability prediction from ConvE-STF if no temporal correlation is embedded. The samples are generated independently from each time step. This impacts the probability of weather window since it is depending on exceedance probability on a sliding window.

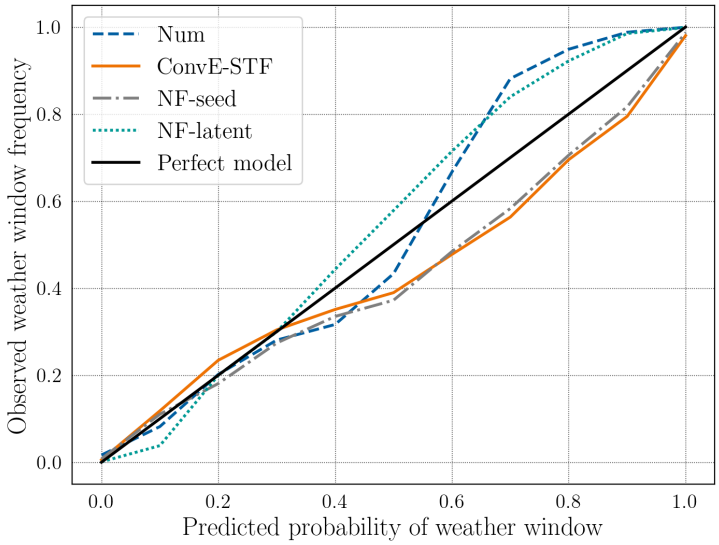


Figure V.19 – Evaluation of the probabilistic weather window forecasting calibration for the numerical model and ConvE-STF model. Models above the 1 : 1 line under-estimate the weather window probability and conversely. The spred of models is higher for weather windows with probability above 50%.

to the 1 : 1 line. For the Num-copula model, the difference is higher for high probability weather windows, which are widely under-estimated. For ConvE-STF-copula, the probability of weather is over-estimated for most windows, with a maximum for marginal situations around 50% chance of weather windows. The NF-seed model has a very similar trend to ConvE-STF, with very slightly better calibration, while the NF-latent model tend to largely under-estimate the weather window probability, showing that the sampling with latent copula creates samples with artificially higher spread compared to the predicted samples per lead time.

We therefore consider the separable Gaussian copula and seed-based sampling in latent space two competitive methods for generating probabilistic scenarios. Further research would be needed to explore temporal embedding methods for normalizing flow. Research directions are described in Section V.3.2.2.

The models behaviours are very different, though they are not well discriminated by the ES of the BS. The asymmetry in the consequences of FP and FN events are to be considered to be representative of the operational value of probabilistic metocean forecasts.

Table V.5 – Value metrics of implemented forecast models. The best obtained scores are show in bold. The scores are given as mean and standard deviation over the 6 splits. DT is the expectation of the downtime before an operation successfully starts. EFM and RFM are the expectation of the overhead cost due to FN and FP, without and with risk considerations.

Model	DT [hours]	EFM [€]	RFM [€]
Num-det	$\mathbf{4.5 \pm 1.1}$	1130 ± 230	1885 ± 1500
Num	4.7 ± 0.8	1140 ± 220	$\mathbf{1220 \pm 260}$
ConvE-STF	4.7 ± 0.8	1020 ± 150	1474 ± 540
ConvE-STF - NF - seed	4.7 ± 1.0	$\mathbf{990 \pm 130}$	1370 ± 290
ConvE-STF - NF - latent	5.3 ± 1.0	1280 ± 310	1980 ± 1280

V.4.3 Operational value

The operational value of the model forecast is evaluated using the metrics described in Section V.3.3.3, that try to reproduce a realistic operational framework to evaluate the real impact of using the different models. In particular, for offshore operations decision-making, the probability of weather window is required. Unless stated otherwise, the decision-making is based on the P50 of the forecast, i.e. $p_{critical} = 0.5$. The deterministic decision-making from the point forecast of numerical models is taken as a reference and called «Num-det».

The operational metrics are shown in Table V.5. The mean downtime DT to be expected when planning an operation at the forecast issue time is the lowest for deterministic decision-making with numerical model. This shows how the integration of forecast uncertainty impacts the weather window estimation and hence the decision-making. The downtime can be understood as the mean waiting time from forecast issue time before an operation is carried out successfully.

For a given operation, the use of generated scenarios to compute the weather window probability will naturally tend to increase the number of FN events, due to the added stochasticity with the copulas sampling. For a mean value right under the operability limit, less than 50% of the samples will be above limitations at a certain lead time. But integrated over the window length, with stochasticity introduced with the temporal correlation, the number of scenario crossing the operability at any lead time during the weather window will be above 50%, hence resulting to a 'No window' forecast.

The EFM penalizes FP and FN events asymmetrically to represent their associated

costs. According to this metrics, the ConvE-STF-NF-seed is the most efficient for operational planning. However, adding a risk penalty for conditions above limits during operations adds a strong penalization of FP, and none for FN, which changes the results. The ConvE-STF-copula being over-confident in its weather window prediction as illustrated in Figure V.19, it is more penalized than the Num-copula model by the risk term addition. This shows the importance of integrating a penalization for risky situation, and corroborates the conclusions of [60] that the forecast selection with EFM for economical optimization tends to increase the risk of performing operations in risky situations.

As discussed in Section V.3.3.3, when considering the EFM, the decision-making is to be optimal when consider a critical threshold $p_{critical}$ for the weather window probability to perform an operation. This threshold is shown in [7] to be equal to

$$p_{critical} = \frac{C_{FP}}{C_{FN} + C_{FP}} \tag{V.40}$$

hence balancing the number of FN and FP events depending on their associated costs. In this example, the critical threshold is equal to $p_{critical} = 0.31$. We can vary the threshold between 0 and 1 by varying the cost C_{FP} between 0 and $+\infty$ and plot the associated EFM for different thresholds values. The further away from $p_{critical} = 0.5$, the bigger the incentive to use probabilistic forecasts for decision-making. This is illustrated in Figure V.20, where the EFM normalized by the sum of C_{FN} and C_{FP} is plotted as function of the critical threshold. This graphs is proposed by [7] and helps understanding the effects of the decision-making based on the optimal threshold of the EFM. It shows that the ConvE-STF-copula model with its over-estimation of weather window probability is sub-optimal compared to the numerical model baselines at the critical threshold for thresholds between 0.5 and 0.7. Both curves are situated under the curve of the deterministic decision-making from numerical models point forecast, which emphasizes the importance of considering the forecast uncertainty in the decision-making under cost asymmetry.

For this case study, the obtained EFM and RFM based on P50 and $p_{critical}$ based decision-making are plotted with error bars for all models in Figure V.21. It can be noted that the $p_{critical}$ decision-making improves the EFM for all models, which shows that the weather window probability calibration is acceptable. It furthermore emphasizes the importance of balancing the decision-making based on the cost asymmetry of FP and FN. By lowering the probability threshold from which the operation will be planned, the number of FN is lowered and the number of FP is increased, which ends up reducing the EFM. It is intuitive that a high stake operation with low vessel mobilization cost should

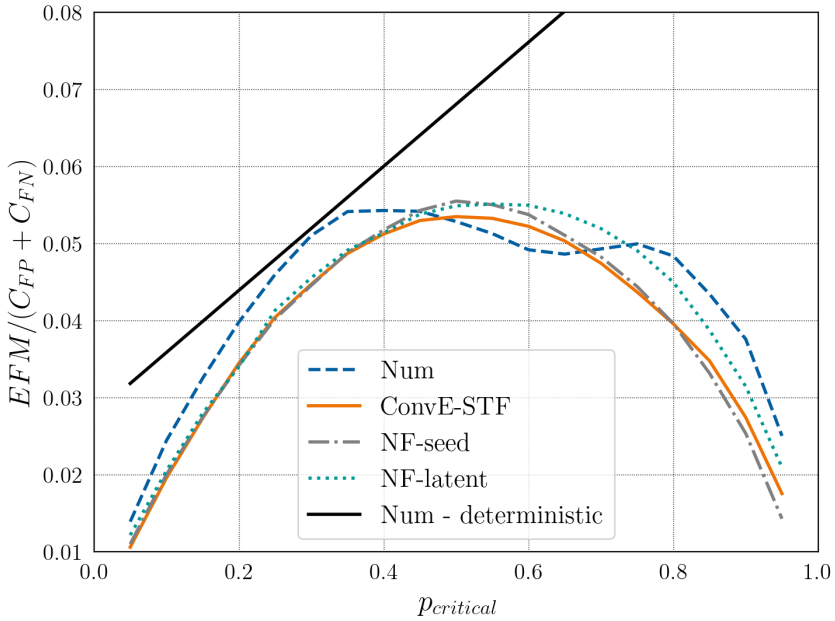


Figure V.20 – Comparison of resulting normalized EFM as function of the critical threshold for decision-making. The solid black line represents deterministic decision making. The added-value of probabilistic decision making is higher for high and low critical thresholds.

be tempted even with low success probability to be economically optimal. On the contrary, a low-stake operation with high vessel costs should be tempted only in favourable conditions with high weather window probability. In this case study, we consider a corrective maintenance with high production losses requiring a single crew transfer to the platform, which in economic terms should be tempted even for uncertain weather, which is illustrated by the value of the optimal critical threshold $p_{critical} = 0.31$.

However, as illustrated with the proposed RFM that includes a risk penalization, this economic optimum implies a higher operational risk. The increase of FP implies that more operations are planned under marginal conditions with observed wind and wave values above operability limits. While this is only penalized by the cost of vessel and crew in the EFM, it is penalized as function of the observed conditions in the RFM. Though the risk function proposed in this work is not realistic, the higher above limits the conditions, the higher the risk of dangerous events. It emphasizes that the risk taken by offshore crew will increase with the pressure of economic considerations. This work should serve as a basis for a more comprehensive inclusion of human factors and operational risks in the forecast value estimation.

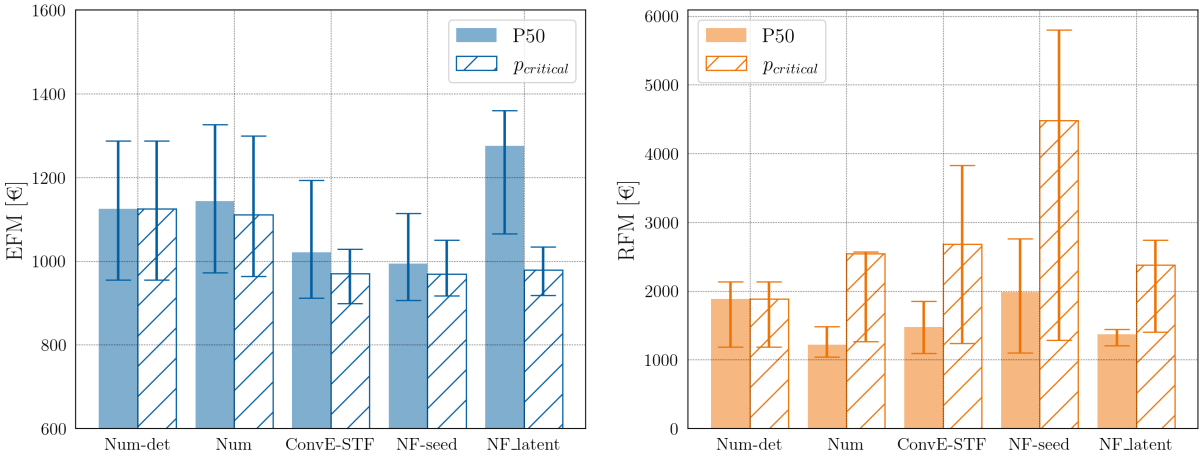


Figure V.21 – Comparison of EFM and RFM for a decision-making strategy based on a P50 estimation, and a decision-making strategy based on the critical threshold $p_{critical}$.

V.5 Conclusions and perspectives

The work on the Planier dataset experiment connects all the previous chapter of this thesis. The proposed ConvE-STF architecture, and its normalizing flows based extension are applied to a high quality dataset to evaluate their operational value for offshore wind energy maintenance.

The Planier LIDAR associated with the CANDHIS buoy measurements proves to be a very high quality dataset for offshore wind energy research. The AROME wind speed prediction at hub height is very consistent and proves to be a competitive baseline for post-processing model development. The WW3 wave forecast as a higher error and can be improved by post-processing. The length of the dataset is however an issue, and this highly advocates for longer measurement campaigns and easier access to model and in-situ measurements from weather services.

The ConvE-STF likelihood training is performant for estimating the forecast uncertainty. The estimation of the joint uncertainty of wind and wave parameters is important since these variables are correlated. The relaxing of the Gaussian posterior assumption with normalizing flows greatly improves the reliability of the forecast, showing the relevance of non-Gaussian characteristics of the posterior distribution, especially when dealing with positive and highly correlated environmental variables.

Operational decision-making requires an estimation of the weather window probability. It can be estimated from the generation of weather scenarios. The use of Gaussian copulas

with a separable covariance matrix is a competitive tool for generating trajectories. It allows for the use of a conditional covariance between variables.

Two approaches are proposed to generate scenarios with normalizing flows which could in theory relax the Gaussian copula assumption on the posterior. By numerically generate samples with seeding, scenarios can be generated that are then transformed through the flow. Doing so maintains the latent distribution per lead time. Though the temporal relationship between lead time has no reason to be realistic using this technique, it showed the best quality and performances compared to Gaussian copula assumption. This approach is compared to a model generating samples in the latent space using an empirical Gaussian copula. Its results are of poorer quality, probably due to non-realistic samples in the latent space. Several other possibilities are presented, in connection to recent work in the field of normalizing flows, that should be thoroughly studied for conditional scenario generation. RNN-conditioned normalizing flows, and full scenario latent space appear like appealing solutions, while preliminary work proved that they imply important numerical instabilities.

The operational value of metocean forecast models is to be considered in the forecast evaluation. Two metrics are proposed in addition to the literature Economic Forecasting Metric, the mean downtime and a Risk Forecasting Metric that includes probability of dangerous events. The probabilistic decision-making is shown to be beneficial in economic terms due to the cost asymmetry. However, we show that it can drastically increase the risk taken during the operation by pushing crews to operate in marginal and potentially dangerous conditions. A realistic risk penalization based on real costs and numerical results for operational limitations and dangerous events should be constructed. By doing so, a new critical threshold that would include human and material risk in the optimization could be constructed. We put the stress on the importance of modelling the human factor and risk taken by crew in the evaluation of forecast model, since operational decision-making eventually relies on the captain's decision for vessel safety.

Conclusion

Results summary

Motivated by industrial considerations, this thesis was dedicated to improving metocean forecasting for the safe and efficient planning of marine operations. The scope of work has been defined together with industrial and academic partners to specifically answer a technical challenge with innovative forecasting methods. In line with the recent developments in learning-based tools for weather forecasting, several data-driven methods were explored to answer the identified research questions and industrial needs.

Operational decision-making under uncertainty has been identified as one of the key cost reduction factor for offshore wind energy operational expenditures. It requires very short-term metocean forecasts of a range of limiting parameters, with uncertainty quantification, to be efficient. Current industrial standards degrade operability limits with semi-empirical factors to account for forecast uncertainty, which leads to excessive conservatism. The generation of reliable and high quality multivariate probabilistic forecasts of limiting parameters is a prerequisite for a more comprehensive uncertainty consideration in the decision-making. It would allow for the use probabilistic thresholds in weather window estimation.

Many studies apply learning-based methods for wave and weather forecast emulation or post-processing. For offshore studies, the availability of in-situ measurement data is the main obstacle to the development of such techniques. From this observation, a method based on the unsupervised clustering of numerical weather prediction data is proposed in Chapter III. It proved efficient compared to state-of-the-art data analysis tools to define a sensors network on offshore areas to optimally sample the wind resource. This study gives recommendations for the deployment of floating LIDAR networks in the French waters for the development of future offshore wind farms. It was used by Météo France for the production of a recommendation report for the French government on this very topic.

Taking advantage of a recent open source weather dataset in France, Chapter IV explores deep-learning based architecture for the probabilistic forecasting of the wind components at an island-based weather station. The use of numerical weather prediction

and neighbouring in-situ measurements as explanatory variables proved beneficial for the weather forecast post-processing. Convolutional architecture are then essential to handle the large amount of input data while keeping a reasonable computational cost. It is shown that the amount of input data that can be used as input is a key factor in the improved performances of deep learning architecture compared to a learning-based non-linear regression framework based on gradient boosting. The integration of neighbouring in-situ measurements with deep learning furthermore paves the way to forecast updates near real time, which could be a game changer for maintenance operations execution. A generative non-parametric assumption and a Gaussian assumption are compared for the shape of the posterior distribution. Though normalizing flows bring marginal gain in reliability in this case, they show important non-linear capabilities and could be of great use for modelling complex relationships between e.g. wind and waves parameters.

Building upon Chapter IV results, and benefiting from co-located offshore measurements of wind and waves at an offshore site representative of future offshore wind development areas, a real case study implementation is proposed in Chapter V. An implementation of the model built in Chapter IV is made for the joint wind and wave probabilistic forecasting 24 hours ahead. For this dataset, the offshore LIDAR based measurements are much less perturbed. The wind speed numerical predictions then prove very accurate, and are hardly improvable in terms of mean error. The deep learning model however reliably predicts the joint forecast uncertainty. Given the complex relationship between variables, the use of normalizing flows significantly improves model calibration, which is a strong argument for non-Gaussian posterior distributions.

To bridge the gap with the operational constraints of decision-making, scenarios (temporally sound samples) need to be generated from probabilistic forecasts. This increases significantly the dimension of the problem and requires the use of copulas to model the dependency. A Gaussian copula based on a semi-empirical separable covariance matrix is proposed as a baseline representative of the state-of-the-art. Two innovative methods are proposed for normalizing flows based scenario generation, that embed a temporal relationship in the latent space. Such methods appear promising to relax assumptions on the temporal correlation of the posterior distributions, that should be explored further as a potential alternative to complex covariance matrices for Gaussian copulas.

Eventually, the models are evaluated in a new operational value framework. Building upon the recent specified literature, it proposes a penalization of the risk taken during operation to better represent the human factor in decision-making. This point is an im-

portant conclusion of the discussions held with industrial partners, and results show that it can drastically influence decision-making. This penalization should counterbalance the incentives induced by the search of economic optima.

Perspectives

Larger and longer measurement campaigns for offshore research

The two experiments set up in Chapters IV and V highlight the lack of available offshore measurement campaigns in the French future offshore development areas. Though the development of data-driven models appears inevitable for a wide range of forecast applications for offshore wind, they require several years of environmental monitoring to be correctly trained and tested without data representativity issues. When widely deployed and maintained for long periods such as the FINO network in Germany, offshore measurement stations allow for numerous ambitious research studies. The Planier weather station used in Chapter V should then be maintained and diversified and would then permit significant research advances for offshore wind energy. The proposed sampling strategy in Chapter III could be augmented with additional environmental parameters to monitor such as waves, bio-chemical monitoring or biodiversity observation. A long-term multi-parameter monitoring network would be beneficial for understanding the complex impact of offshore wind on the environment, and would be crucial for monitoring climate change effects.

Thorough comparison of models reliability with operational ensemble predictions

The evaluation of the forecast uncertainty prediction was made throughout the thesis with the probabilistic metrics introduced in Chapter II. However, a fair comparison with actual state-of-the-art probabilistic forecast models which are the ensemble forecasts of numerical models is lacking. These forecasts are certainly limited in the number of ensemble members (16 for AROME in France), but their uncertainty estimation carries physical information. The proposed deep learning models can capture complex relationships given enough training data, including complex uncertainty patterns. However, the kind of uncertainty that ensemble forecasts provide is very different in nature. It would be crucial

that those two approaches are compared, especially in uncertain weather situations that can output diverging scenarios. Unfortunately we were not able to access ensemble forecasts from Météo France for this project. The ensembles are not accessible online due to the amount of data that it represents. Archives of ensemble forecasts for AROME only date back a few years ago, which was not applicable for our case study. This comparison is a priority for future work.

Forecast improvement with additional explanatory variables

In this thesis, numerical prediction and in-situ measurements of weather and wave from neighbouring measurement stations are used as input. They contain explanatory variables that condition the forecast of metocean parameters at an offshore target. The deep learning architectures have the advantage of being able to accommodate a large amount of input data, and this should be put to profit to look for explanatory variables from different data sources. Exogenous variables such as the sea surface temperature, convective potential energy, or indirect measurements of sea surface roughness from satellite imagery for example could be of great help for explaining some variance in the dataset, and especially uncertainties in the forecast. Furthermore, the meteorological situation in the Gulf of Lion is very complex. The choice of input data, spatial masks for numerical prediction input and explanatory variables could certainly be optimized in collaboration with weather experts of the area. In the proposed models, expert-oriented data pre-processing could be a great addition. For example, the Ligurian current East of the Gulf of Lion is a driver of local climate and should probably be better integrated in the input data. Seemingly, the Tramontane wind blowing in the Western Gulf of Lion and influencing the strong North-Western events in the study area should probably somehow be part of the input data. In this thesis, we have been limited by the amount of data that we could gather. Both for the offshore target and the input data. Through collaboration between weather institutes, offshore wind experts and forecast specialists, there is in our opinion a great room for improvement of the presented work. The case study have been constructed on the Eastern Gulf of Lion, a very complex meteorological area, and are to be tested on different maritime facades, with different wind and wave regimes. The generalization of the method is an important perspective of this thesis, with already gathered LIDAR measurements on all the next tender areas for offshore wind.

Flexible conditional multivariate scenario generation with generative approaches

The scenario generation introduced in Chapter V is of great importance for temporal dependant decision-making. It is the case for maintenance operations planning and execution, but also for a wide range of energy related applications including power production forecast, market participation, energy systems balancing etc. The toolbox for generating scenarios is often limited to copulas, which carry strong assumption on the temporal correlation. Gaussian copulas which have been studied in this work can be implemented easily, but they rely on the estimation of a covariance matrix. For likelihood based training such as the one proposed, proposing a parametrized covariance matrix is very problematic for numerical stability due to determinant computation and matrix inversion. These can be numerically unstable in high dimensions. The use of Gaussian copulas in a multivariate framework also implies fixing the dependency between variables. In this work, by using a separable covariance matrix between fixed empirical temporal correlation and conditional cross-variable covariance, we have been able to partially maintain some conditional correlation between variable. By introducing two novel methods using normalizing flows, we propose a research direction that unlocks the assumption-free scenario generation with generative methods. As shown in Chapter V, a naive assumption with a numerical trick for sampling in the latent space can bring improvements to the scenario generation. Based on recent literature, we advocate for further research on this topic, for example using scenarios latent space, or recurrent network conditioners for iterative sampling.

Operational evaluation framework with real maintenance data

In Chapter V, an evaluation framework for the probabilistic forecasting of weather window for a specific operation is presented. In addition to the economic metric proposed by [6], we proposed a risk forecasting metric that is to penalize the risk taken during operations. However, the values used both for the economic metric and the risk metric are not realistic. The industrial reality of operations is complex and prone to intense competition which prevents data sharing. Though the collaborative nature of the FLOWTOM project helps with the integration of industrial constraints, we regret the difficulties in accessing data. Data sharing is the cornerstone of research and development in offshore wind development. The rate at which new installations should be built is unprecedented

to efficiently fight against climate change. With a market maturity still in its infancy, offshore technologies would benefit from a wider collaboration between different actors. In addition to more realistic operability limits, durations and costs, it would be very interesting to compare decision-making strategies and forecast models on a real operational sequence. This would require monitoring and sharing data of a full maintenance sequence (e.g. planned inspection, blade replacement or installation), and playing a posteriori scenarios of decision-making based on different forecasts, both numerically and expert-based.

Offshore prediction from onshore measurements: other relevant metrics for offshore wind

This thesis tackled the joint wind and wave forecasting for operations planning, but offshore wind energy is sensitive to a wider range of phenomenon. The introduced methods should be evaluated for different other applications. Market participation or active control of wind turbines for example would require the forecast of other atmospheric or oceanic variables. In a research project led by France Energies Marines called NEMO, the impact of turbulence on offshore wind turbines and the possibility to characterize it with LIDAR measurements is being studied. A continuation of this thesis is planned for 2025 with the building of data-driven offshore turbulence forecast models from onshore measurements. The application of probabilistic forecasts to other applications such as power production requires the development of dedicated value evaluation frameworks. The value of the forecast model will depend on these evaluation framework and might differ from application to application. The models introduced in this work are eventually evaluated on the prediction of an exceedance probability on a sliding window. Their ability to predict extreme events such as ramp events is not assessed at all in this framework. However ramp events are crucial for power systems balancing and should be specifically evaluated for power production forecasts.

BIBLIOGRAPHY

[1] *Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the Intergovernmental Panel on Climate Change [Core writing team, H. Lee and J. Romero (eds.)]* Tech. rep. IPCC, 2023, pp. 35–115.

[2] RTE. *Panorama de l'électricité renouvelable au 31 décembre 2021.* Tech. rep. RTE, 2021.

[3] DNV. *DNV-OS-H101: Marine Operations, General.* Tech. rep. Det Norske Veritas, 2011.

[4] Ciaran Gilbert, Jethro Browell, and David McMillan. « Probabilistic access forecasting for improved offshore operations ». In: *International Journal of Forecasting* 37 (2021), pp. 134–150.

[5] Tomas Gintautas and John Dalsgaard Sørensen. « Improved methodology of weather window prediction for offshore operations based on probabilities of operation failure ». In: *Journal of Marine Science and Engineering* 5 (2017), p. 20.

[6] VM Catterson et al. « An economic impact metric for evaluating wave height forecasters for offshore wind maintenance access ». In: *Wind Energy* 19.2 (2016), pp. 199–212.

[7] James W. Taylor and Jooyoung Jeon. « Probabilistic forecasting of wave height for offshore wind turbine maintenance ». In: *European Journal of Operational Research* 267 (2018), pp. 877–890.

[8] Mengning Wu. « Uncertainty of Machine Learning-Based Methods for Wave Forecast and its Effect on Installation of Offshore Wind Turbines ». Ph.D. thesis. NTNU, 2021.

[9] Peter Bauer, Alan Thorpe, and Gilbert Brunet. « The quiet revolution of numerical weather prediction ». In: *Nature* 525.7567 (2015), pp. 47–55.

[10] Guillaume Couairon et al. « ArchesWeather: An efficient AI weather forecasting model at 1.5 {\deg} resolution ». In: *arXiv preprint arXiv:2405.14527* (2024).

[11] Remi Lam et al. « Learning skillful medium-range global weather forecasting ». In: *Science* 382.6677 (2023), pp. 1416–1421.

[12] Simon Lang et al. « AIFS - ECMWF's data-driven forecasting system ». In: *arXiv preprint arXiv:2406.01465* (2024).

[13] Kaifeng Bi et al. « Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast ». In: *arXiv preprint arXiv:2211.02556* (2022).

[14] Matteo Zambra et al. « Learning-based temporal estimation of in-situ wind speed from underwater passive acoustics ». In: *IEEE Journal of Oceanic Engineering* (2023).

[15] Jincheng Zhang and Xiaowei Zhao. « Spatiotemporal wind field prediction based on physics-informed deep learning and LIDAR measurements ». In: *Applied Energy* 288 (2021), p. 116641.

[16] Maximiliano A Sacco et al. « On-line machine-learning forecast uncertainty estimation for sequential data assimilation ». In: *Quarterly Journal of the Royal Meteorological Society* (2023).

[17] Clément Brochet et al. « Multivariate emulation of kilometer-scale numerical weather predictions with generative adversarial networks: A proof of concept ». In: *Artificial Intelligence for the Earth Systems* 2.4 (2023), p. 230006.

[18] Mengning Wu, Zhen Gao, and Yuna Zhao. « Assessment of allowable sea states for offshore wind turbine blade installation using time-domain numerical models and considering weather forecast uncertainty ». In: *Ocean Engineering* 260 (2022), p. 111801.

[19] Gwennaëlle Larvor et al. *MeteoNet, An Open Reference Weather Dataset by Meteo-France.* 2020.

[20] Tyler Stehly, Patrick Duffy, and Daniel Mulas Hernando. *2022 Cost of Wind Energy Review.* 2023.

[21] Helene Seyr and Michael Muskulus. « Decision support models for operations and maintenance for offshore wind farms: a review ». In: *Applied Sciences* 9.2 (2019), p. 278.

[22] Wilson Guachamin Acero et al. « Methodology for assessment of the operational limits and operability of marine operations ». In: *Ocean Engineering* 125 (2016), pp. 308–327.

[23] Rahul Chitteth Ramachandran et al. « Floating wind turbines: marine operations challenges and opportunities ». In: *Wind Energy Science* 7.2 (2022), pp. 903–924.

[24] Cleveland Abbe. « The physical basis of long-range weather forecasts ». In: *Monthly Weather Review* 29.12 (1901), pp. 551–561.

[25] Vilhelm Bjerknes. « Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik ». In: *Meteor. Z.* 21 (1904), pp. 1–7.

[26] Lewis F Richardson. *Weather prediction by numerical process*. University Press, 1922.

[27] Yann Seity et al. « The AROME-France convective-scale operational model ». In: *Monthly Weather Review* 139.3 (2011), pp. 976–991.

[28] Fabrice Ardhuin and Alejandro Orfila. « Wind waves ». In: *New Frontiers in Operational Oceanography* (2018), pp. 393–422.

[29] Said Obakrim et al. « Learning the spatiotemporal relationship between wind and significant wave height using deep learning ». In: *Environmental Data Science* 2 (2023), e5.

[30] Scott C James, Yushan Zhang, and Fearghal O'Donncha. « A machine learning framework to forecast wave conditions ». In: *Coastal Engineering* 137 (2018), pp. 1–10.

[31] Aurélien Callens et al. « Using Random forest and Gradient boosting trees to improve wave forecast at a specific location ». In: *Applied Ocean Research* 104 (2020), p. 102339.

[32] Marceau Michel et al. « Deep learning for statistical downscaling of sea states ». In: *Advances in Statistical Climatology, Meteorology and Oceanography* 8.1 (2022), pp. 83–95.

[33] Ravi Kumar Pandit, Athanasios Kolios, and David Infield. « Data-driven weather forecasting models performance comparison for improving offshore wind turbine availability and maintenance ». In: *IET Renewable Power Generation* 14.13 (2020), pp. 2386–2394.

[34] Georgios Leontaris, Oswaldo Morales-Nápoles, and A. R. M. (Rogier) Wolfert. « Probabilistic scheduling of offshore operations using copula based environmental time series – An application for cable installation management for offshore wind farms ». In: *Ocean Engineering* 125 (2016), pp. 328–341.

[35] Christos N Stefanakos and Konstandinos A Belibassakis. « Nonstationary stochastic modelling of multivariate long-term wind and wave data ». In: *International Conference on Offshore Mechanics and Arctic Engineering*. Vol. 41960. 2005, pp. 225–234.

[36] Sébastien Fouques, Dag Myrhaug, and Finn Gunnar Nielsen. « Seasonal modeling of multivariate distributions of metocean parameters with application to marine operations ». In: *J. Offshore Mech. Arct. Eng.* 126.3 (2004), pp. 202–212.

[37] Valérie Monbet, Pierre Ailliot, and Marc Prevosto. « Survey of stochastic models for wind and sea state time series ». In: *Probabilistic engineering mechanics* 22.2 (2007), pp. 113–126.

[38] Mengning Wu et al. « Prediction of short-term wind and wave conditions for marine operations using a multi-step-ahead decomposition-ANFIS model and quantification of its uncertainty ». In: *Ocean Engineering* 188 (2019), p. 106300.

[39] Philippe Drobinski et al. « HyMeX: A 10-year multidisciplinary program on the Mediterranean water cycle ». In: *Bulletin of the American Meteorological Society* 95.7 (2014), pp. 1063–1082.

[40] T. Mitchel. *Machine Learning*. McGraw-Hill Education, 1997.

[41] Kurt Hornik. « Approximation capabilities of multilayer feedforward networks ». In: *Neural networks* 4.2 (1991), pp. 251–257.

[42] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[43] Tilmann Gneiting and Adrian E Raftery. « Weather forecasting with ensemble methods ». In: *Science* 310.5746 (2005), pp. 248–249.

[44] Jonathan Dumas et al. « A deep generative model for probabilistic energy forecasting in power systems: normalizing flows ». In: *Applied Energy* 305 (2022), p. 117871.

[45] Pierre Pinson. « Estimation of the uncertainty in wind power forecasting ». Ph.D. thesis. École Nationale Supérieure des Mines de Paris, 2006.

[46] Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. « Probabilistic forecasts, calibration and sharpness ». In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 69.2 (2007), pp. 243–268.

[47] Tilmann Gneiting and Matthias Katzfuss. « Probabilistic forecasting ». In: *Annual Review of Statistics and Its Application* 1 (2014), pp. 125–151.

[48] Tilmann Gneiting et al. « Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds ». In: *TEST* 17 (2008), pp. 211–235.

[49] Tilmann Gneiting and Adrian E Raftery. « Strictly proper scoring rules, prediction, and estimation ». In: *Journal of the American Statistical Association* 102 (2007), pp. 359–378.

[50] Jakob W. Messner et al. « Evaluation of wind power forecasts—An up-to-date view ». In: *Wind Energy* 23 (2020), pp. 1461–1481.

[51] Ricardo J Bessa et al. « 'Good'or 'bad' wind power forecasts: A relative concept ». In: *Wind Energy* 14.5 (2011), pp. 625–636.

[52] B. Yildirim, C. Chryssostomidis, and G. E. Karniadakis. « Efficient sensor placement for ocean measurements using low-dimensional concepts ». In: *Ocean Modelling* 27.3-4 (2009), pp. 160–173.

[53] Krithika Manohar et al. « Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns ». In: *IEEE Control Systems Magazine* 38.3 (2018), pp. 63–86.

[54] B. W. Brunton et al. « Sparse sensor placement optimization for classification ». In: *SIAM Journal on Applied Mathematics* 76.5 (2016), pp. 2099–2122.

[55] A. P. Dempster, N. M. Laird, and D. B. Rubin. « Maximum likelihood from incomplete data via the EM algorithm ». In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22.

[56] Redouane Lguensat et al. « The analog data assimilation ». In: *Monthly Weather Review* 145 (2017), pp. 4093–4107.

[57] Paul Platzer. « Forecasts of dynamical systems from analogs: applications to geophysical variables with a focus on ocean waves ». Ph.D. thesis. Ecole nationale supérieure Mines-Télécom Atlantique, 2020.

[58] P. Pinson and R. Girard. « Evaluating the quality of scenarios of short-term wind power generation ». In: *Applied Energy* 96 (2012), pp. 12–20.

[59] Kashif Rasul et al. *Multivariate Probabilistic Time Series Forecasting via Conditioned Normalizing Flows*. 2021. arXiv: 2002.06103.

[60] C Gilbert, J Browell, and D McMillan. « Visualisation of probabilistic access forecasts for offshore operations ». In: *Journal of Physics: Conference Series*. Vol. 1222. 1. 2019, p. 012040.

[61] *World Energy Outlook 2023*. Tech. rep. IEA, 2023.

[62] *Renewables 2023*. Tech. rep. IEA, 2023.

[63] IEA. *Offshore wind outlook 2019*. Tech. rep. 2019.

[64] Ciaran Gilbert. « Topics in high dimensional energy forecasting ». Ph.D. thesis. Strathclyde University, 2021.

[65] Julia Slingo and Tim Palmer. « Uncertainty in weather and climate prediction ». In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1956 (2011), pp. 4751–4767.

[66] Martin G Schultz et al. « Can deep learning beat numerical weather prediction? » In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200097.

[67] Cristina L Archer et al. « Meteorology for coastal/offshore wind energy in the United States: Recommendations and research needs for the next 10 years ». In: *Bulletin of the American Meteorological Society* 95.4 (2014), pp. 515–519.

[68] Yun Wang et al. « A review of wind speed and wind power forecasting with deep neural networks ». In: *Applied Energy* 304 (2021), p. 117766.

[69] Alireza Shadmani et al. « A review of machine learning and deep learning applications in wave energy forecasting and WEC optimization ». In: *Energy Strategy Reviews* 49 (2023), p. 101180.

[70] Dang Ahn et al. « Comparative evaluation of different offshore wind turbine installation vessels for Korean west–south wind farm ». In: *International Journal of Naval Architecture and Ocean Engineering* 9.1 (2017), pp. 45–54.

[71] Zhengru Ren et al. « Offshore wind turbine operations and maintenance: A state-of-the-art review ». In: *Renewable and Sustainable Energy Reviews* 144 (2021), p. 110886.

[72] Katharina Fischer and Diego Coronado. « Condition monitoring of wind turbines: State of the art, user experience and recommendations ». In: *Fraunhofer-IWES, Bremerhaven* 7 (2015), pp. 51–6.

[73] Brian Jenkins et al. « Limiting wave conditions for the safe maintenance of floating wind turbines ». In: *Journal of Physics: Conference Series*. Vol. 2018. 1. 2021, p. 012023.

[74] Matti Scheu et al. « Human exposure to motion during maintenance on floating offshore wind turbines ». In: *Ocean Engineering* 165 (2018), pp. 293–306.

[75] Det Norske Veritas AS (DNV). *DNVGL-ST-N001 - Marine operations and marine warranty*. 2016.

[76] Peter Lynch. « The origins of computer weather prediction and climate modeling ». In: *Journal of computational physics* 227.7 (2008), pp. 3431–3444.

[77] Edward N Lorenz. « Atmospheric predictability experiments with a large numerical model ». In: *Tellus* 34.6 (1982), pp. 505–513.

[78] Andrew C Lorenc. « Analysis methods for numerical weather prediction ». In: *Quarterly Journal of the Royal Meteorological Society* 112.474 (1986), pp. 1177–1194.

[79] Philippe Courtier, J-N Thépaut, and Anthony Hollingsworth. « A strategy for operational implementation of 4D-Var, using an incremental approach ». In: *Quarterly Journal of the Royal Meteorological Society* 120.519 (1994), pp. 1367–1387.

[80] William Shaw et al. « Scientific challenges to characterizing the wind resource in the marine atmospheric boundary layer ». In: *Wind Energy Science Discussions* 2022 (2022), pp. 1–47.

[81] Jens Tambke et al. « Forecasting offshore wind speeds above the North Sea ». In: *Wind Energy* 8 (2005), pp. 3–16.

[82] Mike Optis et al. « Quantifying sensitivity in numerical weather prediction-modeled offshore wind speeds through an ensemble modeling approach ». In: *Wind Energy* 24 (2021).

[83] Yelena L. Pichugina et al. « Assessment of NWP forecast models in simulating offshore winds through the lower boundary layer by measurements from a ship-based scanning doppler lidar ». In: *Monthly Weather Review* 145 (2017), pp. 4277–4301.

[84] J. A. Sward, T. R. Ault, and K. M. Zhang. « Spatial biases revealed by LiDAR in a multiphysics WRF ensemble designed for offshore wind ». In: *Energy* 262 (2023), p. 125346.

[85] Feng Ye et al. « AIRU-WRF: A physics-guided spatio-temporal wind forecasting model and its application to the US Mid Atlantic offshore wind energy areas ». In: *Renewable Energy* 223 (2024), p. 119934.

[86] Alan J Geer. « Learning earth system models from observations: machine learning or data assimilation? » In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200089.

[87] Ronan Fablet et al. « End-to-end physics-informed representation learning for satellite ocean remote sensing data: Applications to satellite altimetry and sea surface currents ». In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3 (2021), pp. 295–302.

[88] Maxime Beauchamp et al. « End-to-end learning of variational interpolation schemes for satellite-derived SSH data ». In: 2021, pp. 7418–7421.

[89] Matteo Zambra. « Multi-modal AI methods in the context of heterogeneous oceanic observations and multisensor maritime surveillance ». Ph.D. thesis. Ecole nationale supérieure Mines-Télécom Atlantique, 2024.

[90] Redouane Lguensat et al. « Learning generalized quasi-geostrophic models using deep neural numerical models ». In: *arXiv preprint arXiv:1911.08856* (2019).

[91] Fabrice Ardhuin et al. « Observing sea states ». In: *Frontiers in Marine Science* 6 (2019), p. 124.

[92] The Wamdi Group. « The WAM model — A third generation ocean wave prediction model ». In: *Journal of physical oceanography* 18.12 (1988), pp. 1775–1810.

[93] Fabrice Ardhuin et al. « Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation ». In: *Journal of Physical Oceanography* 40.9 (2010), pp. 1917–1941.

[94] Said Obakrim. « Statistical downscaling and climate change in the coastal zone ». Ph.D. thesis. Université de Rennes, 2022.

[95] Ravi Pandit et al. « Sequential data-driven long-term weather forecasting models' performance comparison for improving offshore operation and maintenance operations ». In: *Energies* 15.19 (2022), p. 7233.

[96] Valerie Monbet and Marc Prevosto. « Bivariate simulation of non stationary and non Gaussian observed processes: Application to sea state parameters ». In: *Applied Ocean Research* 23.3 (2001), pp. 139–145.

[97] Arto Niemi and Frank Sill Torres. « Application of synthetic weather time series based on four-dimensional copula for modeling of maritime operations ». In: *OCEANS 2021: San Diego–Porto*. 2021, pp. 1–8.

[98] Yi Yang et al. « Multivariate joint distribution of five-dimensional wind and wave parameters in the sea-crossing bridge region using Hierarchical Archimedean Copulas ». In: *Journal of Wind Engineering and Industrial Aerodynamics* 247 (2024), p. 105684.

[99] Jack Paterson et al. « Assessing marine operations with a Markov-switching autoregressive metocean model ». In: *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment* 234.4 (2020), pp. 785–802.

[100] Ruben de Nie et al. « Offshore infrastructure planning using a vine copula approach for environmental conditions: an application for replacement maintenance of tidal energy infrastructure ». In: *Structure and Infrastructure Engineering* 15.5 (2019), pp. 600–617.

[101] V Valamanesh, AT Myers, and SR Arwade. « Multivariate analysis of extreme metocean conditions for offshore wind turbines ». In: *Structural Safety* 55 (2015), pp. 60–69.

[102] Oleg Gaidai et al. « SEM-REV offshore energy site wind-wave bivariate statistics by hindcast ». In: *Renewable energy* 156 (2020), pp. 689–695.

[103] Cristina L. Archer et al. « Meteorology for coastal/offshore wind energy in the united states: Recommendations and research needs for the next 10 years ». In: *Bulletin of the American Meteorological Society* 95 (2014), pp. 515–519.

[104] David A Smith et al. « Wind lidar evaluation at the Danish wind test site in Høvsøre ». In: *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 9.1-2 (2006), pp. 87–93.

[105] Julia Gottschall et al. « Floating lidar as an advanced offshore wind speed measurement technique: current technology status and gap analysis in regard to full maturity ». In: *Wiley Interdisciplinary Reviews: Energy and Environment* 6.5 (2017), e250.

[106] Maxime Thiébaut et al. « Experimental evaluation of the motion-induced effects for turbulent fluctuations measurement on floating lidar systems ». In: *Remote Sensing* 16.8 (2024), p. 1337.

[107] Liyun He et al. « Learning-based emulation of sea surface wind fields from numerical model outputs and SAR data ». In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.10 (2015), pp. 4742–4750.

[108] Jia Liu et al. « A Spatial downscaling approach for windsat satellite sea surface wind based on generative adversarial networks and dual learning scheme ». In: *Remote Sensing* 14.3 (2022), p. 769.

[109] Aurélien Colin et al. « Reduction of rain-induced errors for wind speed estimation on SAR observations using convolutional neural networks ». In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2023).

[110] Jooyoung Jeon and James W Taylor. « Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation ». In: *International Journal of Forecasting* 32.3 (2016), pp. 991–1004.

[111] I Dinwoodie, VM Catterson, and D McMillan. « Wave height forecasting to improve off-shore access and maintenance scheduling ». In: *2013 IEEE Power & Energy Society General Meeting*. 2013, pp. 1–5.

[112] R Rainaud et al. « Characterization of air-sea exchanges over the Western Mediterranean Sea during HyMeX SOP1 using the AROME-WMED model ». In: *Quarterly Journal of the Royal Meteorological Society* 142 (2016), pp. 173–187.

[113] Claude Millot. « The gulf of Lions' hydrodynamics ». In: *Continental shelf research* 10.9-11 (1990), pp. 885–894.

[114] Amandine Schaeffer et al. « Influence of high-resolution wind forcing on hydrodynamic modeling of the Gulf of Lions ». In: *Ocean Dynamics* 61 (2011), pp. 1823–1844.

[115] Hiba Omrani et al. « Spatial and temporal variability of wind energy resource and production over the North Western Mediterranean Sea: Sensitivity to air-sea interactions ». In: *Renewable energy* 101 (2017), pp. 680–689.

[116] Fabrice Ardhuin et al. « Comparison of wind and wave measurements and models in the Western Mediterranean Sea ». In: *Ocean Engineering* 34.3-4 (2007), pp. 526–541.

[117] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. « Data clustering: a review ». In: *ACM computing surveys (CSUR)* 31.3 (1999), pp. 264–323.

[118] Suman Ravuri et al. « Skilful precipitation nowcasting using deep generative models of radar ». In: *Nature* 597.7878 (2021), pp. 672–677.

[119] Danilo Rezende and Shakir Mohamed. « Variational inference with normalizing flows ». In: 2015, pp. 1530–1538.

[120] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. *Density estimation using Real NVP*. 2017.

[121] Laurent Dinh, David Krueger, and Yoshua Bengio. « Nice: Non-linear independent components estimation ». In: *arXiv preprint arXiv:1410.8516* (2014).

[122] George Papamakarios, Theo Pavlakou, and Iain Murray. « Masked autoregressive flow for density estimation ». In: *Advances in neural information processing systems* 30 (2017).

[123] Thomas Müller et al. « Neural importance sampling ». In: *ACM Transactions on Graphics (ToG)* 38.5 (2019), pp. 1–19.

[124] Conor Durkan et al. « Cubic-spline flows ». In: *arXiv preprint arXiv:1906.02145* (2019).

[125] Conor Durkan et al. « Neural spline flows ». In: vol. 32. 2019.

[126] Mathieu Germain et al. « Made: Masked autoencoder for distribution estimation ». In: *International conference on machine learning*. 2015, pp. 881–889.

[127] Katherine Haynes et al. « Creating and evaluating uncertainty estimates with neural networks for environmental-science applications ». In: *Artificial Intelligence for the Earth Systems* 2.2 (2023), p. 220061.

[128] Moritz N Lang et al. « Bivariate Gaussian models for wind vectors in a distributional regression framework ». In: *Advances in Statistical Climatology, Meteorology and Oceanography* 5.2 (2019), pp. 115–132.

[129] Maximiliano A. Sacco et al. « Evaluation of machine learning techniques for forecast uncertainty quantification ». In: *Quarterly Journal of the Royal Meteorological Society* 148 (2022), pp. 3470–3490.

[130] Mousa Afrasiabi et al. « Advanced deep learning approach for probabilistic wind speed forecasting ». In: *IEEE Transactions on Industrial Informatics* 17 (2021), pp. 720–727.

[131] Amandine Pierrot and Pierre Pinson. « Adaptive generalized logit-normal distributions for wind power short-term forecasting ». In: *2021 IEEE Madrid PowerTech*. 2021, pp. 1–6.

[132] Sándor Baran and Sebastian Lerch. « Log-normal distribution based Ensemble Model Output Statistics models for probabilistic wind-speed forecasting ». In: *Quarterly Journal of the Royal Meteorological Society* 141.691 (2015), pp. 2289–2299.

[133] Pierre Pinson, Gordon Reikard, and J-R Bidlot. « Probabilistic forecasting of the wave energy flux ». In: *Applied Energy* 93 (2012), pp. 364–370.

[134] Hao Zhang et al. « Improved deep mixture density network for regional wind power probabilistic forecasting ». In: *IEEE Transactions on Power Systems* 35.4 (2020), pp. 2549–2560.

[135] Mucun Sun et al. « A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization ». In: *Applied energy* 238 (2019), pp. 1497–1505.

[136] Mucun Sun, Cong Feng, and Jie Zhang. « Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation ». In: *Applied Energy* 256 (2019), p. 113842.

[137] Thomas Muschinski et al. « Cholesky-based multivariate Gaussian regression ». In: *Econometrics and Statistics* (2022).

[138] Tilmann Gneiting, William Kleiber, and Martin Schlather. « Matérn cross-covariance functions for multivariate random fields ». In: *Journal of the American Statistical Association* 105.491 (2010), pp. 1167–1177.

[139] Jethro Browell, Ciaran Gilbert, and Matteo Fasiolo. « Covariance structures for high-dimensional energy forecasting ». In: *Electric Power Systems Research* 211 (2022), p. 108446.

[140] Frederick N Fritsch and Ralph E Carlson. « Monotone piecewise cubic interpolation ». In: *SIAM Journal on Numerical Analysis* 17.2 (1980), pp. 238–246.

[141] Sky McKinley and Megan Levine. « Cubic spline interpolation ». In: *College of the Redwoods* 45.1 (1998), pp. 1049–1060.

[142] Yaoyao He et al. « Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression ». In: *Neurocomputing* 430 (2021), pp. 121–137.

[143] Runmin Zou et al. « Deep non-crossing probabilistic wind speed forecasting with multi-scale features ». In: *Energy Conversion and Management* 257 (2022), p. 115433.

[144] Pierre Pinson et al. « Non-parametric probabilistic forecasts of wind power: required properties and evaluation ». In: *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 10.6 (2007), pp. 497–516.

[145] Ricardo J Bessa. « On the quality of the Gaussian copula for multi-temporal decision-making problems ». In: *2016 Power Systems Computation Conference (PSCC)*. 2016, pp. 1–7.

[146] Pierre Pinson and Henrik Madsen. « Ensemble-based probabilistic forecasting at Horns Rev ». In: *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 12.2 (2009), pp. 137–155.

[147] Pierre Pinson. « Adaptive calibration of (u,v)-wind ensemble forecasts ». In: *Quarterly Journal of the Royal Meteorological Society* 138 (2012), pp. 1273–1284.

[148] Thomas Muschinski et al. « Cholesky-based multivariate Gaussian regression ». In: *Econometrics and Statistics* 29 (2022), pp. 261–281.

[149] M Sklar. « Fonctions de répartition à n dimensions et leurs marges ». In: *Annales de l'ISUP*. Vol. 8. 3. 1959, pp. 229–231.

[150] David Salinas et al. « High-dimensional multivariate forecasting with low-rank gaussian copula processes ». In: *Advances in neural information processing systems* 32 (2019).

[151] Pierre Pinson et al. « From probabilistic forecasts to statistical scenarios of short-term wind power production ». In: *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 12.1 (2009), pp. 51–62.

[152] Mingjian Cui et al. « A copula-based conditional probabilistic forecast model for wind power ramps ». In: *IEEE Transactions on Smart Grid* 10.4 (2018), pp. 3870–3882.

[153] Andrew J Patton. « Modelling asymmetric exchange rate dependence ». In: *International economic review* 47.2 (2006), pp. 527–556.

[154] Jean-David Fermanian and Marten H Wegkamp. « Time-dependent copulas ». In: *Journal of Multivariate Analysis* 110 (2012), pp. 19–29.

[155] Andrew J Patton. « A review of copula models for economic time series ». In: *Journal of Multivariate Analysis* 110 (2012), pp. 4–18.

[156] Hans Hersbach. « Decomposition of the continuous ranked probability score for ensemble prediction systems ». In: *Weather and Forecasting* 15.5 (2000), pp. 559–570.

[157] Michael Scheuerer and Thomas M Hamill. « Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities ». In: *Monthly Weather Review* 143.4 (2015), pp. 1321–1334.

[158] Allan H Murphy et al. « Repetitive decision making and the value of forecasts in the cost-loss ratio situation: a dynamic model ». In: *Monthly Weather Review* 113.5 (1985), pp. 801–813.

[159] Pierre Pinson, Christophe Chevallier, and George N Kariniotakis. « Trading wind generation from short-term probabilistic forecasts of wind power ». In: *IEEE transactions on Power Systems* 22.3 (2007), pp. 1148–1156.

[160] Pierre Pinson. « Wind energy: Forecasting challenges for its operational management ». In: *Statistical Science* 28.4 (2013), pp. 564–585.

[161] Corinna Möhrlen, Ricardo J Bessa, and Nadine Fleischhut. « A decision-making experiment under wind power forecast uncertainty ». In: *Meteorological Applications* 29.3 (2022), e2077.

[162] A Kolios et al. « Effect of weather forecast uncertainty on offshore wind farm availability assessment ». In: *Ocean Engineering* 285 (2023), p. 115265.

[163] Petros Papadopoulos, David W Coit, and Ahmed Aziz Ezzat. « Seizing opportunity: Maintenance optimization in offshore wind farms considering accessibility, production, and crew dispatch ». In: *IEEE Transactions on Sustainable Energy* 13.1 (2021), pp. 111–121.

[164] David Rowell, David McMillan, and James Carroll. « Offshore wind H&S: A review and analysis ». In: *Renewable and Sustainable Energy Reviews* 189 (2024), p. 113928.

[165] George EP Box et al. *Time series analysis: forecasting and control*. Ed. by San Francisco: Holden Bay. 1976.

[166] Ergin Erdem and Jing Shi. « ARMA based approaches for forecasting the tuple of wind speed and direction ». In: *Applied Energy* 88.4 (2011), pp. 1405–1414.

[167] Jose Luis Torres et al. « Forecast of hourly average wind speed with ARMA models in Navarre (Spain) ». In: *Solar energy* 79.1 (2005), pp. 65–77.

[168] Pierre Ailliot. « Modèles autorégressifs à changements de régimes markoviens. Applications aux séries tempo-relles de vent ». Ph.D. thesis. Université Rennes 1, 2004.

[169] Hui Liu and Chao Chen. « Data processing strategies in wind energy forecasting models and applications: A comprehensive review ». In: *Applied Energy* 249 (2019), pp. 392–408.

[170] SN Singh, Abheejeet Mohapatra, et al. « Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting ». In: *Renewable energy* 136 (2019), pp. 758–768.

[171] Lilin Cheng et al. « Ensemble recurrent neural network based probabilistic wind speed forecasting approach ». In: *Energies* 11.8 (2018), p. 1958.

[172] Da Liu et al. « Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm ». In: *Renewable energy* 62 (2014), pp. 592–597.

[173] Ioannis P Panapakidis, Constantine Michailides, and Demos C Angelides. « A data-driven short-term forecasting model for offshore wind speed prediction based on computational intelligence ». In: *Electronics* 8.4 (2019), p. 420.

[174] Aditya N Deshmukh et al. « Neural-network-based data assimilation to improve numerical ocean wave forecast ». In: *IEEE Journal of Oceanic Engineering* 41.4 (2016), pp. 944–953.

[175] Hao Zhang et al. « Uncertain accessibility estimation method for offshore wind farm based on multi-step probabilistic wave forecasting ». In: *IET Renewable Power Generation* 15.13 (2021), pp. 2944–2955.

[176] Norden E Huang et al. « The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis ». In: *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454.1971 (1998), pp. 903–995.

[177] Neeraj Bokde et al. « A review on hybrid empirical mode decomposition models for wind speed and wind power prediction ». In: *Energies* 12.2 (2019), p. 254.

[178] Bharat Kumar Saxena, Sanjeev Mishra, and Komaragiri Venkata Subba Rao. « Offshore wind speed forecasting at different heights by using ensemble empirical mode decomposition and deep learning models ». In: *Applied Ocean Research* 117 (2021), p. 102937.

[179] Zheyong Jiang, Jinxing Che, and Lina Wang. « Ultra-short-term wind speed forecasting based on EMD-VAR model and spatial correlation ». In: *Energy Conversion and Management* 250 (2021), p. 114919.

[180] Dawei Geng, Haifeng Zhang, and Hongyu Wu. « Short-term wind speed prediction based on principal component analysis and LSTM ». In: *Applied sciences* 10.13 (2020), p. 4416.

[181] Yagang Zhang et al. « A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting ». In: *Energy Conversion and Management* 195 (2019), pp. 180–197.

[182] Paul Platzer et al. « Probability distributions for analog-to-target distances ». In: *Journal of the Atmospheric Sciences* 78 (2021), pp. 3317–3335.

[183] Juan Ruiz et al. « Analog data assimilation for the selection of suitable general circulation models ». In: *Geoscientific Model Development Discussions* 2022 (2022), pp. 1–30.

[184] Alex Ayet and Pierre Tandeo. « Nowcasting solar irradiance using an analog method and geostationary satellite images ». In: *Solar Energy* 164 (2018), pp. 301–315.

[185] Deockho Kim and Jin Hur. « Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method ». In: *Energy* 157 (2018), pp. 211–226.

[186] Jerome H. Friedman. « Greedy functionf approximation: A gradient boosting machine ». In: *The Annals of Statistics* 29 (2001), pp. 1189–1232.

[187] Mark Landry et al. « Probabilistic gradient boosting machines for GEFCom2014 wind forecasting ». In: *International Journal of Forecasting* 32.3 (2016), pp. 1061–1066.

[188] Tao Hong et al. *Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond*. 2016.

[189] Alireza Zendehboudi, M Abdul Baseer, and R Saidur. « Application of support vector machine models for forecasting solar and wind energy resources: A review ». In: *Journal of cleaner production* 199 (2018), pp. 272–285.

[190] Jason Ansel et al. « PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation ». In: 2024.

[191] Martín Abadi et al. *TensorFlow, Large-scale machine learning on heterogeneous systems*. 2015.

[192] Bryan Lim and Stefan Zohren. « Time-series forecasting with deep learning: a survey ». In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200209.

[193] Kazutoshi Higashiyama, Yu Fujimoto, and Yasuhiro Hayashi. « Feature extraction of NWP data for wind power forecasting ising 3D-convolutional neural networks ». In: *Energy Procedia* 155 (2018). 12th International Renewable Energy Storage Conference, IRES 2018, 13-15 March 2018, Düsseldorf, Germany, pp. 350–358.

[194] Florian Dupuy et al. « ARPEGE cloud cover forecast postprocessing with convolutional neural network ». In: *Weather and Forecasting* 36.2 (2021), pp. 567–586.

[195] Ashesh Chattopadhyay et al. « Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving deep spatial transformers ». In: *arXiv preprint arXiv:2103.09360* (2021).

[196] Ying-Yi Hong and Thursy Rienda Aulia Satriani. « Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network ». In: *Energy* 209 (2020), p. 118441.

[197] Ceyhun Yildiz et al. « An improved residual-based convolutional neural network for very short-term wind power forecasting ». In: *Energy Conversion and Management* 228 (2021), p. 113731.

[198] Huai-zhi Wang et al. « Deep learning based ensemble approach for probabilistic wind power forecasting ». In: *Applied energy* 188 (2017), pp. 56–70.

[199] Pradeep Hewage et al. « Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station ». In: *Soft Computing* 24 (2020), pp. 16453–16482.

[200] Zhenhao Gan et al. « Temporal convolutional networks interval prediction model for wind speed forecasting ». In: *Electric Power Systems Research* 191 (2021), p. 106865.

[201] Dan Li et al. « Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks ». In: *Energy* 238 (2022), p. 121981.

[202] Xin Liu, Jun Zhou, and Huimin Qian. « Short-term wind power forecasting by stacked recurrent neural networks with parametric sine activation function ». In: *Electric Power Systems Research* 192 (2021), p. 107011.

[203] Mehdi Neshat et al. « A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the Lillgrund offshore wind farm ». In: *Energy conversion and management* 236 (2021), p. 114002.

[204] Jean-François Toubeau et al. « Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets ». In: *IEEE Transactions on Power Systems* 34.2 (2018), pp. 1203–1215.

[205] Yaoran Chen et al. « 2-D regional short-term wind speed forecast based on CNN-LSTM deep learning model ». In: *Energy Conversion and Management* 244 (2021), p. 114451.

[206] Ashish Vaswani et al. « Attention is all you need ». In: *Advances in neural information processing systems* 30 (2017).

[207] Bryan Lim et al. « Temporal fusion transformers for interpretable multi-horizon time series forecasting ». In: *International Journal of Forecasting* 37.4 (2021), pp. 1748–1764.

[208] Ian Goodfellow et al. « Generative adversarial networks ». In: *Communications of the ACM* 63.11 (2020), pp. 139–144.

[209] Diederik P Kingma and Max Welling. « Auto-encoding variational bayes ». In: *arXiv preprint arXiv:1312.6114* (2013).

[210] Zhong Zheng et al. « Generative probabilistic wind speed forecasting: A variational recurrent autoencoder based method ». In: *IEEE Transactions on Power Systems* 37 (2022), pp. 1386–1398.

[211] Kashif Rasul et al. « Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting ». In: *International Conference on Machine Learning*. 2021, pp. 8857–8868.

[212] Jaesung Jung and Robert P Broadwater. « Current status and future advances for wind speed and power forecasting ». In: *Renewable and Sustainable Energy Reviews* 31 (2014), pp. 762–777.

[213] Stephan Rasp et al. « WeatherBench: a benchmark data set for data-driven weather forecasting ». In: *Journal of Advances in Modeling Earth Systems* 12.11 (2020), e2020MS002203.

[214] P.R. Shukla et al. *Climate Change 2022: Mitigation of Climate Change. Working Group III Contribution to the IPCC Sixth Assessment Report*. Tech. rep. IPCC, 2022.

[215] K. S. R. Murthy and O. P. Rahi. « A comprehensive review of wind resource assessment ». In: *Renewable and Sustainable Energy Reviews* 72 (2017), pp. 1320–1342.

[216] Julia Gottschall et al. « Floating lidar as an advanced offshore wind speed measurement technique: current technology status and gap analysis in regard to full maturity ». In: *WIREs Energy and Environment* 6.5 (2017), e250.

[217] Russ E. Davis. « Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean ». In: *Journal of Physical Oceanography* 6.3 (1976), pp. 249–266.

[218] David WJ Thompson and John M. Wallace. « Annular modes in the extratropical circulation. Part I: Month-to-month variability ». In: *Journal of climate* 13.5 (2000), pp. 1000–1016.

[219] Ed Hawkins and Rowan Sutton. « Variability of the Atlantic thermohaline circulation described by three-dimensional empirical orthogonal functions ». In: *Climate Dynamics* 29.7 (2007), pp. 745–762.

[220] G. W. K. Moore, I. A. Renfrew, and Robert S. Pickart. « Multidecadal mobility of the North Atlantic oscillation ». In: *Journal of Climate* 26.8 (2013), pp. 2453–2466.

[221] Yanwu Zhang and James G. Bellingham. « An efficient method of selecting ocean observing locations for capturing the leading modes and reconstructing the full field ». In: *Journal of Geophysical Research: Oceans* 113.C4 (2008).

[222] Alejandro Castillo and Arturo Roman Messina. « Data-driven sensor placement for state reconstruction via POD analysis ». In: *IET Generation, Transmission & Distribution* 14.4 (2019), pp. 656–664.

[223] Xiu Yang et al. « EOF-based constrained sensor placement and field reconstruction from noisy ocean measurements: Application to Nantucket Sound ». In: *Journal of Geophysical Research: Oceans* 115.C12 (2010).

[224] Emily Clark, J. Nathan Kutz, and Steven L. Brunton. « Sensor selection with cost constraints for dynamically relevant bases ». In: *IEEE Sensors Journal* 20.19 (2020), pp. 11674–11687.

[225] Sundeep Prabhakar Chepuri and Geert Leus. « Continuous sensor placement ». In: *IEEE Signal Processing Letters* 22.5 (2014), pp. 544–548.

[226] Thomas L. Mohren et al. « Neural-inspired sensors enable sparse, efficient classification of spatiotemporal data ». In: *Proceedings of the National Academy of Sciences* 115.42 (2018), pp. 10564–10569.

[227] Kai Fukami et al. « Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning ». In: *Nature Machine Intelligence* 3.11 (2021), pp. 945–951.

[228] J. Annoni et al. « Sparse-sensor placement for wind farm control ». In: *Journal of Physics: Conference Series*. Vol. 1037. 2018, p. 032019.

[229] Naseem Ali, Marc Calaf, and Raúl Bayoán Cal. « Clustering sparse sensor placement identification and deep learning based forecasting for wind turbine wakes ». In: *Journal of Renewable and Sustainable Energy* 13.2 (2021), p. 023307.

[230] CEREMA. *Eoliennes en mer en France*. 2022. URL: `https://www.eoliennesenmer.fr/`.

[231] Yann Seity et al. « The AROME-France convective-scale operational model ». In: *Monthly Weather Review* 139.3 (2011), pp. 976–991.

[232] Piet Termonia et al. « The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1 ». In: *Geoscientific Model Development* 11 (2018), pp. 257–281.

[233] PHILIPPE Courtier, J.-N. Thépaut, and Anthony Hollingsworth. « A strategy for operational implementation of 4D-Var, using an incremental approach ». In: *Quarterly Journal of the Royal Meteorological Society* 120.519 (1994), pp. 1367–1387.

[234] Kelly Cohen, Stefan Siegel, and Tom McLaughlin. « Sensor placement based on proper orthogonal decomposition modeling of a cylinder wake ». In: vol. 4259. 2003.

[235] Steven J. Leon, Åke Björck, and Walter Gander. « Gram-Schmidt orthogonalization: 100 years and more ». In: *Numerical Linear Algebra with Applications* 20.3 (2013), pp. 492–532.

[236] Douglas A. Reynolds. « Gaussian mixture models. » In: *Encyclopedia of biometrics* 741 (2009), pp. 659–663.

[237] C. F. Jeff Wu. « On the Convergence Properties of the EM Algorithm ». In: *The Annals of Statistics* 11.1 (1983), pp. 95–103.

[238] Gideon Schwarz. « Estimating the dimension of a model ». In: *The annals of statistics* (1978), pp. 461–464.

[239] R. Rainaud et al. « Characterization of air–sea exchanges over the Western Mediterranean Sea during HyMeX SOP1 using the AROME–WMED model ». In: *Quarterly Journal of the Royal Meteorological Society* 142 (2016), pp. 173–187.

[240] Robin Marcille et al. « Convolutional encoding and normalizing flows: a deep learning approach for offshore wind speed probabilistic forecasting in the Mediterranean Sea ». In: *Artificial Intelligence for the Earth Systems* 1.aop (2024).

[241] Eric P. James, Stanley G. Benjamin, and Melinda Marquis. « Offshore wind speed estimates from a high-resolution rapidly updating numerical weather prediction model forecast dataset ». In: *Wind Energy* 21 (2018), pp. 264–284.

[242] Julia Slingo and Tim Palmer. « Uncertainty in weather and climate prediction ». In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369 (2011), pp. 4751–4767.

[243] Stéphane Vannitsem et al. « Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world ». In: *Bulletin of the American Meteorological Society* 102 (2021), E681–E699.

[244] Ioannis K. Bazionis and Pavlos S. Georgilakis. « Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research ». In: *Electricity* 2 (2021), pp. 13–47.

[245] Gwennaëlle Larvor et al. *MeteoNet, An Open Reference Weather Dataset by METEO-FRANCE. 2020.* 2020.

[246] Robin Marcille et al. « Gaussian mixture models for the optimal sparse sampling of offshore wind resource ». In: *Wind Energy Science Discussions* 2022 (2023), pp. 1–24.

[247] Takuya Akiba et al. « Optuna: A next-generation hyperparameter optimization framework ». In: 2019, pp. 2623–2631.

[248] James Bergstra et al. « Algorithms for hyper-parameter optimization ». In: *Advances in neural information processing systems* 24 (2011).

[249] Olivier Talagrand. « Evaluation of probabilistic prediction systems ». In: *Workshop Proceedings" Workshop on Predictability", 20-22 October 1997, ECMWF, Reading, UK.* 1999.

[250] Cristobal Gallego et al. « Influence of local wind speed and direction on wind power dynamics - Application to offshore very short-term forecasting ». In: *Applied Energy* 88.11 (2011), pp. 4087–4096.

[251] Peter Grönquist et al. « Deep learning for post-processing ensemble weather forecasts ». In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200092.

[252] Alexis A Mouche et al. « On the use of Doppler shift for sea surface wind retrieval from SAR ». In: *IEEE Transactions on Geoscience and Remote Sensing* 50.7 (2012), pp. 2901–2909.

[253] Zamani Ahmadreza et al. « Learning from data for wind–wave forecasting. » In: *Ocean Engineering* 35.10 (2008), pp. 953–62.

[254] Robin Marcille. *rmarcille/conve_stf_meteonet: v1.1.0.* Version review_version. 2024.

[255] Conor Durkan et al. *nflows: normalizing flows in PyTorch*. v0.14. 2020.

[256] Robin Marcille. *ConvE-STF dataset*. 2024.

[257] Hendrik L Tolman et al. « User manual and system documentation of WAVEWATCH III ». In: *Technical note, MMAB contribution* 276.220 (2009).

[258] Alice Dalphinet et al. « On the use of a high resolution wind forcing in the operational coastal wave model WW3 ». In: *14th International Workshop on Wave Hindcasting and Forecasting 5th Coastal Hazard Symposium*. 2015.

[259] Julija Tastu, Pierre Pinson, and Henrik Madsen. « Space-time trajectories of wind power generation: Parametrized precision matrices under a Gaussian copula approach ». In: *Modeling and stochastic learning for forecasting in high dimensions*. 2015, pp. 267–296.

[260] Mario Arrieta-Prieto and Kristen R Schell. « Spatio-temporal probabilistic forecasting of wind power for multiple farms: A copula-based hybrid model ». In: *International Journal of Forecasting* 38.1 (2022), pp. 300–320.

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Méthodes d'apprentissage pour la prévision de variables météo-océaniques : une opportunité pour l'optimisation des opérations de maintenance de l'éolien en mer.

Mot clés : Prévision probabiliste ; Caractérisation météo-océanique ; Éolien en mer ; Opérations de maintenance ; Apprentissage profond ; Mesures en mer.

Résumé : Les opérations de maintenance de l'éolien en mer sont sensibles aux incertitudes des prévision météo-océaniques. Les modèles de prévision numérique sont limités par leur coût de calcul pour l'estimation des incertitudes, ce qui pousse au développement de méthodes basées sur l'apprentissage profond.

L'importance des mesures in-situ en mer est mis en exergue par les résultats de cette thèse. Une méthode basée sur le clustering non supervisé de données de modèle numérique est proposée pour la définition d'un réseau de capteurs optimal pour la reconstruction de la ressource en vent.

Des méthodes d'apprentissage profond sont proposées pour la prévision météo-océaniques probabiliste. Nous montrons leur intérêt pour assimiler un grand nombre de données d'entrée. Une hypothèse de postérieur Gaussien et une approche générative utilisant les flots normalisants sont comparées. Ceux-ci permettent de relâcher les hypothèses sur la distribution postérieure, maintenant une capacité d'échantillonnage et de calcul exact de la vraisemblance.

Un cas d'étude réaliste est construit sur une zone représentative pour l'éolien en mer en France. Pour la prévision jointe du vent et des vagues, les propriétés non-Gaussiennes des flots normalisants se sont montrées bénéfiques à la calibration de la prévision. Un cadre d'évaluation représentatif des opérations en mer est proposé incluant la génération de scénarios et mesurant l'impact économique et le risque lié à la prise de décision. Nous montrons qu'il est crucial de prendre en compte le risque dans la sélection et l'évaluation des modèles de prévision.

Title: Learning-based forecasting of metocean variables: a path to maintenance operations optimization for offshore wind energy

Keywords: Probabilistic forecast; Metocean characterization; Offshore wind energy; Maintenance operations; Deep Learning; Offshore in-situ measurements.

Abstract: Offshore wind energy maintenance operations are highly sensitive to forecast uncertainty. Numerical weather prediction are limited by their computational cost for the uncertainty estimation and the update frequency, which is an argument for the development of data-driven methods.

The importance of offshore measurements is highlighted by the results. A method for designing an optimal sensors network is proposed using unsupervised clustering. This method has been used by the French weather service to define future networks of floating LIDAR for offshore wind.

Deep learning models for the joint probabilistic forecasting of metocean parameters are proposed. Their relevance for assimilating a large amount of input data is demonstrated. A Gaussian posterior and a generative approach using normalizing flows are compared. It is shown that the use of normalizing flows can relax any assumption on the shape of the forecast probability density while maintaining sampling and likelihood computation capabilities.

A real case study dataset is built on a relevant area for offshore wind. The probabilistic models are adapted for joint wind and wave forecasting, for which the non-Gaussian properties of the normalizing flows is beneficial for forecast reliability. An evaluation framework dedicated to offshore operations is proposed, including the generation of probabilistic scenarios and the measure of decision-making economic impact. It is shown that the search for an economic optimum in the probabilistic decision-making leads to higher risk during operations, and this should be taken into account for forecast selection and evaluation.