UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

Ocean Surface Current Uncertainty Quantification

via Drift Modeling and Spaceborne Retrieved Radial Velocities

Victor Cesar Martins de Aguiar

A dissertation for the degree of Philosophiae Doctor

November 2024

Omnia mutantur, nihil interit

“"We absolutely must leave room for doubt or there is no progress and there is no learning. There is no learning without having to pose a question. And a question requires doubt. People search for certainty. But there is no certainty. People are terrified — how can you live and not know?"”
–Richard. P. Feynman

Abstract

Ocean surface current is an essential ocean state variable in operational oceanography. They govern the movement of pollutants, microorganisms, heat, and salt across the ocean. Their dynamics shape marine ecosystems and influence climate patterns. Monitoring and accurately predicting ocean surface currents are essential for managing offshore pollution and its potential impacts on the shoreline, safeguarding marine biodiversity and ensuring maritime safety.

Ocean and atmospheric models are simplified representations of a highly complex, multi-scale flow system where approximations are unavoidable. Combined with the limited and inexact nature of available observations, predictions inherently present a degree of uncertainty. Ensemble modeling addresses these uncertainties by providing multiple future outcomes rather than a single solution. This approach helps quantify the likelihood of specific events, offering a more probabilistic understanding of the ocean behavior and enhancing decision-making in emergency situations, e.g, oil spills. Although widely employed and investigated by the atmospheric community, operational ensemble prediction systems are still emerging in oceanographic forecast centers.

In this thesis, we evaluate how ocean and atmospheric ensemble operational prediction systems address ocean surface current uncertainties through short-term drift modeling and ocean current retrieval using satellite observations. The first part of this work provides background information about ocean dynamics as a multi-scale problem, modeling, remote sensing, and uncertainty. The second part presents the three research papers produced during the Ph.D. project.

We introduce novel approaches to analyzing ensemble performance, with metrics specifically tailored for trajectory modeling, encompassing both delineated oil slicks and drifter trajectories. Our research sheds new light on the effects of horizontal resolution and wind forcing on short-term oil slick drift prediction using operational ensemble models. Furthermore, we present the first estimates of how wind field uncertainty affects ocean radial velocities, using remote sensing as the source of ocean current data.

By bridging remote sensing, numerical modeling, and quantitative ensemble performance metrics, the main results of the three research articles in this compendium showed that (Paper I) wind forcing has little impact on short-term trajectory prediction spread, and lower resolution ensemble models provide trajectory predictions as skillful as those provided by higher resolution deterministic models; (Paper II) model error directly impacts the rank histogram, leading to misinterpretations in assessing the model's spread; and (Paper III) the uncertainty in ocean radial velocity caused solely by the wind field can be as significant as that of the source data itself.

Sammendrag

Overflatestrøm som en nøkkelvariabel i operasjonell oseanografi. Via strømmene transporteres forurensning, mikroorganismer, varme og salt i havene. Havstrømmene er med på å forme marine økosystemer, samt en viktig del av klimasystemet. Overvåking og nøyaktig prediksjon av havstrømmene er viktig for blant annet å beskytte marin biodiversitet, håndtere og iverksette tiltak ved akutt forurensing, og for sikre maritime operasjoner.

Hav- og atmosfæremodeller er forenklede representasjoner av svært komplekse systemer med dynamiske prosesser på ulike tids- og romskalaer, og det er nødvendig med visse forenklinger for å beskrive systemene. Prediksjoner basert på modeller har dermed en iboende usikkerhet, som i tillegg kan påvirkes av begrensede eller unøyaktige observasjoner. Såkalte ensemblesystemer tar høyde for disse usikkerhetene ved å beskrive utfallsrommet fra flere modeller med litt forskjellig utgangspunkt i stedet for én enkelt løsning. Denne tilnærmingen bidrar til å kvantifisere sannsynligheten for spesifikke hendelser, og gir en mer probabilistisk forståelse av havets dynamikk. Dette kan forbedre beslutningsgrunnlaget i nødsituasjoner, som for eksempel ved oljesøl i havet. Innenfor meteorologi er ensemblesystemer allerede et godt etablert verktøy, mens det fortsatt er relativt lite utnytteti operasjonelle oseanografiske varslingssystemer.

I denne avhandlingen ser vi på hvordan operasjonelle ensemblebaserte prediksjonssystemer for hav- og atmosfære behandler usikkerhet i havoverflatestrømmer gjennom korttidsvarsling av drift samt henting av havstrømdata ved bruk av satellittobservasjoner. Den første delen av dette arbeidet gir bakgrunnsinformasjon om havdynamikk som et multi-skala problem, modellering, fjernmåling og usikkerhet. Den andre delen presenterer de tre forskningsartiklene som ble produsert i løpet av Ph.D.-prosjektet.

Vi introduserer nye tilnærminger til å analysere ensemblesystemets ytelse, i form av metrikker spesielt tilpasset trajektoriemodellering, som dekker både oljeflak og drivbaner. Vår forskning belyser effektene av horisontal oppløsning og vindpådrag på korttidsvarsler for oljedrift ved bruk av operasjonelle ensembler. Videre presenterer vi de første estimatene av hvordan usikkerheten

i vindfeltet påvirker havstrømmer, gjennom bruk av fjernmåling som kilde til havstrømdata.

Gjennom å forene fjernmåling, numerisk modellering og kvantitative metrikker, viser hovedresultatene fra de tre forskningsartiklene i dennee avhandlingen at (1) vindpådrag har liten innvirkning på spredningen i korttidsvarsler for trajektorier, og ensemblemodeller med lavere oppløsning gir trajektoriprediksjoner som er like treffsikre som de fra deterministiske modeller med høyere oppløsning; (2) modellfeil påvirker direkte ensemblespredningen, noe som fører til feiltolkninger; og (3) usikkerheten i havstrømobservasjoner som utelukkende skyldes vindfeltet, kan være av samme størrelsesorden som strømmene

Acknowledgements

I once believed that being a scientist was a solitary journey, but I couldn't have been more mistaken. Many people have supported and guided me along the way, making it possible for me to be here now, writing this document.

Firstly, I would like to thank my supervisors, Dr. Malin Johansson, Dr. Johannes Röhrs, and Dr. Torbjørn Eltoft, for all the knowledge, availability, and enthusiasm, for allowing my creativity to flourish and for the trust they had in me. I also thank the CIRFA project for funding my research and providing many opportunities that I could never imagine I would one day experience.

I would also like to express my gratitude to Dr. Harald Johnsen, Dr. Artem Moiseev, and Dr. Jean Rabault for our enthusiastic discussions and invaluable assistance. My sincere thanks go to Equinor and the European Space Agency for providing the essential datasets that supported my research. I am also deeply thankful to the CAGE group for the opportunity to join their research cruise.

A warm thank you to my colleagues in Tromsø, with whom I shared my days for almost three years. Although not working with sea ice, you all made me realize how it can be beautiful. My special thanks to Truls, for teaching me the beauties of the northern Norway dialect; To Henrik, for introducing me to the amusing German humor and punctuality (it's 11:57, still not lunch time); To Cat, for always trying to drag me to the climbing gym; to Ed, for his lame lunch topics; and to Laust, for showing me that I can understand Danish. Your friendship made this journey more delightful.

I also dedicate this work to my dear friends from MET Norway. In Bergen, Lars Robert Hole, and Knut-Frode Dagestad for the friendship and support they deposited on me even before I came to Norway. Martin, Ole, Joan, Leilane, and Torunn for their friendship and support over these five years. In Oslo, to Silje and Jozef for showing that distance means little when friendship means so much, Martina for the scientific discussions and turbofolk song sharing, and Edel and Marta for the technical support and goodwill. I am deeply grateful to you all.

I would like to thank my beloved family in Brazil, who have supported me for thirty years, showing that love and care can span across a whole ocean. Last but not least, a huge thank you to my sweetheart Hilde, who is by my side on the bright and dark days. Te amo.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation

The development of contemporary society – whether economically, historically, or culturally – has profound roots in the ocean. Promptly recognized as a resourceful environment, we established ourselves on its margins; societies flourished in the coastal environments, trading routes were established, and close links between overseas societies were created. Knowledge of the sea became a valuable legacy, passed through generations of mariners who learned to read the skies to anticipate storms, navigate complex currents, and map treacherous waters. The ocean, in many ways, became both a canvas for human ambition and a conduit for shared wisdom, binding societies through the common pursuit of discovery and survival on the open seas.

Approximately 40% of the current world population is estimated to reside within 100 km of the coastlines, and ocean-based goods and service industries profit USD 2.5 trillion per year worldwide [1]. In Norway, around 30% of the country's gross domestic product is directly related to the exploration of marine resources, specifically the exploration of crude oil and natural gas (28%, [2]) and fisheries (3%, [3]).

Given the significant human presence along coastlines and at sea, continuous monitoring of coastal regions and the continental shelf is of utmost importance to support onsea decision-making. The ultimate objective of operational oceanography is to establish an efficient process chain where (1) observational data is acquired and (2) processed at prediction centers, (3) assimilated into the numerical systems, and (4) model outputs (e.g. nowcasts, extended range forecasts, and hindcasts) are validated and delivered to users. Modeled products may be (5) further converted into value-added products for predicting aquaculture and fishery-related activities, coastal flooding, navigability support (e.g. BarentsWatch portal), oil spill preparedness planning and recovery, and search-and-rescue operations, to name a few. As deliverables are at the endpoint of such a complex workflow, they inherit all four preceding steps' errors and uncertainties. Except for point (3), this work covers all the remaining four steps, emphasizing number 4.

Predicting the state of the ocean is the main activity of operational oceanographic centers. For most of these daily activities, like in weather forecasting, the chaotic nature of ocean currents (or the wave field, sea surface temperature) may often be disregarded. Nevertheless, vis-à-vis the Great Storm of 1987 in the United Kingdom, uncertainties matter in extraordinary situations, and our numerical prediction systems should be able to foresee such events. These were the words of the BBC Anchor man Michael Fish the morning after the Great Storm to the on-duty forecaster Ian McCaskill [4]:

"*Well Ian, you chaps were a fat lot of good last night – if you can't forecast the worst storms for several centuries three hours before they happen, what are you doing!*."

As we shall see in detail later in Chapters 2 and 4, on-duty forecaster promptly noticed that the numerical systems did not present consistent skillful weather predictions, *i.e.*, sometimes predictions were accurate and at other times they largely disagreed from observations. The unfortunate misforecast example was the final reason for establishing an operational probabilistic forecast system in 1992 by the European Centre of Medium-Range Weather Forecast (ECMWF). Implementing similar systems by independent research centers has shown the benefits of ensemble modeling for extending weather predictability and as a necessary tool for risk analysis [4]. In the near future, the ensemble model is expected to become ECMWF's only global weather forecasting product.

Ensemble modeling has essentially three goals [5]: (I) increase the predictability horizon, (II) provide an indication of the reliability of the forecast, and (III) provide a quantitative overview of the forecast probability of a given variable of interest. By shifting the deterministic nature of geophysical flows to a probabilistic framework, ensemble modeling acknowledges the inherent imperfections in observational data used for initialization and model parameterizations. Physically meaningful perturbations are applied to generate multiple forecasts, with each ensemble member representing an equally probable scenario that reflects the potential future states of the ocean or atmosphere. This approach captures the uncertainty in the system, providing a more comprehensive view of possible outcomes. In retrospect to the Great Storm of 1987, warnings could have been issued four days in advance if an ensemble prediction system was available [6].

Extraordinary situations also happen at the sea daily. Around three thousand acute cases are reported to the Norwegian Joint Rescue Coordination Centre (Hovedredningssentralen) every year [1], of which almost 40% of them represent the deployment of emergency position-indicating locators, man overboard situations and missing boats. Although not all cases progress to emergency response operations, responders may ask for scientific support, and this often involves providing drift trajectory forecasts of the object of interest. Achieving successful outcomes in forecast modeling requires oceanic and atmospheric prediction systems to provide accurate information about future states within a short time frame. But just like the atmosphere, the ocean is also turbulent and chaotic. A single deterministic prediction might not represent the real ocean state, and its outputs may yield an unreliable estimation of the object of interest's drift and, consequently, misleading decisions.

1. https://www.hovedredningssentralen.no/dokumenter/

Ensemble modeling in oceanography has been studied for nearly three decades [7], and uncertainty quantification is a key factor to be considered in oil slick drift prediction[8, 9, 10]. Although still in its early stages, operational ensemble ocean modeling is rapidly advancing, driven by the clear demand for such tools in real-world marine applications. This work contributes to this demand by quantifying the uncertainty of ocean surface currents in an operational ocean ensemble prediction system through drift modeling. The uncertainty quantification is verified by employing several ensemble performance metrics using observed oil slicks and drifter trajectories as validation sets. Additionally, we also assess how uncertainties in the wind field from an atmospheric ensemble model impact the retrieval of ocean surface currents.

Figure 1.1 illustrates the main components of this work. Oil slicks and drifter instruments travel around the ocean surface due to wind stress and many other multi-scale phenomena (symbols). Drifters provide their position information, but we detect oil slicks using satellites. The operational ocean model (grid) aims to represent these ocean phenomena accurately so skillful predictions of, for example, oil slicks can be made. Ensemble members are generated using ensemble modeling, and we can quantify how certain we are about our predictions using this method. We focus here on ocean surface currents within the constant flux layer. Currents are intense and rapidly changing in this layer, and turbulence is a predominant feature [11].

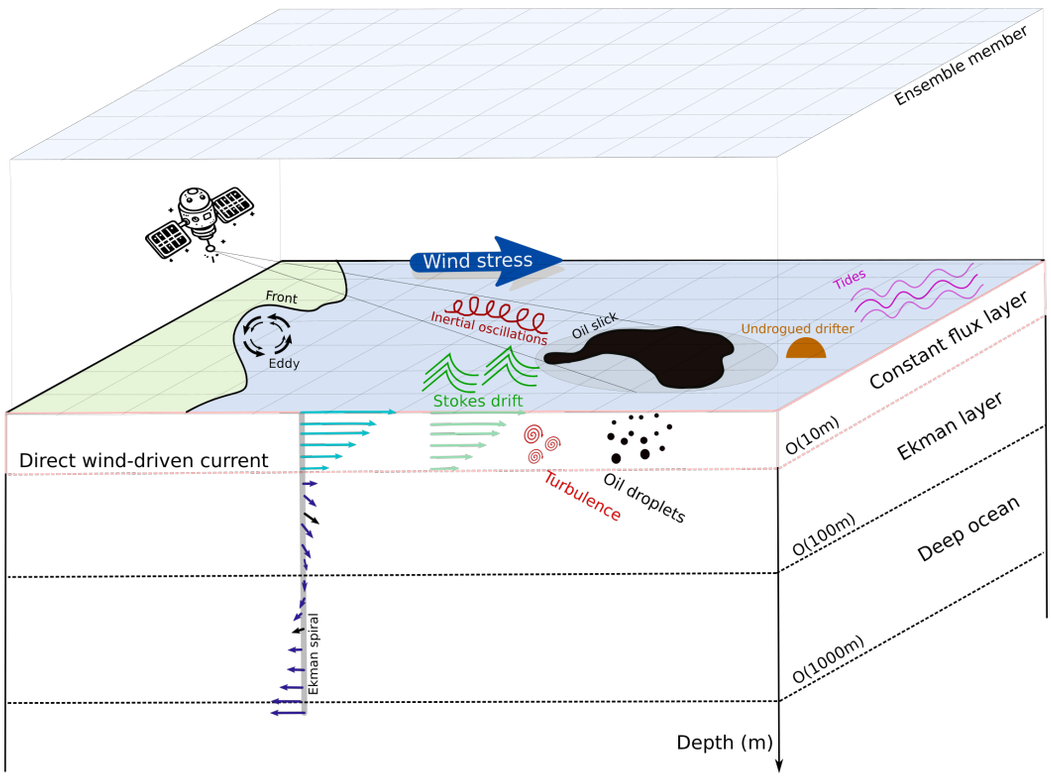

Figure 1.1: Illustration of the main components of this work. Different ocean phenomena are represented by symbols, and their corresponding names are displayed: eddy, front, inertial oscillation, tides, turbulence, and wind stress. Oil slick, observed with a satellite, and drifters (half-spheres), are two of our observation sources. The ocean model and its ensemble member (afloat) are displayed as grids. This work focuses on ocean surface currents within the constant flux layer. Adapted from [11].

1.2 Objectives

The scientific contributions in this study combine operational ocean and atmospheric ensemble prediction systems with in-situ and satellite image-based information to explore uncertainty in ocean surface currents through drift simulations and data retrieval. Three main objectives are defined as:

1. Investigate the short-term trajectory prediction capability and ensemble performance of an operational ocean ensemble prediction system using drifter data and remotely sensed oil slicks.

2. Explore the applicability of ensemble prediction systems for oil slick drift modeling.

3. Investigate the uncertainty in ocean surface currents derived from Synthetic Aperture Radar imagery using an operational atmospheric ensemble prediction system.

Determining uncertainties in ocean currents through various data sets, models, and regions allows us to enhance value chains and scientific services in operational centers.

1.3 Thesis Outline

This thesis is structured as follows: Chapter 2 describes the background information about the main ocean circulation features, including the areas investigated in the three research articles. The chapter also describes the main concepts of numerical modeling, both hydrodynamic and Lagrangian drift. The final part of the chapter highlights the importance of observations for model initialization and validation. Chapter 3 covers the remote sensing part of the thesis, focusing on synthetic aperture radar. It provides an overview of imagery acquisition, oil slick detection using the backscatter information, and how ocean currents can be retrieved using the Doppler centroid anomaly information. Chapter 4 covers the uncertainty topic, presenting the conceptual basis, ensemble generation, and ensemble performance metrics. Chapter 5 summarizes the research articles developed during the project. Chapters 6 to Chapter 8 contains the three articles included in this work. Finally, Chapter 9 provides the main conclusions of the studies and recommendations for future work.

/2

Ocean Surface Currents and Prediction

> “Let us, in this chapter, set out with the postulate that the sea, as well as the air, has its system of circulation and that this system, whatever it be, and wherever its channels lie, whether in the waters at or below the surface, is in obedience to physical laws.”
>
> — Matthew Fontaine Maury, *The Physical Geography of the Sea (1855)*

Ocean surface currents are the large-scale movements of water that flow at the top layer of the ocean, driven primarily by winds and affected by the Earth's rotation. These currents are a vital part of the Earth's climate system by redistributing heat from one region to another. Additionally, surface currents impact human activities such as navigation, fishing, and the movement of pollutants. Understanding ocean surface currents is essential for environmental management and maritime operations.

Processes at a wide range of spatial and temporal scales permeate the ocean surface, and understanding their dynamics is fundamental to predicting the transport of energy and pollutants using numerical models. This chapter introduces the general aspects of ocean surface currents, their major forcings, and their typical scales. It also introduces the four regions investigated in the three research articles. Since numerical modeling is an important component in this thesis, we provide an overview of the main concepts around ocean modeling, trajectory modeling, and the computational tools employed in Papers I and II. We finalize this chapter by describing the importance of observations for model initialization and validation.

2.1 General Aspects of Ocean Surface Currents

Our understanding of ocean dynamics has significantly advanced since Benjamin Franklin provided the first sketch of the Gulf Stream in the 1700s. Figure 2.1 shows the Gulf Stream sketch[2] overlaying the modeled ocean speed field output from NEMO-eNATL60[3], a high-resolution ocean model. It is evident that our modern understanding of the ocean and its complexity has significantly evolved from earlier perceptions.

2. https://www.loc.gov/resource/g9112g.ct000136/?r=0.144,-0.116,0.926,0.822,0
3. https://github.com/ocean-next/eNATL60

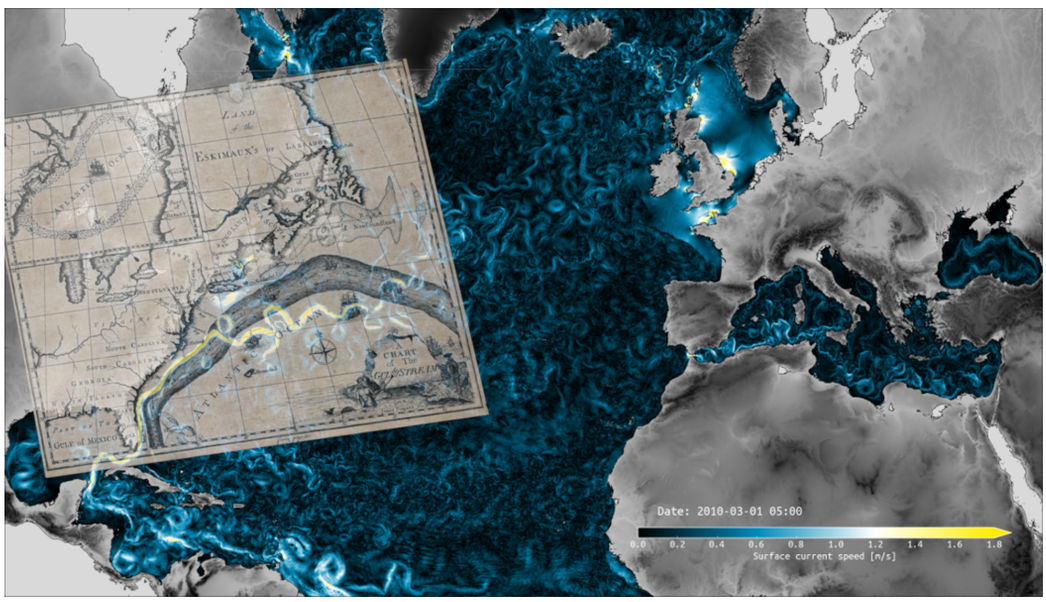

Figure 2.1: Modeled ocean current speed (m/s) from NEMO-eNATL60[3] overlayed by the Gulf Stream sketch created by Benjamin Franklin[2].

The contributions of HMS Endeavour and HMS Challenger provided an extraordinary overview of the ocean environment. Still, the sampling strategy was not designed – for they did not know – to capture transient phenomena at the sea. In 1909, Björn Helland-Hansen and Fridtjof Nansen identified '*vortex-movements*' and '*boundary-waves*' in Norwegian waters, which caused oscillations of isopycnals in the water column. However, the authors emphasized that undersampling limited their ability to draw firm conclusions about these phenomena [12]. Like Matthew Fountaine Maury and Henrik Mohn, they were fully aware that the onset of ocean currents ought to be elucidated through physical laws.

The rapid advances in observational and theoretical physical oceanography between the 1920s and 1970s showed that the ocean is neither static nor simply perennial, but rather a myriad of multi-scale phenomena happening concomitantly in a continuum spectrum (see Figure 2.2). At the one side of the spectra, the onset of large-scale [$O(10^6\text{m})$] ocean surface currents is primarily due to the kinetic energy input from winds, barotropic tides, and solar heating [13]. At the other end of the spectrum [$O(10^{-3}\text{m})$], energy dissipation occurs. The transference of energy is promoted by transient motions and turbulence in the so-called inertial range [13].

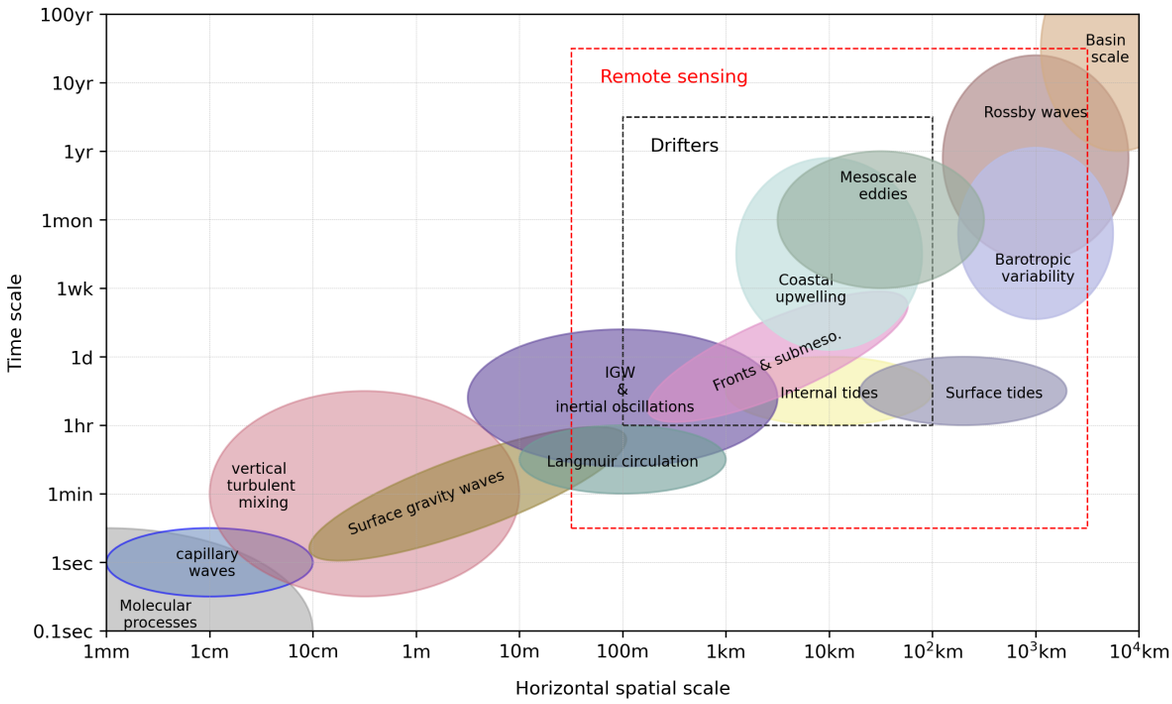

Figure 2.2: Temporal and horizontal spatial scales of several ocean processes. The red (black) dashed box represents the phenomena observed with remote sensing (drifters). Adapted from [14].

2.1.1 Wind Driven

Wind forcing represents the direct transference of momentum from the atmosphere to the ocean surface. Together with the influence of the Earth's rotation, wind stress exerted by the large-scale atmospheric circulation is the main driver of major ocean current systems and gyres [15, 16]. These, in turn, represent the low-frequency background where equally important wind-driven transient phenomena are embedded. Under the additional influence of viscosity, ocean surface currents do not necessarily follow a linear down-wind drift motion, but rather steered and oscillatory.

A classic example is given by the Ekman solution [17], where an ageostrophic surface current veered 45° to the right of the wind direction (in the north hemisphere) arises under steady wind conditions and constant eddy viscosity profile. The steady-state assumption holds for periods of 20 days or longer [18], but such conditions are nevertheless rarely observed. Accurate description and short-term prediction of ocean surface currents, and consequently the drift of objects, therefore requires a detailed description of time-dependent responses of the ocean relative to winds.

The Ekman problem, or more generally wind-driven ocean surface currents, falls into the field of boundary layer flows induced by shear turbulence. It can be shown that the Ekman solution also accepts a time-dependent solution, where the ageostrophic surface velocities (u′ and v′) forced by the wind stress (τ) are modified by the Earth rotation (f) and eddy viscosity (ν_E), with a given fluid density (ρ):

$$\frac{\partial u}{\partial t} - fv = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{1}{\rho}\frac{\partial \tau_x}{\partial z} \tag{2.1}$$

$$\frac{\partial v}{\partial t} + fu = -\frac{1}{\rho}\frac{\partial p}{\partial y} + \frac{1}{\rho}\frac{\partial \tau_y}{\partial z} \tag{2.2}$$

where p represents the horizontal pressure field, and (u,v) are the total horizontal velocity components as u = u′ + U_G and v = v′ + V_G. Invoking geostrophic balanced currents as lower boundary conditions (U_G and V_G), considering $\tau_x = \rho\nu_E\frac{\partial u}{\partial z}$, $\tau_y = \rho\nu_E\frac{\partial v}{\partial z}$, and setting the boundary conditions at the surface (u′, v′ = τ_x, τ_y at z = 0), one can find the following solution for u and v at the surface (z = 0):

$$u(z=0,t) = \underbrace{U_G}_{\text{Geostrophic}} + \underbrace{\frac{|\tau|}{2\rho\sqrt{\nu_E f}}[cos(-ft-\phi+\frac{\pi}{4})]}_{\text{Ageostrophic}} \tag{2.3}$$

$$v(z=0,t) = \underbrace{V_G}_{\text{Geostrophic}} + \underbrace{\frac{|\tau|}{2\rho\sqrt{\nu_E f}}[sin(-ft-\phi+\frac{\pi}{4})]}_{\text{Ageostrophic}} \tag{2.4}$$

Where ϕ is the phase angle of the stress vector. It is possible to see that the final solution is composed of the background geostrophic currents and an oscillating component. Figure 2.3 presents the solution at 80°N as a progressive vector diagram integrated over 24 hours.

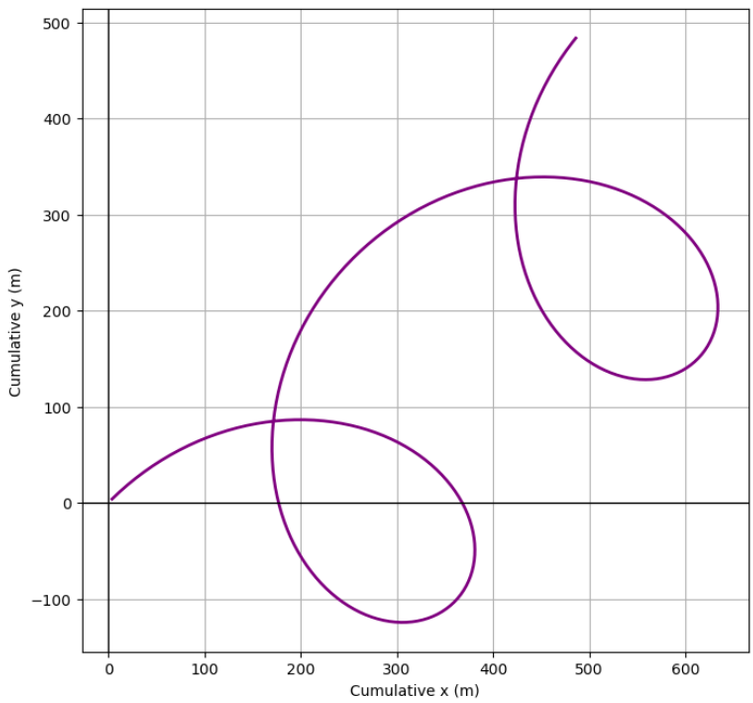

Figure 2.3: Progressive vector diagram for the solutions presented in Eq. 2.4 computed at 80°N and 24 hours of integration.

The cycloidal motion of the time-dependent solution remarkably resembles inertial oscillations often observed in drift trajectories and, to a lesser extent, in oil slicks. Inertial oscillations represent the free drift of the ocean currents, already set in motion, balanced by the Earth's rotation. In its most simple form, the analytical expressions for the time-dependent velocity components u(t) and v(t) are given by

$$u(t) = u_0 cos(ft) + v_0 sin(ft) \tag{2.5}$$

$$v(t) = v_0 cos(ft) - u_0 sin(ft) \tag{2.6}$$

Figure 2.4 shows three oil slicks detected in the Barents Sea (approx. 75°N and 31°E) using RADARSAT-2. The loops indicate inertial oscillations acting on the slick drift, and given the latitude, we can estimate that the oil was at the ocean surface for about 12 hours (local inertial period).

These two highly idealized examples show how ocean surface currents behave in a non-trivial manner when rotation is taken into account. The problem, and therefore the solutions, get substantially more difficult when three other

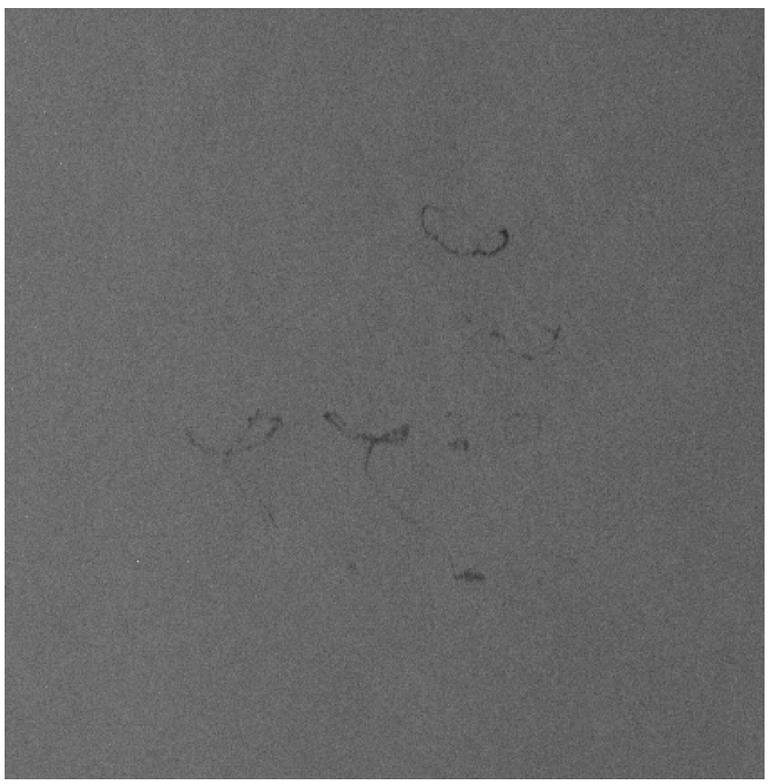

Figure 2.4: Oil slicks observed with RADARSAT-2 in the Barents Sea (approx. 75°N and 31°E), 2022. The cycloids are indications of inertial oscillations acting on the slick drift. Copyright raw data CSA, 2022, provided by NSC/KSAT 2022.

factors are included: depth-dependent turbulence, stratification, and horizontal gradients [19]. Turbulence ultimately describes how momentum is transmitted to the ocean, how it is modulated (dampened or enhanced) by the stratification, and how it dissipates. Predicting wind-driven ocean surface currents relies on the accurate representation of the ocean interior besides the atmospheric field and its history, and this can be provided by an ocean model.

2.1.2 Meso- and Submesocale

Stratification and horizontal gradients give rise to larger-scale, rotating phenomena: meanders and eddies. Figure 2.5 shows an algal bloom observed in August 2021, with Sentinel-3 Ocean and Land Colour Instrument (OLCI) in true color in the Barents Sea where the coexistence of eddies and filaments can be seen. Considered as mesoscale [$O(10^5)$m] and submesoscale features [$O(10^2$ - $10^4)$m], they are embedded in the inertial range and play also a central role on the dispersion of energy and solutes through turbulence. The seemingly linear and direct energy cascade from *big whirls so on to viscosity* is nevertheless far from simple, with evidence of bidirectional cross-scale energy cascading [13, 20, 21].

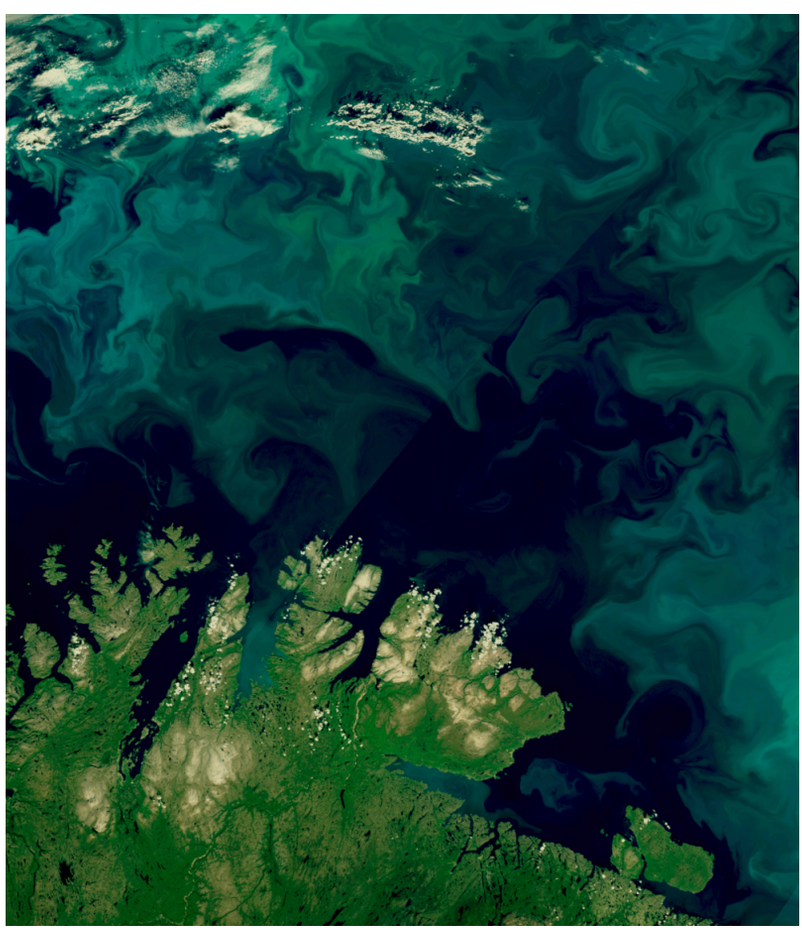

Figure 2.5: Sentinel-3 Ocean and Land Colour Instrument (OLCI) True Color acquired over the Barents Sea on 08.08.2021 during an algal bloom event. Copernicus Sentinel-3 data 2021.

Mesoscale eddies, often referred to as the "weather patterns" of the ocean,

are coherent rotating structures with horizontal scales of tens to hundreds of kilometers and lifespans ranging from weeks to several months. As the Earth's rotational effect depends on the latitude due to the Coriolis term, the radii of eddies decrease as we move polewards. These eddies can propagate long distances, transporting physical properties such as heat, salt, and nutrients within their cores, thereby influencing the biogeochemical and dynamical characteristics of the regions they traverse. The formation of mesoscale eddies is predominantly driven by the growth of instabilities in the background flow, with baroclinic and barotropic instabilities being the primary mechanisms.

Baroclinic instabilities occur in regions where there is a significant vertical shear in the horizontal currents and a pronounced density gradient (e.g., at fronts and in the presence of sloping isopycnals), leading to the conversion of available potential energy into eddy kinetic energy. Conversely, barotropic instabilities arise in areas with strong horizontal shear, where variations in velocity across the flow drive the transfer of kinetic energy from the mean flow to perturbations. These mechanisms are influenced by factors such as bathymetric features, wind-driven circulation, and interactions between large-scale currents, making the generation of mesoscale eddies a complex interplay of dynamical processes. Mesoscale eddies were reported to trap, detain, and deliver surface oil in an oil spill accident in the Black Sea [22]. During the Deepwater Horizon blowout in the Gulf of Mexico in 2010, locally and remotely generated mesoscale eddies impacted the distribution of oil following the incident [23, 24, 25]. Being an important component of the ocean surface circulation, accurate modeling of these phenomena is a key goal in operational ocean prediction systems.

These eddies are nevertheless considered two-dimensional rotating bodies, meaning that their angular velocity exceeds the vertical component in orders of magnitude. Even smaller phenomena are embedded in them, such as filaments and spirals, where the Earth's rotation force diminishes and equalizes the role of buoyancy and non-linear advection (in scaling arguments). These submesoscale features have life spans of hours to weeks and serve as a direct path to energy dissipation. Due to their three-dimensional properties, they typically possess enhanced vertical velocities and play an important role in vertical mixing in the ocean interior. Submesoscale motions are often associated with processes such as frontogenesis, mixed-layer turbulence, and the influence of surface gravity waves, which can lead to the formation of Langmuir circulation cells. Although mesoscale eddies control the pathways of pollutants at the sea surface, acting as a dynamic transport barrier, submesoscale motions can promote leakage of tracers outside these barriers [26]. A great review of submesoscale dynamics is provided by [27] and [28].

2.1.3 Turbulence and Mixing

The previous sections explained some important concepts about the energy input at the ocean and how different phenomena at multi-scale ranges emerge and impact the transport of objects at the ocean surface. This last part of the section describes the final path of our 'energy budget': dissipation.

The ocean is in a perpetual turbulent state [29], and it is closely related to the entangling existence of multi-scale phenomena and dispersion. A turbulent motion can not be formulated employing a single equation that describes its evolution, as it rather represents a random displacement from one point to another. Mixing, on the other hand, represents the combination of dispersion and diffusion [30]. If you (cautiously) pour milk into coffee and softly stir it, the initial liquid will create filaments and increase the gradient between the coffee and the milk, whereby diffusion becomes more efficient to homogenize them. Therefore turbulence enhances diffusion, promoting quicker homogenization than simply allowing the liquids to sit undisturbed (where molecular diffusivity is the only mechanism at work). Mixing is irreversible. In our energy budget example, the energy contained in the submesoscale filaments and eddies will be dissipated into heat, which can not be restored to the system.

Drifters and oil slick at the ocean surface are also subject to turbulence, the latter even into a 'three-dimensional' aspect. If one deploys drifters at the ocean, let's consider a pair separated by a small distance L, the rate at which these instruments depart from each other indicates how turbulent the environment is. If features smaller than this initial separation distance are predominant and energetic, the drifters disperse rapidly. Otherwise, they drift coherently for a longer period. The rate at this separation occurs is scale dependent, and Richardson found its diffusion K to be proportional to $K(L) \approx L^{4/3}$, with an average separation growth scaled in time as $\langle L^2 \rangle \approx t^3$.

Turbulence also occurs in the vertical plane, for instance, when waves break into an oil slick, promoting the entrainment of oil droplets within the ocean interior. Therefore, one must estimate both horizontal and vertical eddy diffusivities in a numerical model to represent the drift of oil slicks as the latter also impacts its horizontal trajectory [31, 32]. The theory beyond this point is intricate and beyond the scope of the current work, but the reader is referred to [30] and [33]. We finish this section by stressing that determining the eddy diffusivity value through observations requires a significant amount of data [34, 35], and its inclusion in numerical models requires solving a closure problem.

2.2 Regional Circulation

The previous section provided an overview of the different processes at the ocean surface on a multi-scale range. In practice, the impact of the different phenomena can manifest to a greater or lesser extent depending on the region of interest. We investigated four areas (Figure 2.6): the Skagerrak (red square, Paper III), the Norwegian Sea (green square, Paper I), the Barents Sea (blue square, Paper II), and the Fram Strait (blue square, Paper II). We shall introduce them now.

The Skagerrak Sea is located between Norway, Sweden, and Denmark. It is estimated that 40,000 to 60,000 large vessels sail in the region every year, with numerous shipwrecks still containing bulk oil and hazardous materials [36]. The general ocean surface circulation there is maintained by the relative fresh-water input from the Baltic Sea (BO), and this is balanced by the Norwegian Coastal Current (NCC) into the North Sea on its eastern flank. There, 67 oil and gas platforms are operational. The Jutland current, originating in the North Sea and flowing along the Danish coast, also contributes to the inflow into the Skagerrak. The NCC is the main circulation feature in the area. Originated as an extension of the Baltic inflow and driven by the local wind field [36], the current flows along the shelf break, off the Southern Norwegian tip, and heads polewards through the Norwegian Sea. Eddies and meanders are often observed during the NCC adjustment due to changing wind direction [37]. Observations and high-resolution numerical models are thus key elements for monitoring the region [38]. Paper III in this thesis explores the uncertainty of ocean current retrieval using remote sensing techniques.

In the Norwegian Sea, the NCC is deflected closer to the shore due to the influence of another prominent flow, the Norwegian Atlantic Slope Current (NwASC), which follows the steep contours of the continental slope. The complex topography in this region causes the NwASC to become unstable, leading to intense mesoscale activity characterized by the formation of eddies and other small-scale ocean features [39]. There, 23 oil and gas platforms are operational, with an increasing aquaculture sector onshore in the fjords and along the coastline, and fishing activities offshore. Due to its dynamic complexity, the presence of oil platforms, and natural resources, uncertainty quantification of short-term drift predictions in the region is necessary to ensure a fast response in case of an oil spill incident.

Both currents flow along the Norwegian coastline until around the latitude of 72°N, where a branch of the NwASC heads towards the west coast of Svalbard (West Spitsbergen Current, WSC) and a secondary ramification enters into the relatively shallow Barents Sea (North Cape Current, NCaC) together with the NCC. The encounter of the warmer waters carried by the NwASC, and the

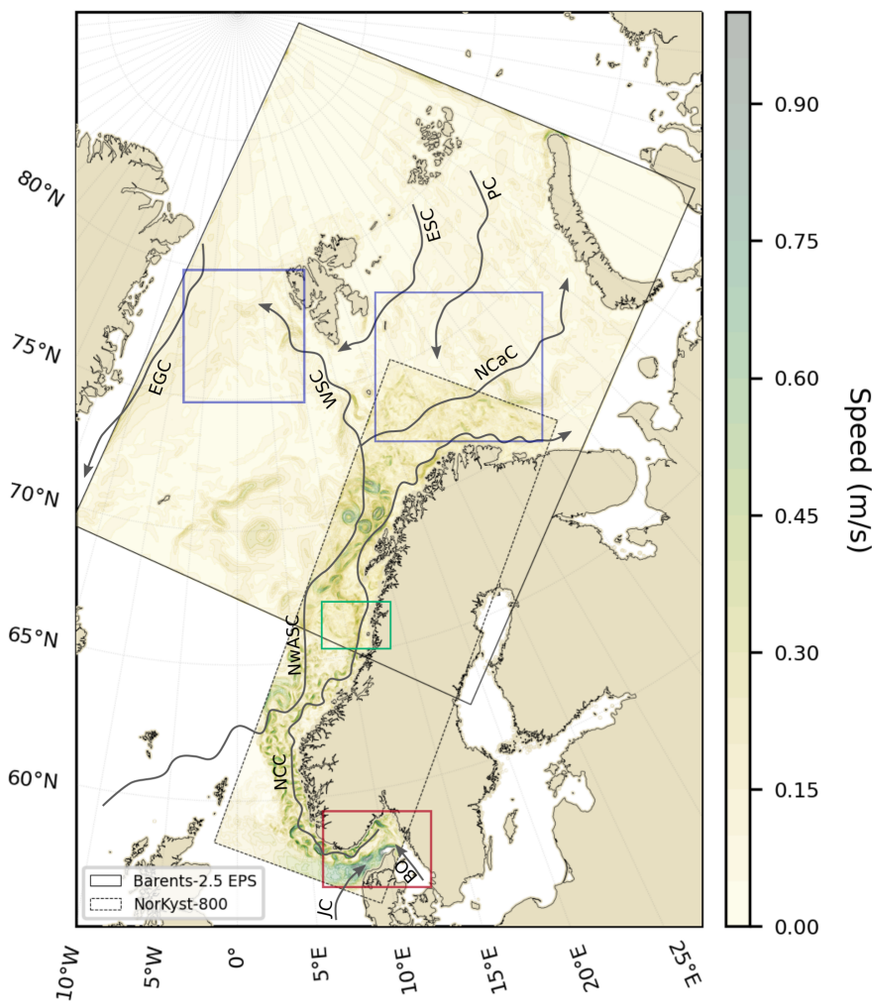

Figure 2.6: Representation of the regions of interest: Skagerrak (red square), Norwegian Sea (green square), Barents Sea, and Fram Strait (blue squares). The model domains of NorKyst-800 (black dashed domain) and Barents-2.5 EPS (black solid domain) are also displayed, with the background map representing ocean current speed (m/s) on 10.03.2024. Arrows depict the mean ocean currents in the region with their associated acronyms: EGC - East Greenland Current, WSC - West Spitsbergen Current, ESC - East Spitsbergen Current, PC - Persey Current, NCaC - North Cape Current, NwASC - Norwegian Atlantic Slope Current, NCC - Norwegian Coastal Current, BO - Baltic Outflow, and JC - Jutland Current.

cold currents (e.g. Persey Current, PC) from the northern Barents Sea create a hydrographically complex environment at the central portion of the sea. The inflow of warm Atlantic waters into the Barents Sea regulates the extent of

sea ice formation, making the region a focal point for climate research [40, 41, 42].

However, existing studies primarily address subsurface and long-term transport processes, leaving a gap in the understanding of short-term dynamics of surface currents. The region is also of economic interest, as it currently hosts two oil and gas exploration fields, while the Norwegian Offshore Directorate estimates that over 50% of the remaining unexplored petroleum reserves are located there[4]. The combination of extensive shipping traffic and resource exploration underscores the need for accurate, well-calibrated operational ocean models to support ongoing and future activities.

The Fram Strait is considered a gateway connecting the Arctic and the Atlantic by the WSC on its eastern side and the East Greenland Current (EGC) on the west. Differently from the Barents Sea, where wind-driven and tidal currents are the main forcings of surface currents, the Fram Strait is known by the all-year-round strong mesoscale activity [43], especially due to eddy shedding and detachment from the WSC towards the western side of the basin. These eddies have an average radius and lifetime of approximately 5 km and 10 days, respectively [44]. The East Spitsbergen Current originated at the upper-north Barents Sea, circulates around Svalbard, and inflows between its west coast and the WSC. As we will explore in the next chapter, accurately modeling small-scale phenomena requires a fine mesh grid [45], which remains a challenge in operational oceanography due to the high computational costs and the demand for high-resolution observations. The different dynamic characteristics between the Barents Sea and the Fram Strait motivated the evaluation of trajectory ensemble predictions in these regions.

4. https://www.norskpetroleum.no/en/petroleum-resources/resources-per-sea-area/

2.3 Ocean General Circulation Modeling

The previous section provided an overview of the ocean surface currents, their scales, and how some solutions can be found analytically. The Navier-Stokes equation is a set of prognostic partial differential equations, meaning that predictions for variables of interest can be made given initial and boundary conditions. The analytical solutions allow us to investigate a range of phenomena, from wind-driven ocean-basin scale ocean currents to geostrophic turbulence and beyond, but these are compartmentalized sets. Integrating all phenomena within the kinetic energy spectrum has still no analytical solution.

As weather prediction follows similar principles, Cleveland Abbe [46] and Vilhelm Bjerknes [47] idealized the basic ideas of numerical weather prediction already in the early 20th century. Lewis Fry Richardson developed further the concepts and his work culminated in the now recognized ground-breaking book, the "*Weather Prediction by Numerical Process*" published in 1922 [48]. Both atmospheric and ocean models have drastically advanced since the first weather numerical forecast set by Jule Charney in 1950 [49]. A series of advances in computational resources and theoretical work supported by observation systems allows us to perform ocean and weather predictions for the whole globe in a feasible period.

Ocean circulation models rely on the numerical discretization of the primitive equations that govern momentum, mass, heat, and salinity conservation. This discretization process involves approximating these partial differential equations using finite differences, applying incremental steps in both time (t) and space (x, y, z). By transitioning from a continuous to a discretized space, solutions are computed at specific grid points, with the increments determining where variables like velocity and temperature are calculated. We refer to this type of numerical modeling as performed in an Eulerian frame of reference.

Horizontal grids, which are essential for defining spatial structure, can be categorized into two main types based on grid-point spacing: regular and irregular grids. Regular grids feature equal spacing between grid points, providing uniform horizontal resolution across the model domain, while irregular grids use non-uniform spacing, allowing for finer resolution in regions of interest, such as coastal areas. The choice of grid spacing directly influences the horizontal resolution of the model and, thus, determines the smallest scales of phenomena that can be captured accurately by the model. Higher resolution grids, for instance, enable the simulation of finer-scale processes like coastal eddies and fronts, while lower-resolution grids are typically limited to large-scale dynamics, such as basin-wide currents and planetary waves. The effective resolution of an ocean model is between 4 [50] and 10 [51] grid points, meaning that for a

global ocean model with 25 km x 25 km horizontal resolution, motions smaller than 200 km (e.g. mesoscale and submesoscale) are poorly resolved. In other words, the presence of phenomena smaller than 200 km must be taken into account through parameterizations as they are not explicitly resolved by the model.

As described in the previous section, stratification and vertical movements also influence the transport of substances at the ocean surface. Therefore, ocean models must also represent the ocean interior and how overlying layers interact with each other. There are three main types of vertical grids used in ocean modeling: z-level (geopotential) grids, terrain-following (sigma) grids, and isopycnal (density-following) grids. Z-level grids divide the water column into horizontal layers of constant depth, offering simplicity and efficiency but struggling to represent steep topography accurately. Terrain-following (sigma) grids, on the other hand, adapt their layers to the shape of the seafloor, allowing for a more accurate representation of bottom boundary processes and sloping terrains, although they can introduce numerical artifacts in regions with strong stratification. Lastly, isopycnal grids align layers along surfaces of constant density, making them ideal for simulating large-scale ocean circulation and the movement of water masses with minimal numerical diffusion. However, they may be less effective in capturing dynamics near the surface or mixed-layer processes.

Another basic ingredient when developing ocean models is defining the initial and boundary conditions. The initial conditions define the initial state of the model, in order words, values of a given variable prescribed on each (x,y,z) point at t = 0. In operational systems, initial conditions are typically obtained from previous forecasts, a process known as a warm start, and then are updated through data assimilation techniques that incorporate real-time observations to improve accuracy. When previous forecasts are not available, initialization is performed using alternative data sources, such as climatologies, reanalysis products, hindcasts, or any other relevant fields that represent the desired state. After setting the initial conditions, the model is run (or integrated) over a specified period until it reaches a stable, balanced state, a process known as model spin-up. This spin-up phase ensures that any artificial transients arising from the initial state dissipate, allowing the model to establish realistic dynamics.

Boundary conditions, on the other hand, define how variables behave at the boundary of the domain, including surface, bottom, and lateral. Surface conditions dictate how e.g. heat, moisture, and momentum are transferred from the atmosphere to the ocean surface. These conditions are essential for driving surface currents and influencing processes such as evaporation, vertical mixing, and stratification. Bottom boundary conditions describe interactions with the

seafloor, including bottom friction, sediment transport, and topographical influences. These conditions can be further categorized into slippery or non-slippery conditions: slippery bottom conditions allow for frictionless flow parallel to the seafloor, minimizing drag, while non-slippery (or no-slip) conditions enforce a zero velocity at the bottom, making them crucial for simulating the effects of turbulence and bottom shear stresses on flow dynamics.

Lateral boundary conditions are applied at the edges of the model domain and can vary based on the desired behavior of the simulation. Open boundary conditions enable momentum and tracers to flow freely in and out of the domain, which is ideal for modeling interactions with adjacent basins or representing connections to the open ocean. These can be implemented using methods like radiation or relaxation schemes to prevent artificial reflections of waves or currents at the boundary. Reflective boundary conditions, on the other hand, act as solid walls that prevent any flow across the boundary, useful for closed basins or idealized experiments where confinement is desired. In addition, slippery or non-slippery reflective conditions can be applied depending on whether the boundary allows tangential flow or enforces zero velocity, respectively. More complex approaches can include periodic boundary conditions, where flow is reflected on the boundary domain, or sponge layers, which gradually dampen disturbances to minimize numerical artifacts near the boundaries. The careful selection of these boundary conditions is crucial for ensuring that the model behaves realistically and accurately simulates the physical exchanges and dynamics at the limits of the domain.

A few remarks should be nevertheless highlighted. The numerical discretization of equations must satisfy three fundamental criteria: consistency, stability, and convergence. Consistency ensures that as the grid spacing and time step approach zero, the discrete equations accurately represent the original differential equations. Stability requires that numerical errors do not grow uncontrollably over time. Convergence guarantees that, with sufficiently fine resolution and smaller time steps, the solution of the discrete equations approaches the true solution of the continuous equations. Additionally, the relationship between time step and grid spacing must adhere to the Courant-Friedrichs-Lewy (CFL) condition, which dictates that the time step must be small enough to ensure that the information from wave propagation does not exceed one grid cell per time step. This condition is crucial to maintaining numerical stability and preventing unrealistic oscillations or errors. For these reasons, different numerical approximations are employed for different terms of the primitive equation at different time steps.

Finally, because numerical models are only approximations of the real ocean, they rely heavily on parameterizations to represent small-scale processes that cannot be resolved directly. Parameterizations are simplified equations or

empirical formulas used to approximate phenomena smaller than two grid points, known as subgrid-scale processes such as turbulence, mixing, convection, and air-sea interactions. Physically, turbulence is represented as deviations from the average. Rather than solving the primitive equations, one has to solve an expanded version of it, the Raynolds-Averaged Navier-Stokes (RANS) equations, where its terms are split into an average and fluctuations. If one tries to solve it, extra unknown terms appear, leading to an undetermined system. Closure in this sense approximates the unknown terms to other known physically meaningful terms, usually through relations, and must also obey some other criteria [52]. Defining its closure scheme and magnitude can significantly impact oil drift simulations [53, 54] and vary spatially [55].

Forecasting at the Norwegian Meteorological Institute

This brief overview of the numerical discretization of partial differential equations sets the fundamental basis of current atmospheric and ocean models. This section aims to describe the state-of-the-art prediction systems, focusing on the Regional Ocean Modeling System [ROMS 56]. ROMS is a hydrostatic ocean model designed to simulate regional-scale ocean circulation. It uses a split-explicit time-stepping algorithm to solve the three-dimensional RANS equations, which are formulated on an Arakawa-C grid. It employs a staggered finite-difference approach with terrain-following sigma vertical coordinates, allowing high resolution in both horizontal and vertical dimensions. Its horizontal pressure gradient scheme minimizes errors in steep bathymetry, making it suitable for complex topographies like continental shelves and coastal regions. The model utilizes a variety of advection schemes, such as third-order upstream and fourth-order centered options, to resolve scalar and momentum transport, while the vertical mixing is handled using advanced closure models like K-profile-parameterization (KPP) or the Generic Length Scale (GLS) turbulence models [57].

The Norwegian Meteorological Institute (MET-Norway) uses ROMS as the setup which in-house regional ocean models (e.g. NorKyst-800, NorShelf, and Barents-2.5 EPS) are built upon. Each model has different spatial resolutions (0.8 km, 2.4 km, and 2.5 km, respectively), and different configurations. NorKyst-800 [38] was developed in a joint effort between MET-Norway and the Institute of Marine Research (IMR) to support private and research activities along the Norwegian coast and continental shelf. The model provides a single forecast in time, i.e., a deterministic model.

The Barents-2.5 EPS is an ensemble prediction system (EPS) developed and maintained at MET-Norway. Its numerical grid covers the Norwegian Sea, part of the North Atlantic, the Barents Sea, and the Fram Strait. The EPS represents one

of the few operational regional ensemble models currently available; another is the TOPAZ model developed by the Nansen Environmental and Remote Sensing Center (NERSC). Barents-2.5 EPS is a singular tool for ocean current uncertainty quantification, and it was the main model under investigation in Papers I and II. In Paper I, trajectory simulations forced with Barents-2.5 EPS and NorKyst-800 were intercompared to verify whether the higher resolution, deterministic model (NorKyst-800) can provide more skillful results relative to the lower-resolution EPS. The EPS performance was evaluated against observations using several metrics (some described in Section 4.2.2) in both research articles.

2.4 Trajectory Modeling

A numerical modeling set aims to reproduce and predict the environmental variables in a numerical grid, often performed in an Eulerian framework. For drifting objects at sea, such as oil slicks and drifters, the most natural frame of choice is the one where the object of interest follows the underlying flow, i.e., the Lagrangian frame of reference. This section introduces the concepts of Lagrangian trajectory modeling and the link between the two frameworks.

A particle P released in a continuum media and subject to external forces will, after some time, trace out a trajectory $\mathbf{X}^P(\mathrm{t})$ in space. As trajectories are dependent on their location at the initial time, $\mathbf{X}^P(\mathbf{a}, \mathrm{t})$ where $\mathbf{a} = \mathbf{X}^P(\mathrm{t} = \mathrm{t}_0)$. For a discrete trajectory, the along-track Lagrangian velocity ($\mathbf{V}_L$) between any two consecutive data points is therefore given by

$$\mathbf{V}_L = \left.\frac{d\mathbf{X}^P(\mathbf{a}, t)}{dt}\right|_{\mathbf{a}} \tag{2.7}$$

As Eulerian and Lagrangian frameworks describe the same fluid, a given property F (e.g. temperature, salinity, or velocity) must have the same value in both descriptions at the same position **a** and time t. In other words,

$$\underbrace{F[\mathbf{X}^P(\mathbf{a}, t)]}_{\text{Lagrangian}} = \underbrace{F(\mathbf{x}, t)}_{\text{Eulerian}} \tag{2.8}$$

where **x** represents the three-dimensional cartesian coordinates X_1, X_2, and X_3. If one observer seeks to express the rate of change of F at the fixed point **a**, Eq. 2.8 can be written as:

$$\frac{\partial}{\partial t}F[\mathbf{X}^P(\mathbf{a},t)]\Big|_{\mathbf{a}} = \frac{\partial F}{\partial t} + \frac{\partial F}{\partial X_i}v_i = \frac{\partial F}{\partial t} + (\nabla F)\cdot\mathbf{v} \equiv \frac{D}{Dt}F(\mathbf{x},t) \tag{2.9}$$

For i = 1, 2, 3. Put into words, the rate of change of F in the Eulerian frame of reference is expressed by a local, time-dependent term ($\frac{\partial F}{\partial t}$), plus an advective term ($(\nabla F)\cdot\mathbf{v}$) representing the advection of F due to e.g., currents from a different location [58].

Observations of the drift of floating objects have regarded that $\mathbf{V}_L$ can be approximated at first order as depending solely on the wind speed [59], resulting in the long-standing 3% rule-of-thumb. Despite physically sounding by also invoking the air-water drag force balance [60], the approach was somehow discredited due to its oversimplification and neglect of the ocean currents and waves. More recent investigations have shown that using just the wind field as forcing does not produce skillful drift forecasts [61], and our Paper I revealed that for short-term predictions, its contribution to drift uncertainty was minimal.

For an object residing at the ocean surface, it is currently well accepted that the velocity of an object at the ocean surface is composed by a linear combination of ocean current $\mathbf{V}_O$, a fraction of the wind ($\alpha\mathbf{V}_W$) and Stokes drift velocities ($\mathbf{V}_S$) as in:

$$\mathbf{V}_L = \mathbf{V}_O + (\alpha\mathbf{V}_W) + \mathbf{V}_S \tag{2.10}$$

When performing a trajectory modeling prediction, the input fields are often obtained from geophysical models. In other words, virtual particles are released in a numerical grid and atmospheric-ocean-wave modeled velocity fields are interpolated in space and time to their position. If we define the velocity vector in the Eulerian frame of reference as $\mathbf{v}$(x,t), the position update from t to t+Δt is then given by [62]:

$$\underbrace{\mathbf{X}^P(t+\Delta t)}_{\text{Next position}} = \underbrace{\mathbf{X}^P(t)}_{\text{Current position}} + \underbrace{\int_t^{t+\Delta t}\mathbf{v}(\mathbf{x}(\tau),\tau)d\tau}_{\text{Eulerian velocity}} \tag{2.11}$$

Where τ is time in the Eulerian frame. Equation 2.11 takes the form of a numerical solution of an ordinary differential equation, and different schemes can be employed to solve the integration problem. We shall briefly discuss three of them, namely the forward Euler, improved Euler, and Runge-Kutta

4th order schemes. For simplicity, consider the following ordinary differential equation and initial condition:

$$\begin{aligned} y'(t) &= f(t, y(t)), \\ y(0) &= y_0 \end{aligned} \tag{2.12}$$

Although the solution of the ordinary differential equation is unknown, we can use Taylor expansion to approximate y(t) at a given point $y(t_n) = y_n$ as follows:

$$y(t) = y(t_n) + y'(t_n)(t - t_n) + \frac{y''(t_n)}{2!}(t - t_n)^2 + \frac{y'''(t_n)}{3!}(t - t_0)^3 \ldots \tag{2.13}$$

where $y^{(n)}$ represents the order of the derivative. Considering $\Delta t = (t\text{-}t_n)$, Eq. 2.13 can be rewritten as:

$$y(t_n + \Delta t) = y(t_n) + y'(t_n)(\Delta t) + \frac{y''(t_n)}{2!}(\Delta t)^2 + \frac{y'''(t_n)}{3!}(\Delta t)^3 \ldots \tag{2.14}$$

Truncating the expansion at the first derivative, we get an expression fairly similar to our initial problem in Eq. 2.11:

$$\begin{aligned} y(t_n + \Delta t) &\approx y(t_n) + y'(t_n)(\Delta t) + R, \\ y(t_n + \Delta t) &\approx y(t_n) + f(t_n, y_n)(\Delta t) + R \end{aligned} \tag{2.15}$$

Where R is the highest order of truncation error, in this case $O(\Delta t)^2$. This is known as the forward Euler's method, a first-order explicit numerical scheme for solving ODEs. Notice this approach is equivalent to using the quadrature rule:

$$\begin{aligned} y(t + \Delta t) - y(t_n) = \int_{t_n}^{t} y'(t)dt &= \int_{t_n}^{t} f(t, y(t))dt, \\ y(t + \Delta t) - y(t_n) \approx (t - t_n)f(t_n, y(t_n) + R &\equiv (\Delta t)f(t_n, y_n) + R, \\ y(t + \Delta t) &\approx y(t_n) + f(t_n, y_n)(\Delta t) + R \end{aligned} \tag{2.16}$$

Better approximations for the solution can be obtained for our initial problem in Eq. 2.11 by refining the solution for the integral, for instance using its average value between the two endpoints f(t_n, y_n) and f($t_{t+\Delta t}$, $y_{t+\Delta t}$). This is known as the improved Euler's method (or Heun's method), and it involves refining our integration method by employing the trapezoidal rule:

$$y(t+\Delta t) \approx y(t_n) + \frac{\Delta t}{2}(f(t_{t+\Delta t}, y_{t+\Delta t}), f(t_n, y_n)) + R \tag{2.17}$$

Notice that to advance the equation in time, the formula also requires knowing the function already one step ahead, f($t_{t+\Delta t}$, $y_{t+\Delta t}$), which transforms it into an implicit method. We can nevertheless use our known forward formula (Eq. 2.16) to substitute it into $y_{t+\Delta t}$ and get back again an explicit integration:

$$y(t+\Delta t) \approx y(t_n) + \frac{\Delta t}{2}\left(f(t_{t+\Delta t}, y(t_n) + f(t_n, y_n)(\Delta t)), f(t_n, y_n)\right) + R \tag{2.18}$$

The key advantage of using this method is its reduced error ($O(\Delta t)^3$) relative to the forward Euler's method. Our final method is widely used in Lagrangian particle models, and it was employed in both Papers I and II to perform the simulations. The Runge-Kutta method is a class of implicit and explicit methods, to which both Euler's methods belong, where weighted average estimations of the function are taken in the $t_n < t < t+\Delta t$ interval. An explicit fourth-order scheme, or the 4th order Runge-Kutta (RK4), thus estimates four points within this time interval:

$$y(t+\Delta t) = y(t_n) + (\Delta t)\left(\frac{k_{n1} + 2k_{n2} + 2k_{n3} + k_{n4}}{6}\right) \tag{2.19}$$

Where the coefficients are given by [63]:

$$\begin{aligned}
k_{n1} &= f\left(t_n, y_n\right) \\
k_{n2} &= f\left(t_n + \frac{\Delta t}{2}, y_n + \frac{\Delta t}{2}k_{n1}\right), \\
k_{n3} &= f\left(t_n + \frac{\Delta t}{2}, y_n + \frac{\Delta t}{2}k_{n2}\right), \\
k_{n4} &= f\left(t_n + \Delta t, y_n + \Delta t k_{n3}\right)
\end{aligned}$$

The derivation is long for the scope of this work, so we reserve to mention that the approximation error is of the order $O(\Delta t)^4$ [63]. The choice of the

numerical discretization scheme might introduce large errors in the estimation of the future position of the particle, as exemplified by the initial condition problem [63]:

$$\begin{aligned} y'(t) &= 1 - t + 4y, \\ y(0) &= 1 \end{aligned} \tag{2.20}$$

Which presents the analytical solution y(t) = $\frac{1}{4}t$ - $\frac{3}{16}$ + $\frac{19}{16}e^{4t}$. The three methods are shown in Figure 2.7, with the analytical solution displayed as a solid black line (True) and numerical solutions computed with $\Delta t = 0.1$.

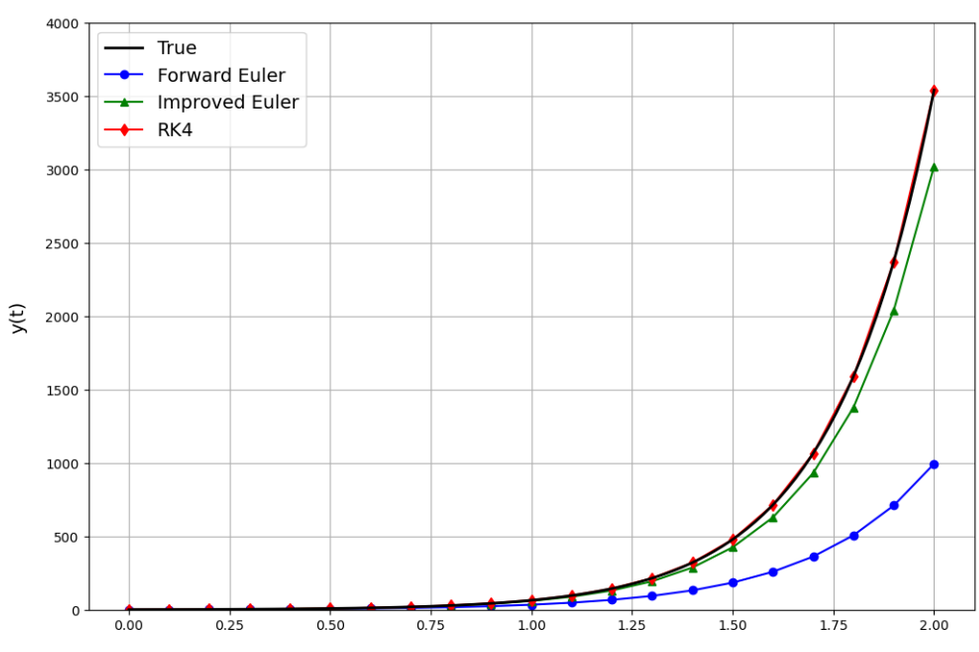

Figure 2.7: Numerical and analytical solutions (black line) for Eq. 2.20. Blue dotted line: Forward Euler's method; Green line with triangle: Improved Euler's method, and red dotted line: RK4.

One can see that the forward Euler's method underestimates the true solution. Many other schemes involving multi-step and variable time steps approaches can be employed, but RK4 generally provides sufficient accuracy and efficiency for Lagrangian particle tracking for simulations not involving long integration times (months to years) [64]. The RK4 numerical scheme was employed in our set of simulations of Papers I and II.

Online and offline modes

The prediction problem posed at the beginning of this chapter can be assumed as integrating a set of virtual particles in time forced by background velocity fields (ocean currents, winds, and Stokes drift) usually defined in an Eulerian grid and interpolated in space and time onto the virtual particles. The procedure can be carried out both in online and offline modes.

In Lagrangian particle tracking, the distinction between online and offline modes hinges on the timing and computational strategy used for advecting particles through a flow field. In the online mode, particles are tracked concurrently with the flow simulation, using the velocity field at each simulation time step to update particle positions. This approach is memory efficient, as it eliminates the need to store the entire flow dataset over time, making it advantageous for high-resolution simulations with limited storage resources. However, the tight coupling with the solver can increase computational demands and restrict flexibility, as re-tracking particles or exploring different initial configurations requires re-running the entire simulation. In contrast, offline mode separates the particle tracking from the flow computation, performing it as a post-processing step using stored velocity fields. This enables extensive flexibility for re-initializing particles, testing multiple configurations, and conducting detailed analyses.

Another advantage of the offline mode, overall for oil slick drift predictions on demand, is that one does not depend on a specific model but can rather make use of, in theory, any operational system available at the region and time of interest given that a suitable reader built in the offline software exists. Several offline packages tailored to oil slick drift and weathering are currently available [see 10], but we shall briefly describe Opendrift, a Python-based open-source package [65] used in Papers I and II.

OpenDrift is a generic modular framework designed to be employed in different applications (ocean drift, oil drift, iceberg drift). Geophysical models and observations in different formats and projections can be used as forcing, and outputs are stored in netCDF files. The driving advection equation, here Eq. 2.10, uses the provided modeled fields, and the position of the virtual particles is updated using either the forward Euler's scheme or RK4 described above. In Papers I and II, RK4 was used as a numerical scheme and readers were modified to include ensemble fields. Opendrift's oil drift and weathering module, OpenOil [32], uses the ADIOS Oil Database [66] to obtain the physical and chemical properties of more than one thousand different oil types.

Oil undergoes a series of physical (evaporation, emulsification, dispersion, and dissolution) and chemical processes (oxidation) when released into the marine

environment. Additionally, other transport pathways at the surface (spreading), and vertical migration (entrainment), can also affect the horizontal advection of oil slicks [32]. Although OpenOil resolves some of these processes, including turbulent mixing, we considered them to a lesser extent in Paper I for the proposed numerical experiment because they would increase the number of degrees of freedom in the uncertainty estimation and deviate from our main goal, *i.e.*, uncertainty estimation of ocean surface currents. Similarly, we did not fine-tune the horizontal eddy diffusivity in Papers I and II.

2.5 Prediction and Validation

The preceding sections introduced the concepts of geophysical fluid dynamics from analytical and numerical perspectives. We have seen so far the complexity of translating a multi-scale, rotating, turbulent, and stratified fluid system into a discretized Eulerian model, and ultimately into trajectory predictions through Lagrangian modeling. Whether solving the equations analytically or numerically, initial conditions must be provided so the prognostic equations, i.e., those time-dependent terms, can perform the predictions. Model initialization, and therefore forecasting, rely on the use of observations.

Although well-known since the early 20th century, when Richardson wrote his book, initialization and forecasting were done manually. For the laboring work, Richardson himself worked for six weeks on his forecast [48], weather prediction was a time-demanding process. For this reason, using analogues was a quick way to initialize models. Analogues are similar representations of the current ocean or atmospheric state at previous times. It assumes that the atmospheric dynamics is, to a certain extent, periodic. Therefore, for predicting tomorrow's weather, one could search in the archive of past observations that resemble today's state and its evolution would be similar to what was observed back then. As we shall see in Chapter 4, it was promptly noticed that nonlinearities are sensitive to errors in the initial conditions [67], acting mainly on the deterioration of the predictions' skill [68].

In operational terms, model (re)initialization is achieved through data assimilation. During this process, observations and forecast variables are integrated by solving a least-squares optimization, with the model's equations serving as constraints [69]. Various data assimilation techniques, such as 3D-Var, 4D-Var, and the Ensemble Kalman Filter, can be utilized in a prediction system. These methods differ in how they minimize the forecast-observation misfit by finding the minimum of an objective cost function. The minimization problem incorporates the probability density functions of both the model forecast and observations, particularly focusing on their first two statistical moments

(mean, variance, and covariance). Of particular relevance to this work is the evolution of error covariances, which governs how uncertainty is propagated in an ensemble prediction system.

Observations are also employed at another important step in modeling, namely model validation. Model validation concerns assessing the prediction quality relative to a set of observations, ideally independent. These can be further divided into in-situ, drifters, and remotely sensed observations.

Drifters represent a natural way of tracking ocean surface and sub-surface currents. Typically, a drifter consists of a waterproof container that houses a GPS device, which transmits its position via satellite. Inside the container, batteries and microelectronics power the system, enabling a sampling rate of typically every hour and lasting from months to years. Whether drogued or undrogued, they are used in a range of applications to improve the trajectory drift modeling, e.g., to represent the drift of oil slicks. The drifters, therefore, serve as an indirect ground truth of the model's capabilities and deficiencies in predicting the pollution displacement at the ocean. This can be performed, e.g., by initializing the forecast at a given position and time reported by the drifter device, performing the prediction over a time span, and verifying the forecast and observed positions using error metrics [e.g. 70, 71, 72]. Combined with remote sensing and hydrodynamic modeling, drift trajectories can also provide fine-scale details of complex dynamic current systems and their impact on oil residence time [73], dynamic barriers acting on the oil slicks drift [74], and differences between mineral and biogenic oil resurfacing [31]. A great review about the use of drifters can be found in [75]. As we hinted before, drifter trajectories were employed in Paper II in this compendium to validate drift trajectory forecasts forced by an ocean ensemble but also to check whether the model was able to capture the observed spectral energy correctly.

Remote sensing technologies, such as satellites, can be used to retrieve ocean properties either directly (e.g., sea surface temperature) or indirectly (e.g., velocity fields derived from nadir-looking altimeters). Different from drifters and other in-situ instruments, remote sensing has the great advantage of sampling large portions of the ocean, providing an instantaneous two-dimensional view of it, and at the same time not disturbing the area. Remote sensing explores the electromagnetic spectrum to retrieve observations of the ocean surface, having three different sources of radiation to communicate the information from there to a sensor operating in a satellite orbiting around the Earth. These are (1) solar radiation reflected at the ocean surface or near the surface, (2) thermal radiation emitted by the ocean surface, and (3) electromagnetic pulses emitted from the satellite towards the Earth, reflected by the ocean surface, and received back at the sensor. The first two are passive devices, working respectively at the optical band and infra-red band (1), and microwave band

(2), whereas (3) represents an active device [76]. A great review about the topic can be found in [77, 78, 79].

The latter represents the class of radar instruments, and it is, up to now, the main workhorse for estimating ocean surface currents. Nadir-looking altimeter radars have been employed for almost 30 years in oceanography, providing unique information about the ocean's energy spectrum down to mesoscale and widely used as data sources in data assimilation in operational centers [80]. Despite its mature technology and processing, interpolation issues [81, 82] and low along-track sampling rate often underestimate the ocean kinetic energy and current speed in coastal regions [83, 84]. The horizontal resolution of nadir-looking derived products is usually coarser than the first baroclinic Rossby radius in high latitudes, hindering direct comparison to investigations focused on small-scale, short-term predictions. For this reason, we opted to employ synthetic aperture radar information in Papers I and III. As explained in the incoming chapter, we used the observations as a passive (through oil slick delineation) and active (ocean current retrieval) tool to address the ocean surface current uncertainty.

/3

Synthetic Aperture Radar and its Applications

> “In the silence of space, the Earth speaks in colors and shadows, and remote sensing lets us hear its voice in data and light.”
>
> — Marcia McNutt

Since the emergence of the satellite era in the early 1960s, space-borne remote sensing has proven to be an invaluable component of Earth observation systems. It has revolutionized our ability to monitor and understand a wide range of environmental and atmospheric phenomena. Satellite-based observations have become an essential research tool for resource management and disaster response.

This thesis focuses on using data from Synthetic Aperture Radar (SAR) satellites, particularly the ESA satellites Sentinel-1A/B. In Paper I, oil slicks observed with SAR were used as a validation dataset for quantifying ocean current uncertainties, and in Paper III, an ensemble atmospheric model was employed to quantify the uncertainty in ocean current retrievals using SAR. This chapter begins with an introduction to key theoretical concepts of the SAR system, followed by a discussion of applications relevant to these two topics.

3.1 SAR Concepts

As with any other radar system, SAR is based on emitting and receiving self-generated electromagnetic (EM) pulses. As an active sensor operating at the microwave band (see Table 3.1), it does not depend on day-light conditions, and it can be employed under 'all-weather' conditions as long as the carrier frequency is located within the atmospheric spectral windows.

Table 3.1: Overview of the standard microwave bands used (or soon to be used) onboard satellite sensors [85].

Band	Frequency [GHz]	Wavelength [cm]
Ka	27 - 40	0.8 - 1.1
Ku	12 - 18	1.7 - 2.4
X	8 - 12	2.4 - 3.8
C	4 - 8	3.8 - 7.5
S	2 - 4	7.5 - 15
L	1 - 2	15 - 30
P	0.3 - 1	30 - 100

3.1.1 SAR Imaging Geometry

Whether mounted on an aircraft or an orbiting satellite, the SAR sensor will follow an orbit at a determined altitude (h) at a certain speed. As the radar sensor moves along its flight direction, pulses are sent in the orthogonal direction on a well-defined illuminated area on the Earth's surface called the

antenna footprint, and successive footprints result in a swath, defining the SAR's imaging area (see Figure 3.1). SAR instruments are side-looking sensors meaning that the two received echoes, delayed by a time interval (δt), ought to stem from two different objects (or different parts of the object) in the footprint, separated by a distance (d).

The amount of power received back at the antenna is characterized by the normalized radar cross-section (NRCS), or backscatter. It depends on the dielectric properties of the surface objects, the properties of the surface's roughness (rms-height and correlation length) relative to the emitted wavelength, and the imaging geometry (e.g. incidence angle: IA, ϕ). The IA is the angle between the radar signal's line of sight and the perpendicular (normal) to the Earth's surface at the point where the signal strikes. The IA increases across the footprint, ranging from a lower angle near the satellite track (near range) to a higher angle farther from the satellite track (far range). Moreover, the backscatter decreases with increasing incidence angle.

The backscattered power is one of the components in the radar equation [86], and a key variable for oil slick detection. The retrieval of ocean surface current, which is treated in Paper III, is based on another important radar parameter, namely the Doppler shift.

The slant range (blue dashed line in Figure 3.1 is the direct distance between the radar antenna and a target on the Earth's surface, measured along the radar's line-of-sight. Since a SAR is a side-looking radar, the slant range direction is actually the direction the pulse propagates towards the target, and the backscattered signal is received back from it.

3.1.2 Spatial Resolution

The ability to distinguish two different targets within the footprint is called (ground) resolution, and it can be described in range (or cross-track) and azimuth (or along-track) direction. For a sensor with pulse bandwidth B, its range resolution is given by:

$$\text{Range resolution} = \frac{c}{2Bsin(\phi)} \tag{3.1}$$

where c is the speed of light (m/s), and B is the bandwidth associated with the pulse. Hence, the larger the bandwidth of the pulse, the better the resolution. Large bandwidth is often achieved using a linear frequency modulation of the pulse (chirp). For the azimuth direction, the targets' relative position to the

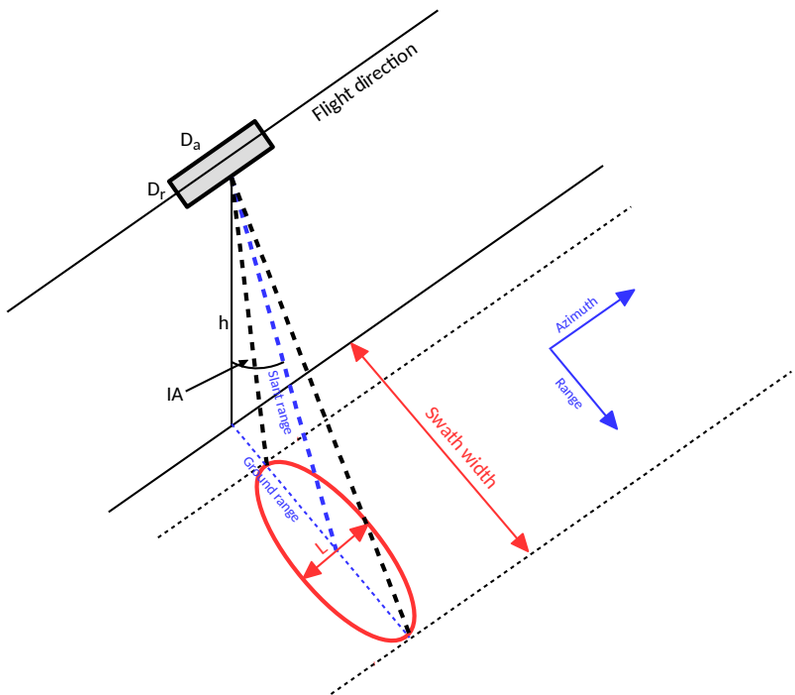

Figure 3.1: Simplified SAR imaging geometry, reproduced with the permission of the author [87]. (Da x Dr): antenna dimensions, h: satellite altitude, IA: incidence angle (ϕ), L: synthetic aperture length. The ellipse represents the antenna footprint.

platform's flight vector results in a Doppler shift. A SAR synthetically simulates the aperture of a much larger real antenna by recording the echoes from a target, as long as it is within the antenna's beamwidth. Using a compression filter that takes advantage of the target's instantaneous Doppler shift, a high azimuth resolution can be achieved, given by:

$$\text{Azimuth resolution} = \frac{D_a}{2} \tag{3.2}$$

where D_a is the antenna length (see Figure 3.1). Both resolutions determine the pixel size of the sampled image, but the real spatial resolution in the ground range, or pixel spacing, is approximately twice that due to theoretical sampling reasons (Nyquist-Shannon sampling theorem). The pixel spacing for Sentinel-1 IW mode is 5 meters (azimuth) x 20 meters (range) for single-look images, which is the data used in this thesis.

3.1.3 Temporal Resolution

Temporal resolution refers to how often a satellite can image the same area, determined by the revisit and repeat cycles. The revisit cycle is the interval between consecutive observations of the same area, and this depends on the satellite's orbit, the swath width, and the target location. For polar-orbiting satellites, the revisit period is more frequent at higher latitudes than in the equatorial region. The repeat cycle is the time it takes for the satellite to image the same position with the exact same viewing geometry. As an example, 41 Sentinel-1A and Sentinel-1B scenes acquired between April and December 2022 were used in Paper I to delineate produced water slicks. If we sought images with the same viewing geometry, less data would be available for the period of interest due to Sentinel-1 A/B combined 6-day repeat cycle.

Polarization

SAR polarization refers to the orientations of the radar wave's electric and magnetic fields, which are always orthogonal to each other. The oscillation direction of the electric component defines the polarization state of an EM wave. It can be oriented either horizontally (H) or vertically (V). In single polarization (single-pol) mode, the radar transmits and receives in the same polarization (co-pol), which in the case of linear polarization is denoted HH or VV. In dual polarization mode, the radar transmits in one polarization and receives in two channels, often one co-pol (e.g. HH) and one cross-pol (e.g. HV). Full or quad polarization systems provide the most detailed information about a target by transmitting and receiving in both horizontal and vertical polarizations (HH, HV, VH, VV). For ocean applications, a dual polarization mode is commonly used due to the higher return signal in VV compared to HH the VV & VH combination is preferred. This is also the combination used by Sentinel-1, over marine areas away from the polar regions (where HH & HV combination is used), whose data is central to this work.

Sentinel-1

Sentinel-1 is a European Space Agency (ESA) satellite mission consisting of a constellation of C-band satellites. The first two satellites, Sentinel-1A and Sentinel-1B, were launched in April 2014 and April 2016, respectively. At the moment, only Sentinel-1A is still operational. These SAR sensors are right-looking radars, operating at the center frequency of 5.405 GHz and can be employed in four different acquisition modes: Stripmap Mode (SM), Interferometric Wide Swath Mode (IW), Extra Wide Swath Mode (EW) and Wave Mode (WM) (See

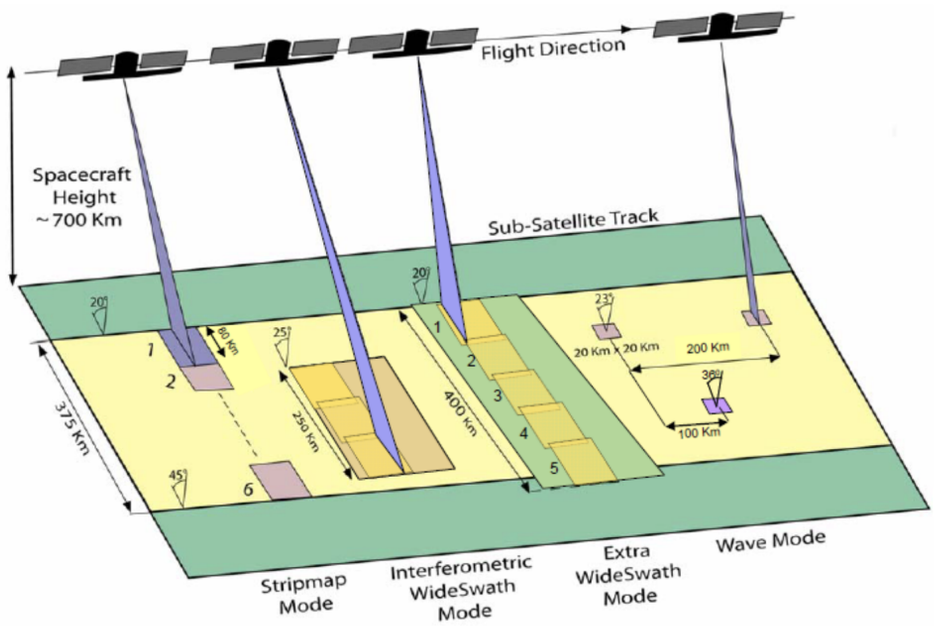

Figure 3.2: Illustration of the different acquisition modes for Sentinel-1. Figure retrieved from ESA [88].

Figure 3.2).

The IW mode was used in Papers I and III. It comprises a large swath width (250 km), with a spatial resolution of 5 m by 20 m. In our case, three sub-swaths (IW1, IW2, and IW3) were used, acquired with the Terrain Observation with Progressive Scans SAR (TOPSAR) technique, which operates by electronically steering the antenna beam both in azimuth and range directions. In this mode, single (HH, VV) and dual-polarizations (HH+HV, VV+VH) are available.

3.2 Surface Scattering

Scattering broadly refers to how an EM wave interacts with a target. In microwave radar applications, one often distinguishes between three types of scattering mechanisms: surface, volume, and double-bounce scattering. Depending on the type of environment, whether it is the ocean surface, urban areas, or forests, one scattering mechanism may dominate, though a combination of all is common.

Surface scattering, or single-bounce scattering, is the primary scattering contrib-

utor at the ocean's surface. This occurs when an incoming EM wave interacts with the interface between two media just once. The roughness of the interface influences the amount of backscatter which is returned to the sensor. Whether a surface appears rough or smooth depends on the roughness height compared to the wavelength of the observing sensor and the length of the roughness. Suppose the roughness of a randomly uneven surface is defined by the standard deviation of height deviations H over the mean height $\overline{h}$. In that case, the question of how large H must be for the surface to appear rough to an observing SAR system can then be addressed. Following the Fraunhofer criterion (Eq. 3.3, a surface is considered rough if the height deviation, H, exceeds:

$$H > \frac{\lambda_R}{32cos(\phi)} \tag{3.3}$$

where λ_R is the wavelength of the observing sensor. For Sentinel-1 at $\phi = 35°$, H > 5.6 mm. Figure 3.3 shows a representation of three different conditions: smooth (a), rough (b), and very rough (c) surfaces. In the first scenario, the incoming EM wave (red arrow) is reflected away from the sensor (specular reflection), and little backscatter is returned to the sensor. In the case of a rough surface (b), specular and diffuse reflection occur thus providing a higher degree of returned signal to the sensor. For very rough surfaces (H > λ_R/2), the backscattered signal is completely diffuse [89].

Sea Surface Scattering

To first order, scattering from the ocean surface and at moderate incidence angles (20° - 45°) can be modeled by the so-called Bragg scattering model. It predicts that only roughness elements, capillary waves in the case of the ocean surface, resonant at the order of the radar wavenumber, contribute to the scattering process [90]:

$$k_B = 2k_R sin(\phi), \tag{3.4}$$

where k_B is the Bragg wavenumber, and k_R is the radar wavenumber. The two-dimensional spectral density of the ocean roughness evaluated at the Bragg wavenumber directly impacts the NRCS (σ^0_{pp}) [90],

$$\sigma^0_{pp} = 4\pi k_R^4 cos^4(\phi)\Gamma_{pp} W(2k_B sin(\phi), 0) \tag{3.5}$$

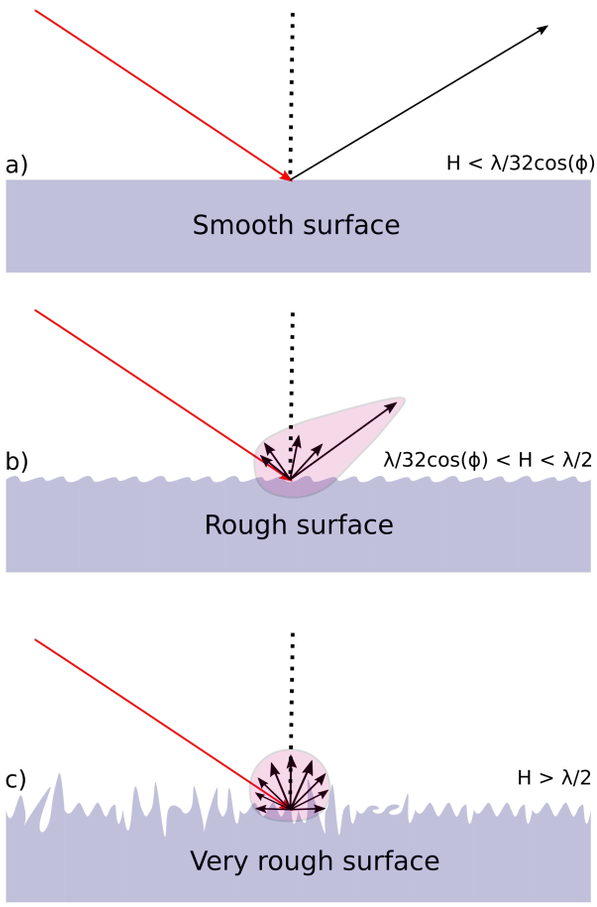

Figure 3.3: Illustration of different scattering surfaces depending on their relative roughness (increasing from a to c) relative to the satellite's wavelength. The red arrow represents the incident EM and the black arrow the reflected EM.

where Γ_{pp} denotes the scattering coefficients for single-pol (pp: HH or pp: VV) configurations. For Sentinel-1, the corresponding Bragg wavelength is 4.83 cm at $\phi = 35°$.

Long waves are also present on the ocean surface, but the SAR cannot directly observe them due to their larger scale relative to the satellite's wavelength. Incorporating modulation effects caused by longer waves has led to the development of a more refined surface backscatter model, known as the two-scale model or composite surface model [90, 91]. Longer waves are observed in SAR images due to their modulation of the Bragg-scale waves through three different mechanisms: tilt modulation, hydrodynamic modulation, and velocity bunching. Tilt modulation represents the change in the local incidence

angle by the slope of the longer wave. Hydrodynamic modulation causes an uneven distribution of the capillary waves along the longer wave, whereas velocity bunching, which is a highly non-linear effect depending on the wave propagation direction relative to the range direction, is the result of a Doppler shift modulation caused by variations of the orbital velocities along the longer wave.

3.3 Doppler Centroid

In SAR processing, the Doppler centroid (DC) plays a critical role in determining the quality of the SAR image. The Doppler frequency shift results from the relative motion between the radar platform and the target, with the Doppler centroid specifically referring to the Doppler frequency at the center of the radar antenna beam. Therefore, in terms of SAR focusing, the DC determines the location of the azimuth beam center and is also used to derive the location of azimuth ambiguities, which can be described as echoes overlapping in frequency from previous or future target positions due to side lobe artifacts. Accurate determination of the Doppler centroid impacts the SAR image processing, specifically during range and azimuth compression. Ultimately, azimuth ambiguities can lead to biases and degradation of radiometric and interferometric information [86]. Different algorithms can be employed to determine the DC [92, 93, e.g.], but the current method used by ESA for Sentinel-1 employs a two-step approach: absolute and fine DC estimation [94]. The absolute DC estimation is based on the satellite's geometric properties (orbit and attitude) and is performed for each range and azimuth block. Such estimation is nevertheless sensitive to small errors in the antenna pointing, for example, an 0.01° yaw angle error can result in a near-range offset of 25 Hz, while the same error in the pitch angle causes an offset of 50 Hz [95]. The fine DC estimation is based on the acquired data, and it undergoes several other steps beyond the scope of this work.

In a side-looking SAR system, where the radar antenna is oriented perpendicular to the platform's direction of motion, the Doppler centroid for a stationary target is ideally zero. However, factors such as target movement toward the sensor or antenna squint (a slight deviation of the antenna beam from its perpendicular orientation) can introduce a non-zero Doppler centroid, known as the Doppler Centroid Anomaly (DCA). The DCA represents the difference between the observed Doppler centroid and the theoretical value for a stationary target, and it plays a significant role in extracting geophysical information using SAR [96]. Research from the early 1980s established a link between Doppler shifts and centroid anomalies, demonstrating their potential for tracking ocean currents [97]. However, limitations in accurately knowing the platform's atti-

tude and antenna-pointing parameters hindered the precision of ocean current retrieval. Recent advances in SAR observation recalibration [e.g., 98] and Doppler-induced geophysical component removal [e.g., 99, 100, 101, 102] have since enhanced ocean current estimation. Building on these advancements, we examined uncertainties in ocean current retrieval related to the wind field in Paper III.

3.4 Ocean SAR Applications

In this this thesis two different ocean applications using SAR were investigated. In the first, oil slicks delineated in SAR images are used to validate oil slick drift modeling, and in the second, the radial current velocity estimated from SAR Doppler measurements is studied with respect to uncertainties in the auxiliary wind data inputs. Both applications benefit from the large swath width offered by SAR satellites and the all-day and night data imaging capability. These two applications are explained in more detail here.

3.4.1 Marine Oil Slick Detection

Oil spill releases into the marine environment arose as a major problem in the 1960s as a result of the development of supertanker ships, offshore installations, and oil exploration over the continental shelves [103]. Since then, major accidents (*e.g. Amoco Cadiz*, France - 1978; *Deepwater Horizon*, USA - 2010) have released massive amounts of oil in the ocean, directly impacting fish stocks, bird colonies, and human resources.

The OSPAR Convention (Convention for the Protection of the Marine Environment of the North-East Atlantic) reported that more than 1500 offshore operational installations were present in the maritime zone under their jurisdiction in 2023 [104]. The same commission reported that 4119 tonnes of oil were discharged and spilled in the North Sea in 2017, where Norway was responsible for approximately 40% of these releases [Table 5e in 105]. Worldwide, the National Research Council estimates 343,200,000 gallons of oil are released every year into the sea [106]. Fifteen oil spills of 1000 gallons (4000 L) or larger are reported every day in the USA [107].

Oil slick detection is a key component in marine pollution surveillance, especially due to illegal discharges and accidents. SAR-based observations provide crucial information for continuous environmental monitoring due to their ability to be used in all-weather conditions as well as during the night. States and private agencies, such as the Office of Satellite and Product Operations

(OSPO, NOAA, USA), Kongsberg Satellite Services (Norway), and the European Maritime Safety Agency (EMSA), employ different sensors (e.g. Sentinel-1, RADARSAT-2, and TerraSAR-X) to provide near real-time information based on to a wide range of stakeholders.

Marine oil slicks can be detected in SAR images due to their higher viscosity and surface tension relative to water, which dampens the gravity-capillary waves. An oil slick in a real SAR image is then expected to appear darker than the surrounding clean water area provided that some wind roughening occurs on the clean water areas. Such decrease in the NRCS is typically observed to be in the range between 5 and 12 dB [108], but it may vary, depending on the radar's incident angle, the type of oil, and the oil thickness. An oil slick makes the sea surface appear smoother, and specular reflection increases, as illustrated in Figure 3.4. Considering a SAR instrument moving from 1 to 5, the backscatter value drops as it passes over 3.

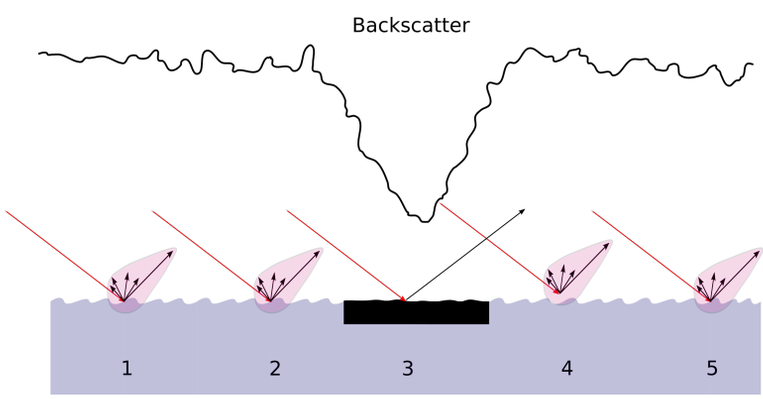

Figure 3.4: Illustration of backscatter signal (black line) for a rough sea surface that also contains a marine oil slick at different times (1 - 5). As the sensor approaches the oil slick (black area in 3), the backscatter power drops and specular reflection increases.

As the smoothness of the surface improves specular reflection but reduces the amount of backscatter received at the sensor, the signal may approach the instrument's noise floor (Noise-Equivalent Sigma Zero, NESZ) for a given image. A low NESZ allows the radar system to detect weaker signals, enhancing sensitivity. Therefore, the backscattered power can be compared to the NESZ in a ratio form, the so-called signal-to-noise ratio (SNR), between the Radar Cross Section (RCS, σ) and the NESZ for a given polarization as follows:

$$SNR_{pq} = \frac{\sigma_{pq}}{NESZ_{pq}} \tag{3.6}$$

where (pq) represents the polarizations of the received and transmitted pulses. The SNR value depends on the sensor properties, e.g. carrier frequency and bandwidth. When monitoring low-backscatter targets like oil slicks, low NESZ and high SNR are desirable to ensure separation between the actual signal and noise. A low SNR decreases the contrast between the slick and surrounding clean sea's backscatter signals, making separation difficult. An SNR value of at least 10 dB is recommended for retrieving reliable information about the oil slick backscatter [109].

Low backscatter areas may also represent other surface features, so-called look-alikes, such as biogenic slicks, low-wind speed areas, wind shadowing, or fish oil. Figure 3.5a shows the VV NRCS (σ^0_{VV}) for a Sentinel-1A scene acquired in the Gulf of Mexico (20.07.2024), with overlaying contours for likely (not confirmed) an oil slick (black), a low-wind speed area (red) and biogenic slicks (blue). NRCS values for different transects (Tr.) over the different features (vertical lines) are displayed according to their color in Figure 3.5b. Such look-alikes also display a drop in the NRCS similar to an oil slick, making the detection and classification of mineral oil slicks in satellite images a challenging task.

In coastal and open water areas away from the polar regions, the VV+VH combination is used as the standard polarization setup by RADARSAT-2 and Sentinel-1, due to the higher returned power over open waters from the VV channel. Using the VV channel means that the SNR is improved over low backscatter targets, such as oil spills, compared to the other polarizations. Oil spill detection is also possible to do using a single channel based on a parameter called "the damping ratio" [110, 111, 112]. Moreover, using a quad polarimetric system, it is possible to characterize slicks and aid separation between mineral oil slicks and biogenic slicks through multi-polarization feature analysis [113].

Oil slick observations are nevertheless scarce. For operational regional model validation, where outputs are time and spatially limited, gathering a sufficient amount of data can be challenging. As an example, out of the 7513 cases monitored by EMSA in 2023, approximately 33% (2527) were verified on-site. Among these, 139 were identified as mineral oil, 219 as fish oil, and 60 as vegetable oil. The remaining cases were categorized as follows: no observable substance (1515), unknown features (396), natural phenomena (67), and other substances, including chemical oil (18), sewage (6), garbage (3), and unknown materials (104) [see reported data at 114].

Since marine surface slicks drift with ocean currents and wind, accurately predicting their future positions becomes crucial in critical or hazardous situations. Achieving this requires prior validation of operational Eulerian and Lagrangian models to ensure the delivery of reliable predictions. In oil drift modeling exper-

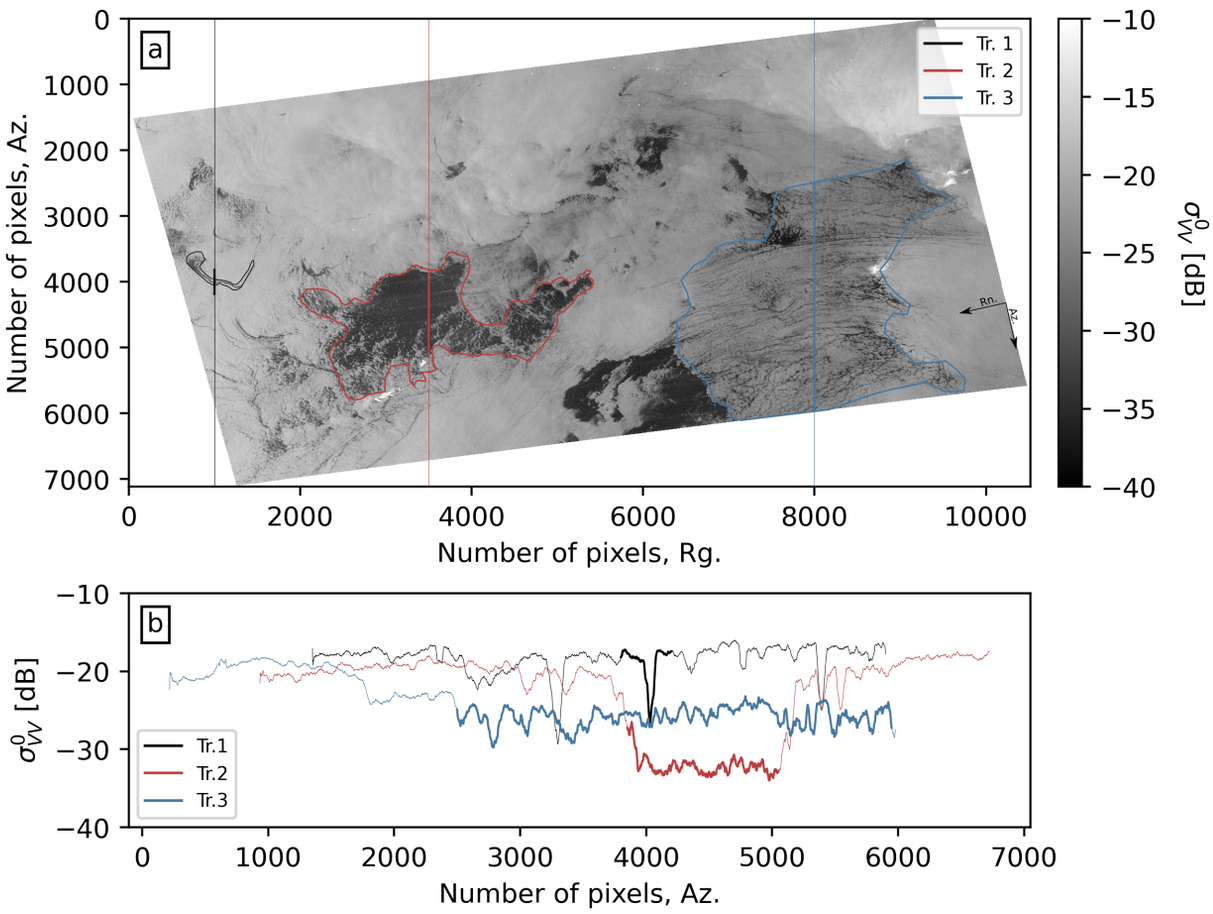

Figure 3.5: Sentinel-1A VV NRCS (dB) acquired in the Gulf of Mexico on 20.07.2024 (a). Contours represent different low-backscatter features, likely (not confirmed) an oil slick (black), low-wind area (red), and biogenic film (blue). Vertical lines are transects along the azimuth direction, with their respective pixel-wise NRCS values shown in (b). Thicker segments represent the approximate location and NRCS over the low-backscatter features. Sentinel-1 data provided by Copernicus 2024.

iments, model performance is commonly validated using delineated oil slicks from major spill incidents [e.g., 71, 115, 116, 117], as these events are often better documented and represent a significant threat of coastal pollution. Controlled field campaigns can provide great insights into the importance of different geophysical forcing on the drift trajectory, the role of in-situ observations for detailed drift description, and how oil-specific properties (e.g. viscosity) can impact horizontal drift predictions due to vertical entrainment and weathering [e.g. 31, 61]. However, these studies are typically limited by the number of observations and may not encompass the diverse environmental conditions

necessary to verify the model's predictive capabilities. A longer observed time series of oil slicks must be compiled to evaluate the model under different ocean and atmospheric conditions.

The primary objective of Paper I was to assess the ensemble and prediction performance of the Barents-2.5 EPS. To achieve this, we compiled a time series of observed produced water (PW) slicks, a byproduct of oil and gas extraction containing a large fraction of saline water compared to oil equivalents, reaching up to 98% of water content in matured fields. As PW slicks can be effectively detected using SAR [118, 119], we monitored an operational floating production storage and offloading (FPSO) unit in the North Sea for 8 months. Sentinel-1 IW VV data was utilized to delineate the legally discharged PW slicks, which served as ground truth for validating Barents-2.5 EPS through short-term oil drift predictions.

3.4.2 Ocean Surface Velocities

Beyond the identification of oil slicks, SAR imagery has been shown to provide important insights into the characterization and retrieval of dynamic variables at the ocean surface, such as eddies and fronts [120], internal waves [e.g. 121], wind vectors, [e.g. 122, 123] and surface currents [e.g. 124]. Paper III focuses on the latter.

Model initialization and validation, as discussed in Section 2.5, depend heavily on observational data. Due to its role in environmental monitoring and climate regulation, ocean currents are also considered an essential climate variable [125]. SAR indirectly estimates ocean surface currents using the Doppler shift information obtained with along-track interferometry [126] and single-antenna observations [96]. The latter method is based on the two-scale decomposition of the ocean surface kinematics, taking into account the contribution of different factors to the Doppler shift. As explained in the *Sea Surface Scattering* subsection, the backscatter in the two-scale model is influenced not only by the Bragg waves, but also by the underlying longer waves and current. [96] demonstrated that a DCA arises from the combined contributions of all moving elements within the dynamic field, whether they are moving toward or away from the satellite. This Doppler shift along the satellite's line-of-sight, known as radial velocity, is just one of the components affecting the overall frequency displacement. In other words, retrieving the underlying ocean current – our variable of interest – depends on removing the contribution of all other sources in the observed signal, both other geophysical sources (wind and waves), and displacements caused by biases in the sensor parameters (non-geophysical). Semi-empirical [e.g. 127, DopRIM] and empirical models [e.g. 122, 128, CDOP] were developed to predict the contribution of the wind field on the observed

Doppler centroid anomaly.

The CDOP model belongs to a family of Geophysical Model Functions (GMFs) neural networks, which are trained on a set of inputs (e.g. wind speed, wind direction, polarization, and incidence angle) to predict the expected Doppler shift for a given wind speed and direction, and radar configurations [80]. In this way, one can remove the impact of the wind on the observation, and the residual value is considered the underlying current contribution to the signal. CDOP was initially designed for the Advanced Synthetic Aperture Radar sensor (ASAR) onboard Envisat, but it has recently been extended to Sentinel-1 [e.g. 98]. Improvements were also made by including the explicit contribution of the wind wave and swell fields in the GMF, together with auxiliary telemetric information for data recalibration. Removing these terms from the total observed Doppler shift has been shown to be of utmost importance in ocean surface current retrieval employing SAR [99, 100, 102, 129, 130]. A gap in the literature nevertheless exists: How does uncertainty in the geophysical components terms impact the retrieved radial velocity field?

In Paper III, we focused on the wind-wave bias estimation (f_{wv}) using CDOP3SiX GMF [100]. Specifically, we explored how uncertainties in the wind speed and direction from an operational atmospheric ensemble model (MEPS, [131]) impact the bias and, consequently, the radial velocity estimations. The total frequency shift can be expressed in mathematical terms as shown in Eq. 3.7:

$$f_{total} = \underbrace{f_{osc} + f_{wv}}_{f_{phys}} + \underbrace{f_{bias}(\beta) + f_{att}(\beta, \phi_{att}(t)) + f_{sca} + \Delta f}_{f_{nphys}}, \tag{3.7}$$

where f_{total} is the observed Doppler shift (Hz), f_{osc} represents the signal contribution from the range component of the ocean surface current, f_{wv} accounts for the wind-wave bias, f_{bias} corresponds to the antenna's electronic mispointing, which varies with the boresight angle β, f_{att} denotes the mispointing error caused by the platform's unstable attitude, $\theta_{att}(t)$, f_{sca} is the scalloping effect in Sentinel-1 TOPSAR mode acquisitions, and Δf includes the remaining unknown residual errors. As CDOP3SiX approximates the contribution of f_{wv}, Eq. 3.7 can be rearranged in terms of f_{osc} as:

$$f_{osc} = f_{total} - f_{nphys} - CDOP3SiX(u_{10}, u_{swell}, u_{sea}, \phi) \tag{3.8}$$

where u_{10} = -Wcos(θ) is the wind range-directed component of a wind vector field with speed W (m/s) and direction (θ, degree) concerning the SAR antenna

look-direction, u_{swell} is the range-directed component of the swell orbital velocity, u_{sea} is the range directed component of the wind sea orbital velocity. Uncertainty comes exclusively from the ensemble wind speed and direction fields (30 ensemble members), *i.e.*, u_{10}^{1}, u_{10}^{2}, ..., u_{10}^{30}, where the superscripts refer to the ensemble member. Thirty resulting estimates of f_{osc} are thus obtained, correspondingly, and each was converted to a radial velocity (RVL) estimate using the following equation:

$$RVL = -\frac{\pi f_{osc}}{k_{Sentinel-1} sin(\phi)} \tag{3.9}$$

In Eq. 3.9, $k_{Sentinel-1}$ represents the Sentinel-1's wavenumber. User and community-guided thresholds [132, 133] were used to determine whether the uncertainty values were significant. The findings in Paper III demonstrated that uncertainties in wind speed and direction contribute to uncertainty in both the wind-wave bias, exceeding 2 Hz, and the retrieved RVL fields, with a standard deviation of up to 0.3 m/s. Understanding these uncertainties, and the role of both wind speed and direction in the retrieval, is a step forward in investigations involving data assimilation of radial velocity fields into ocean operational forecast systems.

4

Uncertainty and Ensemble Modeling

> “For want of a nail, the shoe was lost.
> For want of a shoe, the horse was lost.
> For want of a horse, the rider was lost.
> For want of a rider, the message was lost.
> For want of a message, the battle was lost.
> For want of a battle, the kingdom was lost.
> And all for the want of a horseshoe nail.”
>
> — Unknown author

Uncertainties are inherent features present in any observation and modeling system. Firstly, it represents the precision and accuracy of estimations due to sensor limitations. These, in turn, directly impact how models are initialized and constrained. Secondly, as complex as they are, models are simplifications of the real nature. Errors and parameters in the models also account for how certain we can predict the future state of the ocean or atmosphere. These uncertainties altogether dictate the predictability horizon of our forecasts. Alluding to Lorenz's strange attractor [67], most of the time, the flapping of a butterfly's wings in the Amazon rainforest will not onset a tornado in Texas, but sometimes the unforeseen tiny perturbations at small-scale phenomena may rapidly scale up, leading to large alterations at large-scale features.

This chapter covers the third and final topic of this thesis; uncertainty quantification through ensemble modeling. We start by introducing the background concepts of uncertainty and how ensemble models are developed upon these principles. Next, we describe the metrics used for ensemble performance assessment and conclude the chapter with practical applications with a discussion of research gaps addressed by the scientific contributions presented in this work. Uncertainty in ocean and atmospheric models is the underlying motivation throughout all the articles presented.

4.1 Predictability in a Chaotic System

As pointed out in Subsection 2.5, atmospheric forecasting back in the 1950s, overall for long-range predictions (1 month or longer), was performed using analogues. Edward Lorenz, working as a faculty member at MIT and in statistical-empirical weather prediction, was skeptical of the linear regression methods in vogue at that time. In other words, he did not believe that the atmosphere was periodical as it was thought. Finding a suitable system of equations to verify his hypothesis was the first difficult task. In Lorenz's own words [134]:

"*The difficulties were in finding a suitable system of equations to work with because if I had known exactly what equations to choose in the first place, and exactly what initial states to take in order to get this nonperiodic solution, I probably could have done the whole thing in a couple of months or so with hand computations, which is about the same time it would take to write up the thing afterward for publication. So it wouldn't have taken much more time, but the problem of course is that I had to make many, many tries with many different systems.*"

In communication with Barry Saltzman, who at the time was studying a simplified model with 7 variables representing free convection in the atmosphere

[135], Lorenz noticed that three of these variables did not converge to periodic solutions, even during extended simulations. He worked on the initial convection problem to simplify it towards these three variables, and the solution culminated in his seminal 1963 paper [67]:

$$\begin{aligned} \frac{dX}{dt} &= \sigma(Y - X) \\ \frac{dY}{dt} &= X(\rho - Z) - Y \\ \frac{dZ}{dt} &= XY - \beta Z \end{aligned} \tag{4.1}$$

Where $\mathbf{X} = (X,Y,Z)$ is the state vector of the convection system. Physically, X represents the convective overturning on the x-plane, while Y and Z describe the temperature variation in the horizontal and vertical planes, respectively [136]. The parameters σ, ρ, and β are positive values, and they represent the Prandtl number, Rayleigh number, and the ratio between the width to the height of the plane [136]. This set of equations revolutionized our understanding of atmospheric and oceanic modeling by demonstrating that even a simple three-variable system, highly sensitive to initial conditions, imposes limits on how further in time we can predict the atmosphere and the ocean.

Using the same values for the parameters as in the original paper (σ=10, ρ=8/3, and β=28), we can integrate Eq. 4.1 in time and display the system state in a phase diagram. Figure 4.1 (top row) shows the two-dimensional view of it (only Z(t) and X(t) are displayed). The gray lines depict the orbit of a particle, representing the state of the system, integrated thousands of times from a given initial position. This particle will circle around these two lobes indefinitely, randomly moving from one to the other. But how does this explain its chaotic behavior related to uncertainties in the initial conditions?

Consider a given initial condition for $X(t_0) = x_0$, $Y(t_0) = y_0$, and $Z(t_0) = z_0$. Let's assume that these values correspond to the mean of a Gaussian distribution, with a standard deviation of, say, 0.003. This is your uncertainty value, and you sample the distribution many times (250 times, for example) to obtain your uncertain initial conditions at time = 0 (green dots). We can perform an ensemble simulation by integrating Eq. 4.1 for each of these 250 particles, with their final position at time = T represented as red dots in Figure 4.1. Figure 4.1a, Figure 4.1b, and Figure 4.1c represent the same system, with the same uncertainty radius but initialized at different positions. The trajectory traced by each of the random samples, also known as an ensemble member, is shown in the lower panels for the X(t) variable. Figure 4.1d corresponds to Figure 4.1a

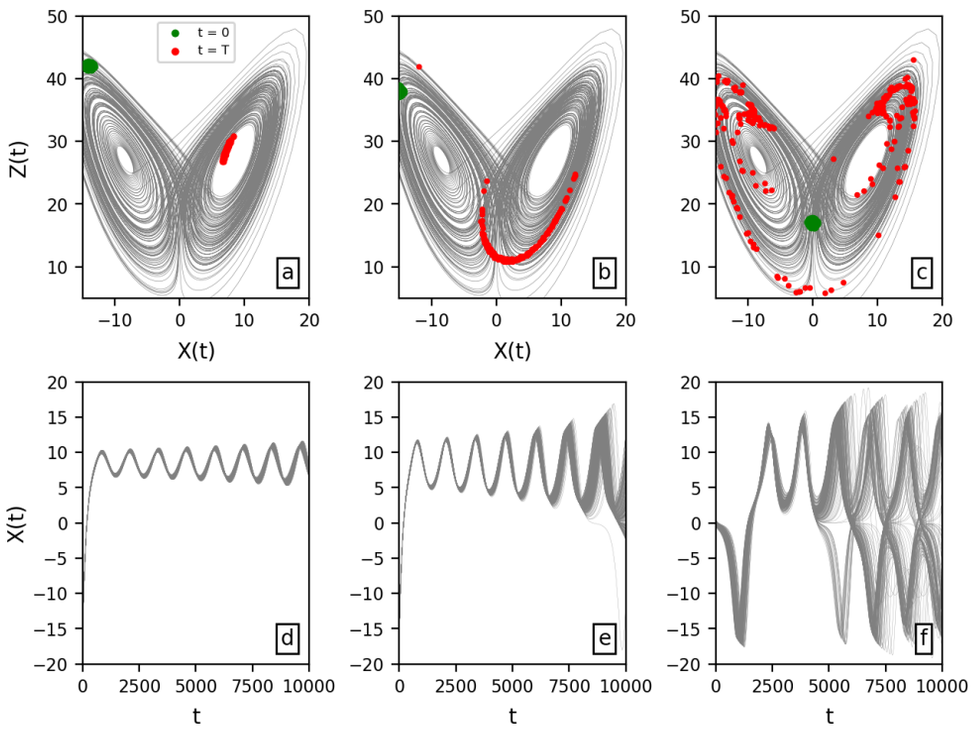

Figure 4.1: Top row: Two-dimensional (Z(t) and X(t)) representation of the Lorenz 1963 attractor. Green dots (250) represent the initial position of the Lorenz system with a 0.003 uncertainty, and the red dots represent their final position at time = T. Bottom row: the trajectory performed by each element for the X(t) variable corresponding to the panel directly above it.

and so on.

These three examples highlight the essence of chaos in dynamical systems: sensitivity to initial conditions in the presence of uncertainty and the divergence of predictions based on initial states (flow-dependency). This is the so-called butterfly effect. In Figure 4.1a, the cloud of red dots did not present substantial distortion relative to their initial positioning, also seen by the closely aligned trajectories in Figure 4.1d. In a real-world application, this reflects a scenario where the atmosphere or ocean exhibits high predictability, allowing a single deterministic model to provide a relatively accurate forecast. Despite inherent complexity and chaos, the system's behavior would remain within predictable bounds. In Figure 4.1b, the cloud of red dots presents a higher degree of distortion, indicating that divergence among the ensemble members starts to set in, and solutions are present in both lobes of the attractor. This represents a case with lower predictability, and a single deterministic model would possibly misforecast the true solution. If the outlier point (upper-left) could represent

the Great Storm event, it is likely that the model would not capture this possible solution. Figure 4.1c represents the case where the final solutions are spread around the attractor, and one can clearly see how the solutions rapidly diverge in time (Figure 4.1d). In this case, the atmosphere or the ocean is in a state of very low predictability, and a completely mismatching prediction is expected from the deterministic model. These cases exemplify, in a general perspective, why models do not present constant skillful predictions.

Lorenz's three-component model provided significant insights into the chaotic nature of nonlinear systems, particularly in Earth sciences. However, it overlooked a critical aspect of geophysical flows: the interaction of processes across different scales. In his 1969 paper [137], Lorenz introduced a more sophisticated model simulating a turbulent atmosphere, where small perturbations were introduced at various wavelengths. His findings revealed that errors originating from small-scale phenomena, such as cloud formation, rapidly grow over time and propagate upward to affect larger-scale systems. For reference, Lorenz estimated that the predictability horizon for global-scale systems (spanning tens of thousands of kilometers) is about two weeks, while for weather patterns (around a thousand kilometers), it is only a few days. Beyond these estimations, Lorenz provided yet another breakthrough: reducing the observational error does not extend these limits.

This brief overview of chaos and predictability establishes the foundational framework of this thesis. As discussed in Chapter 2, atmospheric and ocean modeling relies on accurately defining the initial conditions for the numerical integration of a set of discretized equations. In a multi-scale flow, viz. Section 2.1, not all phenomena are resolved by the models. Indefinitely reducing the model's spatial resolution will not solve the problem, as besides enormous computational costs, current observational systems are unable to constrain operational models down to a few kilometers, especially in the ocean. Chaos is present, and flow-dependency underpins why time series are necessary for model performance evaluation, as performed in all publications contained herein.

The key problem with the deterministic predictability horizon found by Lorenz is that it is a statistical measure by itself. After all, it represents the average predictability limit that a deterministic model can achieve [4]. In practice, however, one does not know where the current atmospheric state is located in the attractor, or how strongly the butterfly will flap her wings. Numerical models, computer power, and knowledge have greatly advanced since the 1960s, but chaos anyway persists. If one seeks to create a prediction system on which future decisions can be made upon its outputs, other methods are needed to provide how reliable these estimations are. We move forward to ensemble modeling now.

4.2 Ensemble Modeling

Edward Lorenz's study of chaotic behavior and error growth in atmospheric models revealed the existence of a limit to how far we can accurately predict the future. His work demonstrated that, even with the best available observations, the forecast will eventually diverge from reality, creating a limited predictability horizon. Acknowledging that errors and uncertainties are unavoidable, the field of ensemble generation developed to account for these limitations within models.

Ensemble modeling is a numerical forecasting technique that generates multiple forecasts by introducing small perturbations in the initial conditions, surface or boundary conditions, or model parameters. Rather than producing a single deterministic forecast, ensemble modeling generates a range of possible outcomes, providing a probabilistic understanding of future conditions[138]. By running these multiple simulations, the method helps quantify the uncertainty inherent in the model predictions. Ensemble forecasts are essential in operational forecast centers, where they improve decision-making processes by indicating the likelihood of different scenarios [5].

4.2.1 Ensemble Generation

Based on Lorenz's findings [67], Edward Epstein proposed in 1969 a stochastic dynamic prediction [139]. In summary, his approach relied on the derivation of the continuity equation assuming a probability density function for the initial conditions, approximated the evolution of only the first two statistical moments (mean and covariance) for the Lorenz 1963 model [67], and compared the approximate solutions to the 'true' distributions obtained from a Monte Carlo ensemble simulation (500 ensemble members). Although positive results were achieved, Epstein promptly noted a drawback in his approach: the consumption of computer power. For the state-of-the-art prediction models containing 10^6 to 10^{10} degrees of freedom [140], predicting the evolution of the state probability in such a stochastic way becomes unfeasible.

As mentioned in the Introduction, ensemble generation ought to introduce meaningful perturbations in the state vector, in a system where the sampled distribution is appreciably smaller than its degrees of freedom. This concept leads to ensemble generation through perturbations that amplify most rapidly over a period of time, e.g. Singular Vectors [138] and Breeding [141]. These two methods are currently employed at the National Centers for Environmental Prediction (NCEP) and ECMWF global atmospheric ensemble models (ECMWF-ENS).

A simpler ensemble generation was promoted already in 1983 using lagged average forecasting [142]. The approach combines the forecast at the analysis time with previous forecasts at times -6, -12, or -24 hours. The ensemble is generated by perturbing the forecast errors, in this sense the error between the initial and average lagged forecasts, and ensemble variance by an amplifying time-dependent factor. Despite its simple implementation, weights must be correctly addressed to each lagged forecast depending on its 'age' and the authors define this as the main difficulty. If one sought a large number of ensemble members, one would need to include excessively 'old' forecasts, and the predictions would be tainted by these if optimal weights were not defined.

State-of-the-art atmospheric ensemble systems rely heavily on observational data to ensure physically meaningful perturbations within the system. For instance, the ECMWF processes around 800 million observations daily, the vast majority of which are collected via approximately 90 satellites. After quality control, this number is reduced to about 60 million data points [143]. Of these, only a small portion pertains to the ocean, and an even smaller fraction is assimilated into the ocean and wave models [143]. The amount of available data greatly reduces for observations acquired in the ocean interior.

The ocean is consequently undersampled, requiring ocean ensemble systems to adopt generation strategies distinct from those used in atmospheric modeling. One approach is to generate the ensemble current field using a time-lagged method, similar to the one previously described. Alternatively, ocean ensembles can be generated by incorporating uncertainties in the surface forcing, obtained from atmospheric ensembles. For instance, in the Barents-2.5 EPS system, each atmospheric ensemble member (from ECMWF and MET-Norway) produces slightly different wind-driven ocean currents. The EPS employs additionally an Ensemble Kalman Filter to regulate the ensemble growth and retain its identity.

4.2.2 Ensemble Performance Assessment

Model validation is a critical process used to assess the performance and reliability of numerical models that simulate physical processes within the atmosphere and oceans. It involves comparing model outputs against observational data, such as satellite and in-situ data, to evaluate how accurately the model represents real-world conditions. By ensuring that models can reproduce observed phenomena under varying conditions, we can enhance confidence in the model's ability to forecast future scenarios and support decision-making.

Error and Spread

Ensemble models generate multiple forecasts by varying the initial conditions, parameters, or model configurations to capture a range of possible outcomes. The goal of validation shifts to evaluating the probabilistic nature of the forecasts rather than a single deterministic output. As ensemble validation uses metrics that focus on the distribution of the ensemble fields, an error metric alone can not determine whether the uncertainty of the ensemble prediction system is well-calibrated or not. In other words, its spread must also be considered.

Error (ϵ) and spread (s) are defined as the root mean square error of the ensemble mean with respect to the observed value and the square root of the ensemble variance, respectively [144]. For a set of one-dimensional time series with N ensemble members, X(t,N), the ensemble mean at each time step t is defined as:

$$\overline{X}(t) = \frac{1}{N}\sum_{n=1}^{N} X_{t,N} \tag{4.2}$$

Whereas error and spread are defined as [145]:

$$\begin{aligned} \epsilon &= \sqrt{\left(\frac{N}{N+1}\right)\frac{1}{T}\sum_{t=1}^{T}(\overline{X}(t) - O_t)^2} \\ s &= \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N-1}\sum_{n=1}^{N}(X_{N_t} - \overline{X}(t))^2\right)} \end{aligned} \tag{4.3}$$

Where T is the length of the segment, and O_t is the observed value. In a well-calibrated system, the ensemble mean should align closely with the observed value, and the spread should represent the expected uncertainty around that mean. Thus, both metrics should reflect similar scales of error and uncertainty.

We can exemplify this concept using a synthetic ensemble simulator based on random walks (see Figure 4.2). The initial distribution is assumed to be Gaussian with a prescribed mean and standard deviation value. From this initial distribution, one random walk was initialized to represent the observation (blue line). Forty ensemble random walk members (red lines) were initialized

with the same mean value, but half of the standard deviation (Figure 4.2a), the same standard deviation value (Figure 4.2b), and double the standard deviation (Figure 4.2c). They ought to represent an underdispersive, well-calibrated, and overdispersive ensembles, respectively. Running this simulator multiple times, say 5000 times, and calculating the error and spread for each simulation, the obtained error-to-spread ratio was 0.5, 1.1, and 1.8. These values represent the underdispersive (overconfident), well-calibrated, and overdispersive (underconfident) cases, respectively.

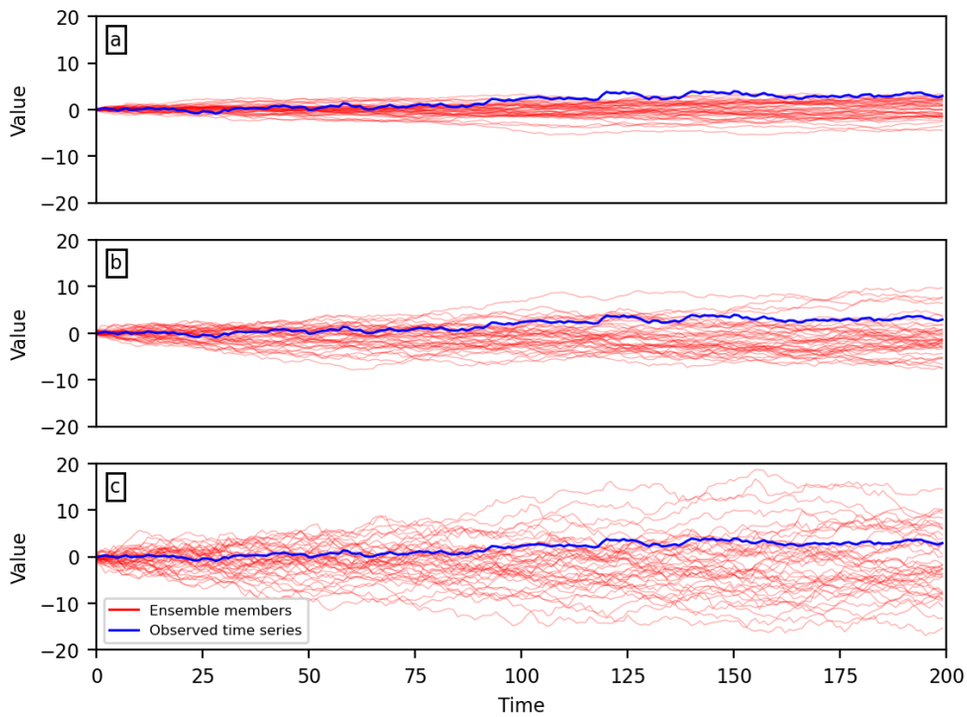

Figure 4.2: Representation of underdispersive (a), well-calibrated (b), and overdispersive (c) synthetic ensembles generated through a random walk scheme. Observation and 40 ensemble members are depicted as blue and red lines, respectively.

In the context of oil slick drift modeling, the calibration of ensemble models is critical for effective decision-making. An overconfident ensemble model, which underestimates the spread of the possible positions of the oil slick, might result in a narrow predicted location. This implies that outlier conditions may not be present in the set of possible solutions, increasing the risk of missing critical areas that need containment or cleanup. Conversely, an underconfident (overdispersive) ensemble model, which overestimates the spread, may result in an excessively large prediction area. This can lead to unnecessary deployment of resources and operational inefficiencies, as responders would need to cover

a wider area than necessary. Therefore, achieving well-calibrated ensembles is essential to balance response strategies, optimize resource use, and minimize environmental and economic impacts.

When dealing with oil slick drift or drifter trajectories, a position vector needs to be considered, and this was done in Paper I and Paper II. For the former, we showed that short-term drift simulations forced solely by a wind ensemble are underdispersive, indicating that including the ocean drift is necessary for properly addressing uncertainties for oil slick drifts. In Paper II, error and spread were used to explain that underdispersion in rank histograms (described below) might stem from model error rather than lack of spread.

Rank histogram

Rank histogram is another diagnostic employed for assessing ensemble modeling performance. The rank histogram is a visual tool, where observed values are ranked within the ensemble predicted values, including outside its range [146].

Differently from error-to-spread, a rank histogram assesses the distribution of the observations relative to the ensemble members, indicating whether the ensemble is well-calibrated, under/overdispersed, or biased. In a well-calibrated ensemble prediction system, if observations and ensemble members are random samples from the same distribution, then the true value is equally likely to fall within (and outside of) the ensemble ranked values [146]. In this case, the rank histogram should be a uniform distribution. In the case of an underdispersive ensemble model, observations are likely to fall outside the ensemble prediction range, creating a convex-shaped histogram. Conversely, for an overdispersive ensemble, observations tend to fall in the center of the ensemble predictions, creating a concave-shaped histogram. Using the synthetic ensemble simulator introduced before, Figure 4.3 shows the rank histograms for the underdispersive case (Figure 4.3a), well-calibrated (Figure 4.3b), and overdispersive case (Figure 4.3c).

Rank histograms can also display whether the ensemble model is biased or not, seen as a sloped rank. While they offer a straightforward diagnostic of ensemble performance, interpreting them in isolation can lead to misleading conclusions [146]. We exemplify this using one of the findings reported in Paper II, where two distinct regions, the Barents Sea and the Fram Strait, were evaluated. Rank histograms for both regions showed a convex shape, indicating underdispersive behavior in the ensemble model. The rank histogram for the Fram Strait simulations exhibited a more pronounced convexity, suggesting stronger underdispersion there and hinting at a narrower model spread com-

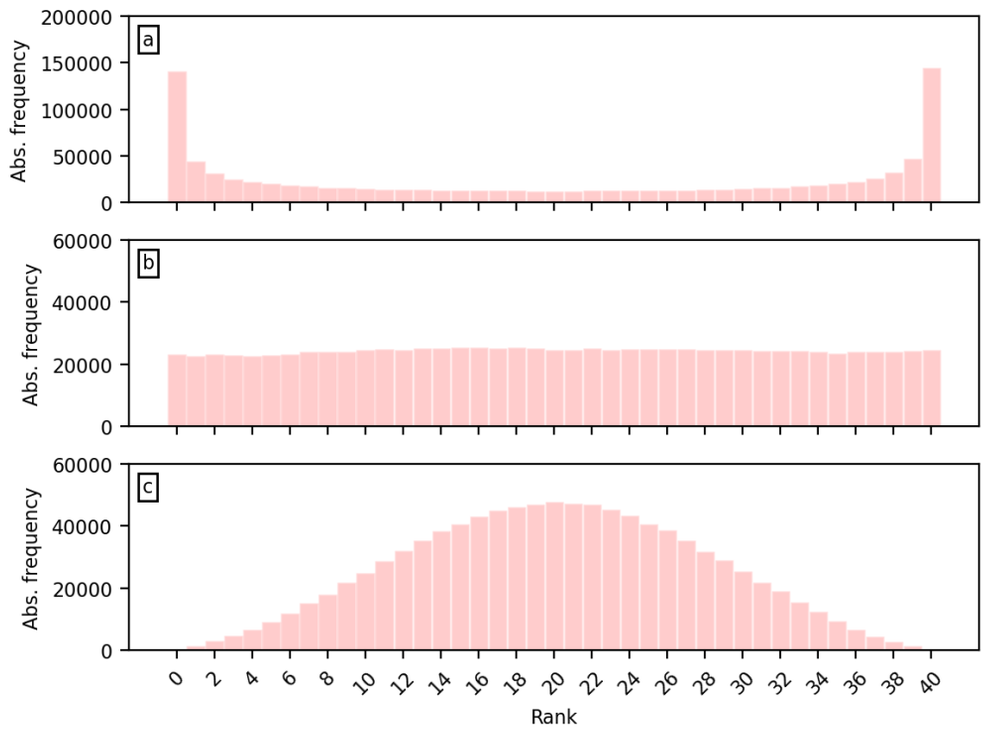

Figure 4.3: Rank histograms for underdispersive (a), well-calibrated (b), and overdispersive (c) ensembles based on the synthetic ensemble time series. Notice the different scale at the ordinate axis in (a).

pared to the Barents Sea. However, while the spread itself was found to be similar across both regions, the ensemble mean error in the Fram Strait was nearly twice as large as in the Barents Sea.

Figure 4.4 shows this finding. Two synthetic ensemble simulations with approximately the same standard deviation were set. For one of them (Figure 4.4b), we added a smooth trend to represent a divergence of the ensemble members relative to the observation at the early integration time. After 5000 iterations, their spread is approximately the same (2.1), but the additional trend increased the ensemble mean error to 4.6 (no trend: 2.7). In simple terms, improving the ensemble model's performance in the Fram Strait should focus on reducing model errors rather than increasing its spread.

Different diagnostic tools provide complementary information about the ensemble prediction system, and as shown with this brief example based on our findings in Paper II, they must be employed in conjunction to assess the model's performance. As for the error-to-spread ratio, we have also extended rank histograms for 2D variables to assess the ensemble predictions in Paper I

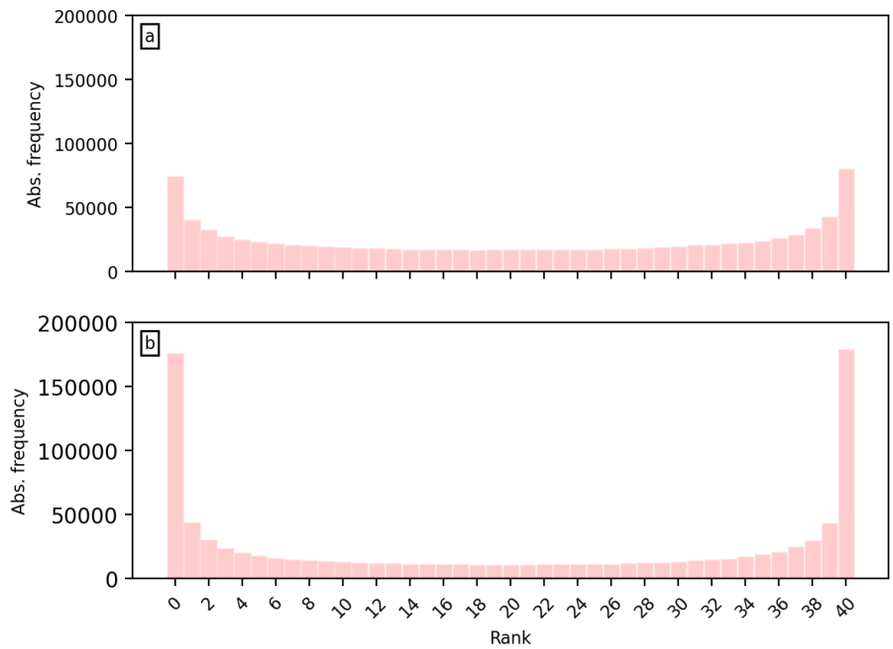

Figure 4.4: Rank histograms generated with the synthetic ensemble simulator (5000 iterations). Both simulations had the same initial standard deviation but an additional trend was added in (b). Note the more pronounced convex shape of (b).

and Paper II.

Reliability

Reliability in ensemble forecasting measures how accurately the forecast probability of an event matches its observed frequency of occurrence. In this method, a binary event (A) is defined, e.g. deployed drifters reach the coastline in 6 hours. Suppose N drifters were deployed in a given region, and they were tracked for a certain period. The relative frequency F of the event A happening can be defined as F = M/N, where M is the number of occurrences satisfying the event. Suppose an ocean ensemble model is sought to be validated against this set of observations through trajectory forecasting. In order to assess its reliability, probability bins (e.g. 0-0.0.5, 0.0.5-0.1, ..., 0.95-1) are defined. A reliable ensemble should, therefore, provide the same probability of occurrence for the event as its observed frequency. In other words, if none of the drifters reached the coastline in 6 hours, the predictions should also output a null probability of occurrence.

When evaluated over many cases, the average probability of occurrence ought to match the average observed frequency in each bin, meaning their relationship should lie at the identity line in the case of a reliable ensemble model. Figure 4.5 illustrates reliability diagrams for different ensemble characteristics generated using the synthetic ensemble simulator, with an event $A < 10$ defined on a distance-like measure created for the synthetic time series. As shown in the diagnostic metrics discussed earlier, the same three cases are also displayed here, i.e., underdispersive (Figure 4.5a), reliable (Figure 4.5b), and overdispersive (Figure 4.5c) ensembles. The dashed identity line represents perfect reliability, providing a reference for evaluating each ensemble's calibration. One can observe how the reliability diagrams become S-shaped when the ensemble cases depart from the well-calibrated model (Figure 4.5b). For the underdispersive case (Figure 4.5a), the model is overconfident and tends to assign probabilities that are more extreme than they should be. Lack of variability (too little uncertainty) yields underestimations at the low-probability end ($0 <$ forecast probability < 0.5) and overestimation at the opposite end ($1 >$ forecast probability > 0.5), leading to an increased likelihood of false negatives and false positives, respectively. A similar interpretation can be made for the overdispersive case (Figure 4.5c).

Reliability diagrams were used in Paper II to assess whether the ensemble prediction system can realistically represent the distance traveled by the observed trajectories. The results also hint at the predictability of the drift simulations, which we found to be different in the two regions of interest. An overview of the three metrics is presented in Table 4.1.

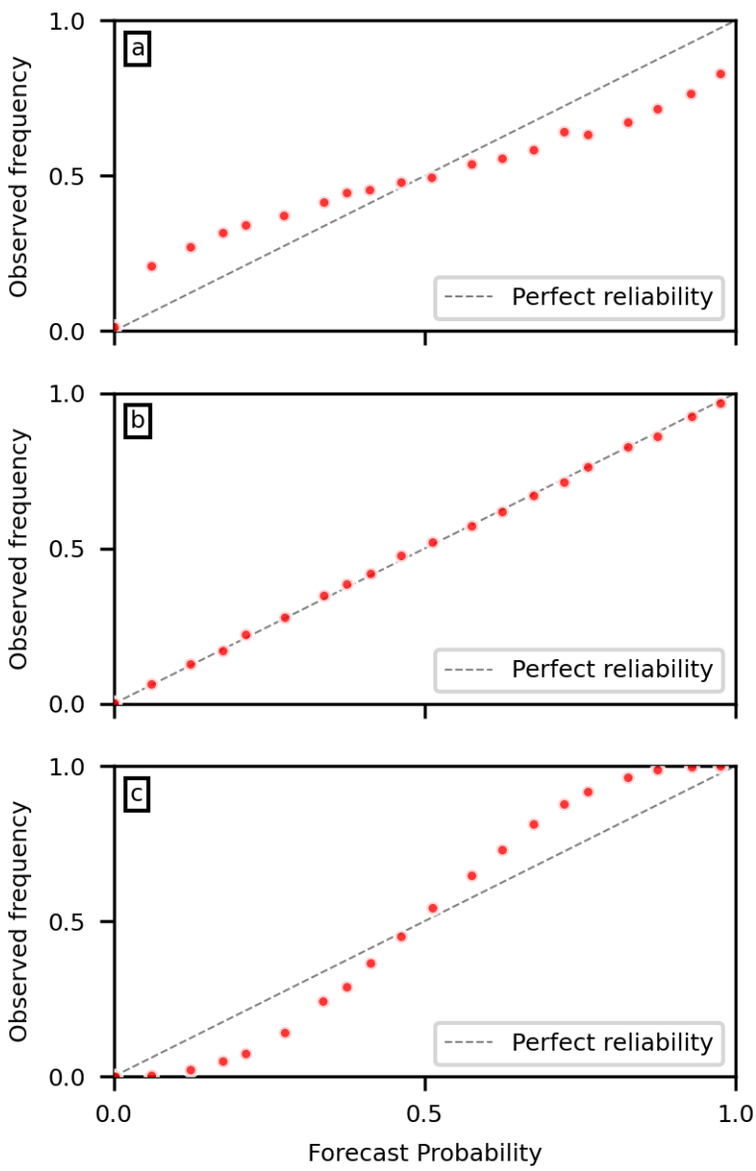

Figure 4.5: Reliability diagrams for underdispersive (a), reliable (b), and overdispersive (c) ensembles based on the synthetic ensemble time series.

4.3 Uncertainty Quantification in Trajectory Prediction

The previous sections established the motivation, theoretical background, and assessment of ocean surface current uncertainty through ensemble model-

Table 4.1: Comparison of spread-to-error ratio, rank histogram, and reliability diagnostics.

Aspect	**Spread-to-Error Ratio**	**Rank Histogram**	**Reliability**
Definition	Measures how well ensemble spread captures forecast uncertainty relative to forecast error.	Visual tool to evaluate how well observations are represented within the ensemble and to identify biases.	Describes the alignment between forecast probabilities and observed frequencies.
Output	A single value (e.g., close to 1 for a well-calibrated ensemble).	Histogram showing the frequency distribution of observation ranks within the ensemble.	Assessed using reliability diagrams.
Purpose	Quantifies if the ensemble spread accurately represents forecast uncertainty.	Detects under- or over-dispersion and biases based on the histogram shape.	Shows whether predicted probabilities match observed event frequencies.
Interpretation	Values near 1 suggest good balance; <1 indicates under-dispersion, >1 indicates over-dispersion.	Flat histogram suggests good reliability; U-shaped means under-dispersion, dome-shaped means over-dispersion, skewed shapes indicate biases.	Deviations (s-shaped) from the diagonal line indicate under- and overestimations, whereas offsets suggest model biases for each probability class.

ing. We head now to the applications and research gaps found in the literature.

Section 4.1 highlighted that the predictability of weather patterns was initially estimated to last only a few days. Similarly, early predictability assessments

for ocean surface currents—based on geostrophic flow models [147], quasi-geostrophic simulations [148], and models assimilating altimetry data [149] suggested a predictability horizon of approximately 10 days, where forecast errors doubled in under two weeks. In practical scenarios, such as oil slick predictions, this limit is reduced to just 3 days [9]. A prime example is the 2010 Deepwater Horizon oil spill in the Gulf of Mexico, the largest marine oil spill in history, where over 4 million barrels of oil were released, resulting in severe environmental damage and legal costs exceeding $69 billion for BP. During the spill, a multi-model ensemble (MME) approach was employed to predict oil slick drift [150, 151], achieving partial success. However, post-incident analysis revealed two important remarks: (1) rapid error growth within two-day simulations and (2) the complex influence of the Loop Current on oil drift pathways over the continental shelf. These findings underscore the limitations of deterministic models in capturing the chaotic dynamics of oil slick drift, emphasizing the need for improved predictive frameworks in high-stakes environmental crises.

A multi-model ensemble approach, also known as an 'opportunity ensemble' differs from single-model ensemble systems by gathering independent ocean and/or atmospheric models available at the period and area of interest. Due to its practical use on-demand, the multi-model ensemble technique is still widely employed for emergency situations [152, 153, 154]. Although useful results are obtained, an epistemological caveat exists: every single model used in the MME is the best-guess of its corresponding system rather than a sample from the same probability space. That is to say, MME estimates the inter-model structural uncertainty rather than the internal uncertainty associated with chaos. Multi-model ensembles have also been under scrutiny by the atmospheric community for almost 40 years [155]. The current consensus is that MME outperforms single deterministic models [156, 157], but there is ongoing debate about whether it consistently surpasses single-model ensemble prediction systems [e.g. 158, 159, 160, 161]. Additionally, multi-model super-ensembles [162], hyper-ensembles [163, 164] and supermodeling [165, 166] have also been explored.

This detour on MME does not undermine the rationale for employing single-model operational ensemble prediction systems. First, it is challenging to demonstrate the independence among the deterministic models in the MME framework, meaning that they may represent a small portion of the possible solutions. Secondly, the spread among the ensemble members might stem from external factors (e.g. data assimilation schemes, numerical methods, and parameterizations) rather than the environmental state. Finally, super-ensembles, hyper-ensembles, and supermodeling are strongly based on data training and statistical modeling, suitable for extended-range and seasonal forecasts. The lack of trajectory prediction studies employing operational single-model ensemble prediction systems is one of the gaps in the literature, and we target

that in Paper I and Paper II.

A series of publications focusing on the theoretical and practical applications of ocean ensemble prediction systems in the Gulf of Mexico have shown to provide additional insights unknown under a deterministic standpoint, revealing details about the sensitivity due to initial condition and surface forcing perturbations [167, 168], mesoscale variability [169], Lagrangian coherent structures [170], and unconstrained scales in the region [171]. Further efforts were also made in the Mediterranean [172, 173] and North Sea [37, 174], more recently with new approaches employing simplified ocean models (shallow-water equations) for short-term predictions [e.g. 175, 176].

In the context of surface drifter and oil slick trajectory prediction, previous research [e.g. 177, 178] using geophysical model ensembles have focused on the sensitivity analysis rather than on validation of the ensemble prediction system. The research presented here aims to fill this gap in the scientific field. As demonstrated in the first two articles, constraining the uncertainty problem with observations allows us to define the conditions and areas under which the ensemble models are prone to perform well. Without observations or other types of delimiting sets, one can not determine whether the ensemble prediction system can be employed operationally.

The Norwegian Offshore Directorate (NOD) estimates that about half of the total petroleum resources in Norwegian waters are still left to be discovered, and more than 60% of that is located in the Barents Sea [179]. Validation of the Barents-2.5 ensemble prediction system using observed oil slicks and drifter trajectories is, therefore, timely to support further activities in the region on demand. Additionally, understanding the ensemble system's performance may better support the creation of a strategy plan for regional Spill Impact Mitigation Assessments (SIMA). Finally, when the source of the oil slick is unknown but it is observed e.g. stranded at the coast [e.g. 180, 181], the integration of remote sensing, Lagrangian particle tracking, and ocean ensemble prediction systems could provide uncertainty quantification of the potential source of the oil through backtracking.

/5

Overview of Publications

5.1 Paper Summaries

This chapter summarizes the candidate's scientific contributions during the research project. It begins with an overview of the three research articles that form the core of the thesis and concludes by highlighting additional works in which the candidate participated as either first author or co-author.

5.1.1 Paper I

> de Aguiar, V., Röhrs, J., Johansson, A.M. and Eltoft, T. (2023). "**Assessing ocean ensemble drift predictions by comparison with observed oil slicks**". *Frontiers in Marine Science*, 10, 2023. ISSN 2296-7745. doi: 10.3389/fmars.2023.1122192.

This study introduces the use of atmospheric-ocean ensemble models on short-term prediction (6 hours) of the drift patterns for produced water (oil and water mixture) slicks. Produced water slicks were delineated in 41 C-band Sentinel-1A/1B scenes acquired over the Norne oil platform, located offshore of the Norwegian coast. The delineated produced water slicks were used as ground truth and oil drift simulations were performed using two forcing sets: (I) an ensemble prediction system composed of 6 ensemble members and discretized on a 2.5 km regular grid (Barents-2.5 EPS), and (II) a deterministic high-resolution (0.8 x 0.8 km) model (NorKyst-800). For each forcing set, two simulation approaches were further defined: (1) Full setup, where the wind field, ocean surface currents, and Stokes drift fields, and (2) only the wind field.

The design of the two experiments, along with their comparison to a high-resolution model, aimed to achieve two objectives: First, to identify the potential main sources of uncertainty in short-term oil drift predictions. Second, to assess whether the high-resolution model offers more accurate predictions compared to the ensemble prediction system. The simulations were evaluated in three different ways, member-wise, model intercomparison, and ensemble performance.

Our findings indicate that the primary source of uncertainty in short-term oil slick drift predictions arises from ocean surface currents. In contrast, wind field uncertainty has minimal to no impact on the variability of drift forecasts. In terms of model comparison, NorKyst-800 does not offer improved accuracy. In other words, the Barents-2.5 EPS delivers forecasts that are just as accurate as the high-resolution model. The main outcome of this research indicates that an ensemble model provides drift predictions not resolved by the deterministic model without a detrimental impact on its accuracy.

5.1.2 Paper II

> de Aguiar, V., Idžanović, M, Röhrs, J. and Johansson, M. (2024). "**Performance diagnostics for probabilistic Lagrangian drift prediction**". *Journal of Operational Oceanography* (In reveiw).

Here we extend the verification of the ocean ensemble prediction system used in Paper I. Undrogued drifters were deployed in the Fram Strait and the Barents Sea in April and August 2022, respectively. Drift simulations were performed using Barents-2.5 EPS and evaluated against the observed drifter trajectories. Unlike Paper I, where only 6 ensemble members were available for each simulation, we expanded the number of available predictions by adding 42 extra ensemble members from previous bulletins, yielding 48 available ensemble members in total. We aimed to evaluate whether including more ensemble members in the analysis could yield an increased ensemble performance. The drift simulations were evaluated by rotary spectra, error-to-spread, rank histogram, and reliability.

The modeled rotary spectra generally align with the observed spectra in the subinertial range, capturing the inertial peak associated with tides and inertial oscillations well. However, a steeper energy decay in the modeled rotary spectra compared to the observed velocities is evident in the superinertial range. Rank histograms revealed that predictions are underdispersive in both regions, stronger in the Fram Strait. By examining the spread-to-error ratio, we suggested this is related to the model error rather than a smaller spread in the simulations (example in Figure 4.4). Predictions are reliable in the Barents Sea for up to 12 hours, whereas no reliability was found in the Fram Strait. Increasing the number of ensemble members did not impact the trajectory prediction performance. Two outcomes were provided given these findings: (1) the need for higher-resolution models and data assimilation for improving drift forecasts, especially in the Fram Strait, and (2) the further development of perturbation schemes to improve ensemble spread.

5.1.3 Paper III

> de Aguiar, V., Moiseev, A., Johansson, A.M., Röhrs, J., Johnsen, H. and Eltoft, T. (2024). "**Uncertainty Estimation of SAR Current Retrievals using an Atmospheric NWP Ensemble**". *Journal of Geophysical Research: Oceans* (Submitted).

Here we assess the uncertainty propagation from an operational atmospheric ensemble prediction system to ocean surface current retrievals using Sentinel-1A imagery. Ocean surface current velocity along the satellite's line-of-sight

can be retrieved using the echoed Doppler information observed by the SAR satellite. The total Doppler signal contains not only the ocean surface current but also a wind-wave bias term. Removing the latter uses atmospheric and wave models in conjunction with a GMF.

Uncertainty in wind-wave bias and radial velocity was estimated using ensemble wind speed and direction fields from an operational atmospheric prediction system (MEPS) combined with the CDOP3SiX GMF for the first time. The study used 251 Sentinel-1A scenes acquired over the North Sea, Skagerrak, Kattegat, and the Baltic Sea. Community-guided thresholds (2 Hz and 0.1 m/s) were used as baselines.

Two lines of investigation were considered, namely (I) the wind speed and direction uncertainty levels under which the retrievals are expected to surpass the established thresholds and (II) the maximum uncertainty for wind speed and direction that suffices the thresholds. Two approaches were introduced for the latter, (1) an analytical ensemble uncertainty propagation scheme and (2) empirical cumulative distribution functions based on the ensemble outputs.

Radial velocity retrievals tend to become unreliable when the wind speed uncertainty exceeds 20% of the mean field. Both analytical and probabilistic approaches for estimating maximum uncertainty produced similar overall results, though the analytical method lacked finer details. This highlights the importance of ensemble models for uncertainty quantification. While uncertainties in the wind field influence the magnitude of radial velocity retrievals, they have a lesser impact on directionality. We stressed that retrievals under low-wind speed conditions tend to exceed the thresholds and should be treated with caution. Finally, we suggested that wind speed and direction uncertainties should not be disregarded in further studies, especially under along-track wind direction conditions.

5.2 Other Scientific Contributions

Other contributions are provided below, where PR stands for peer-reviewed, P stands for poster, OP stands for oral presentation, and R stands for report.

As first author:

PR: **de Aguiar, V.**, Dagestad, K.-F., Hole, L.R. and Barthel, K. (2022). "Quantitative assessment of two oil-in-ice surface drift algorithms", *Marine Pollution Bulletin*, vol. 175. https://doi.org/10.1016/j.marpolbul.2022.113393.

OP: **de Aguiar, V.**, Idžanović, M., Röhrs, J., Johansson, M. and Eltoft, T.(2023). "Forecasting surface oil slicks and drifter trajectories", *CIRFA Seminar*.

OP: **de Aguiar, V.**, Idžanović, M., Röhrs, J., Johansson, M. and Eltoft, T.(2023). "Surface Circulation in the Fram Strait and Barents Sea: A Drifter Study", *NOOS Annual Meeting*.

OP: **de Aguiar, V.**, Moiseev, A., Johansson, M., Röhrs, J., Eltoft, T and Johnsen, H.(2023). "Quantifying Surface Radial Velocities Uncertainty Using Atmospheric Ensemble", *CHESS Annual Meeting*.

P: **de Aguiar, V.**, Idžanović, M., Röhrs, J., Johansson, M. and Eltoft, T.(2023). "Time-lagged Ensemble Model Verification for Trajectory Prediction", *CIRFA Annual Meeting*.

OP: **de Aguiar, V.**, Idžanović, M., Röhrs, J., Johansson, M. and Eltoft, T.(2023). "Drift Tracking for Oil Spill Monitoring", *CIRFA Annual Meeting*.

P: **de Aguiar, V.**, Idžanović, M., Röhrs, J., Johansson, M. and Eltoft, T.(2024). "Time-lagged Ensemble Model Verification for Short-term Prediction of Drifter Trajectories", *EGU General Assembly*. https://doi.org/10.5194/egusphere-egu24-15796.

As co-author:

PR: Pavlov, V., **de Aguiar, V.**, Hole, L.R. and Pongrácz, E. (2021). "A 30-Year Probability Map for Oil Spill Trajectories in the Barents Sea to Assess Potential Environmental and Socio-Economic Threats", *Resources* 11, no. 1: 1. https://doi.org/10.3390/resources11010001.

P: Quigley, C., Johansson, M., Serov, P. **de Aguiar, V.**, Winsborrow, M. and Mattingsdal, R. (2022). "Satellite detection and drift estimates of naturally occurring oil seeps in the Barents Sea", *Arctic Science Summit Week*.

PR: Sutherland, G. **de Aguiar, V.**, Hole, L.R., Rabault, J., Dabboor, M. and Breivik, Ø. (2022). "Estimating a mean transport velocity in the marginal ice zone using ice–ocean prediction systems", *The Cryosphere*, 16, 2103–2114. https://doi.org/10.5194/tc-16-2103-2022.

R: Serov, P. *et al.* (2022). "CAGE22-6 cruise report: GEO-3144/8144 Teaching Cruise: Geologically controlled hydrocarbon seepage in Hopendjupet and the wider Barents Sea", *Centre for Arctic Gas Hydrate, Environment and Climate Report Series*, 10. https://doi.org/10.7557/cage.6769.

PR: Rabault, J. et al. (2023). "A dataset of direct observations of sea ice drift and waves in ice", *Sci Data*, 10. https://doi.org/10.1038/s41597-023-02160-9.

PR: Pavlov, V., James, N.A., Maden, E.A., **de Aguiar, V.**, Hole, L.R., Liimatainen, H. and Pongrácz, E. (2023). "Impacts of offshore oil spill accidents on island bird communities: A test run study around Orkney and Svalbard archipelagos", *Environmental Pollution*, 334:1. https://doi.org/10.1016/j.envpol.2023.122193.

Finally, the candidate reviewed three manuscripts for different journals.

/6

Paper I: Assessing ocean ensemble drift predictions by comparison with observed oil slicks

de Aguiar, V., Röhrs, J., Johansson, A.M. and Eltoft, T. Front. Mar. Sci., 10, 5335–5355, https://doi.org/10.3389/fmars.2023.1122192, 2023.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 16 May 2023
DOI 10.3389/fmars.2023.1122192

OPEN ACCESS

EDITED BY
Chunbo Luo,
University of Exeter, United Kingdom

REVIEWED BY
Carlo Brandini,
Department of Earth System Sciences and Technologies for the Environment (CNR), Italy
Georgios Sylaios,
Democritus University of Thrace, Greece

*CORRESPONDENCE
Victor de Aguiar
victor.d.aguiar@uit.no

RECEIVED 12 December 2022
ACCEPTED 06 April 2023
PUBLISHED 16 May 2023

CITATION
de Aguiar V, Röhrs J, Johansson AM and Eltoft T (2023) Assessing ocean ensemble drift predictions by comparison with observed oil slicks.
Front. Mar. Sci. 10:1122192.
doi: 10.3389/fmars.2023.1122192

Assessing ocean ensemble drift predictions by comparison with observed oil slicks

Victor de Aguiar[1]*, Johannes Röhrs[2], Anna Malin Johansson[1] and Torbjørn Eltoft[1]

[1]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway, [2]Norwegian Meteorological Institute, Research and Development Department, Oslo, Norway

Geophysical models are cornerstone pieces in marine forecasting of floating objects and pollution, such as marine surface oil slicks. Trajectory forecasts of oil spills inherit the uncertainties from the underlying geophysical forcing. In this work we compare the forecast capabilities of an ocean ensemble prediction system (EPS) to those from a higher resolution deterministic model on the representation of oil slick drift. As reference, we use produced water (PW) slicks detected and delineated from 41 C–band Sentinel-1A/B satellite synthetic aperture radar images between April and December, 2021. We found that the EPS provided at least equivalent member-wise results relative to simulations forced with the deterministic model. Ensemble verification through rank histograms and spread-error relationship showed that including the ocean fields is necessary to address model uncertainties. Whether considering the ocean field or not, the modeled slicks were counterclockwise rotated between 20° and 30° relative to the ones observed in the satellite images, and these were deflected about 45° to the right of the observed wind direction.

KEYWORDS

ensemble modeling, SAR, trajectory prediction, produced water, remote sensing observations

1 Introduction

Ocean, wave and atmospheric model products are routinely used by research institutes and state agencies as backbones in oil drift models for real time prediction and contingency plans (e.g. Sutherland et al., 2020). Despite their paramount importance in modeling oil slick drift, geophysical circulation models exhibit uncertainty and hence the modeled variables might not accurately reflect the environmental state (Jacobs et al., 2021; Röhrs et al., 2023a).

Numerical prediction systems are naturally chaotic, meaning that they are highly sensitive to initial conditions. By slightly perturbing the initial conditions of a deterministic dynamical system, Lorenz (1963) showed that non-periodic solutions are unstable and these can evolve into considerably distinct states. Uncertainties also arise from the model physics, truncation errors and imposed boundary conditions. These intrinsic limitations

present in the prediction process resulted in development of ensemble forecasting, where equally probable ensemble members are generated by perturbing the model configuration and initialization to assess the uncertainty and address the most likely leading scenarios of the system's state (Buizza, 2019).

In trajectory forecasts, ocean currents, waves and the wind fields are necessary to predict the future positions of a target of interest, and as such, modeled trajectories inherit errors of the underlying forcings. The basic oil drift model states that the oil particle velocity (V_{oil}) at the ocean surface is composed as a combination of ocean motions (V_o), in which the tidal currents, geostrophic currents, Ekman transport *etc.* are included, a fraction of the wind speed (αV_w), where α represents the wind drift factor, and the Stokes drift (V_s) as follows:

$$\boldsymbol{V}_{oil} = \underbrace{\boldsymbol{V}_o}_{\text{Ocean}} + \underbrace{(\alpha \boldsymbol{V}_w)}_{\text{Wind}} + \underbrace{\boldsymbol{V}_s}_{\text{Stokes Drift}} \quad (1)$$

Based on Eq. 1, oil drift ensemble modeling can be performed as Monte Carlo simulations to cover ranges of possible scenarios by randomly varying the initial position, the releasing time within a period of interest and the wind drag factor term (Nordam et al., 2017; Röhrs et al., 2018; Villalonga et al., 2020; de Aguiar et al., 2022). Despite its implementation simplicity, this approach is widely used in environmental risk assessments as they might indicate regions potentially impacted and possible oil pathways in case of a real oil spill accident takes place (Olita et al., 2019; Sepp Neves et al., 2020). Nonetheless, as pointed out by Barker et al. (2020), the outputs result essentially in a greater diffusion of the trajectories since the ocean dynamic system is fundamentally the same.

To overcome this drawback, multi-model ensembles have increasingly been used as multiple distinct models are freely available in web-platforms, e.g., the Copernicus Marine Environment Monitoring Service (CMEMS). By integrating existing national models with products available at CMEMS, Zodiatis et al. (2016) developed the Mediterranean Decision Support System for Marine Safety (MEDESS-4MS) service, a decision system designed to support the European and Mediterranean emergency centers in oil spill response. Connected to 14 ocean models, 7 atmospheric models and 7 wave models, De Dominicis et al. (2016) used MEDESS-4MS on the reproduction of an oil slick detected in the Western Mediterranean Sea and trajectories of drifters subject to distinct wind and wave exposure. The authors highlight superior forecast accuracy of higher resolution ocean models. Conversely, Dagestad and Röhrs (2019) reports that mesoscale ocean features resolved by a high-resolution model (2.4 x 2.4 km) might actually be seen as degrading noise due to its low predictive skill. As independent models are used, the multi-model approach provides distinct leading states of the ocean and atmosphere, but perturbed single model ensembles are required to statistically access model uncertainty (Frogner et al., 2022).

Ocean models at submesoscale permitting range (1 -10 km) have been implemented to improve the representation of the kinetic energy cascade, thus theoretically enhancing ocean model results towards more realistic dynamics. Large amounts of data are necessary to constrain circulation at these scales, and as the pace of observation systems' implementation does not follow the rapid advances in computational capability, forecasts provided by such models might actually degrade faster than predictions issued by the coarser ones (Jacobs et al., 2021).

While constraining initial conditions in eddy-resolving ocean models remains a challenge, ensemble models are considered a valuable tool to provide estimates of uncertainty and equally probable ocean leading states. Thoppil et al. (2021) showed that a lower resolution (1/12.5°) ensemble forecast system extends the predictability of ocean mesoscale features with wavelengths greater than 150 km to between 20 and 40 days, outperforming its higher resolution (1/25°), deterministic version. By modeling drifter trajectories, Melsom et al. (2012) found that the ensemble average trajectories are generally more accurate than the corresponding results from a deterministic model. In a similar study, Khade et al. (2017) also highlighted better agreement of the ensemble mean trajectory to particle drifts in comparison to single deterministic simulations, despite that no independent observations were used for validation.

Less attention has been paid to the uncertainties in ocean models compared to wind forecasts due to its perceived secondary role in marine pollutant modelling (Li et al., 2019; Kampouris et al., 2021), though Jones et al. (2016) showed that the oil slick drift modeling was improved when ocean current data retrieved from drogued buoys were included in the drift modelling. Additionally, other studies have also shown that poorly resolved mesoscale features (Jorda et al., 2007) and inertial oscillations (Brekke et al., 2021) impact the modeled drift of slicks, thus misrepresenting the predicted trajectory if not taken into account.

Three research questions are addressed in this work: (1) We compare the forecast capabilities of an operational ocean ensemble prediction system to a higher resolution deterministic model on the simulation of produced water (PW) drift, (2) we assess the ability of EPS to estimate model uncertainties and (3) we investigate the role of the wind field on PW drift.

PW is a byproduct mixture consisting largely of saline water and low concentrations of oil which when being released result in thin mineral oil slicks. Here we utilize continuously released PW from the operational oil extraction platform *Norne*, located in the Norwegian Sea about 100 km offshore. The PW release occurs at 12 m depth and the oil plume rises through the water column within minutes to the surface. PW releases represent the general case of a coherent thin oil film, and thin oil films have been found to make up approximately 90% of the area and volume of accidental oil spill releases (Svejkovsky et al., 2016), which underscores the importance of correctly modeling slicks for oil spill risk assessments. Such films are difficult to model given their ephemeral nature on the sea surface, and development of improved models is currently hampered by limited experimental data under a wide range of oceanic conditions. Skrunes et al. (2019) and Johansson et al. (2021) showed that PW slicks may be satisfactorily detected and characterized using Synthetic Aperture Radar (SAR) for wind speeds ranging between 1 - 12 ms^{-1}.

This work is structured as follows: Section 2 describes the PW slicks observations, geophysical models, the Lagrangian particle tracking model, simulation setups and the metrics used to assess the simulations. Section 3 presents the results obtained in this work

and these are discussed in Section 4. Finally, Section 5 summarizes the key results.

2 Data and methods

2.1 SAR images and PW masks

SAR remote sensed data is a cornerstone in environmental monitoring for surface oil detection due to its all-weather capabilities, *i.e.*, neither lack of daylight nor cloud coverage inhibit its data acquisition. Due to its viscous and dielectric properties, oil slicks might be identified in satellite SAR images, as its presence at the ocean surface attenuates capillary waves and consequently results in a decrease of the backscatter energy received by the radar (normalized radar cross section, NRCS) within the slick area relative to the surrounding water. Thus the oil slicks is observed as a dark region in SAR images.

NRCS damping is typically observed in the range between 5 and 12 dB (Alpers et al., 2017), but varies depending on the radar incident angle and its polarization (HH, VV, HV and VH), where the letters indicate the transmitted and received signal orientation, respectively. It should be highlighted that the reduction of NRCS does not necessarily indicate the presence of a mineral oil slick since the observation of dark regions in SAR images might also be related to look-alikes, *i.e.*, biogenic films and areas of low wind speed. In addition, oil slick detection by SAR is also hindered at high wind speeds due to the increase of the ocean roughness, thus essentially limiting its detection for wind speeds within the range between 1.5 - 2 ms $^{-1}$ and 10 - 14 ms $^{-1}$ (Fingas and Brown, 2018).

Covering the period between April and December 2021, 41 Sentinel-1 scenes obtained in Interferometric Wide (IW) Ground Range Detected (GRD) mode (VV/VH) were used here. The Sentinel-1 mission is composed by two polar-orbiting satellites, Sentinel-1A (2014-ongoing) and Sentinel-1B (2016-2022), operating in C-band (5.405 GHz), with pixel spacing of 10 m x 10 m in the azimuth and range directions, and incidence angles between 32.9° - 43.1° in this acquisition mode. The scenes were calibrated, speckle-noise filtered and georeferenced in the European Space Agency's (ESA) SNAP toolbox. The PW slicks were manually delineated using the processed scenes in QGIS, see Figure 1. The width of the slicks varied between 85 m and 200 m, while their length ranged from 1 km to 15 km. Slicks were detected in all of the considered months apart from June, when the platform was not operational due to maintenance procedures. August had the highest number of observed PW slicks (12), and November and December the lowest, with 1 slick each.

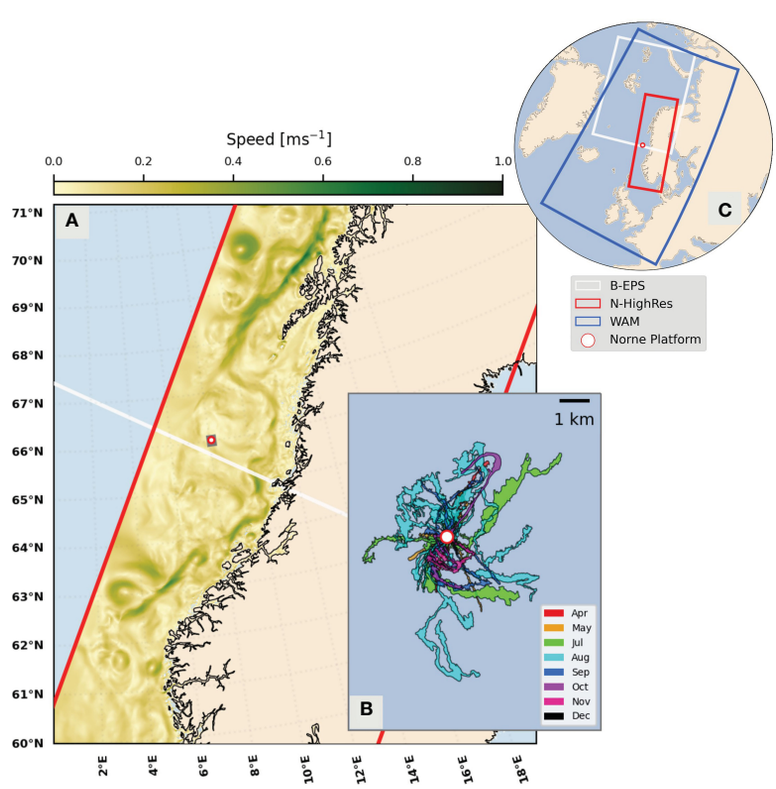

FIGURE 1
Overview of the area of study. **(A)** Location of the Norne platform (dot) offshore the Norwegian coast. The colormap in represents the N-HighRes ocean current speed (ms $^{-1}$) averaged between the 1st and 14th of April, 2021. The grey polygon around the platform is zoomed-in in **(B)**, with the delineated PW slicks for each month. Red and white lines depicts the grid domain of N-HighRes and B-EPS. Their full extent, including the WAM domain (blue), can be seen in **(C)**.

2.2 Produced water and *in situ* wind data

According to the national regulations, PW discharge on the Norwegian shelf is legal provided that the monthly average oil concentration does not exceed 30 gm^{-3} (Miljørapport, 2019). When the platform is operational PW is continuously released from a sub-surface pipe from which the oil rises to the surface. The release volume and concentration vary over time, though the variability is on a bi-weekly to monthly time scale. Release data from the Norne platform in 2021 was provided by the operator and the average daily PW oil release was 253 kg, average water release was 19754 m^3, with an average oil concentration of ~ 10.85 gm^{-3} and an average flux of 0.23 m^3 s^{-1}. The oil volume percentage of 0.013% means that it is classified as an oil-in-water type of emulsion (Lu et al., 2020), though such low concentrations put the releases close to the oil free water surfaces. Damping ratio, relative thickness, estimates of produced water slicks indicates low variability and hence confirm the assessment of thin surface slicks (Skrunes et al., 2019; Johansson et al., 2021).

Wind speed and direction are recorded every 3 hours at Norne. Observations between April and December 2021 were used to assess the modeled winds of Barents 2.5-EPS and NorKyst-800 through the Root Mean Square Error (RMSE), bias and Willmott skill score (Willmott, 1981, WS). The alignment between observed wind direction and the PW bearing angle was also verified in order to evaluate the role of winds on the transport of the PW slicks. The findings are presented in Section 3.1.2.

2.3 Geophysical forcing

Norne is located in the Norwegian continental shelf, bordered by the Norwegian Coastal Current (NCC) flowing northward alongshore and by the Norwegian Atlantic Slope Current (NwASC) along the continental slope. The region is generally considered mesoscale eddy rich, with stable dynamic boundaries on the continental slope preventing cross-shelf transport (Dong et al., 2021). The abundance of eddies can also be observed in Figure 1A. The existence of density fronts in the region is known (Dong et al., 2021), and turbulent baroclinic instabilities might affect the PW drift.

2.3.1 Ocean model ensemble

Ocean currents are provided by two regional setups of the Regional Ocean Modeling System (ROMS, (Shchepetkin and McWilliams, 2005)): The Barents 2.5-EPS (B-EPS) ocean model provides a 6-member ensemble with 2.5 km horizontal resolution (B-EPS), and the NorKyst-800 (N-HighRes) ocean model provides deterministic forecasts at higher resolution of 800 m (Figure 1). B-EPS is a coupled ocean-ice system for the Barents Sea and Northern Norway and N-HighRes is a regional model for the entire coast of Norway. The model domains overlap at the experiment site and provide daily forecasts with 66 hrs lead time.

B-EPS is initialized each day from the forecast of the previous day, with update increments provided by an Ensemble Kalman filter data assimilation scheme (Sakov and Oke, 2008). While sea surface temperature and *in-situ* hydrography are assimilated to constrain the density fields, no data that directly constrains the velocity fields is being assimilated. Details on the configuration for B-EPS model are given in (Röhrs et al., 2023b). In principle, ensemble spread in B-EPS is maintained by forcing of the ocean model with an atmospheric ensemble. The first is forced by a high-resolution weather forecast model AromeArctic (Müller et al., 2017a) (2.5 km x 2.5 km), while the subsequent members are forced by random members of the ECMWF-ENS forecast (ECMWF, 2012). In addition, each ensemble member retains their identity from the previous forecast cycle. Consequently, as the current field develops independently in each member the B-EPS ensemble permits various realizations of mesoscale circulation across the members. While ensemble spread of SST is validated in Röhrs et al. (2023b), spread in surface circulation and the ability of the ensemble to represent uncertainties in currents are the subject of ongoing investigations.

N-HighRes is a regional ocean model covering the whole Norwegian coast at an eddy-resolving resolution. The model is forced by the atmospheric forecast model Arome-MEPS (Müller et al., 2017a). Details of N-HighRes are provided in Asplin et al. (2020) and references therein.

While B-EPS assimilates sea surface temperature and sparse *in-situ* data, neither of the two ocean models exhibits constraint of mesoscale ocean circulation that directly provides forecast skill in surface currents. A degree of predictive skill in ocean surface currents may however arise by the relatively high degree of predictability in the used wind forcing.

Both model setups include the major tidal constituents provided by TPXO 7.2, and receive lateral boundary conditions from the TOPAZ4 hydrodynamic model (Xie et al., 2017). TOPAZ4 has a horizontal resolution of about 10 km, it is based on the Hybrid Coordinate Ocean model and is part of the Copernicus Marine Environment Monitoring Service for the Arctic domain.

2.3.2 Atmospheric forcing ensemble

While each of the used ocean model representations receives different atmospheric forcing, the wind from the respective model is also a direct input for the oil drift simulations, as surface slicks become direct subject to atmospheric drag.

AROME-Arctic and Arome-MEPS (hereinafter AROME) are the regional high-resolution weather prediction models for the Barents Sea and Norway, respectively. Both systems issue weather forecasts 66 hrs ahead with multiple update cycles each day. AROME is based on the HARMONIE-AROME model configuration of the ALADIN-HIRLAM system (Bengtsson et al., 2017), with boundary conditions obtained from ECMWF's IFS.

2.3.3 Wave prediction model

A prognostic model for the evolution of the wave energy spectrum, WAM (Komen et al., 1994), is used to calculate Stokes drift and wave entrainment rate of surface oil. The used implementation of WAM at MET Norway has a horizontal resolution of 4 km, resolving the wave energy spectrum in 36 directions and 36 frequencies (Gusdal and Carrasco, 2012). The

Stokes drift is calculated from significant wave height, peak period and the wave direction according to Breivik et al. (2014), hence taking into account both wind sea and swell contributions to the depth-dependent Stokes drift profile. While surface oil slick is advected by the surface Stokes drift, submerged oil particles are moved by the Stokes drift at the respective particle depth.

2.4 Oil drift model - OpenOil

The open-source oil transport and weathering Lagrangian framework, OpenOil, is used to perform the drift simulations (Röhrs et al., 2018). Embedding tabulated oil information provided by the Norwegian Clean Seas Association for Operating Companies (NOFO) and linked to the ADIOS oil library (https://github.com/NOAA-ORR-ERD/PyGnome), OpenOil is part of the OpenDrift Lagrangian particle tracking model (Dagestad et al., 2018).

Oil surface transport in OpenOil is formulated as presented in Eq. 1. The wind drift factor (α) was fixed at 3%. The releasing point is located at 12 m depth and it takes around four minutes for the PW to reach the ocean surface. We initialise the drift simulations at the ocean surface.

OpenOil distinguishes between submerged droplets and elements that are part of a surface slick. Redistribution between these reservoirs is controlled by buoyancy, wave entrainment and vertical mixing (Nordam et al., 2019). Buoyancy depends on oil droplet density and the particle size. The droplet radius distribution depends on the viscosity, interfacial tension and wave dissipation as parameterized according to Li et al. (2017). Entrainment of surface particles is parameterized as function of wave energy dissipation from the wave model, using empirical relationships provided by Li et al. (2017). Further details on the implementation of surface interaction of particles and their behavior in the turbulent water column are given in Röhrs et al. (2018).

2.5 Simulation setup and evaluation

For each PW delineated from the SAR scenes, simulations were started 6 hrs before the acquisition time and ran for this same time interval. The ensemble simulations were performed using the ocean and atmospheric fields provided by the B-EPS, while N-HighRes forcing was used for the deterministic runs. In both cases, Stokes drift was imported from WAM, where the *e*-folding depth decay is calculated according to (Breivik et al., 2016). Two setups were defined in this work: (Setup 1) Simulations including ocean forcing, Stokes drift and the wind fields, and (Setup 2) Simulations forced solely by the wind fields. Weathering processes were not considered in the simulations.

The evaluation is divided into three different approaches, namely Member-wise Assessment, Model Comparison and Ensemble Verification. These are described below.

2.5.1 Member-wise assessment

The 656 model results (41 scenes, 8 simulations per scene and 2 setups) were evaluated against their respective PW observations (member-wise assessment). As these are simple snapshots and not time series of PW displacement, commonly used skill scores (*e.g.* the normalized cumulative Lagrangian separation (SS), Liu and Weisberg (2011)) are not applicable. We therefore considered two recently proposed metrics by Dearden et al. (2022) for instantaneous observations, the Centroid Skill Score (C_{SS}) and Area Skill Score (A_{SS}):

$$C_I = \frac{\Delta x}{L_{OBS}} \tag{2}$$

where Δx is the distance between predicted and observed centroids at a given time instance and L_{OBS} is the length of the observed oil spill. The C_{SS} is defined as:

$$\mathcal{C}_{SS} = \begin{cases} 1 - \frac{C_I}{C_{thr}}, & \text{for} \quad C_I < C_{thr} \\ 0, & \text{for} \quad C_I > C_{thr} \end{cases} \tag{3}$$

where C_{thr} is a tolerance threshold defined by the user. The A_{SS} is defined similarly as:

$$A_I = \frac{|A_{PR} - A_{OBS}|}{A_{OBS}} \tag{4}$$

which is simply the absolute difference between the predicted and observed oil spill areas, normalized by the observed area. The area skill score is then defined as:

$$\mathcal{A}_{SS} = \begin{cases} 1 - \frac{A_I}{A_{thr}}, & \text{for} \quad A_I < C_{thr} \\ 0, & \text{for} \quad A_I > C_{thr} \end{cases} \tag{5}$$

A C_{thr} value of 1, adopted in Eqs. 3 and 5, means that the model to present any skill, the predicted parameter must not exceed the magnitude of its observed counterpart. One would obtain a perfect skill score if both C_{SS} and A_{SS} = 1.

Setting the tolerance threshold is not trivial and it was previously shown to being sensible to the forecast horizon (Révelard et al., 2021) and region under investigation (de Aguiar et al., 2022). For this reason, two other metrics were applied, namely the centroid distance error (Δ*x* in Eq. 2) and offset angle (OA). The latter is the angle between modeled and observed centroids, ranging from 0° to ± 180°.

Figure 2 illustrates the verification approach. Each color represents trajectories forced by one of the eight different forcing fields, and the highlighted points are their respective centroids. Considering N-HighRes output as an example, the distance error between observed (red dot) and modeled (white) centroids is depicted as the solid black line (Δ*x*, centroid distance error) whereas the angle between the two orange lines represents the offset angle.

2.5.2 Model comparison and ensemble verification

The member-wise metrics (C_{SS}, A_{SS}, centroid distance error and offset angle) do not provide information of the quality of the ensemble system. Confidence ellipses (CE) are often used in ensemble modeling as a proxy of spread. We therefore consider CEs as follows: the cloud of particles follow a two-dimensional

10.3389/fmars.2023.1122192

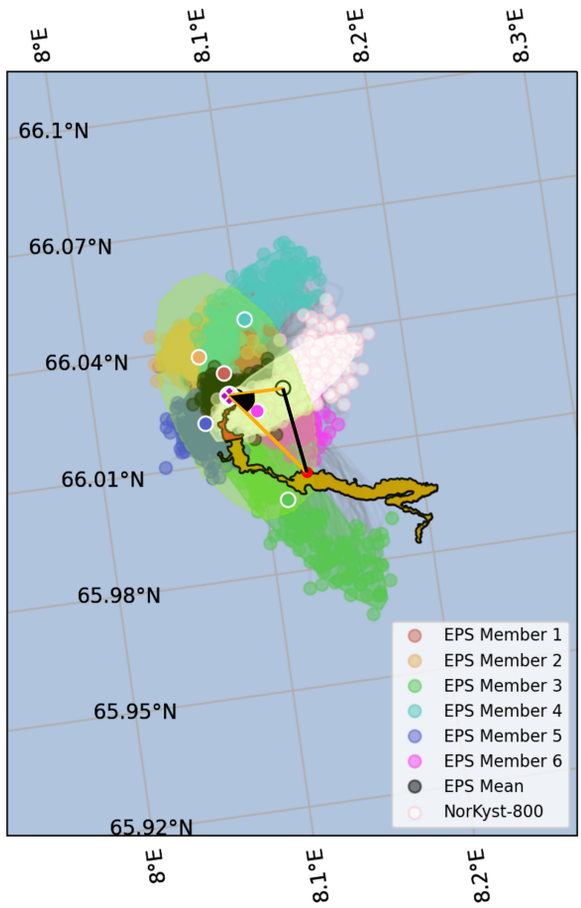

FIGURE 2
Example of PW slick simulation. The solid yellow area represents the SAR derived PW slick. Outputs forced by the different ocean-atmospheric models are shown with different colors (see legend). The white ellipse represents the 2 std confidence ellipse (CE) for N-HighRes, and modeled centroids are shown as highlighted dots. The mean B-EPS CE is shown as the pale green ellipse. The solid black line represents the distance error between the PW centroid (solid red dot) and the N-HighRes centroid (white dot), and the offset angle between them is represented by the solid black sector. Simulations are initialized at the Norne location (white-cross purple dot). Note that the N-HighRes covers only a segment of the delineated PW slick (solid red shape).

Gaussian distribution. Two-standard deviation ellipses were fitted for each of the model outputs, and the area of these were used in A_{SS} (Eqs. 4 and 5). An additional CE was fitted considering all ensemble members (mean EPS). This is illustrated in Figure 2 as ellipses enclosing each forcing with their respective colors, and a pale-green ellipse representing the mean EPS.

As the mean ensemble CE can be inflated to the point that it will likely cover the observed PW slick, we also evaluated the performance of the EPS by counting the number of observed centroids that fall within the mean EPS CE domain (hit). The relation between the area of the latter and the separation distance between modeled and observed centroids was also inspected.

For a model with statistical skill, the forecast results and the observed true state – the verification – ought to be independent draws from the same probability distribution, which may be evaluated in terms of a model and observation value histogram. In addition, the spread in forecast results among a N-member EPS is supposed to reflect the actual uncertainty in forecasts. Rank histograms and the spread-error relation are used to determine the ensemble spread. By definition, a rank histogram is built on $N + 1$ bins (Hamill, 2001). For each observed centroid position, its longitude and latitude is verified against the ranked ensemble members coordinates. A well-calibrated ensemble system evaluated at many independent cases ought to provide a uniform rank histogram, while overdispersive (underdispersive) systems will present a concave (convex) distribution. For 2D rank histograms, these distributions are seen as homogeneous, centered and clustered in the corners, respectively.

Spread (s) and error (ϵ) are hereby defined as the square root of the average ensemble variance and the root mean square error of the ensemble mean ($\bar{X}$) (Eckel and Mass, 2005), respectively. The latter is simply the ensemble mean position and it is defined as:

$$\bar{X} = \left(\frac{1}{N_{x,y}}\right)\sum_{n=1}^{N} X_{x,y} \tag{6}$$

where N is the number of ensemble members and x,y represents all the longitude and latitude positions of the virtual particles (X). The spread s is given by:

$$s = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{N-1}\sum_{n=1}^{N}(\bar{X}_{N_{x,y}} - \bar{X})^2\right)} \tag{7}$$

where $\bar{X}_{N_{x,y}}$ is the mean position of a given ensemble member and M is the number of observations. The error (ϵ) is then defined as:

$$\epsilon = \sqrt{\left(\frac{N}{N+1}\right)\frac{1}{M}\sum_{m=1}^{M}(\bar{X} - PW_{x,y})^2} \tag{8}$$

where $\mathrm{PW}_{x,y}$ is the observed PW centroid. Similarly to the rank histogram, the spread-error relationship also indicates if the EPS addresses uncertainties correctly ($\epsilon \approx s$), if it overestimates ($\epsilon << s$) or underestimates them ($\epsilon >> s$). To examine the role of the wind field on the modeled PW slick drift, we conducted the same analysis on simulations solely wind-forced (Setup 2). The work design is illustrated in Figure 3.

3 Results

3.1 PW drift simulations

3.1.1 Observed and modeled overall distributions

Heatmaps of the observed slicks, and contours representing simulations for Setup 1 (magenta) and Setup 2 (black), for B-EPS (a) and N-HighRes (b), respectively, are shown in Figure 4. Their latitudinal (right sub-panel) and longitudinal (top sub-panel) probability densities are also displayed. One can notice that the B-EPS for Setup 1 presents the highest spread and its distributions fit well to the observed. The results also indicate that differences between the two set of simulations for N-HighRes are not as pronounced as for the ensemble system.

The mean latitude and longitude positions were centered around the releasing point, as expected. The ratio between

10.3389/fmars.2023.1122192

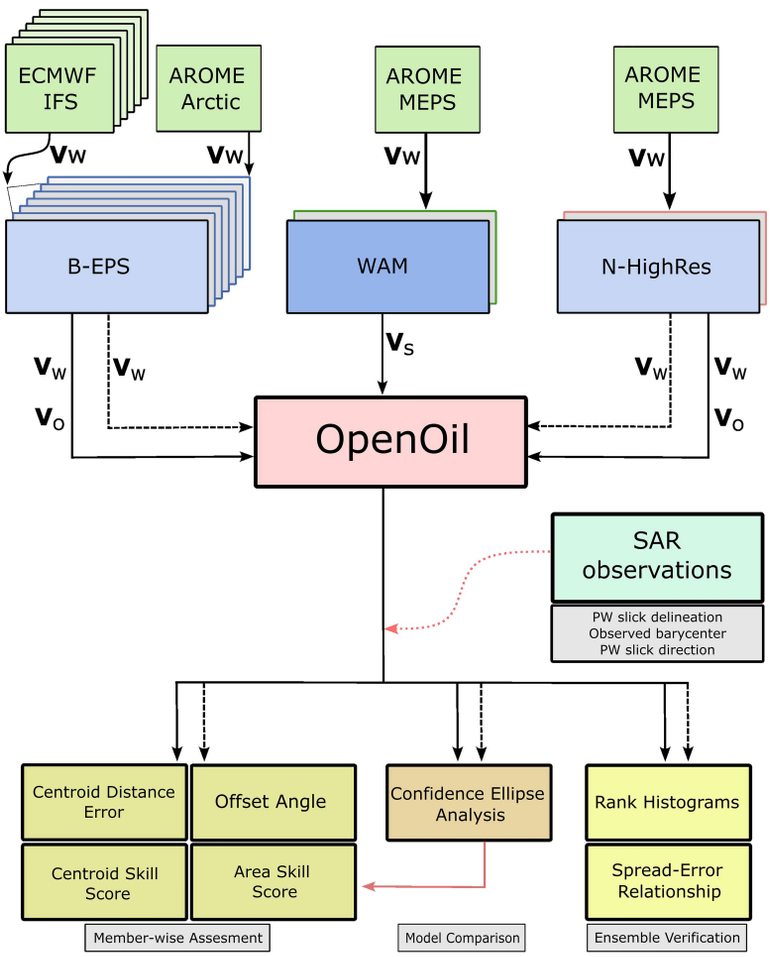

FIGURE 3
Flow chart representing the work design. Blue blocks: the three model products considered in this study as inputs. Dashed arrows represent simulations wind forced (V_w, Setup 2) while the solid lines symbolize a combination of ocean currents (V_o), Stokes drift (V_s) and the wind field (Setup 1). Red block: OpenOil, the Lagrangian modeling framework. Green block: Observed PW slicks, their centroids and drift direction obtained from Sentinel-1A/1Bscenes. Yellow blocks: Evaluation (Member-wise Assessment, Model Comparison and Ensemble Verification).

observed (Obs.) and modeled (Mod., B-EPS and N-HighRes) standard deviations of the probability distributions (P) for Setup 1 and Setup 2 are shown in Table 1. The results indicate that virtually no difference exists between B-EPS and N-HighRes in Setup 2. Including the ocean fields, increased the standard deviation in both models, with B-EPS presenting slightly better overall performance.

3.1.2 Member-wise assessment of modeled PW drift

The member-wise assessment for Setup 1 are shown in Table 2 and for Setup 2 in Table 3. N-HighRes presented slightly better performance than the EPS members in every metric but A_{SS}. One can further notice that all models have negative OA values, meaning that the modeled centroid is located to the left of the observed ones. Setup 2 provided better results for all metrics considered relative to Setup 1, with lower variability in the outputs. As for Setup 1, the modeled slicks presented a counterclockwise rotation relative to the observations.

3.2 PW slick drift and wind direction

The modeled slicks, whether including ocean forcing or not, are predominantly rotated anticlockwise relative to the observations. This finding indicates that the atmospheric models exhibit a consistent bias. We show in Figure 5 the polar histograms of speed and direction errors between modeled (B-EPS member 1 (a); N-HighRes (b)) and observed winds. Both models slightly overestimate the wind speed, presenting a mean error ($\mu_{Diff_{spd}}$) of 0.92 ms $^{-1}$ and 1.23 ms $^{-1}$, respectively. Modeled winds are rotated clockwise relative to the observations, having N-HighRes a lower mean wind direction error (9.8°) compared to B-EPS member 1 (21.3°). About 44 % and 50 % of the observations fall within the bounding boxes, respectively. A summary of the evaluation can be seen in Table 4.

The mismatch between observed wind direction and PW bearing angle is shown in Figure 6A. Due to gaps in the wind time series between April and June, 28 scenes were assessed instead of 41. Twenty-three of these presented deflection greater than 0°

10.3389/fmars.2023.1122192

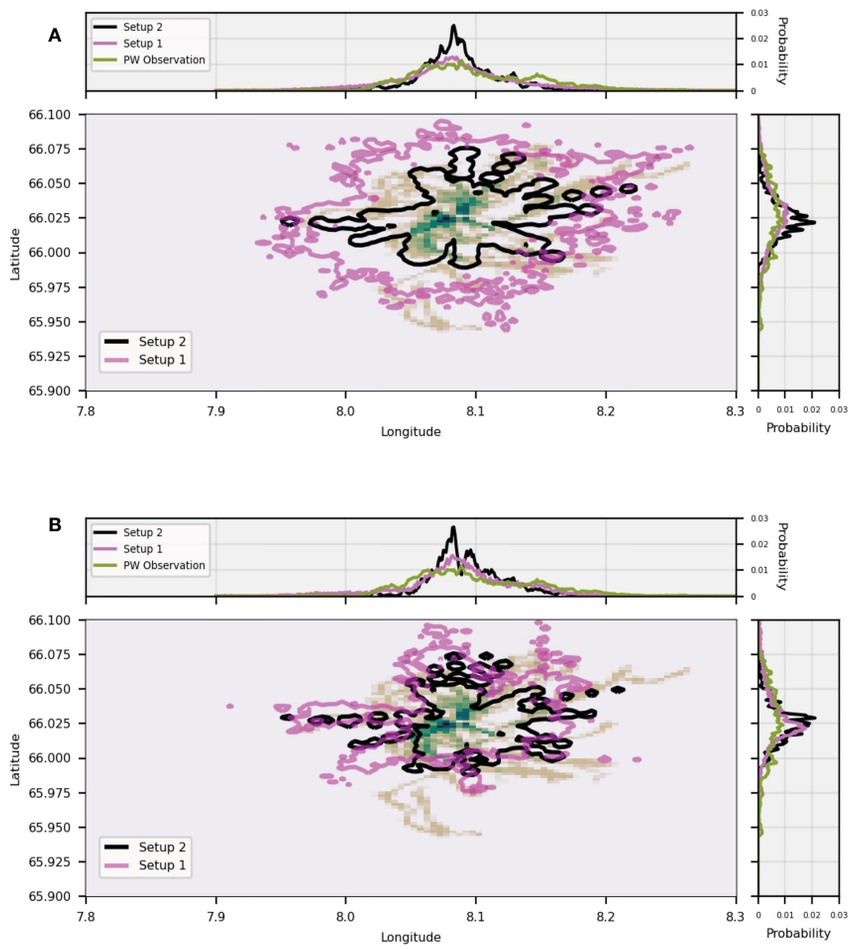

FIGURE 4
Distribution of observed (heatmap) and modeled (contours) slicks. Magenta (black) represents Setup 1 (Setup 2) simulations. Outputs in panel **(A)** were obtained with B-EPS and **(B)** with N-HighRes. The top and right sub-panels show the longitudinal and latitudinal distributions following the color legend.

relative to the observed wind direction, with the majority being concentrated between 30° and 45°. An illustrative representation of the results is show in Figure 6B.

3.2.1 Model comparison

The area covered by the mean EPS CE gives an indication of the members spread. Tables 5, 6 show the average (μ), standard deviation (σ) of the CE areas, and the number of hits (Hits) of each model product for Setup 1 and Setup 2, respectively. The results clearly show an increase in the average extent of all CEs with the inclusion of ocean currents and the Stokes drift in the simulations (Setup 1). One can further notice that the Setup 2 mean B-EPS CE shows virtually no difference of μ and σ values relative to the B-EPS members, in contrast to Setup 1 (about four-times higher).

TABLE 1 Standard deviation (σ) ratio between observed (Obs.) and modeled (Mod.) for longitudinal and latitudinal probability distributions (P) in Figure 4.

Model	$\frac{\sigma_{Lon}(P_{Obs.})}{\sigma_{Lon}(P_{Mod.})}$	$\frac{\sigma_{Lat}(P_{Obs.})}{\sigma_{Lat}(P_{Mod.})}$
B-EPS Setup 1	0.9	1
B-EPS Setup 2	1.6	1.9
N-HighRes Setup 1	1	1.3
N-HighRes Setup 2	1.6	1.9

The average (std) ratio between the mean EPS and the members CE's areas for Setup 2 was estimated to 0.99 (0.05), and 0.26 (0.17) for Setup 1. Additionally, the ensemble members had an average area about 30 % (0.01 %) higher (smaller) than N-HighRes for Setup 1 (Setup 2).

The mean EPS CE for Setup 1 also presented the highest number of hits (36), 46% higher than N-HighRes. This could indicate that the CEs are simply inflated to the point that their area are more likely to encompass the observations, but the average distance error between the confidence ellipse center and the observed PW centroid was 2 km. For Setup 2, only 2 (4) hits were registered for N-HighRes (mean B-EPS).

3.2.2 Ensemble verification

Figure 7 shows the rank histograms of Setup 1 (A) and Setup 2 simulations (B), respectively. Panels (C–E) are examples of

TABLE 2 Mean (μ) and standard deviation (σ) of the Centroid Skill Score (C $_{SS}$), Centroid Distance Error (CDE, km), Area Skill Score (A $_{SS}$) and Offset Angle (OA, °) for Setup 1.

Model	μ, σ C $_{SS}$	μ, σ CDE [km]	μ, σ A $_{SS}$	μ, σ OA [°]
B-EPS #1	0.19, 0.26	3.4, 1.9	0.07, 0.23	-33.7, 69.0
B-EPS #2	0.22, 0.26	3.0, 1.4	0.08, 0.25	-11.0, 75.0
B-EPS #3	0.25, 0.27	3.1, 1.9	0.07, 0.23	-23.5, 68.2
B-EPS #4	0.25, 0.28	3.0, 1.9	0.04, 0.14	-15.3, 67.2
B-EPS #5	0.28, 0.31	2.6, 1.3	0.11, 0.28	-19.8, 74.0
B-EPS #6	0.24, 0.27	2.8, 1.3	0.09, 0.24	-35.5, 46.6
B-EPS mean	0.33, 0.31	2.2, 1.3	0.10, 0.26	-20.7, 49.8
N-HighRes	0.35, 0.33	2.1, 1.1	0.05, 0.17	-28.2, 38.9

overdispersive, consistent and underdispersive rank histograms created from synthetic data randomly sampled from Normal distributions with the same mean (0) and decreasing standard deviation (10, 1, 0.1). Comparing these to the obtained results, it is possible to notice the resemblance between Setup 2 rank histograms (B), and the underdispersive case (E), and Setup 1 (A) with panel (D).

The ratio between error (ϵ) and spread (s) (Eqs. 7 and 8) for an ideal ensemble system should be approximately 1. The previous findings are supported by $\frac{\epsilon}{s} \approx 1.3$ for Setup 1 and $\frac{\epsilon}{s} \approx 69$ for Setup 2.

4 Discussion

4.1 Representation of PW slick drift using ocean forecast models

The results presented in the previous section indicate that the EPS reproduced the overall variability of PW slick drift (Figure 4) and performs similarly to N-HighRes in a member-wise level (Tables 2, 3). The skill scores obtained for the latter are on average slightly higher than for B-EPS, though the difference is small, especially when considering the ensemble mean. It is also worth noticing that the Centroid Skill Score and Area Skill Score did not present sensitivity relative to 'bad' simulations (*e.g.*, Figures 8C, F), whereas the unskilled predictions were addressed as skillful (not shown). We stress that our work did not attempt to model the shape of the slicks as they are too narrow (between 85 m to 200 m) to be resolved by the operational models.

For their short spatial and temporal scales, PW slicks are highly influenced by higher frequency, evanescent small scale features with life spans from less than 5 hrs (Kirincich, 2016) to shorter than he inertial period (Callies et al., 2020). These phenomena are unconstrained by current observation systems, resulting in lower predictive skill of higher resolution models (Sandery and Sakov, 2017; Jacobs et al., 2021).

Eddy-resolving ocean models may directly improve the statistics of Lagrangian currents due to their better representation of the kinetic energy spectrum at small scales. However, the use of high-resolution ocean models may not translate into improved predictive skill due to misrepresentation of the circulation features in space or time (Révelard et al., 2021). Neither B-EPS nor N-HighRes are expected to exhibit predictive skill for surface currents beyond wind-driven and bathymetry-constrained flow, but rather represent uncertainties in such unconstrained scales.

TABLE 3 Same as Table 2, but for only wind forced simulations (Setup 2).

Model	μ , σ C $_{SS}$	μ, σ CDE [km]	μ, σ A $_{SS}$	μ, σ OA [°]
B-EPS #1	0.40, 0.30	1.7, 0.96	0.12, 0.25	-22.8, 45.5
B-EPS #2	0.41, 0.29	1.7, 0.96	0.12, 0.24	-22.7, 45.4
B-EPS #3	0.40, 0.30	1.7, 0.96	0.13, 0.26	-22.7, 45.3
B-EPS #4	0.41, 0.29	1.7, 0.96	0.12, 0.24	-22.8, 45.5
B-EPS #5	0.41, 0.29	1.7, 0.96	0.12, 0.25	-23.7, 45.5
B-EPS #6	0.41, 0.29	1.7, 0.97	0.12, 0.23	-23.8, 46.7
B-EPS mean	0.40, 0.30	1.7, 0.96	0.13, 0.25	-22.7, 45.4
N-HighRes	0.40, 0.31	1.7, 0.91	0.13, 0.26	-24.3, 47.5

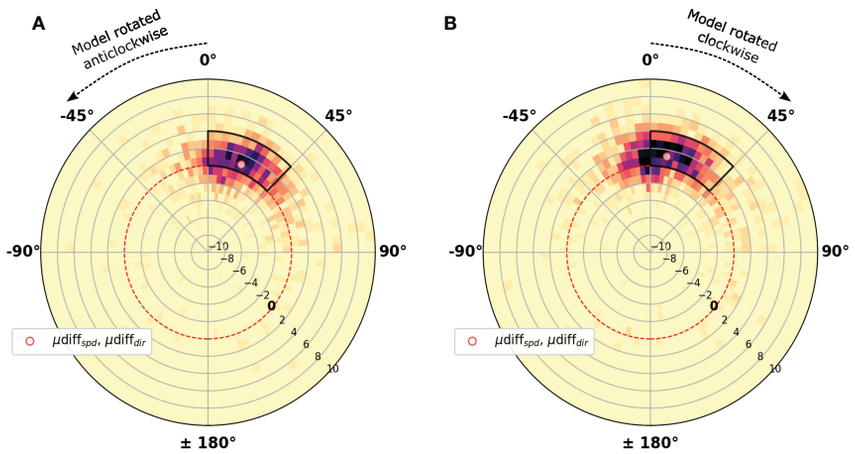

FIGURE 5
Polar heatmaps of wind speed (radius, m s $^{-1}$) and direction (angle) errors between model (V $_{wm}$) and observation (V $_{wo}$), Apr - Dec 2021. **(A)** for B-EPS member 1 and **(B)** for N-HighRes. Positive angles represent modeled winds rotated clockwise relative to the observed wind direction and positive values in radius indicate V $_{wm}$ > V $_{wo}$. The red dashed line indicates where V $_{wm}$ = V $_{wo}$, the white dot indicates the mean speed and mean direction errors for each model and the black polygon the region of highest concentration.

4.2 Modeling oil drift using ensemble forecasting

Confidence ellipses have been used as an indirect tool by the atmospheric community for the assessment of ensemble spread for more than a decade. In our analysis, we found a 36/41 (≈87.8%) hit ratio for the ensemble simulations in Setup 1. This is fairly close to the theoretical value of 86.5 % associated to the considered two standard deviations confidence interval (Wang et al., 2015). The discrepancy of μ between B-EPS members and the mean B-EPS CE areas for Setup 1 (see Tables 5) entails from diverging trajectories, otherwise, the average mean B-EPS CE area should be roughly similar to those of the ensemble members. We therefore stress that the mean EPS CE areas are not arbitrarily inflated, but their growth rather reflects the B-EPS spread induced by the uncertainties in the ocean field. Converging to Melsom et al. (2012) and Sandu et al. (2020) findings, we also observed slightly better accuracy of the mean EPS CE (2 km) relative to N-HighRes (2.1 km).

Despite showing better performances in all member-wise metrics, Setup 2 results are indistinguishable among the distinct model products (see Tables 1, 3 and 6). Such lack of variability stems from higher homogeneity of atmospheric horizontal flows in comparison to the ocean (Figure 2), *i.e.*, the simulated trajectories experience essentially alike wind conditions. Additionally, due to horizontal (500 km) and temporal (6 hrs) decorrelation scales adopted on ECMWF-ENS's perturbation schemes, the ensemble spread of winds is generally low at the spatial (1-15 km) and temporal (1-6 hrs) ranges investigated here. We acknowledge that uncertainties in the wind field are the main sources of variability in large and persistent oil slick scenarios (e.g. Li et al., 2019; Kampouris et al., 2021), but we have evidences that it is overridden by the ocean currents on short scales.

These findings corroborate with the rank histograms (Figure 7). The uniform- and cornered-like patterns obtained for Setup 1 and Setup 2, respectively, fits the artificial examples of consistent and underdispersive cases. A uniform histogram is a necessary, but not sufficient condition for determining the reliability of an ensemble system due to conditional biases, nonrandom sampling or low number of samples (Hamill, 2001). For finite size ensemble systems, the ideal error-spread ratio is altered by an adjusting

TABLE 4 Willmott Skill (WS), Bias and Root Mean Square Error (RMSE) for wind speed (spd) and direction (dir). Each atmospheric product was assessed relative to wind observations at Norne between April and December, 2021.

Model	WS $_{spd}$	WS $_{dir}$	Bias $_{spd}$	Bias $_{dir}$	RMSE $_{spd}$	RMSE $_{dir}$
B-EPS # 1	0.90	0.81	0.92	21.2	2.2	89.6
B-EPS # 2	0.90	0.81	0.77	21.1	2.1	88.7
B-EPS # 3	0.91	0.81	0.73	20.7	2.1	88.6
B-EPS # 4	0.90	0.80	0.77	19.6	2.1	90.0
B-EPS # 5	0.90	0.80	0.70	21.1	2.2	91.6
B-EPS # 6	0.91	0.81	0.70	18.9	2.0	89.7
N-HighRes	0.90	0.84	1.23	9.85	2.3	83.4

10.3389/fmars.2023.1122192

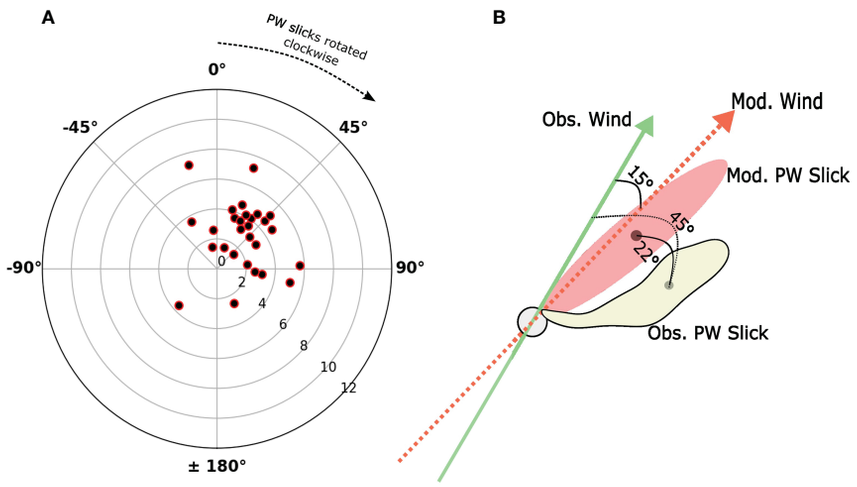

FIGURE 6
(A) Offset angle between observed PW centroid and observed wind directions. Radius values indicate the observed wind speed (ms $^{-1}$). **(B)** Schematic illustration of the slick deviation, where modeled slicks (*Mod. PW Slick*) veered counterclockwise (≈22°) relative to the observed slicks (*Obs. PW Slick*). The latter were deflected about 45° to the right of the observed wind (*Obs. Wind*). Modeled winds (*Mod. Wind*) were veered clockwise around 15° relative to *Obs. Wind*.

factor $\frac{\epsilon}{s} = 1 + \sqrt{\frac{2}{N-1}}$, ≈ 1.2 for $N = 6$ (Eckel and Mass, 2005). The error-spread ratio analysis confirmed that Setup 2 simulations presented a value ($\frac{\epsilon}{s} \approx 69$) overly above the desired ($\frac{\epsilon}{s} \approx 1.2$). Taking into account the ocean currents and the Stokes drift greatly improved the spread among the EPS members, resulting in $\frac{\epsilon}{s} \approx 1.3$. Despite the proximity of the releasing location to the boundary domain (≈ 30 km), these results show that the ensemble spread in surface currents is readily developed at the study region.

Due to the number of possible future states that EPS provide, often exceeding 20 members, ensemble modeling has been used by weather prediction centers to forecast the strength, probability of occurrence and regions impacted by hazardous events. Figure 8 shows examples of PW drift simulations with different degrees of ensemble spreading. Panels (B, C) and (E, F) illustrate cases where particles follow a well-defined path and responders can have higher confidence to base their decision plan. Note that low spread does not necessarily imply accurate predictions as all the members can converge to a wrong predictions (Figures 8C, F). Decision making is hindered in the case that simulations diverge or present bifurcations, as shown in Figures 8A, D. In this case, the drift uncertainty is higher and this leads to a more problematic recovery/rescue planning.

TABLE 5 Average (μ, km ^{2}) and standard deviation (σ, km ^{2}) of the confidence ellipse area (CE) of each model product and the mean B-EPS for Setup 1.

Model	μ CE Area [km ^{2}]	σ CE Area [km ^{2}]	Hits
B-EPS # 1	11.9	7.0	15
B-EPS # 2	9.8	5.3	10
B-EPS # 3	10.6	6.1	15
B-EPS # 4	10.1	5.7	12
B-EPS # 5	7.8	4.2	11
B-EPS # 6	10.0	5.3	17
B-EPS mean	7.2	3.0	12
N-HighRes	8.7	4.3	19
mean B-EPS	44.7	21.9	36

The number of hits represents how many times the observed centroid fell within the respective CE domain (Hits).

4.3 The misalignment between observed slicks and the wind direction

An offset in the drift direction between modeled and observed slicks was found for both Setup 1 and 2. Such mismatch can result from circulation features not resolved by the ocean model, but also from errors in the description of Ekman transport in the boundary layer schemes for both the ocean and the atmosphere models (Sandu et al., 2020) The latter is evidenced in Figure 5 and Table 4 as bias in wind speed and direction, leading to the observed systematic offset in drift directions for the oil slick simulations

Stokes drift (Röhrs et al., 2012), Langmuir circulation (Yang et al., 2014) and wave-current interactions (Staneva et al., 2021) were shown to impact drift trajectories and improve predictive skill when included. While the Stokes drift is explicitly included in the trajectory simulations, the used ocean and wave model setups are not coupled, meaning that wave-current interactions are not represented in the ocean model (Geernaert, 1993; Chen et al., 2020). The influence of these phenomena are small when considered independently, but introduce further errors on short-term predictions not accounted for in operational models. The Stokes drift has been shown to have an important contribution to

10.3389/fmars.2023.1122192

TABLE 6 Same as Table 5, but for Setup 2.

Model	μ CE Area [km ^{2}]	σ CE Area [km ^{2}]	Hits
B-EPS # 1	1.5	0.81	2
B-EPS # 2	1.6	0.82	3
B-EPS # 3	1.6	0.84	3
B-EPS # 4	1.6	0.89	3
B-EPS # 5	1.6	0.81	2
B-EPS # 6	1.6	0.86	3
B-EPS mean	1.6	0.83	2
N-HighRes	1.7	0.91	4
mean B-EPS	1.6	0.83	2

surface slick motion, as discussed in Röhrs et al. (2012); Jones et al. (2016); van den Bremer and Breivik (2018).

Trajectory models are at the endpoint of a complex chain of observation systems and forecast models. Beyond the impact of internal and external sources of uncertainties (Barker et al., 2020), drift modeling is also sensible to releasing time (Li et al., 2019). Although released constantly, the residence time of PW slick at the ocean surface is unknown, hence the considered forecast time (6 hrs) is arbitrary. Another drawback of this study is related to the delineation process as this depends on the operator experience and it is subjective. We recommend that further investigations on drift ensemble modeling should link remotely-sensed observations and *in-situ* data, preferably oil droplet diameter and drifter trajectories. It would also be useful to test the influence of perturbations in the model physics, boundary conditions and wave-ocean interactions on the drift modeling.

5 Concluding remarks

Here we assessed the capability of an ocean ensemble model (B-EPS) to represent oil slick drift and its uncertainty, compared to a higher resolution deterministic ocean model (N-HighRes). The simulated trajectories were conducted with the open-source Lagrangian framework OpenDrift, and forced by these two ocean models in addition to wind forcing and Stokes drift (Setup 1). The predictions were evaluated against Produced Water slicks delineated from 41 Sentinel-1A/1B scenes over the Norne platform between April and December, 2021. Three approaches were considered for the verification: Member-wise Assessment, Model Comparison and Ensemble Verification. The importance of the wind field on the modeled trajectories was also investigated by forcing the virtual particles only with atmospheric models (Setup 2).

Produced water slicks are thin films with low oil concentrations, generally observed as narrow stripes and lasting for a short period. For short range predictions as considered here (6 hrs), we showed that simulations forced solely by the wind fields are underdispersive.

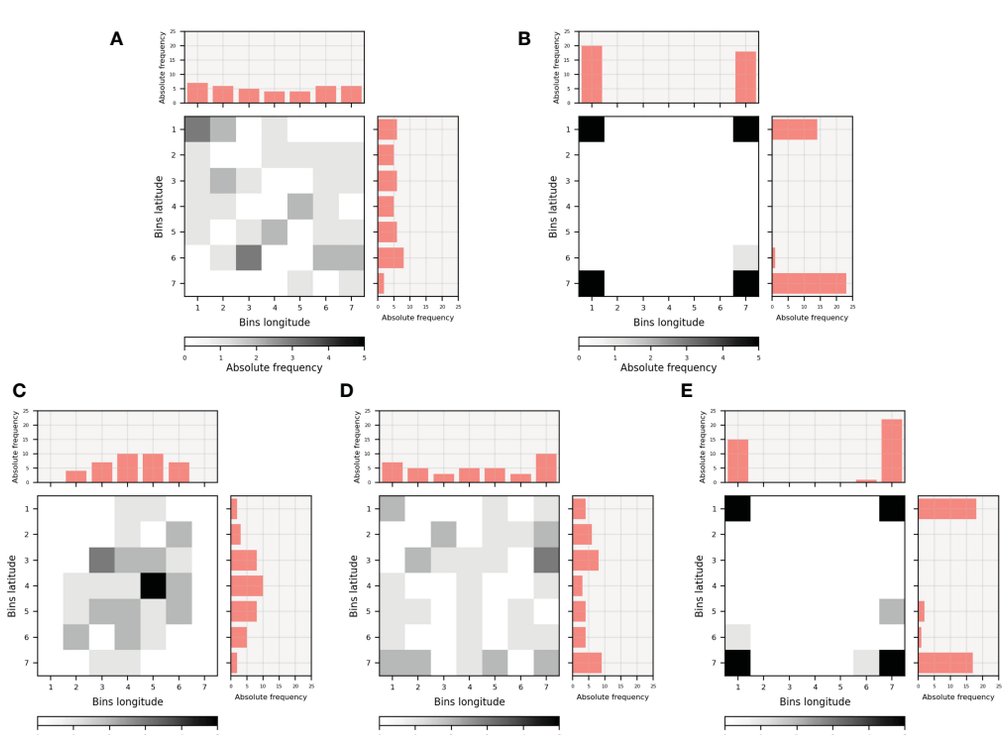

FIGURE 7
Two-dimensional rank histograms of B-EPS for Setup 1 **(A)** and Setup 2 **(B)** in absolute frequency. Panels **(C–E)** represent overdispersive, consistent and underdispersive rank histograms created from synthetic data randomly sampled from Normal distributions with the same average (0) and decreasing std (10, 1, 0.1), respectively. Top and right sub-panels show the longitudinal and latitudinal rank histograms.

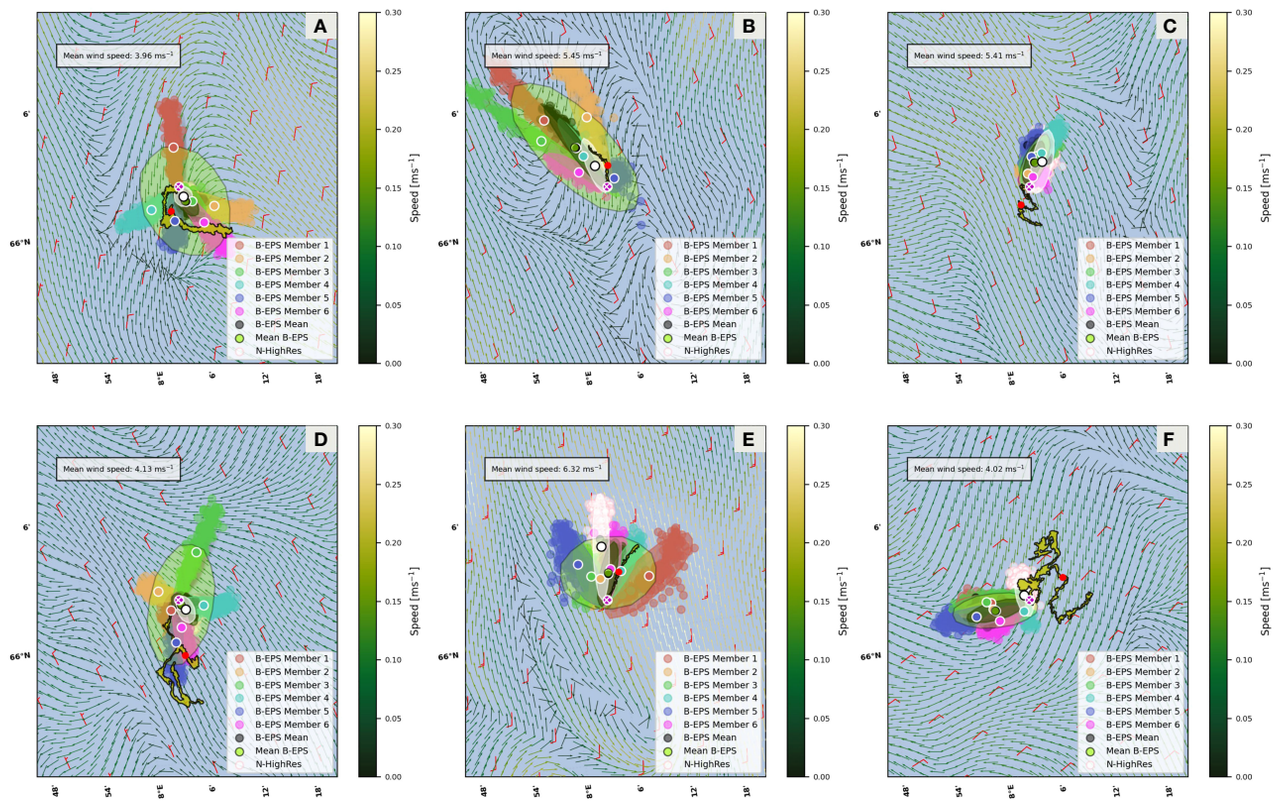

FIGURE 8
Example of high spread, partially accurate **(A, D)**; low spread, accurate **(B, E)** and low spread, inaccurate **(C, F)** simulations. The white-crossed, magenta dot represents Norne. Wind barbs (red) and ocean current vectors with superimposed speed (colormap, ms $^{-1}$) were obtained from N-HighRes. The modeled mean wind speed (ms $^{-1}$) is also shown on the top left box. Respective Figures for all 41 PW slick drift cases are provided as electronic supplement to this document.

For larger and more persistent slicks, the wind forcing may still be considered the main source of variability.

Despite its coarser horizontal resolution, the ensemble prediction system, and especially its mean field, presented similar member-wise results relative to the deterministic ocean model. We have also shown that the observed PW drift presented a consistent clockwise deviation from modeled and observed wind directions, possibly resultant from systematic errors in the description of boundary layer schemes and wave-current interactions not resolved in the used models. The analysis of confidence ellipses and distributions of drift patterns show that realistic drift modeling requires to include both the ocean and wind forcing, and analysis of ensemble spread indicates that including the currents is necessary to properly address uncertainties in drift modeling.

Data availability statement

All geophysical forcing data is available at https://thredds.met.no/thredds/catalog.html and the level 1 Sentinel-1A/Sentinel-1B data are available at https://creodias.eu/. Wind measurements are freely available at https://seklima.met.no/. OpenDrift can be downloaded at https://opendrift.github.io/. Produced water slick masks can be provided upon request.

Author contributions

Study design and conception: VA, JR, AJ, and TE. Organization of SAR database: VA. Trajectory simulations and analysis: VA. Produced water data set and analysis: AJ. Manuscript initial draft: VA. Manuscript editing: all authors. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Research Council of Norway (RCN) project Center for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) (237906).

Acknowledgments

Copernicus Sentinel data 2021 for Sentinel data. Equinor is gratefully acknowledged for providing data about produced water releases from the Norne platform. We thank the reviewers for the constructive comments and suggestions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1122192/full#supplementary-material

References

Alpers, W., Holt, B., and Zeng, K. (2017). Oil spill detection by imaging radars: challenges and pitfalls. *Remote Sens. Environ.* 201, 133–147. doi: 10.1016/j.rse.2017.09.002

Asplin, L., Albretsen, J., Johnsen, I. A., and Sandvik, A. D. (2020). The hydrodynamic foundation for salmon lice dispersion modeling along the Norwegian coast. *Ocean Dynamics* 70, 1151–1167. doi: 10.1007/s10236-020-01378-0

Barker, C., Kourafalou, V., Beegle-Krause, C., Boufadel, M., Bourassa, M., Buschang, S., et al. (2020). Progress in operational modeling in support of oil spill response. *J. Mar. Sci. Eng.* 8. doi: 10.3390/jmse8090668

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., et al. (2017). The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Monthly Weather Rev.* 145, 1919–1935. doi: 10.1175/MWR-D-16-0417.1

Breivik, Ø., Bidlot, J.-R., and Janssen, P. A. (2016). A Stokes drift approximation based on the Phillips spectrum. *Ocean Model.* 100, 49–56. doi: 10.1016/j.ocemod.2016.01.005

Breivik, O., Janssen, P. E. A. M., and Bidlot, J.-R. (2014). Approximate Stokes drift profiles in deep water. *J. Phys. Oceanogr.* 44, 2433–2445. doi: 10.1175/JPO-D-14-0020.1

Brekke, C., Espeseth, M., Dagestad, K.-F., Röhrs, J., Hole, L. R., and Reigber, A. (2021). Integrated analysis of multisensor datasets and oil drift simulations — a free-floating oil experiment in the open ocean. *J. Geophysical Research: Oceans* 126, e2020JC016499. doi: 10.1029/2020JC016499

Buizza, R. (2019). Introduction to the special issue on "25 years of ensemble forecasting. *Q. J. R. Meteorol. Soc.* 145, 1–11. doi: 10.1002/qj.3370

Callies, J., Barkan, R., and Garabato, A. (2020). Time scales of submesoscale flow inferred from a mooring array. *J. Phys. Oceanogr.* 50, 1065–1086. doi: 10.1175/JPO-D-19-0254.1

Chen, S., Qiao, F., Xue, Y., Chen, S., and Ma, H. (2020). Directional characteristic of wind stress vector under swell-dominated conditions. *J. Geophysical Research: Oceans* 125, e2020JC016352. doi: 10.1029/2020JC016352

Dagestad, K.-F., and Röhrs, J. (2019). Prediction of ocean surface trajectories using satellite derived vs. modeled ocean currents. *Remote Sens. Environ.* 223, 130–142. doi: 10.1016/j.rse.2019.01.001

Dagestad, K.-F., Röhrs, J., Breivik, Ø., and Ådlandsvik, B. (2018). Opendrift v1.0: a generic framework for trajectory modelling. *Geoscientific Model. Dev.* 11, 1405–1420. doi: 10.5194/gmd-11-1405-2018

de Aguiar, V., Dagestad, K.-F., Hole, L., and Barthel, K. (2022). Quantitative assessment of two oil-in-ice surface drift algorithms. *Mar. pollut. Bull.* 175, 113393. doi: 10.1016/j.marpolbul.2022.113393

Dearden, C., Culmer, T., and Brooke, R. (2022). Performance measures for validation of oil spill dispersion models based on satellite and coastal data. *IEEE J. Oceanic Eng.* 47, 126–140. doi: 10.1109/JOE.2021.3099562

De Dominicis, M., Bruciaferri, D., Gerin, R., Pinardi, N., Poulain, P., Garreau, P., et al. (2016). "A multi-model assessment of the impact of currents, waves and wind in modelling surface drifters and oil spill," in *Deep Sea research part II: topical studies in oceanography*, vol. 133. , 21–38. doi: 10.1016/j.dsr2.2016.04.002

Dong, H., Zhou, M., Hu, Z., Zhang, Z., Zhong, Y., Basedow, S. L., et al. (2021). Transport barriers and the retention of calanus finmarchicus on the Lofoten shelf in early spring. *J. Geophysical Research: Oceans* 126, e2021JC017408. doi: 10.1029/2021JC017408

Eckel, F., and Mass, C. (2005). Aspects of effective mesoscale, short-range ensemble forecasting. *Weather Forecasting* 20, 328–350. doi: 10.1175/WAF843.1

ECMWF (2012). "IFS documentation CY38R1, part VII: ECMWF wave model," in *ECMWF Model Documentation, European Centre for Medium-Range Weather Forecasts*.

Fingas, M., and Brown, C. (2018). A review of oil spill remote sensing. *Sensors* 18. doi: 10.3390/s18010091

Frogner, I.-L., Andrae, U., Ollinaho, P., Hally, A., Hämäläinen, K., Kauhanen, J., et al. (2022). Model uncertainty representation in a convection-permitting ensemble–SPP and SPPT in HarmonEPS. *Monthly Weather Rev.* 150, 775–795. doi: 10.1175/MWR-D-21-0099.1

Geernaert, G. (1993). Characteristics of the magnitude and direction of the wind stress vector over the sea. *J. Mar. Syst.* 4, 275–287. doi: 10.1016/0924-7963(93)90014-D

Gusdal, Y., and Carrasco, A. (2012). "Validation of the operational wave model WAM at met.no - report 2011," in *Tech. rep., Norwegian Meteorological Institute, Oslo. Report No. 23/2012 gusdal validation2012*.

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Rev.* 129, 550–560. doi: 10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2hamillinterpretation2001

Jacobs, G., D'Addezio, J. M., Ngodock, H., and Souopgui, I. (2021). Observation and model resolution implications to ocean prediction. *Ocean Model.* 159, 101760. doi: 10.1016/j.ocemod.2021.101760

Johansson, A., Skrunes, S., Brekke, C., and Isaksen, H. (2021). "Multi-mission remote sensing of low concentration produced water slicks," in *EUSAR 2021; 13th European Conference on Synthetic Aperture Radar*. 1–6 johansson2021.

Jones, C., Dagestad, K.-F., Breivik, Ø., Holt, B., Röhrs, J., Christensen, K., et al. (2016). Measurement and modeling of oil slick transport. *J. Geophysical Research: Oceans* 121, 7759–7775. doi: 10.1002/2016JC012113

Jorda, G., Comerma, E., Bolaños, R., and Espino, M. (2007). Impact of forcing errors in the CAMCAT oil spill forecasting system. a sensitivity study. *J. Mar. Syst.* 65, 134–157. doi: 10.1016/j.jmarsys.2005.11.016

Kampouris, K., Vervatis, V., Karagiorgos, J., and Sofianos, S. (2021). Oil spill model uncertainty quantification using an atmospheric ensemble. *Ocean Sci.* 17, 919–934. doi: 10.5194/os-17-919-2021

Khade, V., Kurian, J., Chang, P., Szunyogh, I., Thyng, K., and Montuoro, R. (2017). Oceanic ensemble forecasting in the gulf of Mexico: an application to the case of the Deep Water Horizon oil spill. *Ocean Model.* 113, 171–184. doi: 10.1016/j.ocemod.2017.04.004

Kirincich, A. (2016). The occurrence, drivers, and implications of submesoscale eddies on the Martha's Vineyard inner shelf. *J. Phys. Oceanogr.* 46, 2645–2662. doi: 10.1175/JPO-D-15-0191.1

Komen, G. J., Cavaleri, L., Doneland, M., Hasselmann, K., Hasselmann, S., and Janssen, P. A. E. (1994). *Dynamics and modelling of ocean waves* (Cambridge: Cambridge University Press).

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

Li, Y., Yu, H., Wang, Z.-y., Li, Y., Pan, Q.-q., Meng, S.-j., et al. (2019). The forecasting and analysis of oil spill drift trajectory during the Sanchi collision accident, East China Sea. *Ocean Eng.* 187, 106231. doi: 10.1016/j.oceaneng.2019.106231

Liu, Y., and Weisberg, R. H. (2011). Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. *J. Geophysical Research: Oceans* 116. doi: 10.1029/2010JC006837

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmospheric Sci.* 20, 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Lu, Y., Shi, J., Hu, C., Zhang, M., Sun, S., and Liu, Y. (2020). Optical interpretation of oil emulsions in the ocean – part II: applications to multi-band coarse-resolution imagery. *Remote Sens. Environ.* 242, 111778. doi: 10.1016/j.rse.2020.111778

Melsom, A., Counillon, F., Lacasce, J., and Bertino, L. (2012). Forecasting search areas using ensemble ocean circulation modeling. *Ocean Dynamics* 62. doi: 10.1007/s10236-012-0561-5

Miljørapport (2019). Olje og gassindustriens miljøarbeid. fakta og utviklingstrekk. *Tech. report Norsk olje gass milio*.

Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M.A.Ø., Noer, G., and Korosov, A. (2017a). Characteristics of a convective-scale weather forecasting system for the European Arctic. *Monthly Weather Rev.* 145, 4771–4787. doi: 10.1175/MWR-D-17-0194.1

Müller, M., Homleid, M., Ivarsson, K.-I., Koltzow, M. A. O., Lindskog, M., Midtbø, K. H., et al. (2017b). AROME-MetCoOp: a Nordic convective-scale operational weather prediction model. *Weather Forecasting* 32, 609–627. doi: 10.1175/WAF-D-16-0099.1

Nordam, T., Dunnebier, D., Beegle-Krause, C., Reed, M., and Slagstad, D. (2017). Impact of climate change and seasonal trends on the fate of Arctic oil spills. *Ambio* 46, 442–452. doi: 10.1007/s13280-017-0961-3

Nordam, T., Nepstad, R., Litzler, E., and Röhrs, J. (2019). On the use of random walk schemes in oil spill modelling. *Mar. pollut. Bull.* 146, 631–638. doi: 10.1016/j.marpolbul.2019.07.002

Olita, A., Fazioli, L., Tedesco, C., Simeone, S., Cucco, A., Quattrocchi, G., et al. (2019). Marine and coastal hazard assessment for three coastal oil rigs. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00274

Révelard, A., Reyes, E., Mourre, B., Hernández-Carrasco, I., Rubio, A., Lorente, P., et al. (2021). Sensitivity of skill score metric to validate Lagrangian simulations in coastal areas: recommendations for search and rescue applications. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.630388

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

Röhrs, J., Dagestad, K.-F., Asbjørnsen, H., Nordam, T., Skancke, J., Jones, C. E., et al. (2018). The effect of vertical mixing on the horizontal drift of oil spills. *Ocean Sci.* 14, 1581–1601. doi: 10.5194/os-14-1581-2018

Röhrs, J., Gusdal, Y., Rikardsen, E., Duran Moro, M., Brændshøi, J., Kristensen, N. M., et al. (2023a). Barents-2.5km v2.0: an operational data-assimilative coupled ocean and sea ice ensemble prediction model for the Barents Sea and Svalbard. *Geoscientific Model. Dev. Discussions* 2023, 1–31. doi: 10.5194/gmd-2023-20

Röhrs, J., Sutherland, G., Jeans, G., Bedington, M., Sperrevik, A., Dagestad, K.-F., et al. (2023b). Surface currents in operational oceanography: key applications, mechanisms, and methods. *J. Operational Oceanogr.* 16, 60–88. doi: 10.1080/1755876X.2021.1903221

Sakov, P., and Oke, P. (2008). A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters. *Tellus A: Dynamic Meteorol. Oceanogr.* 60, 361–371. doi: 10.1111/j.1600-0870.2007.00299.x

Sandery, P., and Sakov, P. (2017). Ocean forecasting of mesoscale features can deteriorate by increasing model resolution towards the submesoscale. *Nat. Commun.* 8, 1–8. doi: 10.1038/s41467-017-01595-0

Sandu, I., Bechtold, P., Nuijens, L., Beljaars, A., and Brown, A. (2020). On the causes of systematic forecast biases in near-surface wind direction over the oceans. doi: 10.21957/wggbl43u

Sepp Neves, A., Pinardi, N., Navarra, A., and Trotta, F. (2020). A general methodology for beached oil spill hazard mapping. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00065

Shchepetkin, A. F., and McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Model.* 9, 347–404. doi: 10.1016/j.ocemod.2004.08.002

Skrunes, S., Johansson, A., and Brekke, C. (2019). Synthetic aperture radar remote sensing of operational platform produced water releases. *Remote Sens.* 11. doi: 10.3390/rs11232882

Staneva, J., Ricker, M., Carrasco Alvarez, R., Breivik, Ø., and Schrum, C. (2021). Effects of wave-induced processes in a coupled wave–ocean model on particle transport simulations. *Water* 13. doi: 10.3390/w13040415

Sutherland, G., Soontiens, N., Davidson, F., Smith, G., Bernier, N., Blanken, H., et al. (2020). Evaluating the leeway coefficient of ocean drifters using operational marine environmental prediction systems. *J. Atmospheric Oceanic Technol.* 37, 1943–1954. doi: 10.1175/JTECH-D-20-0013.1

Svejkovsky, J., Hess, M., Muskat, J., Nedwed, T., McCall, J., and Garcia, O. (2016). Characterization of surface oil thickness distribution patterns observed during the Deepwater Horizon (MC-252) oil spill with aerial and satellite remote sensing. *Mar. pollut. Bull.* 110, 162–176. doi: 10.1016/j.marpolbul.2016.06.066

Thoppil, P., Frolov, S., Rowley, C., Reynolds, C., Jacobs, G., Metzger, E., et al. (2021). Ensemble forecasting greatly expands the prediction horizon for ocean mesoscale variability. *Commun. Earth Environ.* 2, 1–9. doi: 10.1038/s43247-021-00151-5

van den Bremer, T. S., and Breivik, O. (2018). "Stokes drift," in *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences*, vol. 376. , 20170104. doi: 10.1098/rsta.2017.0104

Villalonga, M., Infantes, M., Colls, M., and Ridge, M. (2020). Environmental management system for the analysis of oil spill risk using probabilistic simulations. application at Tarragona Monobuoy. *J. Mar. Sci. Eng.* 8. doi: 10.3390/jmse8040277

Wang, B., Shi, W., and Miao, Z. (2015). Confidence analysis of standard deviational ellipse and its extension into higher dimensional euclidean space. *PloS One* 10, 1–17. doi: 10.1371/journal.pone.0118537

Willmott, C. (1981). On the vaidation of models. *Phys. Geogr.* 2, 184–194. doi: 10.1080/02723646.1981.10642213

Xie, J., Bertino, L., Counillon, F., Lisæter, K. A., and Sakov, P. (2017). Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013. *Ocean Sci.* 13, 123–144. doi: 10.5194/os-13-123-2017

Yang, D., Chamecki, M., and Meneveau, C. (2014). Inhibition of oil plume dilution in Langmuir ocean circulation. *Geophysical Res. Lett.* 41, 1632–1638. doi: 10.1002/2014GL059284

Zodiatis, G., De Dominicis, M., Perivoliotis, L., Radhakrishnan, H., Georgoudis, E., Sotillo, M., et al. (2016). "The Mediterranean decision support system for marine safety dedicated to oil slicks predictions," in *Deep Sea research part II: topical studies in oceanography*, vol. 133. , 4–20. doi: 10.1016/j.dsr2.2016.07.014

/7

Paper II: Performance diagnostics for probabilistic Lagrangian drift prediction

de Aguiar, V., Idžanović, M., Röhrs, J. and Johansson, M. Under review at Journal of Operational Oceanography.

Performance diagnostics for probabilistic Lagrangian drift prediction

Victor de Aguiar[a*]*, Martina Idžanović[b] , Johannes Röhrs[b], Malin Johansson[a]

[a] *Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway*

[b] *Division for Ocean and Ice, Norwegian Meteorological Institute, Oslo, Norway*

October 13, 2024

Abstract

Trajectory forecasting based on geophysical models is a useful tool for contingency and emergency aid at sea. Uncertainties in the used geophysical models, which propagate into the trajectory forecast, can be addressed through ensemble modeling. Here, we evaluate the performance of an operational ensemble prediction system on short-term forecasts for 17 undrogued drifter trajectories deployed in the Barents Sea and Fram Strait in 2022. Predicted and observed trajectories were compared by rotary spectra analysis, rank histogram, reliability diagram, and error/spread to determine the ability of the model to reproduce the observed physical processes and their uncertainties. We found that the physical processes dominating the observed spectra at inertial and subinertial frequencies are accounted for in the modeled trajectories, but that the model underestimates the energy content for higher frequencies (> 0.083 cph) with up to two orders of magnitude. Ensemble underdispersion is linked to model error rather than initial and boundary conditions. Reliability is achieved if the main forcings are accurately reproduced by the geophysical models. For highly dynamic regions, such as the Fram Strait, transient small-scale phenomena representation is critical at the uppermost ocean layer for accurate trajectory forecasting.

**Corresponding author. Email: victor.d.aguiar@uit.no

Keywords: Trajectories, Ensamble prediction systems, Drift modelling, In-situEnsemble prediction, trajectory modeling, surface drift, reliability, model validation

1 Introduction

Predicting the drift of objects and substances at the ocean's surface is pertinent to a wide range of maritime applications, e.g. in search-and-rescue operations (e.g. Breivik and Allen, 2008), pollution tracking (e.g. Röhrs et al., 2018) and biological modeling (e.g. Asplin et al., 2014). Drift predictions require dynamic modeling of the geophysical flow because the oceanic circulation is not readily observed to a degree that allows sufficient spatio-temporal coverage (Röhrs et al., 2023b). As a result of insufficient observation coverage, forecast models exhibit large uncertainties (Jacobs et al., 2021). For operational monitoring of objects and their drift trajectories in real time, no remote sensing system for ocean currents provide sufficient temporal resolution (de Aguiar et al., 2023). Instead, satellite-tracked drifters are widely used for drift prediction studies and following objects in real time (e.g. Jones et al., 2016; Garcia-Pineda et al., 2020).

Geophysical models are the backbones of any ocean drift prediction application (Christensen et al., 2018). The accuracy of the simulated target path heavily depends on how well the geophysical flow is reproduced. The turbulent behaviour of the ocean and atmosphere, the intrinsic chaotic properties, and limitations of the numerical prediction systems hinder forecast accuracy beyond a few days. Previous studies have shown that ocean models struggle to accurately resolve drifter motions within the inertial band (e.g. Brekke et al., 2021; Arbic et al., 2022), and hence modeled trajectories often diverge from observed already in the first hours of simulations.

For a generic drift model, the velocity of a particle at the ocean surface ($\mathbf{V}_p$) is composed of the linear sum of Eulerian ocean currents ($\mathbf{V}_O$), a fraction of the wind speed $\mathbf{V}_W$ and the Stokes drift ($\mathbf{V}_S$) as

$$\mathbf{V}_p\Big|_{z=0} = \mathbf{V}_O\Big|_{z=-h} + (\alpha\mathbf{V}_W)\Big|_{z=10} + \mathbf{V}_S\Big|_{z=0} \tag{1}$$

where α is a wind drift factor that can be a combination of air drag on the object and direct wind-driven currents unresolved by the used ocean model. Modeled ocean surface currents ($\mathbf{V}_O$) represent the depth-integrated motion between the ocean surface and a reference depth (z=-h) that matches the

application (Röhrs et al., 2023b). The value for the wind drift factor varies from 1% to 6% (Huang, 1983; Röhrs and Christensen, 2015), and decreases as the vertical resolution of the used ocean model increases (Callies et al., 2017) due to a more explicit description of the direct wind forcing on upper ocean currents. In operational setups, the wind and Stokes drift fields are obtained from numerical prediction models, and they are often referenced at 10 m height (z=10) and at the ocean free-surface (z=0), respectively.

Addressing the relative importance of each velocity term in Eq. 1 is an active research field as it determines to what extent each phenomena can be parameterized instead of numerically resolved. For short-term simulations (less than 24 hours), ocean currents present the dominant role on the drift variability and uncertainty (Röhrs and Christensen, 2015; Zhang et al., 2020; de Aguiar et al., 2023), shifting to wind-dominant for periods beyond the inertial period ($T = \frac{2\pi}{f}$, where f is the Coriolis parameter in $\mathrm{rad\,s^{-1}}$) (Li et al., 2013, 2019; Kampouris et al., 2021).

Short-term predictions are critical for emergency response and preparedness, and hence uncertainties and errors in the geophysical forcing must be quantified to allow adequate interpretation of model forecasts. Ensemble Prediction Systems (EPS) are widely used to address uncertainties in numerical weather prediction (NWP) models. Instead of providing deterministic forecasts based on one model realisation, an EPS pursues a probabilistic approach and provides the likelihoods of scenarios to users in order to support decision making (Leutbecher and Palmer, 2008; Palmer, 2017). The performance of an EPS is established in terms of reliability, ensemble spread, and skill as outlined in Hamill (2001) and Bröcker and Smith (2007). For drift trajectories however, ensemble generation has been traditionally performed through multi-model ensemble (e.g. Zodiatis et al., 2016) or by perturbations of uncertain parameters such as the wind-drift factor α in Eq. 1. Although the relative mean distance error can be reduced using both methods, such ensembles do not map out the possibility space of ocean circulation realisations (Leutbecher et al., 2017). Instead, ensemble prediction of drift trajectories should be built on EPSs of the underlying geophysical models.

In this work, we assess the performance of the Barents-2.5 ocean EPS using two surface drifter trajectory data sets deployed in 2022 in the Fram Strait and Barents Sea. Short-term trajectory forecasts (five days) are performed from each member of the EPS. Observed and modeled trajectories are evaluated through rotary spectra analysis, and new methodologies are employed for assessing the performance of the Lagrangian drift simulation and ensemble prediction. Possible sources of model error and insufficient

ensemble spread are reasoned. This work is structured as follows: Section 2 introduces the drifter data sets, model products, simulation setups, and study area. The spectral analysis and ensemble quality results are presented in Section 3 and discussed in Section 4. The work is summarized in Section 5.

2 Data and Methods

In this section, we first introduce the drifter data set and used model products. Thereafter, we document analysis methods to interpret the response of drifter motion to geophysical forcing – through rotary co-spectra analysis – and methods to assess the performance of the probabilistic trajectory forecasts in Section 2.3.

2.1 Drifter data and study area

Seventeen surface drifters were deployed during two cruises organized by the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Centre for Arctic Gas Hydrates, Environment and Climate (CAGE). The drifter trajectories are shown in Figure 1. The drifters (OpenMetBuoy-v2021) were built following the design presented in Rabault et al. (2022). The sensors were assembled inside a 12 cm×12 cm×9 cm acrylic box, undrogued such that it drifts partially submerged (approx. 30% in water) at the ocean surface. Geographic positions were recorded every thirty minutes with an accuracy of roughly 2.5 m - 10 m. The sampling period was downsampled to one hour to match the model output frequency.

During the CIRFA cruise (Dierking et al., 2022), starting from Svalbard and crossing the Fram Strait towards the east Greenland coast (Belgica Bank), a total of 11 OpenMetBuoy-v2021 were deployed between 2022-04-22 and 2022-05-09. The CAGE cruise (Serov et al., 2022) took place between 2022-08-02 and 2022-08-20, and went from Svalbard to Tromsø, Norway. Three OpenMetBuoy-v2021 were deployed on August 7, 2022 in Hopendjupet, northern Barents Sea, and another three were released south of Bear Island 11 days later. Appendix A provides an overview of deployment positions, acquisition periods (start and end time), and observed speeds (average and standard deviation).

Both the Fram Strait and the Barents Sea receive inflow water masses from the Norwegian Atlantic Slope Current (NwASC). Flowing along the Norwegian continental slope, the NwASC splits at around 72°N, with one branch (North Cape Current, NCaC) entering the Barents Sea trough the Barents Sea

Opening (BSO) while the other, topographic steered, reaches the Fram Strait alongshore the west coast of Svalbard (West Spitsbergen Current, WSC), and merges with the East Spitsbergen Current (ESC). Together with the Norwegian Coastal Current (NCC), the NCaC heads eastward, as depicted by the southernmost CAGE drifter trajectories. A ramification, however, proceeds towards the central Barents Sea and interacts with the westward Persey Current (PC) originated from the eastern part of the basin, and the NCaC itself around the area of the northernmost CAGE drifter deployment. This location ($\approx$ 75°N) is a critical latitude for the semi-diurnal M_2 tide component, with the amphidromic point roughly located at $\approx$ 75°N and $\approx$ 30°E. On the eastern side of the Fram Strait, where the CIRFA drifters were deployed, the WSC is characterized by high eddy kinetic energy, with maximum (minimum) values in January-February (September-October) (Wekerle et al., 2017), and with average radius and lifetime of 5 km and 10 days, respectively (Wekerle et al., 2020). Another branch of the WSC recirculates into the Fram Strait and merges into the East Greenland Current, flowing southward along the shelf break.

2.2 Geophysical model data

2.2.1 Ocean and atmosphere circulation models

Ocean currents used for the trajectory simulations are obtained from Barents-2.5, an operational ocean EPS for the regional circulation in the Barents Sea and Fram strait (Figure 1) with 24 ensemble members (Röhrs et al., 2023). Barents-2.5 EPS is built on the Regional Ocean Modeling System (ROMS) version 3.7 (Shchepetkin and McWilliams, 2005). ROMS solves the Reynolds averaged, hydrostatic primitive equations using a bottom-following coordinate system with free surface (Shchepetkin and McWilliams, 2005). It has a horizontal resolution of 2.5 km with 42 stretched bottom following vertical layers, where the thickness of the uppermost layer varies between 0.5 m and 1 m. Turbulent mixing is parameterized using prognostic equations for turbulent kinetic energy and a generic length scale (Umlauf and Burchard, 2005). Boundary conditions are provided by the TOPAZ4 ocean-ice coupled model (Xie et al., 2017) and tidal forcing from TPXO 7.2 (Egbert and Erofeeva, 2002). Further details on Barents-2.5 EPS can be found in Röhrs et al. (2023), and validation of ocean currents forecasts is provided in Idžanović et al. (2023).

Wind forcing for the trajectory model, as well as the atmospheric forcing for the ocean model are provided by two separate weather prediction models: Four ensemble members use a time-lagged ensemble

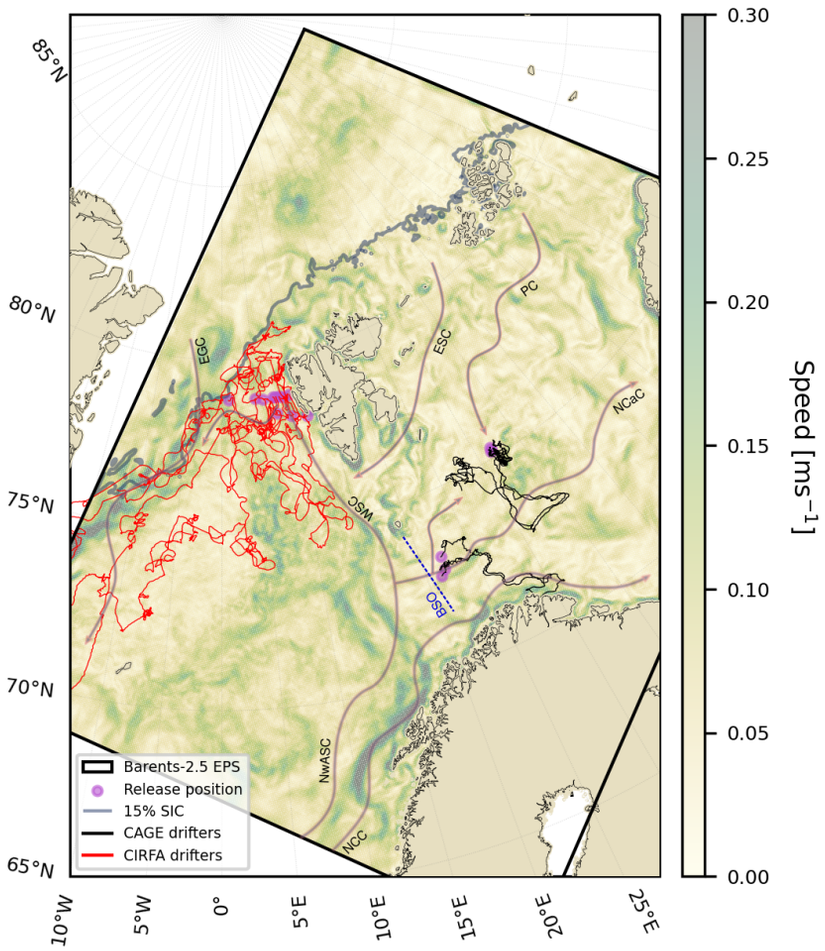

Figure 1: Overview of the study area. The black rectangle represents the Barents-2.5 EPS domain. Red and black solid lines represent the CIRFA and CAGE trajectories, respectively. Arrows depict the mean ocean currents in the region with their associated acronyms (see text in Section 1 for reference). The blue dotted line represents the Barents Sea Opening (BSO), while the background map shows the instant ocean current speed from Barents-2.5 EPS (in yellow-green shading) for 2nd August 2022. The gray contour represents the sea-ice edge for concentrations equal to 15%. Purple dots are the drifters' deployment location.

of the regional AROME-Arctic weather prediction model (Müller et al., 2017) and 20 ensemble members are forced by members of EC-ENS, a global ensemble NWP system from the European Centre for Medium-Range Weather Forecast. While EC-ENS provides 6-hourly forcing at around 10 km resolution, AROME-Arctic delivers hourly forcing on the same grid as Barents-2.5 with 2.5 km resolution. The time resolution of ECMWF-IFS fields are upsampled through linear interpolation to provide hourly input for the trajectory simulations.

2.2.2 Ensemble Prediction System

The model setup of Barents-2.5 EPS consists of 24 unique ensemble members. Six members are produced at four bulletin times during the day (at 00:00, 06:00, 12:00, and 18:00). While each member is forced by various atmospheric ensemble realisations as described above, each member is initialized by the state of the same member from the day before, and adjusted by small analysis increments from a data assimilation scheme (see Röhrs et al., 2023). Each of the 24 members, while integrated throughout subsequent forecast cycles, is hence providing independent realisations of the ocean circulation. The ensemble spread, e.g. the degree by which current forecasts of various members differ, is a consequence of (i) spread in the atmospheric forcing, and (ii) the full integration history of the respective member, as initially small perturbations grow over time to cause unique circulation features.

The data assimilation in Barents-2.5 EPS consists of inclusion of satellite retrievals of sea surface temperature and sea-ice concentration (Rusin et al., 2024) as well as *in-situ* observations of temperature and salinity. Analysis increments for each member are calculated using the Ensemble Kalman Filter data assimilation scheme (Evensen, 1994), as described in Röhrs et al. (2023). In brief, each ensemble member is adjusted for observed sea surface temperature, sea-ice concentration, and in-situ hydrography once a day. The analysis increments are small enough to pertaining ensemble spread.

2.2.3 Stokes drift from a wave prediction model

The wave prediction model MyWaveWam 4 km provides the Stokes drift $\mathbf{V}_S$ as used in Eq. 1 for the trajectory model. MyWaveWam4 is based on WAM Cycle 4.7, providing hourly forecasts of the wave field and Stokes drift (Gusdal and Carrasco, 2012). The model physics is described in Komen et al. (1994). The Stokes drift is computed by integration of the direction surface wave energy spectra F(f,θ) as follows (Janssen and Behrens, 2013):

$$\mathbf{V}_S = \frac{4\pi}{g} \int_0^{2\pi} \int_0^{+\infty} f\mathbf{k}F(f,\theta)\, df\, d\theta \tag{2}$$

where g is the acceleration of gravity (ms^{-2}) and $\mathbf{k}$ is the wave number vector. The herein used MyWaveWAM setup discretizes the wave spectra using 36 directions and 36 frequencies. Table 1 summarizes the model inputs from each geophysical model as used for the trajectory simulations.

Table 1: Summary of geophysical model data used in this study.

Model	**Type**	**number of inputs**	**Resolution horizontal/temporal resolution**	**Variables**[a]
Barents-2.5 EPS	Ensemble	24	2.5 km / 1 h	V_O
MyWaveWAM	Deterministic	1	4 km / 1 h	V_S
Arome-Arctic	Ensemble	4	2.5 km / 1 h	V_W
EC-ENS	Ensemble	20	2.5 km / 1 h	V_W

[a]: as in Eq. 1.

2.3 Methods

Trajectory simulations were performed using the OpenDrift particle tracking model (Dagestad et al., 2018), where modeled outputs were linearly interpolated onto the virtual particle's coordinates. Velocity data is time-integrated using a 4th order Runge-Kutta advection scheme to yield trajectories using the interpolated ocean currents, Stokes Drift, and wind fields as forcing as in Eq. 1. The wind drift factor (α) was set at 0.01 (1%), and it was obtained through optimizing α in Eq. 1 with observed drifter velocities and the modeled fields. This value is low compared to empirical values of 0.02-0.04 to only account for direct wind drag on the drifter, because the Stokes Drift is added explicitly in Eq. 1, and wind induced ocean currents are included in the used current velocities (Röhrs et al., 2012; Wagner et al., 2022). Integration of hourly velocity fields using a 4th order scheme provides a numeric precision that accounts for the spatio-temporal variability of the used forcing data (Nordam and Duran, 2020; Xiong and MacCready, 2024).

The number of available ensemble members used for the trajectory simulations was extended to 48 by incorporating data from the previous forecast cycle of the ocean model. (black box in Figure 2). This approach allows us to exploit the model's full forecast lead time of 66 hours. Applying this window over the period of the observed trajectory data rendered an archive of ocean and atmospheric forecasts used for the trajectory as follows: 24 trajectory ensemble members employ forecast data from the latest ocean and atmosphere ensemble of 24 members. The additional 24 ensemble members (red bars in Figure 2) were initialized 24 (bulletin 4), 30 (bulletin 5), 36 (bulletin 6) and 42 (bulletin 7) hours prior to the analysis (bulletin 0). We refer to this extended set as the first 24 members (42 - 24 hours prior to analysis), whereas the primary set (yellow bars in Figure 2) is referred to as the last 24 members (18 - 0 hours prior to analysis).

The trajectory predictions are integrated for five days for each of the extended 48-member ensemble.

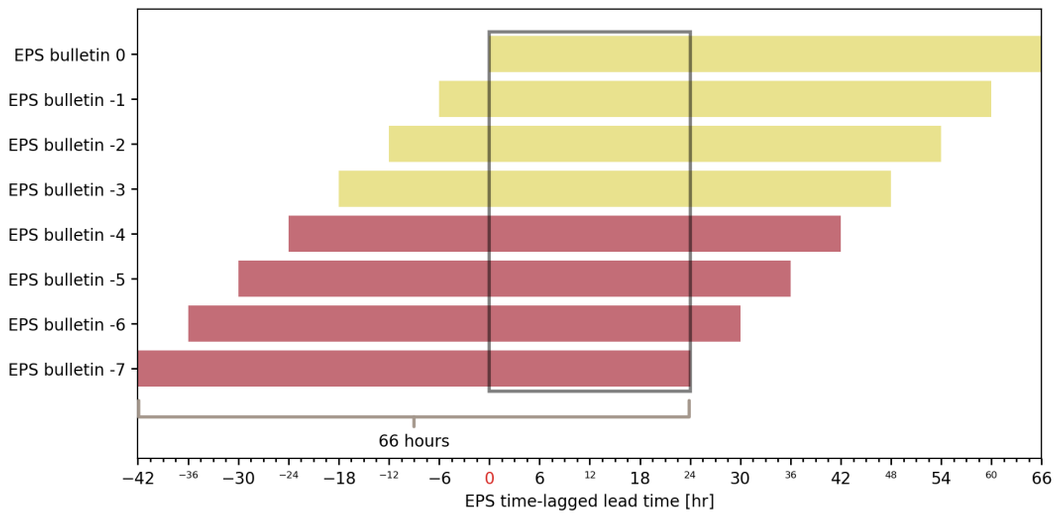

Figure 2: Representation of the extended ensemble approach. The EPS bulletin 0 depicts the analysis time. Yellow bars represent four EPS bulletins (0 to -3) in a forecast cycle, each providing 6 ensemble members. Red bars represent four extra bulletins (-4 to -7) obtained from a previous forecast cycle (42 - 24 hours prior to analysis). These two sets, namely the first and last 24 members respectively, overlap within a 24-hour window (black box) yielding 48 ensemble members. See Section 2.3 for more details.

Simulations are re-initialized every day with the respective drifters observed location and time, yielding overlapping trajectories for the 5-day simulations. As modeled trajectories depart from the observations as time evolves, re-initialization allows comparison of model vs. observed trajectory segments estimations at the same geographical location (Liu and Weisberg, 2011; Dagestad and Röhrs, 2019).

Using the trajectory simulations, we assess (i) spectral characteristics of observed and modeled drifter velocities for 5-day segments and (ii) the performance of the trajectory ensemble to represent uncertainty in the drift predictions for 1-day segments (see Appendix A). The 5-day segments overlap in time, but sample slightly different model trajectories due to the evolving model error. The overlap makes best use of all available data and enhances the spectral energy estimation by providing more spectral estimates.

2.3.1 Spectral analysis

For the purpose of assessing the spectral characteristics, a rotary spectra analysis is performed as follows: the horizontal velocity components of the observed and simulated trajectories are obtained for 5-day segments (120 samples). For each 5-day segment, a Blackman window is applied to the velocity component anomalies with a 50% overlap. The rotary spectra is then estimated following Gonella (1972), in which the velocity component anomalies (u_x, u_y) are represented in the complex form as:

$$u(t) = u_x(t) + iu_y(t), \tag{3}$$

with the associated Fourier transform given by:

$$u_\omega = \frac{1}{d}\int_0^d u(t)e^{-i\omega t}\,dt, \tag{4}$$

d being the duration of the time series and ω the angular velocity. The sign of the latter determines cyclonic (S_+, counterclockwise, positive frequencies) and anticyclonic (S_-, clockwise, negative frequencies) motions with the following energy spectra:

$$\begin{aligned} S_+ &= \frac{1}{2} < u_+ u_+^* > = \frac{1}{8}(P_{u_x u_x} + P_{u_y u_y} + 2Q_{u_x u_y}) \\ S_- &= \frac{1}{2} < u_- u_-^* > = \frac{1}{8}(P_{u_x u_x} + P_{u_y u_y} - 2Q_{u_x u_y}), \end{aligned} \tag{5}$$

where $P_{u_x u_x}$ and $P_{u_y u_y}$ are the autospectra of u_x and u_y, and $Q_{u_x u_y}$ is the quadrature spectrum between the components. The brackets $< \cdot >$ represent block averages of 120 samples and the asterisk denotes the complex conjugate. As the same concept can be expanded for a given pair of complex vector series *u(t)* and *v(t)*, the coherence (Eq. 6) between the modeled and observed spectra can be estimated as:

$$C(\omega) = \frac{P_{uv} - iQ_{uv}}{\sqrt{P_{uu}P_{vv}}}, \tag{6}$$

where P_{uv} is the rotary cross spectra between the vectors.

2.3.2 Performance metrics for modeled ensemble trajectories

The performance of the EPS is assessed through the average model error (ϵ) and ensemble spread (s), two-dimensional (2D) rank histograms, and reliability diagrams. All the performance metrics described below are performed over 24-hour segments after initialization.

Error and ensemble spread are defined as the root mean square error of the ensemble mean with respect to the observations and the square root of the ensemble variance, respectively (Fortin et al., 2014). For a given time step t, the ensemble mean position $\overline{X}$(x,y,t) is calculated as

$$\overline{X}(x,y,t) = \frac{1}{N}\sum_{n=1}^{N} X_{x,y,t}, \tag{7}$$

where N is the number of ensemble members (48), and x and y represent the longitude and latitude position of all virtual particles. The error and ensemble spread are hence defined as:

$$\epsilon = \sqrt{\left(\frac{N}{N+1}\right)\frac{1}{T}\sum_{t=1}^{T}(\overline{X}(t) - O_{x,y,t})^2} \text{ and}$$
$$s = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N-1}\sum_{n=1}^{N}(X_{N_{x,y,t}} - \overline{X}(x,y,t))^2\right)}, \tag{8}$$

where $X_{N_{x,y,t}}$ is the position of a given ensemble member, T is the length of the segment (24 hours), and $\mathrm{O}_{x,y,t}$ is the observed drifter position. We then calculate both error spread as a distance metrics (in kilometers) from the latitude and longitude pairs. The $\frac{\epsilon}{s}$ ratio and the rank histograms determine the ensemble's consistency (Eckel and Mass, 2005). The estimation of $\frac{\epsilon}{s}$ was performed over the first 24 hours after each initialization. For an ideal ensemble spread, the ratio ought to be ≈ 1, whereas underdispersive (overdispersive) behaviour is found for values greater (smaller) than 1. Suitable ensemble spread means that a validating observation is statistically indistinguishable from a random draw of the ensemble forecast.

The ensemble spread is also analysed in terms of a rank histogram built on N+1 bins (Hamill, 2001), where N represents the number of ensemble members. The concept can be expanded to 2D variables (de Aguiar et al., 2023), where geographical positioning (latitude and longitude) is used instead of along-track distance. For a consistent EPS, the 2D rank histogram ideally provides a homogeneous distribution when evaluated over many independent cases. For overdispersive and underdispersive systems, the distributions are centered and clustered at the corners of the 2D plan, respectively.

An EPS can estimate the probability of events, e.g. surpassing a certain threshold of drift displacement within a specified time. Reliability diagrams are used to check the accuracy of these probabilities by comparing them to observed events (Wilks, 2019). The method consists of defining a binary event (it either happened or it did not) and comparing the EPS-estimated probability of the event with the frequency of the event in the observations. Consider the along-track cumulative distance estimated over the first 24 hours after each initialization. An example event is A="distance is less than 10 km", for which in every trajectory point (x, y, t) either occurs or does not occur, i.e. $A(x, y, t) = \{0, 1\}$. The EPS provides the predicted probability P of this event A_{mod} at a given time as $P(x, y, t) = \sum_{n=1}^{N} A_{mod}(x, y, t)$. Within

each probability bin b, the collection is represented by the mean of the forecast probabilities P_b. The probability bins b range from 0 to 1 in steps of 0.1. To make it comparable to observations, we consider the same event, but this time we only look at the N_b points in the model that fall within b represented as $X_b \in x_b, y_b, t_b$. By assessing the event in the observations at the N_b points, we can describe the observed frequency in probability bin b as $F_b = \sum_{b=1}^{N_b} A_{obs}(X_b)/N_b$.

In a reliable EPS, the observed frequency F_b matches the forecast probability P_b for each bin resulting in a straight diagonal line in a reliability diagram. Deviations from the diagonal are due to (i) imperfect ensemble spread and/or (ii) model biases. An EPS with insufficient spread will have higher model probabilities for low observed frequencies, and lower model probabilities for high observed frequencies, resulting in a reversed S-shaped reliability diagram. A normal S-shape indicates excessive model spread. A biased EPS exhibits too low or too high probabilities for all observes frequencies, resulting in an upward or downward shifted curve in the reliability diagram (Bröcker and Smith, 2007).

3 Results

An overview of simulated ensemble trajectories and their along-track cumulative distance relative to the corresponding observed data set is presented in this section. The rotary auto- and cross-spectra analysis is introduced in Subsection 3.2 and the ensemble-performance diagnostic in Subsection 3.3.

3.1 Trajectory simulation

Figure 3 shows two examples for (a) the CAGE and (b) the CIRFA data sets. Observed trajectories are depicted as solid black lines. The yellow dots represent initialization points every 24 hours, and the gray lines represent 1-day predicted trajectories forced by the 48 ensemble members. By visual inspection, observed trajectories are seemingly well represented in the Barents Sea, with the forecast ensemble envelopes located around the observed path. Less accurate trajectories are visible for the CIRFA drifters, overall in the vicinity of loops and meanders. Modeled trajectories are also shorter relative to the observed segments. No perceptible differences between the first and last 24 ensemble members for either of the data sets were noticed. Similar results were found for the other drifters.

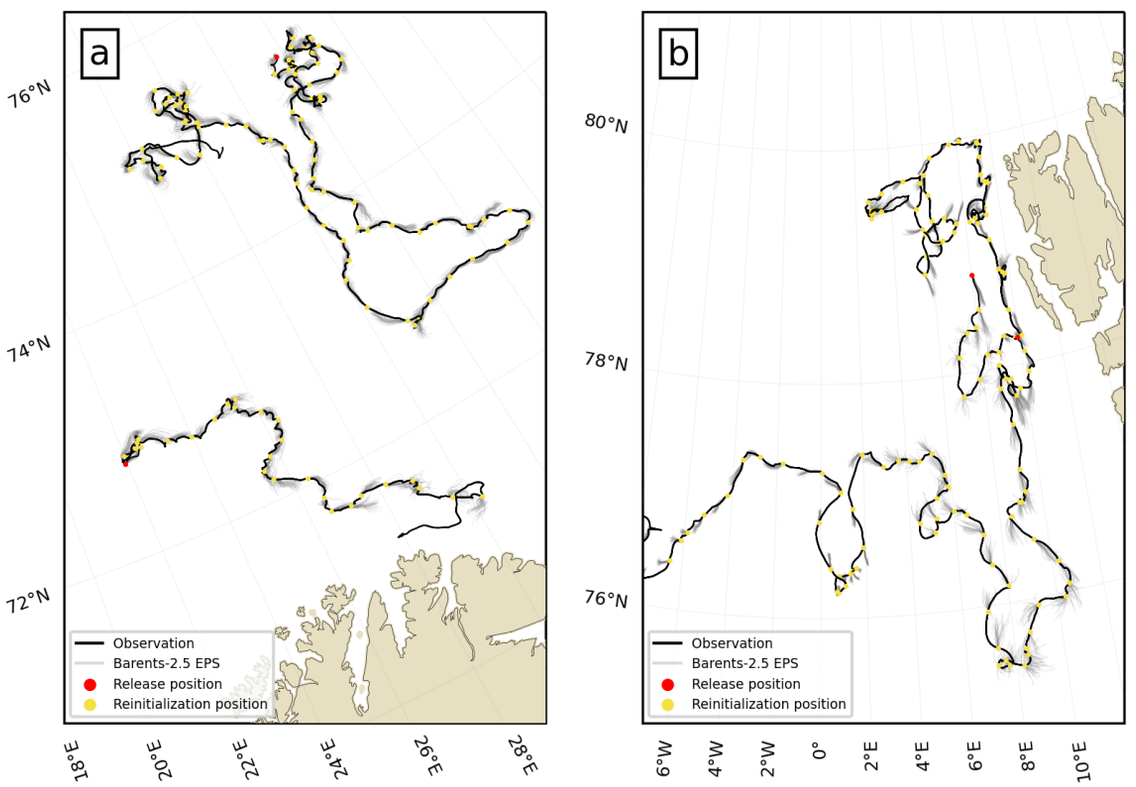

Figure 3: Examples of predicted trajectories for the CAGE (a, drifters 206633 and drifter 206650) and CIRFA (b, drifter 15_waves_LSM and drifter 4) drifter trajectories. Solid black lines represent observed trajectory whereas gray lines are modeled trajectories for 48 ensemble members, 24 hours after initialization. Red dots are positions where the devices where deployed and yellow dots represent the model re-initialization positions (every 24 hours).

Figure 4 shows the distribution of along-track cumulative distances traveled by (a) CAGE and (b) CIRFA drifters over 24 hours segments (green histogram). Simulations forced by the first (black line) and last 24 members (blue line) of Barents-2.5 EPS are also shown. One can first notice that the two sets of ensemble members show essentially no differences. For the CAGE data set, the simulated trajectories represented well the observed distribution, slightly under (over) estimating frequencies for distances shorter (larger) than 15 km. Conversely, overestimation of modeled trajectories for the CIRFA drifters occur for distances smaller than 10 km, presenting also larger deviations relative to the observed distribution.

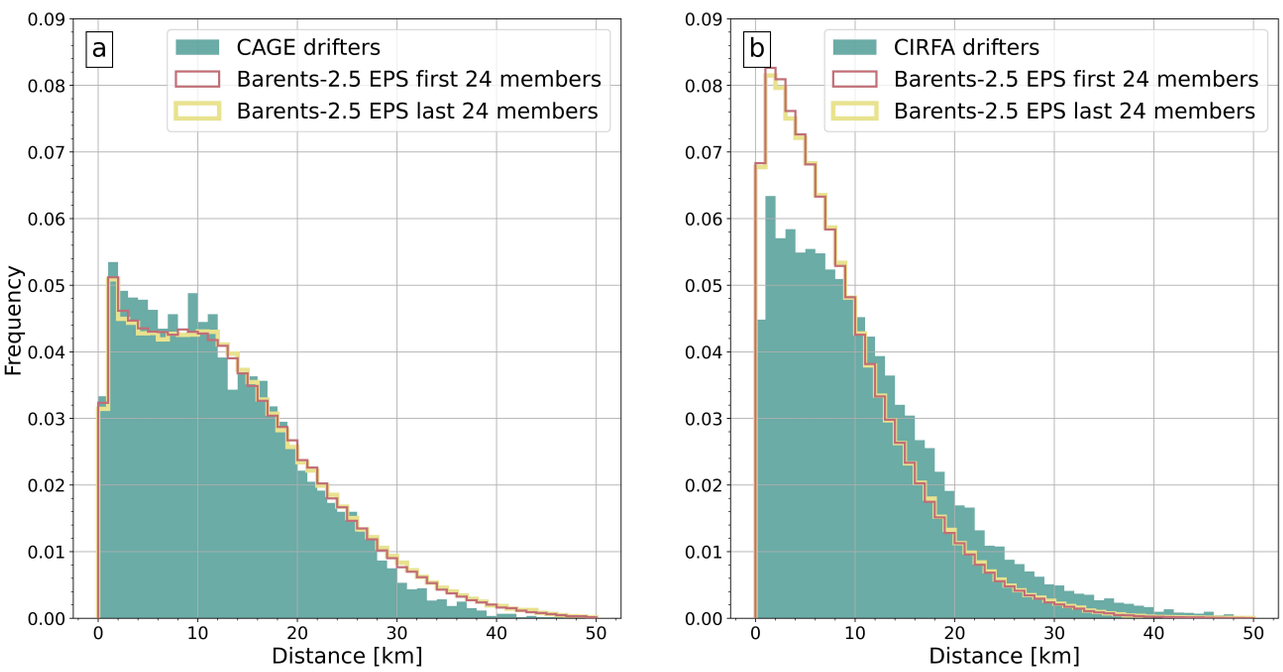

Figure 4: Distribution of along-track cumulative distances for CAGE (a) and CIRFA (b) drifters estimated over the first 24 hours after initialization. Green areas represent observations, whereas red and yellow lines represent modeled cumulative distances forced using the first (42 - 24 hours prior to analysis) and last (18 - 0 hours prior to analysis) 24 members, respectively.

3.2 Spectral properties of observed and modeled drift velocities

In this section, we present an analysis of the energy density of observed and modeled drifter velocities as a function of frequency (cycles per hour, cph) based on 5-day segments. Figure 5 shows the rotary spectra for observed (solid lines) and modeled (dashed lines) drifter velocities for (a) the CAGE and (b) the CIRFA drifters. The observed spectra reveal a dominance of clockwise rotation throughout the evaluated bands, and a qualitative difference of the energy magnitude and spectral shape between the two regions.

The energy magnitude around the inertial frequency is captured well by the model. The modeled velocity spectra present a steeper decay towards the higher frequencies ($|\omega| > 10^{-1}$ cph) and between tidal components K_1, M_2, and M_4. The modeled trajectories overall retained the main spectral features present in the observations. One can nevertheless notice modeled energy underestimation throughout the whole spectra in Figure 5b. A discrepancy at the M_4 tidal band at 6.21 hours is very distinct in the model spectra but not present in the observations.

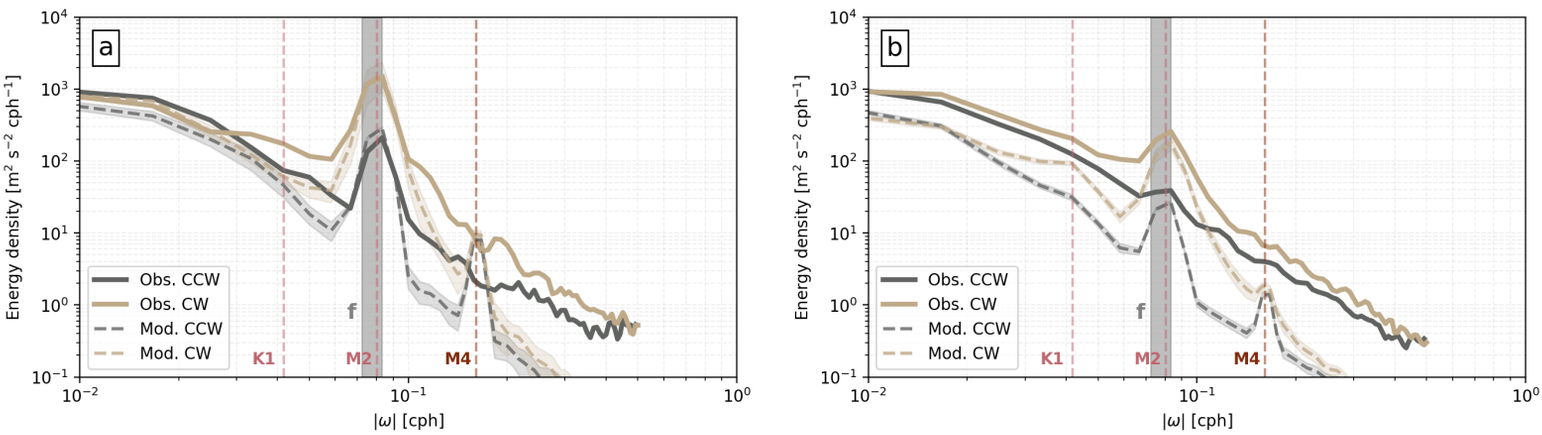

Figure 5: Energy density of CAGE (a) and CIRFA (b) rotary spectra in cycles per hour (cph). Solid lines represent the mean observed clockwise (CW, dark yellow) and counterclockwise (CCW, black) energy, while dashed lines represent the estimated average energy for modeled trajectories. The shaded areas depicts the 95% confidence interval. Tidal components K_1, M_2, and M_4, as well as the range of the Coriolis parameter in the domain (60° to 85°) for the drifter data (vertical gray bar) are also shown.

An analysis of spectral coherence between observed and modeled drifter velocities in Figure 6 indicates that low frequency motions from the model system are to a certain degree in phase with observed motions, in particular for the Barents Sea region. Both data sets show a broad coherence peak below the inertial frequency and around the $\pm M_2$ tidal band. The magnitude of coherence for the CIRFA simulations is half as small compared to the CAGE simulations, but well above the random variations obtained from the ensemble. The broad peak is associated with the low-frequency background wind forcing (see Figures 9 and 10 in Appendix B), while the other two peaks are related to the inertial oscillations (negative ω) and the M_2 tidal component (positive ω).

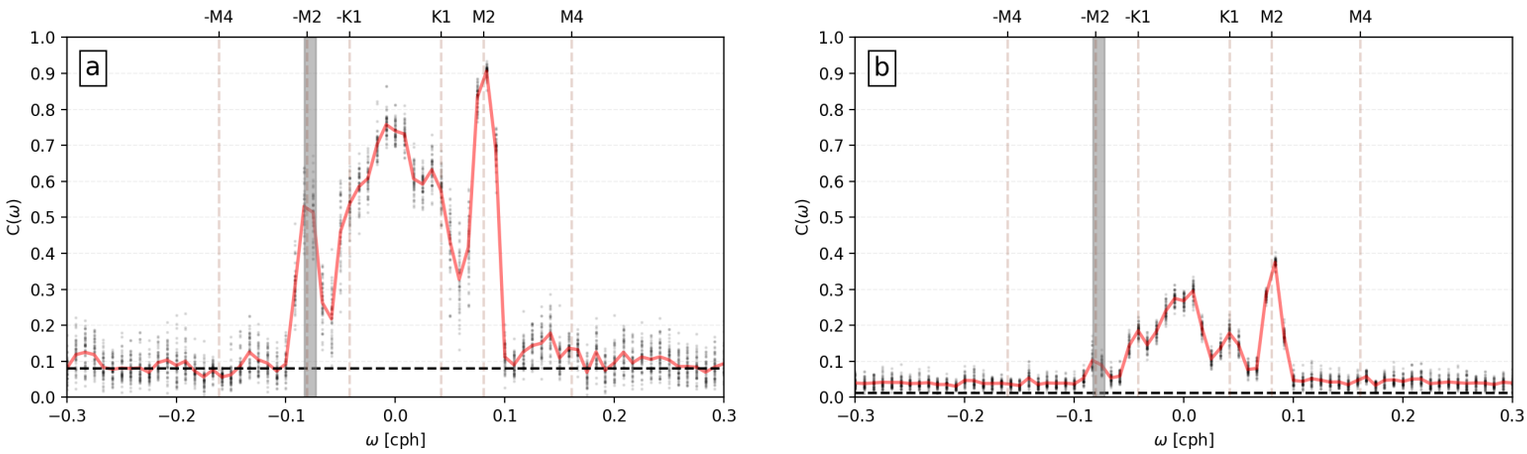

Figure 6: Coherence C(ω) between the observed drifter velocities and modeled trajectory velocities for CAGE (a) and CIRFA (b) estimated via cross rotary spectra analysis. The solid red line represents the average estimation for each frequency. The black dots represent each ensemble member averaged over the respective data set for each frequency bin. Tidal components K_1, M_2, and M_4, as well as the range of the Coriolis parameter in the domain for the drifter data, 60°N to 85°N (vertical gray bar) are also shown. Positive (negative) frequencies correspond to counterclockwise (clockwise) rotations. The horizontal dashed lines represent the 99% confidence limit.

3.3 Ensemble performance diagnostics

The ensemble spread is shown in Figure 7 with a rank histogram for each of the two geographical directions, and as joint heat map for drift displacements. A general tendency of the EPS to underestimate model uncertainty is identified for the lowest and highest absolute frequency values. We note that the rank histograms and the heat map show insignificant skewness and bias towards either side. In addition, despite an underestimation of the most extreme drift displacements, the central part of the histogram is flat, indicating that moderate drift displacements are spread appropriately in the ensemble. These results apply for both study regions, and regardless of the evaluation time after the model initialization, i.e., it consistently preserves this behaviour throughout the forecast lead time (not shown here).

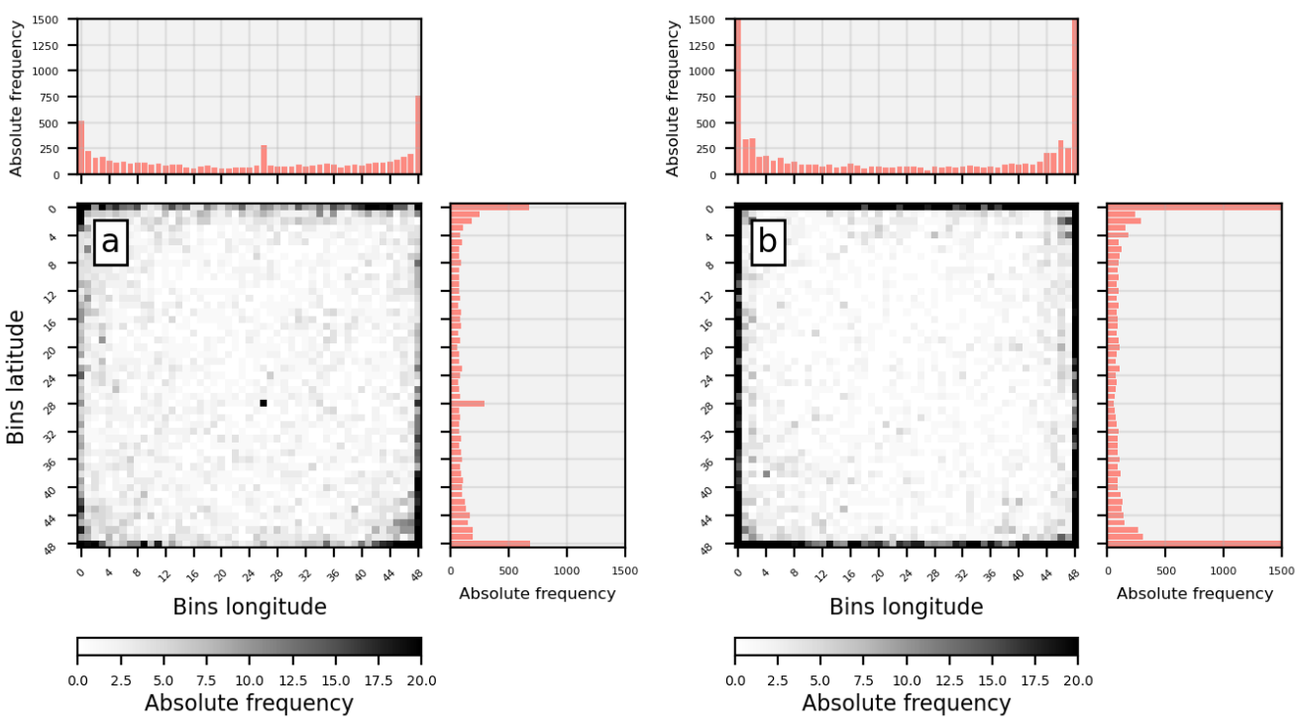

Figure 7: Two-dimensional (2D) rank histograms for CAGE (a) and CIRFA (b) data sets in absolute frequency. Top and right sub-panels show the equivalent one-dimensional (1D) rank histograms for longitude and latitude, respectively.

The underdispersive characteristic of the EPS is further evaluated by the model error-spread ratio ϵ/s. The ratio for each 24 hours reinitialized forecast was bounded between 1 and 2 for the CAGE simulations whereas it varies between 2.5 and 5 for CIRFA simulations. The average and standard deviation values for ϵ and s are shown in Table 2. The spread of the modeled trajectories is similar between the two data sets. However, one can promptly notice that the error for the CIRFA data set nearly doubles with respect to the CAGE simulations. When the two time-lagged ensemble groups are evaluated individually (Table 3), differences between the first and last 24 members do not exceed 15%.

Table 2: Average and standard deviation ensemble mean error (ϵ) and spread (s) of along-track cumulative distance for CAGE and CIRFA simulation sets for all 48 members. Error and spread were calculated over the first 24 hours after initialization using Eq. 8. Average and standard deviation were estimated using all segments.

Data set	**Error [km]** **average, standard deviation**	**Spread [km]** **average, standard deviation**
CAGE	6, 1.2	3.7, 0.55
CIRFA	10.5, 1.6	3.1, 0.9

Table 3: As in Table 2, but the results are split into the first (42 - 24 hours prior to analysis) and last (18 - 0 hours prior to analysis) 24 ensemble members. Values in parenthesis represent (average, standard deviation) in kilometers.

Data set	**First 24 members** **Error, Spread**	**Last 24 members** **Error, Spread**
CAGE	(6, 1.2), (2.6, 0.3)	(6, 1.2), (2.4, 0.4)
CIRFA	(10.9, 1.3), (2.3, 0.5)	(10.8, 1.3), (2.2, 0.5)

A primary objective of ensemble predictions is to provide reliable probabilistic forecasts, i.e. probability of events obtained from the EPS must match observed realisation frequencies. The along-track cumulative distance of observed and modeled trajectories calculated over the first 24 hours after initialization were also used to address the EPS reliability. Figure 8 shows reliability diagrams for four different events: the along-track cumulative distance being (i) less than 10 km, (ii) greater than 10 km and less than 20 km, (iii) greater than 20 km and less than 30 km, and (iv) greater than 30 km.

The EPS performs differently in the two regions. In the Barents Sea (Figure 8a), the EPS is reliable for distances below 20 km. The nearly straight diagonal in the reliability diagram indicates that the EPS can accurately predict events that are unlikely to happen (low probability) and events that are very likely to happen (high probability). The forecast probabilities become less reliable as the distance grows, suggesting that the model loses predictive capability at longer lead times. Taking into account that

the trajectories' hourly displacement ranges between 1 km to 1.6 km, the model manages to accurately reproduce the observed CAGE trajectories up to about one inertial period. For the CIRFA drifters (Figure 8b), overestimation (underestimation) occurs for values below (above) 10 km. These findings corroborate and further detail the observations depicted in Figure 4.

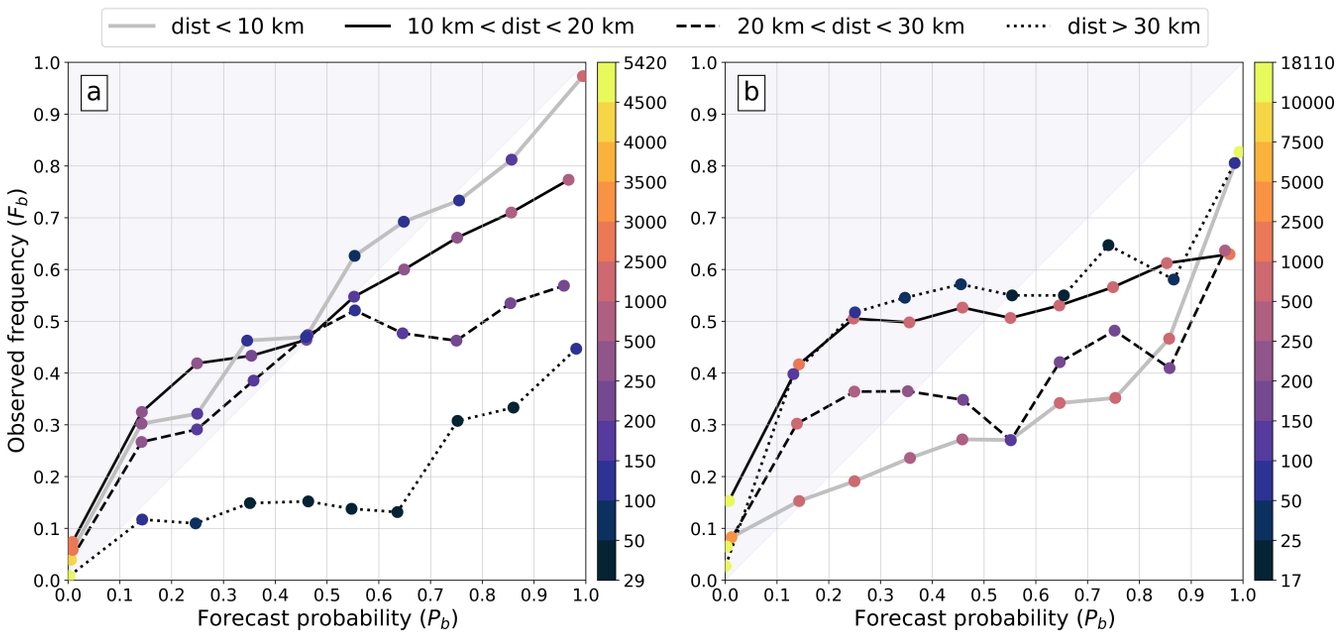

Figure 8: Reliability diagrams for CAGE (a) and CIRFA (b) drifters. Distance (dist) refers to the along-track cumulative distance estimated over 24 hours after initialization. Grey solid, and black solid, dashed and dotted lines show forecast reliability for distances less than 10 km, greater than 10 km (20 km) and less than 20 km (30 km), and greater than 30 km, respectively. The horizontal axis is the forecast probability P_b from Barents-2.5 EPS binned from 0 and 1 in steps of 0.1. The vertical axis indicates the corresponding observed frequency of occurrence F_b. The different colors indicate the number of forecast probabilities that fell into a bin for a given event. Note the different colorbar ranges due to different numbers of segments in the drifter data sets.

4 Discussion

The rotary spectra analysis highlights the distinct dominant mechanisms of the surface circulation in the two study regions, with predominance of inertial oscillations (CW) and the semi-diurnal tidal component (CCW) as forcing for the CAGE drifters in the Barents Sea. These features are also visible for the CIRFA drifters in the Fram Strait-West Spitsbergen area, but less distinct (see Figure 5).

The modeled drift velocities captured the main subinertial spectral features, but underestimated with up to two orders of magnitude the energy density at higher frequencies (> 0.08 cph). The lack of energy at higher frequencies stems from the inability of Barents-2.5 EPS to properly resolve fast, short-wavelength motions due to the lack of horizontal resolution (Müller et al., 2015; Ohlmann et al., 2017; Savage et al.,

2017; Yu et al., 2019). Similar findings for global estimation of surface kinetic energy were reported by Yu et al. (2019) and Arbic et al. (2022), where the authors hint on a deficit of high-frequency energy due to the atmospheric forcing. Klenz et al. (2022) found an understimation higher than one order of magnitude of wind-stress energy spectra at superinertial frequencies for ERA5 reanalysis data when compared to *in-situ* wind measurements obtained with Minimet drifters. It is thus possible that the issues with the atmospheric forcing also contribute to the sharp energy decay at superinertial frequencies and the unrealistic peak at the M_4 frequency seen in Figure 5.

Subinertial drift is largely driven by wind forcing, which is well predicted by the used atmosphere forcing, and as a result, wind driven ocean circulation is expected to be predictable in the ocean model. Both the subinertial drift and in particular the near-inertial response, which is a delicate process that depends on well-described mixed layer dynamics (Röhrs et al., 2023a), is well captured in the drift simulations. As a result, we see distinct coherence between predicted and observed drifts (Figure 6). The semi-diurnal lunar component has predominantly a CCW rotation due to the incursion of Kelvin waves around the Barents Sea (Furevik and Foldvik, 1996), but residual M_2 currents may also display CW rotation in the vicinity around Bear and Hopen islands (Kowalik and Marchenko, 2023). We therefore do not neglect the possibility that M_2 energy might also contribute to the CW component presented in Figure 5 in addition to the inertial oscillations.

Scrutinizing the two regions using the herein presented diagnostic tools allowed us to identify regimes where the model performance is captured well, and regimes that lack predictability. In the case of the CIRFA data set, it is well known that the Fram Strait - West Spitsbergen system is highly dynamic, with an interplay between barotropic (Teigen et al., 2010) and baroclinic instabilities (Teigen et al., 2011). Eddies generated there possess radii predominantly within the submesoscale and lower bound of the mesoscale ranges, i.e. 1 km - 12 km (Wekerle et al., 2020; Bashmachnikov et al., 2020), respectively. The first baroclinic Rossby radius of deformation varies between 2 km (on the continental shelf) to 6 km (on the shelf break) there (Nurser and Bacon, 2014). Considering four (Woodring et al., 2016) to ten grid points (Soufflet et al., 2016) as the model's effective resolution, Barents-2.5 EPS is unable to properly resolve such small-scale phenomena resulting in large errors at both sub- and superinertial scales.

The limited horizontal resolution and lack of energy inhibits the most extreme drift velocities, and modeled trajectories quickly diverge from observations due to unresolved motions. This means that in

the ensemble performance it can be seen as insufficient ensemble variability and unbalanced error/spread. Supporting the findings in Idžanović et al. (2023), the rank histograms (Figure 7) presented here also indicate an underdispersive characteristic of Barents-2.5 EPS, since their shapes with tall outer ranks and flat in the center agrees with the theoretical description provided by Hamill (2001). Considering that the spread between the two simulation sets is similar (see Tables 2 and 3), we suggest that the pronounced U-shaped rank histogram for the CIRFA drifters (Figure 7b) is induced by model error as it nearly doubles with respect to the CAGE simulations. Further perturbations may yield ensemble spread growth, but it should be balanced by a decrease in the model error through the assimilation of dynamical fields.

We highlight that underdispersion does not necessarily imply lack of reliability. Despite the U-shaped rank histogram and error/spread higher than 1, Figure 8 shows that simulations performed for the CAGE data set present straight diagonal lines of along-track cumulative distances up to 20 km, or one inertial period equivalently. Tides and wind-driven inertial oscillations are major driving forces on the drift velocities (Figure 5a), and high reliability is achieved for such processes that are well resolved by the model. Previous research on data assimilation for trajectory forecast have found that skill improvements are barely observed after 12 hours of simulation (Vandenbulcke et al., 2009; Holm et al., 2020, e.g.), suggesting that ocean models in general have limited, and similar, forecast horizon.

Extending the number of available ensemble members does not improve the EPS performance at the scales investigated here. As the ensemble spread is nearly saturated already at its initialization, the EPS's underdispersive performance stems from the ensemble mean error as indicated by the U-shape histogram. The U-shape can also be due to model error (Hamill, 2001), especially in the Fram Strait - West Spitsbergen region. Ensemble mean error reduction may be achieved by assimilating the dynamic fields in the EPS. Constraining initial conditions at spatial scales less than 50 km, particularly in highly dynamic areas, is a current frontier in operational ocean forecasting. Although still under debate (Spencer Jones et al., 2023), observations acquired with the recently launched Surface Water and Ocean Topography (SWOT) mission are expected to enhance our capability to constrain ocean currents at those scales. Beyond its impacts on ocean current's forecasting performance, it is also worthwhile of further investigations how such observations may improve the modeled kinetic energy spectra and ensemble spread.

5 Concluding Remarks

The forecast capabilities of trajectory simulations based on an operational ocean EPS, Barents-2.5, was assessed by comparison with observed drifter trajectories in the Fram Strait and the Barents Sea. Using an ensemble of trajectory simulations, the model was evaluated in terms of wind-response, ensemble spread, and reliability performance for short-term prediction analysis (1 - 5 days). Reliable drift displacement for up to 20 km (about 1 inertial period, $\approx$ 12 hours) can be obtained with the model under well-resolved conditions. Unresolved small-scale phenomena impact both the energy spectra at sub- and superinertial frequencies and ensemble dispersion. The model underdispersive behaviour, in particular in the Fram Strait region, stems from the model error due to insufficient horizontal resolution rather than caused by insufficient spread during model initialisation.

6 Acknowledgements

We acknowledge funding by the Research Council of Norway through grant no. 237906 (CIRFA). The CAGE drifters were provided by the Fram Center through the Arctic Ocean flagship grant (project "Thin-Tec - Thin ice Measurement Technology", PI M. Johansson). Torbjørn Eltoft is gratefully acknowledged for scientific discussions. We thank Dr. Sveinung Olsen, M.Sc. Catherine Taelman, Dr. Jean Rabault, and Dr. Gaute Hope for technical support. We also acknowledge the crews on *R/V* Helmer Hanssen and *R/V* Kronprins Haakon, as well as CAGE and CIRFA researchers for their support onboard and deployment help.

7 Disclosure statement

No potential conflict of interest was reported by the authors.

8 Data availability statement

All geophysical forcing data is available at https://thredds.met.no/thredds/catalog.html. OpenDrift can be downloaded at https://opendrift.github.io/. Observed drifter trajectories can be downloaded at https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/EPB9AG.

A Drifters

Table 4: Drifter deployments during CIRFA (April-May 2022, gray) and CAGE (August 2022) cruises.

Drifter ID	**Deployment position (lat [°N], lon [°E])**	**Start (UTC)**	**End (UTC)**	**Speed (ms^{-1}) (Avg., Std.)**[b]	**# Daily Segments**
drifter 9_waves_ISM	78.1, 11.9	2022-04-22 23:53	2022-06-01 21:00	0.30, 0.15	39
drifter 4	78.2, 09.9	2022-04-23 03:40	2023-03-03 13:30	0.29, 0.19	73
drifter 10_waves_ISM	78.3, 07.8	2022-04-23 06:50	2022-10-20 23:53	0.29, 0.16	149
drifter 11_waves_ISM	78.8, 0.99	2022-04-23 19:55	2022-11-24 20:00	0.25, 0.14	181
drifter 8	78.8, 04.9	2022-05-08 11:41	2022-07-22 07:00	0.24, 0.14	58
drifter 7	78.8, 06.1	2022-05-08 13:50	2022-06-09 09:30	0.29, 0.15	31
drifter 2	78.8, 07.5	2022-05-08 16:02	2022-05-27 11:00	0.27, 0.13	18
drifter 15_waves_LSM	78.8, 07.9	2022-05-08 17:28	2022-07-10 11:30	0.26, 0.13	57
drifter 1	78.8, 08.5	2022-05-08 19:13	2022-07-05 19:53	0.24, 0.13	57
drifter 12_waves_LSM	78.8, 09.0	2022-05-08 22:08	2022-07-04 05:44	0.21, 0.14	55
drifter 5	78.8, 10.0	2022-05-08 21:53	2022-09-02 13:30	0.23, 0.14	86
drifter 206638	75.2, 31.6	2022-08-07 00:30	2022-10-26 07:00	0.26, 0.13	80
drifter 206689	75.2, 31.7	2022-08-07 02:00	2022-08-28 03:00	0.28, 0.15	21
drifter 206633	75.0, 30.8	2022-08-07 05:30	2022-11-01 08:30	0.27, 0.13	86
drifter 206646	73.1, 21.9	2022-08-18 03:30	2022-08-30 20:00	0.29, 0.14	12
drifter 206650	72.8, 21.9	2022-08-18 07:00	2022-09-20 00:00	0.29, 0.16	32
drifter 206665	72.6, 21.4	2022-08-18 09:00	2022-09-17 22:30	0.33, 0.17	30

[b]: Average (avg.) and standard deviation (std.) calculated over the entire time series extent.

B Drifter and modeled wind velocities coherence

Figure 9 shows the coherence (Eq. 6) between observed drifter and modeled wind velocities for CAGE (a) and CIRFA (b) data sets. The same in Figure 10, but for modeled drifter velocities instead.

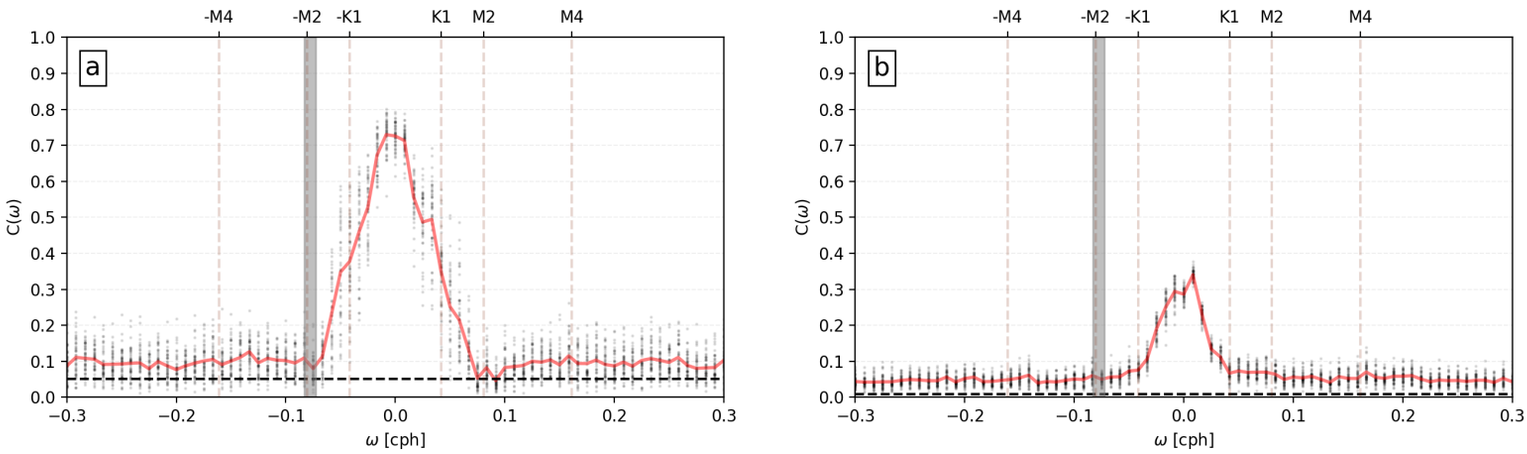

Figure 9: Same as Figure 6, but coherence estimation between observed drifters and modeled wind velocities. The horizontal dashed line represents the 99% confidence limit.

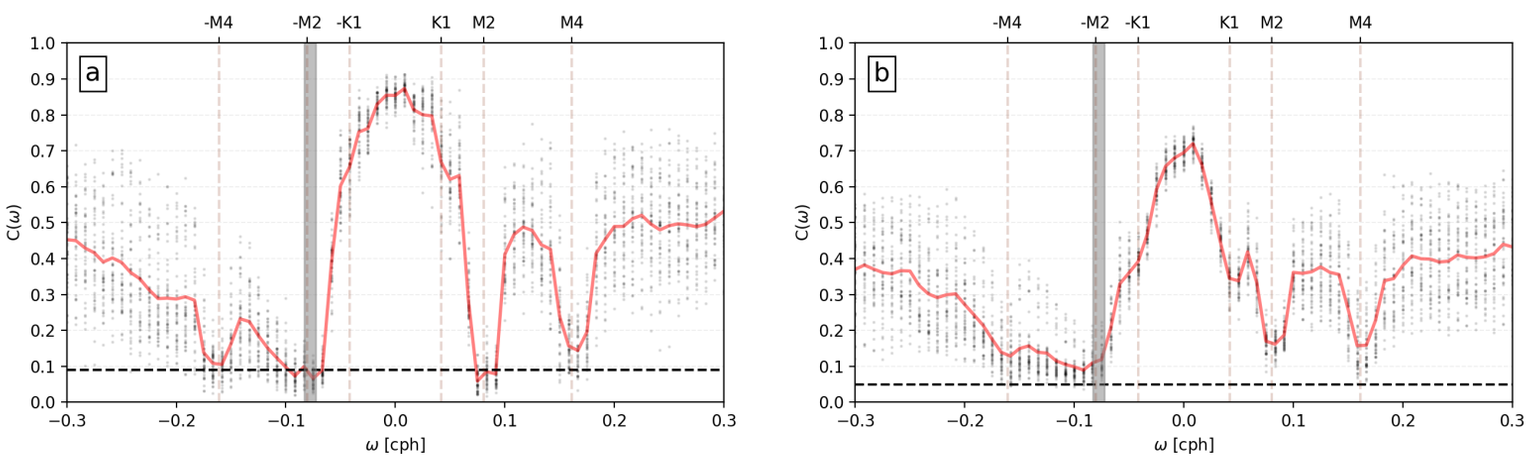

Figure 10: Same as Figure 6, but coherence estimation modeled drifters and modeled wind velocities.

References

Arbic, B., Elipot, S., Brasch, J., Menemenlis, D., Ponte, A., Shriver, J., Yu, X., Zaron, E., Alford, M., Buijsman, M., Abernathey, R., Garcia, D., Guan, L., Martin, P., and Nelson, A. (2022). Near-surface oceanic kinetic energy distributions from drifter observations and numerical models. *Journal of Geophysical Research: Oceans*, 127(10):e2022JC018551. e2022JC018551 2022JC018551.

Asplin, L., Johnsen, I., Sandvik, A., Albretsen, J., Sundfjord, V., Aure, J., and Boxaspen, K. (2014). Dispersion of salmon lice in the Hardangerfjord. *Marine Biology Research*, 10(3):216–225.

Bashmachnikov, I., Kozlov, I., Petrenko, L., Glok, N., and Wekerle, C. (2020). Eddies in the north Greenland sea and Fram strait from satellite altimetry, sar and high-resolution model data. *Journal of Geophysical Research: Oceans*, 125(7):e2019JC015832. e2019JC015832 2019JC015832.

Breivik, Ø. and Allen, A. (2008). An operational search and rescue model for the Norwegian Sea and the North Sea. *Journal of Marine Systems*, 69(1):99–113. Maritime Rapid Environmental Assessment.

Brekke, C., Espeseth, M., Dagestad, K.-F., Röhrs, J., Hole, L. R., and Reigber, A. (2021). Integrated analysis of multisensor datasets and oil drift simulations — A free-floating oil experiment in the open ocean. *Journal of Geophysical Research: Oceans*, 126(1):e2020JC016499.

Bröcker, J. and Smith, L. (2007). Increasing the reliability of reliability diagrams. *Weather and Forecasting*, 22(3):651 – 661.

Callies, U., Groll, N., Horstmann, J., Kapitza, H., Klein, H., Maßmann, S., and Schwichtenberg, F. (2017). Surface drifters in the German Bight: model validation considering windage and stokes drift. *Ocean Science*, 13(5):799–827.

Christensen, K., Breivik, Ø., Dagestad, K.-F., Röhrs, J., and Ward, B. (2018). Short-Term Predictions of Oceanic Drift. *Oceanography*, 31(3):59–67.

Dagestad, K.-F. and Röhrs, J. (2019). Prediction of ocean surface trajectories using satellite derived vs. modeled ocean currents. *Remote Sensing of Environment*, 223:130–142.

Dagestad, K.-F., Röhrs, J., Breivik, Ø., and Ådlandsvik, B. (2018). OpenDrift v1.0: a generic framework for trajectory modelling. *Geosci. Model Dev.*, 11(4):1405–1420.

de Aguiar, V., Röhrs, J., Johansson, A., and Eltoft, T. (2023). Assessing ocean ensemble drift predictions by comparison with observed oil slicks. *Frontiers in Marine Science*, 10.

Dierking, W., Schneider, A., Eltoft, T., and Gerland, S. (2022). CIRFA Cruise 2022. Cruise report.

Eckel, F. and Mass, C. (2005). Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting. *Weather and Forecasting*, 20(3):328 – 350.

Egbert, G. and Erofeeva, S. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology*, 19(2):183 – 204.

Evensen, G. (1994). Inverse methods and data assimilation in nonlinear ocean models. *Physica D: Nonlinear Phenomena*, 77(1):108–129.

462 Fortin, V., Abaza, M., Anctil, F., and Turcotte, T. (2014). Why should ensemble spread match the
463 RMSE of the ensemble mean? *Journal of Hydrometeorology*, 15(4):1708 – 1713.

464 Furevik, T. and Foldvik, A. (1996). Stability at m2 critical latitude in the barents sea. *Journal of*
465 *Geophysical Research: Oceans*, 101(C4):8823–8837.

466 Garcia-Pineda, O., Androulidakis, Y., Le Hénaff, M., Kourafalou, V., Hole, L., Kang, H., Staples, G.,
467 Ramirez, E., and DiPinto, L. (2020). Measuring oil residence time with gps-drifters, satellites, and
468 unmanned aerial systems (UAS). *Marine Pollution Bulletin*, 150:110644.

469 Gonella, J. (1972). A rotary-component method for analysing meteorological and oceanographic vector
470 time series. *Deep Sea Research and Oceanographic Abstracts*, 19(12):833–846.

471 Gusdal, Y. and Carrasco, A. (2012). Validation of the Operational Wave Model WAM at met.no - Report
472 2011. Technical report, Norwegian Meteorological Institute, Oslo. Report No. 23/2012.

473 Hamill, T. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather*
474 *Review*, 129(3):550–560.

475 Holm, H., Sætra, M., and van Leeuwen, P. (2020). Massively parallel implicit equal-weights particle filter
476 for ocean drift trajectory forecasting. *Journal of Computational Physics: X*, 6:100053.

477 Huang, J. (1983). A review of the state-of-the-art of oil spill fate/behavior models. *International Oil*
478 *Spill Conference Proceedings*, 1983(1):313–322.

479 Idžanović, M., Rikardsen, E., and Röhrs, J. (2023). Forecast uncertainty and ensemble spread in surface
480 currents from a regional ocean model. *Frontiers in Marine Science*, 10.

481 Jacobs, G., D'Addezio, J., Ngodock, H., and Souopgui, I. (2021). Observation and model resolution
482 implications to ocean prediction. *Ocean Model.*, page 101760.

483 Janssen, P. and Behrens, A. (2013). Documentation of a web based source code library for WAM.
484 Technical report, Norwegian Meteorological Institute. Reference: MyWave- D1.1.

485 Jones, C., Dagestad, K.-F., Breivik, Ø., Holt, B., Röhrs, J., Christensen, K., Espeseth, M., Brekke, C.,
486 and Skrunes, S. (2016). Measurement and modeling of oil slick transport. *Journal of Geophysical*
487 *Research: Oceans*, 121(10):7759–7775.

Kampouris, K., Vervatis, V., Karagiorgos, J., and Sofianos, S. (2021). Oil spill model uncertainty quantification using an atmospheric ensemble. *Ocean Science*, 17(4):919–934.

Klenz, T., Simmons, H., Centurioni, L., Lilly, J., Early, J., and Hormann, V. (2022). Estimates of near-inertial wind power input using novel in situ wind measurements from minimet surface drifters in the Iceland Basin. *Journal of Physical Oceanography*, 52(10):2417 – 2430.

Komen, G., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., and Janssen, P. (1994). *Dynamics and Modelling of Ocean Waves*. Cambridge University Press.

Kowalik, Z. and Marchenko, A. (2023). Tidal motion enhancement on spitsbergen bank, barents sea. *Journal of Geophysical Research: Oceans*, 128(1):e2022JC018539. e2022JC018539 2022JC018539.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R., Goddard, J., Haiden, T., Hogan, R., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P., Subramanian, A., Vitart, F., Wedi, N., and Weisheimer, A. (2017). Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2315–2339.

Leutbecher, M. and Palmer, T. (2008). Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515–3539. Predicting weather, climate and extreme events.

Li, Y., Yu, H., Wang, Z.-y., Li, Y., Pan, Q.-q., Meng, S.-j., Yang, Y.-q., Lu, W., and Guo, K.-x. (2019). The forecasting and analysis of oil spill drift trajectory during the Sanchi collision accident, East China Sea. *Ocean Engineering*, 187:106231.

Li, Y., Zhu, J., Wang, H., and Kuang, X. (2013). The error source analysis of oil spill transport modeling: a case study. *Acta Oceanologica Sinica*, 32:41–47.

Liu, Y. and Weisberg, R. (2011). Evaluation of trajectory modeling in different dynamic regions using normalized cumulative lagrangian separation. *Journal of Geophysical Research: Oceans*, 116(C9).

Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M., Noer, G., and Korosov, A. (2017). Characteristics of a convective-scale weather forecasting system for the European Arctic. *Monthly Weather Review*, 145(12):4771 – 4787.

Müller, M., Arbic, B., Richman, J., Shriver, J., Kunze, E., Scott, R., Wallcraft, A., and Zamudio, L. (2015). Toward an internal gravity wave spectrum in global ocean models. *Geophysical Research Letters*, 42(9):3474–3481.

Nordam, T. and Duran, R. (2020). Numerical integrators for Lagrangian oceanography. *Geoscientific Model Development*, 13(12):5935–5957.

Nurser, A. and Bacon, S. (2014). The rossby radius in the Arctic Ocean. *Ocean Science*, 10(6):967–975.

Ohlmann, J. C., Molemaker, M. J., Baschek, B., Holt, B., Marmorino, G., and Smith, G. (2017). Drifter observations of submesoscale flow kinematics in the coastal ocean. *Geophysical Research Letters*, 44(1):330–337.

Palmer, T. (2017). The primacy of doubt: Evolution of numerical weather prediction from determinism to probability. *Journal of Advances in Modeling Earth Systems*, 9(2):730–734.

Rabault, J., Nose, T., Hope, G., Müller, M., Breivik, Ø., Voermans, J., Hole, L., Bohlinger, P., Waseda, T., Kodaira, T., Katsuno, T., Johnson, M., Sutherland, G., Johansson, M., Christensen, K., Garbo, A., Jensen, A., Gundersen, O., Marchenko, A., and Babanin, A. (2022). Openmetbuoy-v2021: An easy-to-build, affordable, customizable, open-source instrument for oceanographic measurements of drift and waves in sea ice and the open ocean. *Geosciences*, 12(3).

Röhrs, J. and Christensen, K. (2015). Drift in the uppermost part of the ocean. *Geophysical Research Letters*, 42(23):10,349–10,356.

Röhrs, J., Christensen, K., Hole, L., Broström, G., Drivdal, M., and Sundby, S. (2012). Observation-based evaluation of surface wave effects on currents and trajectory forecasts. *Ocean Dynamics*, 62:1519–1533.

Röhrs, J., Dagestad, K.-F., Asbjørnsen, H., Nordam, T., Skancke, J., Jones, C. E., and Brekke, C. (2018). The effect of vertical mixing on the horizontal drift of oil spills. *Ocean Science*, 14(6):1581–1601.

Röhrs, J., Gusdal, Y., Rikardsen, E. S. U., Durán Moro, M., Brændshøi, J., Kristensen, N. M., Fritzner, S., Wang, K., Sperrevik, A. K., Idžanović, M., Lavergne, T., Debernard, J. B., and Christensen, K. H. (2023). Barents-2.5km v2.0: an operational data-assimilative coupled ocean and sea ice ensemble prediction model for the Barents Sea and Svalbard. *Geoscientific Model Development*, 16(18):5401–5426.

Rusin, J., Lavergne, T., Doulgeris, A., and Scott, K. (2024). Resolution enhanced sea ice concentration: a new algorithm applied to amsr2 microwave radiometry data. *Annals of Glaciology*, pages 1–12.

Röhrs, J., Halsne, T., Sutherland, G., Dagestad, K.-., Hole, L., Broström, G., and Christensen, K. (2023a). Current shear and turbulence during a near-inertial wave. *Frontiers in Marine Science*, 10.

Röhrs, J., Sutherland, G., Jeans, G., Bedington, M., Sperrevik, A., Dagestad, K.-F., Gusdal, Y., Mauritzen, C., Dale, A., and LaCasce, J. (2023b). Surface currents in operational oceanography: Key applications, mechanisms, and methods. *Journal of Operational Oceanography*, 16(1):60–88.

Savage, A., Arbic, B., Alford, M., Ansong, J., Farrar, J., Menemenlis, D., O'Rourke, A., Richman, J., Shriver, J. F., Voet, G., Wallcraft, A., and Zamudio, L. (2017). Spectral decomposition of internal gravity wave sea surface height in global models. *Journal of Geophysical Research: Oceans*, 122(10):7803–7821.

Serov, P., Patton, H., Mazzini, A., Mattingsdal, R., Shephard, G., Cooke, F., de Aguiar, V., Holm, V., Alessandrini, G., Meza Cala, J., and Luerssen, P. (2022). Cage22-6 cruise report: Geo-3144/8144 teaching cruise: Geologically controlled hydrocarbon seepage in hopendjupet and the wider barents sea. *CAGE – Centre for Arctic Gas Hydrate, Environment and Climate Report Series*, 10.

Shchepetkin, A. and McWilliams, J. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9(4):347–404.

Soufflet, Y., Marchesiello, P., Lemarié, F., Jouanno, J. andCapet, X., Debreu, L., and Benshila, R. (2016). On effective resolution in ocean models. *Ocean Modelling*, 98:36–50.

Spencer Jones, C., Xiao, Q., Abernathey, R., and Smith, K. S. (2023). Using Lagrangian filtering to remove waves from the ocean surface velocity field. *Journal of Advances in Modeling Earth Systems*, 15(4). e2022MS003220 2022MS003220.

Teigen, S., Nilsen, F., and Gjevik, B. (2010). Barotropic instability in the West Spitsbergen Current. *Journal of Geophysical Research: Oceans*, 115(C7).

Teigen, S., Nilsen, F., Skogseth, R., Gjevik, B., and Beszczynska-Möller, A. (2011). Baroclinic instability in the West Spitsbergen Current. *Journal of Geophysical Research: Oceans*, 116(C7).

Umlauf, L. and Burchard, H. (2005). Second-order turbulence closure models for geophysical boundary layers. A review of recent work. *Continental Shelf Research*, 25(7–8):795–827.

Vandenbulcke, L., Beckers, J.-M., Lenartz, F., Barth, A., Poulain, P.-M., Aidonidis, M., Meyrat, J., Ardhuin, F., Tonani, M., Fratianni, C., Torrisi, L., Pallela, D., Chiggiato, J., Tudor, M., Book, J., Martin, P., Peggion, G., and Rixen, M. (2009). Super-ensemble techniques: Application to surface drift prediction. *Progress in Oceanography*, 82(3):149–167.

Wagner, T., Eisenman, I., Ceroli, A., and Constantinou, N. (2022). How winds and ocean currents influence the drift of floating objects. *Journal of Physical Oceanography*, 52(5):907 – 916.

Wekerle, C., Hattermann, T., Wang, Q., Crews, L., von Appen, W.-J., and Danilov, S. (2020). Properties and dynamics of mesoscale eddies in Fram Strait from a comparison between two high-resolution ocean–sea ice models. *Ocean Science*, 16(5):1225–1246.

Wekerle, C., Wang, Q., von Appen, W.-J., Danilov, S., Schourup-Kristensen, V., and Jung, T. (2017). Eddy-resolving simulation of the Atlantic Water circulation in the Fram strait with focus on the seasonal cycle. *Journal of Geophysical Research: Oceans*, 122(11):8385–8405.

Wilks, D. (2019). *Statistical methods in the atmospheric sciences*, volume Fourth Edition. Amsterdam, Netherlands, Elsevier.

Woodring, J., Petersen, M., Schmeißer, A., Patchett, J., Ahrens, J., and Hagen, H. (2016). In situ eddy analysis in a high-resolution ocean climate model. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):857–866.

Xie, J., Bertino, L., Counillon, F., Lisæter, K. A., and Sakov, P. (2017). Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013. *Ocean Sci.*, 13(1):123–144.

Xiong, J. and MacCready, P. (2024). Intercomparisons of tracker v1.1 and four other ocean particle-tracking software packages in the regional ocean modeling system. *Geoscientific Model Development*, 17(8):3341–3356.

Yu, X., Ponte, A., Elipot, S., Menemenlis, D., Zaron, E., and Abernathey, R. (2019). Surface kinetic energy distributions in the global oceans from a high-resolution numerical model and surface drifter observations. *Geophysical Research Letters*, 46(16):9757–9766.

595 Zhang, X., Cheng, L., Zhang, F., Wu, J., Li, S., Liu, J., Chu, S., Xia, N., Min, K., Zuo, X., and Li,
596 M. (2020). Evaluation of multi-source forcing datasets for drift trajectory prediction using lagrangian
597 models in the South China Sea. *Applied Ocean Research*, 104:102395.

598 Zodiatis, G., De Dominicis, M., Perivoliotis, L., Radhakrishnan, H., Georgoudis, E., Sotillo, M., Lardner,
599 R., Krokos, G., Bruciaferri, D., Clementi, E., Guarnieri, A., Ribotti, A., Drago, A., Bourma, E.,
600 Padorno, E., Daniel, P., Gonzalez, G., Chazot, C., Gouriou, V., Kremer, X., Sofianos, S., Tintore, J.,
601 Garreau, P., Pinardi, N., Coppini, G., Lecci, R., Pisano, A., Sorgente, R., Fazioli, L., Soloviev, D.,
602 Stylianou, S., Nikolaidis, A., Panayidou, X., Karaolia, A., Gauci, A., Marcati, A., Caiazzo, L., and
603 Mancini, M. (2016). The Mediterranean Decision Support System for Marine Safety dedicated to oil
604 slicks predictions. *Deep Sea Research Part II: Topical Studies in Oceanography*, 133:4–20. Physical,
605 chemical and biological observations and modelling of oil spills in the Mediterranean Sea.

/8

Paper III: Uncertainty Estimation of SAR Current Retrievals using an Atmospheric NWP Ensemble

de Aguiar, V., Moiseev, A., Johansson, A.M., Röhrs, J., Johnsen, H. and Eltoft, T. Submitted to JGR: Oceans.

Uncertainty Estimation of SAR Current Retrievals using an Atmospheric NWP Ensemble

de Aguiar, V. [1], Moiseev, A.[2], Johansson, A.M.[1], Röhrs, J.[3], Johnsen, H.[4]and Eltoft, T.[1]

[1]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway
[2]Nansen Environmental and Remote Sensing Center, Bergen, Norway

[3]Division for Ocean and Ice, Norwegian Meteorological Institute, Oslo, Norway
[4]NORCE Norwegian Research Center AS, Tromsø, Norway

Key Points:

- Modeled wind speed and direction ensembles are used to address the uncertainty on ocean radial velocity (RVL) retrieved with Sentinel-1.
- A relationship between the ensemble variables and RVL uncertainties is established.
- RVL wind-driven uncertainty can be estimated using atmospheric ensemble models, potentially expanded to global products.

Corresponding author: Victor de Aguiar, victor.d.aguiar@uit.no

Abstract

Retrieval of ocean surface current using single-antenna synthetic aperture radar (SAR) relies on the removal of the wind and wave signal (wind-wave bias) from the observed SAR Doppler shit. This is often performed using deterministic atmospheric and wave models input into a geophysical model function, though atmospheric models are chaotic, and their predictability is flow-dependent. 251 Sentinel-1 SAR scenes, obtained in 2023 over Skagerrak and Kattegat, are analyzed to assess the uncertainty in the radial velocity estimation caused by uncertainties in the wind speed and direction provided by MEPS, an operational regional atmospheric ensemble system. Using ensemble surface wind direction and speed from MEPS, we investigate the conditions under which the radial velocity retrieval uncertainties exceed community-guided thresholds. The maximum wind speed and direction values that meet the specified thresholds are also estimated using two approaches. Our findings show that retrievals have a higher degree of uncertainty when the wind speed uncertainty exceeds 20% of the mean field, with larger uncertainties associated with low-wind speed conditions. A strong dependency between satellite antenna-look and maximum uncertainty values is reported. Employing ensemble models on radial velocity uncertainty quantification has promising potential and may be extended to operational global products.

Plain Language Summary

Ocean surface currents are one of the key marine environmental parameters. Satellites operating in the microwave band (for example, Sentinel-1) can directly derive information on the ocean surface currents using Doppler shift information. However, the total signal contains extra contributions that must be removed to provide the current data. Here we focus on the wind-wave bias, a term related to the wind and wave data. Removing the wind-wave bias is traditionally performed by using single-run atmospheric models with tailored Geophysical Model Functions. Using 251 Sentinel-1 images obtained in the Skagerrak and Kattegat, we investigate how uncertainties in the modeled wind data cause errors in the retrieved ocean current. Establishing accuracy thresholds for the ocean current retrieval, we investigate under which conditions the accuracy is not met. Our work shows that low-wind speed conditions tend to provide unreliable estimates and that the wind-wave bias uncertainty is strongly dependent on the wind speed and direction relative to the satellite orbit.

1 Introduction

Ocean surface currents are central in operational oceanography. Several applications require nowcasts of surface currents, such as oil slick drift predictions (e.g., Brekke et al., 2021) and search-and-rescue operations (e.g., Breivik & Arthur Allen, 2008). Observations of ocean surface currents are also needed for model validation and data assimilation to establish model constraints. Spaceborne synthetic aperture radars (SAR) enable retrieval of ocean surface currents, providing greater spatial coverage compared to in-situ and land-based high-frequency radar platforms.

Pioneering work by Chapron et al. (2005) demonstrated that ocean surface currents can be retrieved by single-antenna spaceborne SAR systems using the Doppler shift information from the received echoes. The method was shown to be useful for monitoring regions of both strong (2 m/s) and weak (0.5 m/s) ocean surface current regimes (e.g., Johannessen et al., 2008; Hansen, Collard, et al., 2011; Hansen, Johannessen, et al., 2011). Several spaceborne SAR missions such as TerraSAR-X (Romeiser et al., 2010), Tandem-X (Romeiser et al., 2014), Envisat ASAR (Chapron et al., 2005; Johannessen et al., 2008), and lately Sentinel-1 (S1) have shown the capability to provide measurements of the radial component of mean Lagrangian surface velocity vector utilizing the Doppler shift (Johnsen et al., 2016).

The mean Lagrangian surface velocity radial component represents the geophysical-induced Doppler "signal" received by the SAR antenna. The geophysical signal has contributions of the radial component of the ocean current field (RVL) and, additionally, an influence of the wind field and sea state (wind sea and swell), referred to as the wind-waves bias. Empirical geophysical model functions (GMFs) are commonly used to estimate the wind and sea state contributions to the Doppler shift, where the CDOP3SiX (Moiseev, Johnsen, Johannessen, et al., 2020; Moiseev et al., 2022) was developed for the operational S1 mission. Following its predecessor (CDOP, developed for Envisat ASAR, see Collard et al. (2008); Mouche et al. (2012)), CDOP3SiX relies on the auxiliary wind and the wave fields from atmospheric and wave models for approximation of the wind and sea state Doppler contributions. Despite the convenience of this approach, the use of auxiliary model fields propagates uncertainties from the used wind and wave models into the derived surface current retrievals. Understanding and quantifying these uncertainties is therefore important to develop further downstream applications into operational oceanography, like data assimilation.

Employing a single deterministic model neither guarantees that the modeled field represents the real atmospheric state nor allows us to investigate uncertainties in the RVL retrieval. Atmospheric ensemble models aim to reproduce the possible atmospheric state in the probabilistic space by providing multiple forecasts. Equally probable ensemble members are generated through perturbations in, e.g., the initial conditions and model parameters (Leutbecher & Palmer, 2008), and we can use the ensemble member variability to understand how it affects RVL retrievals.

Advances in the estimation of the non-geophysical component to the total Doppler shift have significantly improved over the years (e.g., Moiseev et al., 2022). As ensemble prediction systems have become common at weather forecast centres, we can now revisit the RVL retrieval obtained with S1 in coastal regions using a probabilistic set rather than one deterministic output. As uncertainty estimations are required for ocean surface current products by users (e.g., World Meteorological Organization, 2024), we investigate here the atmospheric conditions and maximum uncertainty in the wind speed and direction that fulfill exemplary user criteria.

Two-hundred and fifty-one S1 Interferometric Wide (IW) scenes were collected between January and July 2023 over the North Sea, Skagerrak, Kattegat, and the Baltic Sea. Focusing on the Skagerrak and Kattegat Basins, we evaluate how uncertainties in the wind speed and direction obtained from an operational atmospheric ensemble prediction system propagate into CDOP3SiX and consequently on the estimation of RVL currents. The conditions where the interplay between these two variables may induce high uncertainty on the estimations are assessed, as well as their allowed maximum value to secure reliable retrieval based on user-defined thresholds.

This work is structured as follows: section 2 describes the main ocean surface circulation features in Skagerrak. Section 3 introduces the S1 data set, the Doppler centroid anomaly and CDOP3SiX, the operational atmospheric ensemble prediction (MEPS) and wave model (MyWaveWAM), and the two uncertainty quantification approaches used in this work. Section 4 presents the results found in this research. We discuss the results and present final concluding remarks in Section 5

2 Study area

The Skagerrak and Kattegat are located between the Scandinavian countries (see Figure 1). The current circulation system in Skagerrak is primarily cyclonic, with the outflowing Norwegian Coastal Current (NCC) running alongshore southern Norway (solid arrow in Figure 1a) and the inflowing Jutland Current (JC) circulating around northern Denmark (dotted arrow in Figure 1a). Christensen et al. (2018) showed that the surface circulation is mainly locally wind-driven, with a lagged response of the NCC and JC relative to the wind forcing peaking at 7 and 4 days, respectively.

The atmospheric circulation in the region is characterized by a monsoon-like pattern (Sætre et al., 1988), where anti-cyclonic circulation induced by atmospheric high-pressure on land prevails in winter. Westerly-Northwesterly winds are predominant during the summer months, going around the southern tip of Norway into Skagerrak (Gröger et al., 2019). Synoptic polar lows and fronts are frequent between January and May and might reach wind speeds above 20 m/s (Noer et al., 2011). The orientation and strength of the wind field may also cause upwelling (Sætre et al., 1988). Outflow-blocking conditions are also observed in Skagerrak, leading to instability growth at the surface layers during the NCC adjustment for a new wind condition (Melsom, 2005). Under prevailing conditions, the direction of the NCC and JC, as well as the typical atmospheric circulation, are both perpendicular to the satellite's orbit. This arrangement offers an excellent configuration for instantaneous RVL retrievals.

3 Data and Methods

This section describes the S1 SAR data used, the Doppler centroid anomaly derived from the single look complex (SLC) data, and the model products used in the study. We also introduce the CDOP3SiX GMF and the uncertainty estimation approach used in this work.

3.1 Sentinel-1

We used 251 S1 images acquired in ascending and descending passes between January and July 2023 over the Skagerrak and Kattegat areas (Figure 1). The flow of the NCC and JC in the area is approximately perpendicular to the orbit path, creating favorable conditions for resolving dominant ocean surface currents in the SAR Doppler shift. We used the Doppler shift from the S1 Ocean Surface Current RVL products, openly available from the ESA World Ocean Circulation project (NERSC, 2022). The products have a swath width of 250 km and a pixel size of 1x1 km^2, and are available in coastal areas up to two times per day at higher latitudes.

The World Ocean Circulation (e.g. European Space Agency, 2024) data set is based on the standard S1 IW Level-2 (L2) Ocean RVL products derived using S1 L2 Instrument Processing Facility (IPF) processor (Engen & Johnsen, 2015; Hajduch et al., 2020; Johnsen et al., 2016) with additional calibration steps.

3.1.1 Sentinel-1 IW Doppler Centroid Anomaly

The Doppler shift estimated from S1 SLC scenes (f_{total}) contain both geophysical (f_{phys}) and non-geophysical (f_{nphys}) contributions (OceanDataLab Ltd., 2022; Moiseev et al., 2022):

$$f_{total} = \underbrace{f_{osc} + f_{wv}}_{f_{phys}} + \underbrace{f_{bias}(\beta) + f_{att}(\beta, \theta_{att}(t)) + f_{sca} + \Delta f}_{f_{nphys}} \quad (1)$$

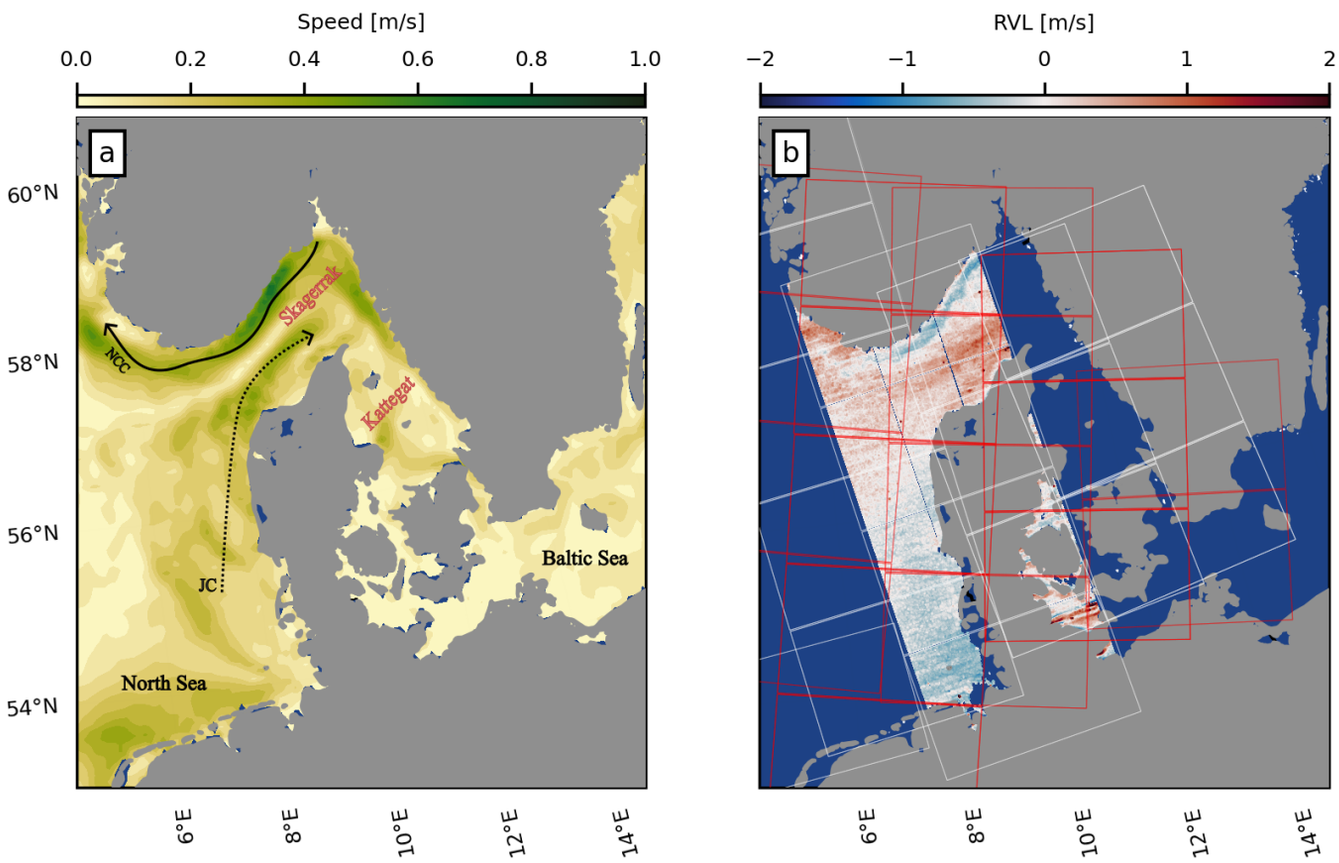

Figure 1. Illustration of the study area. The background map in (a) is the ocean surface current speed averaged between 2023.01.11 and 2023.01.12 using Global Ocean Physics Reanalysis from CMEMS(Copernicus Marine Service, 2023), and overlayed are the NCC and JC indicated as solid and dotted lines. (b) show the ascending (white inset) and descending (red inset) passes of Sentinel-1A. The colormap represents the retrieved RVL for an ascending pass (2023.01.11 11:17 UTC). Blue and red colors represent currents moving towards the satellite antenna-look angle and opposite to it, respectively.

where f_{osc} is the signal induced by range component of the ocean surface current, f_{wv} is the wind-wave bias, f_{bias} is antenna electronic miss-pointing which varies with boresight angle β, f_{att} is miss-pointing error due to unstable platform attitude, $\theta_{att}(t)$, f_{sca} is scalloping in S1 TOPS mode acquisitions, and Δf is residual unknown errors.

A novel post-processing calibration strategy (OceanDataLab Ltd., 2022) based on combining orbits of S1 measurements, orbit data, star tracker configuration, and gyroscope telemetry data was applied to remove f_{nphys}. The calibration is based on two steps: (i) use platform telemetry and orbit information to provide a restituted attitude (AUX_ESTATT) data correction and (ii) use the Doppler shift acquisitions over land within the orbit to provide data-driven antenna bias (AUX_DCBIAS) correction. Moiseev et al. (2022) estimated that the residual uncertainty in the Doppler shift after corrections is about 3.8 Hz which is equivalent to 0.21 - 0.15 m/s in ocean surface RVL from low to high satellite incidence angles.

Rearranging Eq. 1 and approximating f_{wv} using the CDOP3SiX GMF (Moiseev et al., 2022), the Doppler shift contribution of f_{osc} can be explicitly recovered as:

$$f_{osc} = f_{total} - f_{nphys} - CDOP3SiX(u_{10}, u_{swell}, u_{sea}, \theta) \tag{2}$$

where u_{10} = -Wcos(ϕ) is the wind range-directed component of a wind vector field with speed W (m/s) and direction (ϕ, degree) with respect to the SAR antenna look-direction, u_{swell} is the range-directed component of the swell orbital velocity, u_{sea} is the range directed component of the wind sea orbital velocity, and θ is the incidence angle. The RVL can thus be retrieved using f_{osc}:

$$RVL = -\frac{\pi f_{osc}}{k_e sin(\theta)} \tag{3}$$

where k_e is the radar wavenumber $k_e = \frac{2\pi}{\lambda} = \frac{2\pi}{5.54^{-2}m}$.

We show in Figure 2*a* the wind-wave bias estimation using CDOP3SiX at 30° incidence angle for wind speeds (W) between 0 and 20 m/s in 2 m/s intervals, and wind directions (ϕ) in 15° intervals, both equally spaced. The example assumes a fully developed sea for the range-directed wind sea orbital velocities, and the swell influence is absent. The function shows an asymmetric pattern, where values around 0 and 360° (upwind) are higher relative to 180° (downwind).

Figure 2b and Figure 2c show the change of the CDOP3SiX function with respect to W and ϕ in absolute values, respectively. The solid black lines represent gradients equal to 2 Hz, our defined threshold. The reference value is often requested by a user or community, but it induces variations at the RVL retrieval of about 0.1 m/s for C-band SAR at 35 ° incidence angle (Moiseev, Johnsen, Hansen, & Johannessen, 2020). Figure 2b indicates that the 2 Hz threshold is exceeded at upwind and downwind low wind speed conditions (< 4 m/s). For other scenarios, a low sensitivity to changes in the wind speed is indicated. A reversal can be seen in Figure 2c, where the function reveals high sensitivity at along-track angles (90° and 180°). These plots set the ground for our further evaluations.

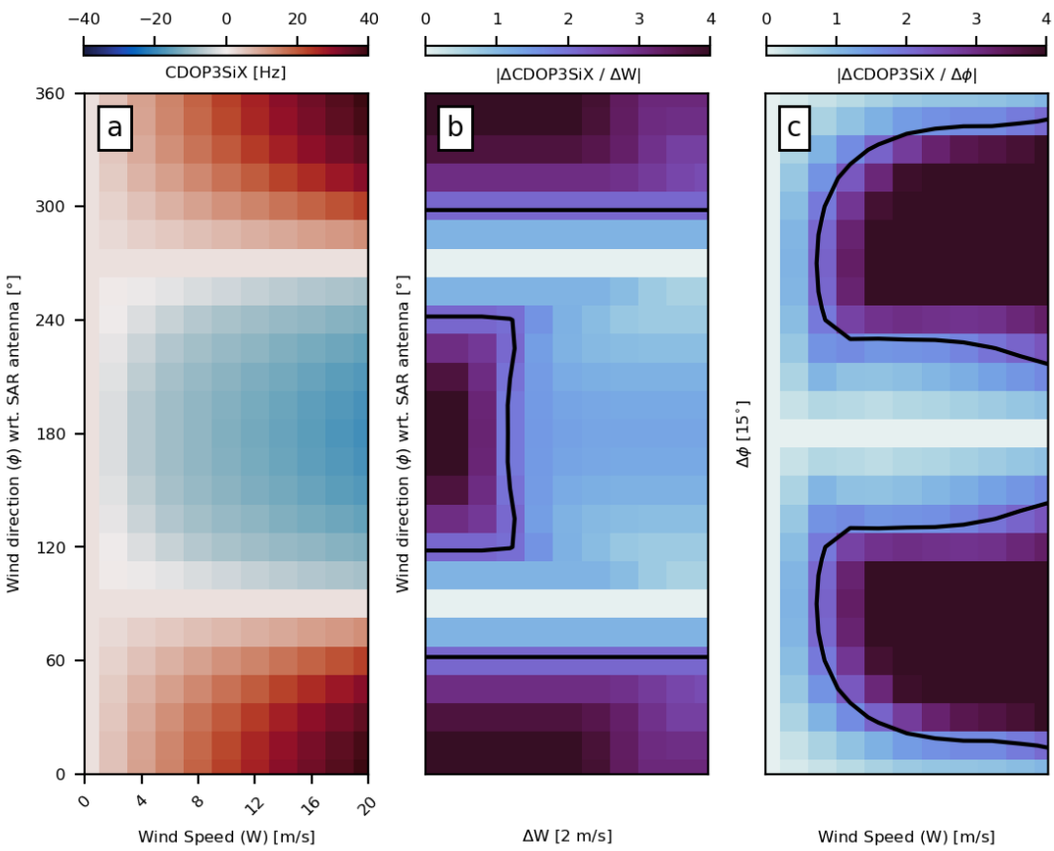

Figure 2. (a) Wind-wave induced Doppler shift predicted using CDOP3SiX GMF at 30° incidence angle for wind speeds (W) between 0 and 20 m/s every 2 m/s, and wind direction (ϕ) with respect to the SAR antenna-look direction every 15°. (b) and (c) show the change of the GMF (ΔCDOP3SiX) with respect to W (ΔW) and ϕ ($\Delta\phi$), respectively. The absolute gradient values are centered at 2 Hz, represented by the solid contours.

3.1.2 Ensemble Wind Fields and Sea State

Wind speed and direction at 10 m height were obtained from the MetCoOp ensemble system (MEPS, Frogner et al., 2019). The operational model has 30 lagged ensemble members, updated hourly and providing forecasts with +66 hours lead time on 2.5 x 2.5 km horizontal resolution. MEPS is based on the HARMONIE-AROME configuration (Bengtsson et al., 2017), with boundary conditions for the control member obtained from the ECMWF High Resolution (ECMWF-HRES) atmospheric model and from ECMWF Ensemble System (ECMWF-ENS) for the perturbed members. Further information on the model setup, including perturbations addressed in the initial conditions and model physics, can be found in (Frogner et al., 2019) and (Frogner et al., 2022).

The sea-state information is obtained from the regional MyWaveWAM wave forecast system version WAM cycle 4.7 (Gusdal & Carrasco, 2012). Wave energy is resolved at 36 directions and frequencies, partitioned between sea and swell waves. The model presents a horizontal resolution of 4 km and provides hourly outputs. The model physics is based on Ardhuin et al. (2010), but it was modified to improve wave growth in high wind conditions following Breivik et al. (2022).

Both wind (MEPS) and wave (MyWaveWAM) model products are interpolated onto each of the S1 scene domains. The wind direction of each MEPS ensemble member is then reprojected relative to the satellite antenna look-direction such as 0° (180°) represents winds flowing towards (opposite) to the SAR antenna look-direction, and 90° or 270° are along-track winds. Range mean orbital wave velocities for swell (u_{swell}) and wind sea (u_{sea}) in Eq. 2 are calculated as $u_{wave} = (H_{wave}/T_{wave})\cos(\psi_{wave})$, where the significant wave height (H_{wave}, in meters), period (T_{wave}, in seconds) and mean direction of propagation (ψ_{wave}, in degrees) are obtained from the MyWaveWAM model. The maximum time difference between the SAR acquisition time and the model outputs is ± 27 minutes.

3.2 Uncertainty Estimation

For each S1 acquisition, the 30 ensemble members from MEPS and the MyWaveWAM wave field are used as input in CDOP3SiX. Thirty wind-wave bias and RVL estimations are thus available for each S1 scene, with ensemble-wise averages and standard deviation obtained at each grid point. Uncertainty in the RVL retrieval should originate from the variability in the wind field, enabling investigations of the role of wind speed and direction uncertainties on the RVL field. The maximum uncertainty of the input variables can be assessed both analytically and probabilistically, as described below.

3.2.1 Maximum Uncertainty Estimation via Uncertainty Propagation

The uncertainty propagation of wind speed and direction to the CDOP3SiX function can be evaluated analytically using the following equation:

$$Z(W,\phi) = W cos(b\phi) + c cos^2(b\phi) + d cos(b\phi) + e \tag{4}$$

we fit b, c, d and e for each W used as input in CDOP3SiX using least-squares regression. The uncertainty propagation is then derived as:

$$\Delta Z^2 = \left(\frac{\partial Z}{\partial W}\Delta W\right)^2 + \left(\frac{\partial Z}{\partial \phi}\Delta\phi\right)^2 \tag{5}$$

where

$$\begin{aligned} \frac{\partial Z}{\partial W} &= cos(b\phi) \\ \frac{\partial Z}{\partial \phi} &= -Wbsin(b\phi) - 2bcsin(b\phi)cos(b\phi) - dsin(b\phi) \end{aligned} \tag{6}$$

The maximum ΔW and $\Delta\phi$ given ΔZ $\leq$ 2 Hz is determined by maximizing the uncertainties such as:

$$\Delta Z = \sqrt{\left(\frac{\partial Z}{\partial W}\Delta W\right)^2 + \left(\frac{\partial Z}{\partial \phi}\Delta\phi\right)^2} \leq 2Hz \tag{7}$$

The method was applied for wind speeds ranging between 0 and 20 m/s and wind directions between 0 and 360°, equally spaced in 2 m/s and 15° respectively. The maximum ΔW and $\Delta\phi$ are evaluated for each combination of the variables. As in Figure 2, we assume a fully developed sea for the range-directed wind sea orbital velocities, and the swell influence is neglected.

3.2.2 Maximum Uncertainty Estimation based on Ensemble Model

The maximum wind speed and direction uncertainties are assessed using the same bin spacing as in the analytical approach. For each data point, we obtain the mean ensemble-wise value for wind speed ($\overline{W}$) and direction ($\overline{\phi}$). The average value determines which (W$_i$, ϕ_i) bin the data data point falls into. For this data point and bin, the standard deviation (σ) of W (m/s), ϕ (degrees), CDOP3SiX (Hz), and RVL (m/s) are also estimated member-wise. The process is done for all the 5.8 million data points available.

The empirical cumulative distribution function (ECDF) of σ for the four variables (σ_W, σ_ϕ, $\sigma_{CDOP3SiX}$ and σ_{RVL}) is then estimated for each bin. Using the thresholds defined above, we obtain the probabilities P($\sigma_{CDOP3SiX} \leq$ 2 Hz) and P($\sigma_{RVL} \leq$ 0.1 m/s) to determine the allowed maximum σ_W and σ_ϕ, i.e., max. σ_W | P($\sigma_{CDOP3SiX} \leq$ 2 Hz).

4 Results

We start this section by providing three examples of retrieved RVL, the auxiliary wind speed and direction information, and their uncertainties in terms of standard deviation across ensemble members. The relationship between uncertainties in the auxiliary wind fields and the wind-wave bias estimated with CDOP3SiX, and the retrieved RVL are introduced in Subsection 4.2. The maximum uncertainty estimation based on the two approaches described in subsection 3.2 are presented in Subsection 4.3.

4.1 Estimated RVL and Wind Field Uncertainties

A set of examples of RVL retrievals represent cases where the atmospheric state exhibits low predictability (Fig. 3), high predictability (Figure 4) and the passing of a cold front (Figure 5). Subfigures a, b and c in each example show the control member of the retrieved RVL (m/s), wind speed (m/s) and wind direction (∘) with respect to the SAR antenna-look angle, respectively. The corresponding ensemble member isobars (in hPa) are also displayed in subfigures b. Subfigures d, e and f show the ensemble-wise standard deviation of the corresponding fields at each grid point. As in Figure 1, blue values in panels a indicate currents flowing towards the satellite antenna-look direction and red values opposite to it.

The RVL field in Figure 3 refers to the S1 ascending pass on 2023.01.16 at 17:17:40 UTC. The NCC signal is distinct in comparison to the examples shown in Figure 4 and Figure 1, displaying rather a meandering-like shape. Additionally, a low atmospheric pressure band (980 hPa) and associated low wind speed (< 5 ṁ/s) can be seen between southern Norway and the Netherlands. Convective cells and low wind speed streaks can also be observed in the S1 backscatter (not shown) along the shear zone visible in Figure 3c. As each ensemble member provides different positioning of the convective band, high variability in the wind speed (Figure 3e) and direction (Figure 3f) yields an RVL uncertainty estimation above 0.1 m/s over the whole domain (Figure 3d).

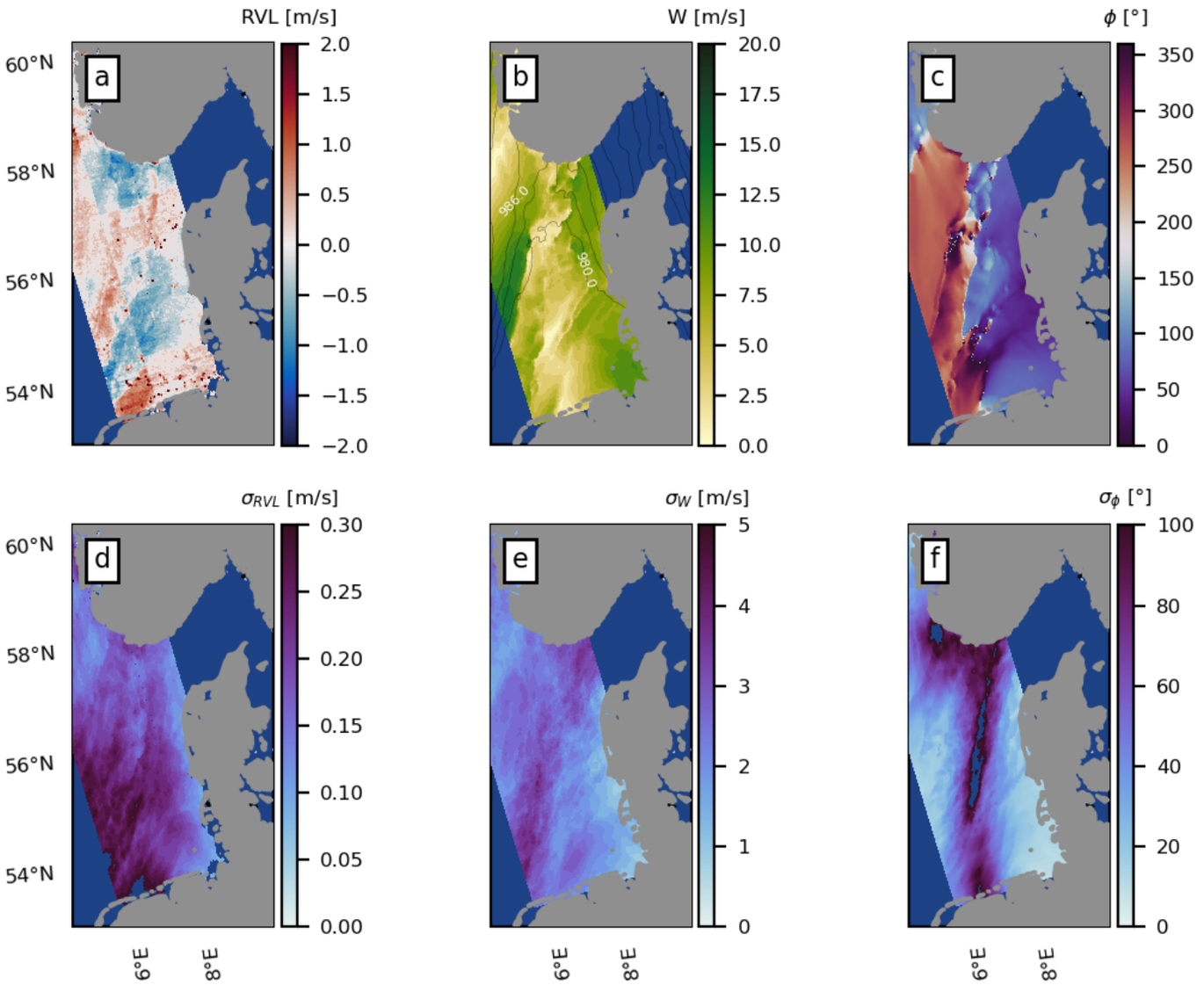

Figure 3. Example of the ocean surface current RVL retrievals with low wind predictability. Ascending pass, 2023.01.16 at 17:17:40 UTC. Panels a, b, and c show the retrieved RVL (m/s), wind speed (m/s), and wind direction with respect to the satellite antenna-look direction (degree), respectively. Panels d, e, and f show the corresponding member-wise standard deviation at each grid point.

The example shown in Figure 4 represents a case where the atmospheric state is seemingly predictable. The NCC signal is visible in Figure 4a. Atmospheric pressure above 1030 hPa is observed over the domain, with wind speeds ranging from 15 m/s in southern Norway to less than 5 m/s off northern Germany (southwest edge, Figure 4b). The wind direction is approximately 180 - 200° (southeasterly winds) with respect to the radar-look direction (descending pass, 2023.02.04 at 05:57:48 UTC). Uncertainties of both wind speed (Figure 4e) and wind direction (Figure 4f) are low over almost the entire domain resulting in reliable estimates of the RVL (Figure 4 d). At the southwestern corner of Figure 4d where σ_{RVL} reaches 0.3 m/s, high σ_W relative to the ensemble member's wind speed (Figure 4e) and σ_ϕ exceeding 100° (Figure 4d) can also be seen.

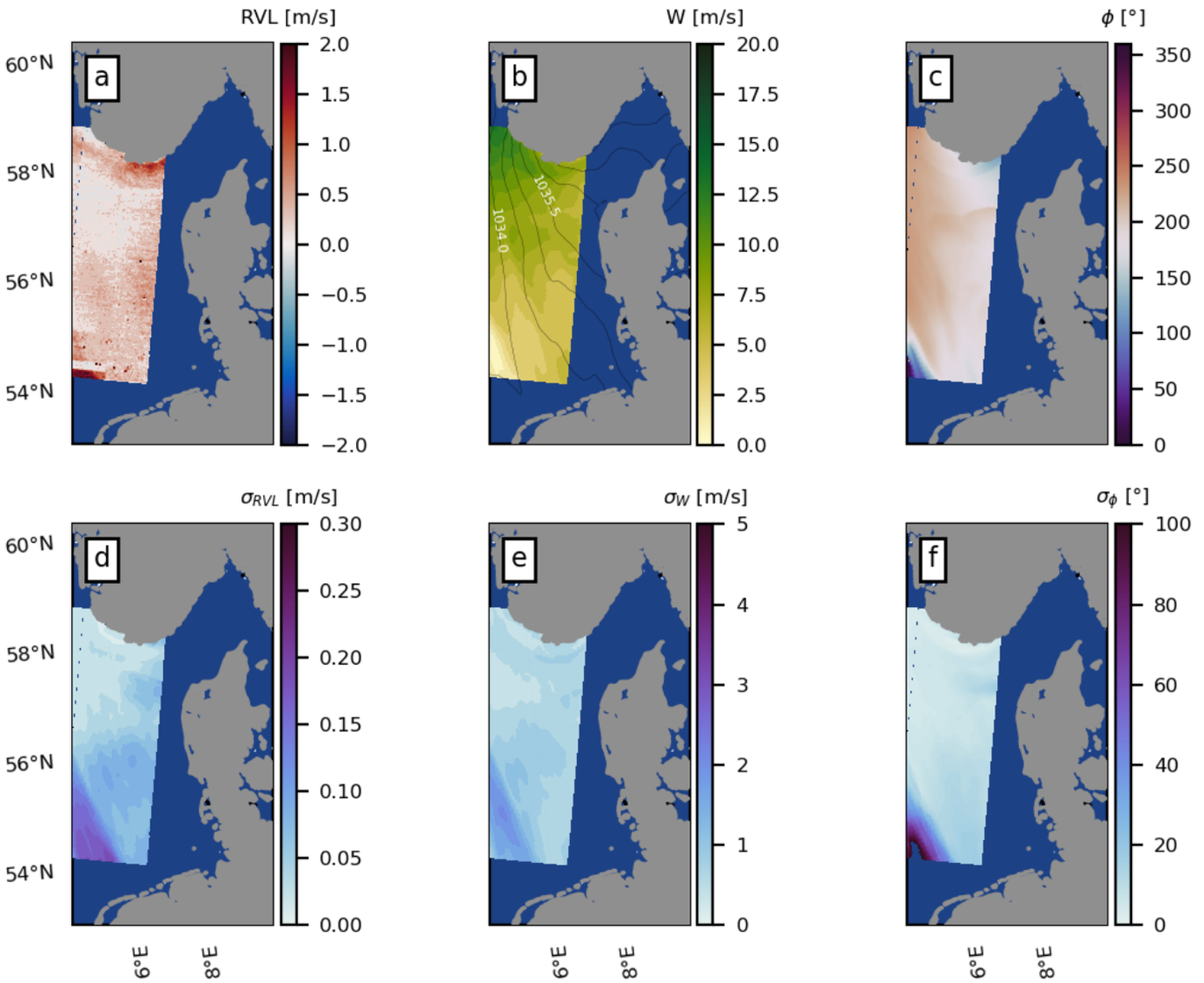

Figure 4. Same as Figure 3, but for a case with high predictability for wind speed and direction. Descending pass, 2023.02.04 at 05:57:48 UTC.

The passing front (Fig. **??**) represents a case (ascending pass, 2023.01.04 at 17:18:55 UTC) where horizontal shear is well defined and limited within a distinguished area (see Figure 5). The red stripe running west-northwest in the RVL field (Figure 5a) is interpreted as an artifact due to the sharp decline of wind speed along the front (Figure 5b) and shift (180° - 360° in wind directions (Figure 5c). South of the front, where variations in both wind speed and direction are small, low uncertainty on the retrieved RVL is achieved (Figure 5d).

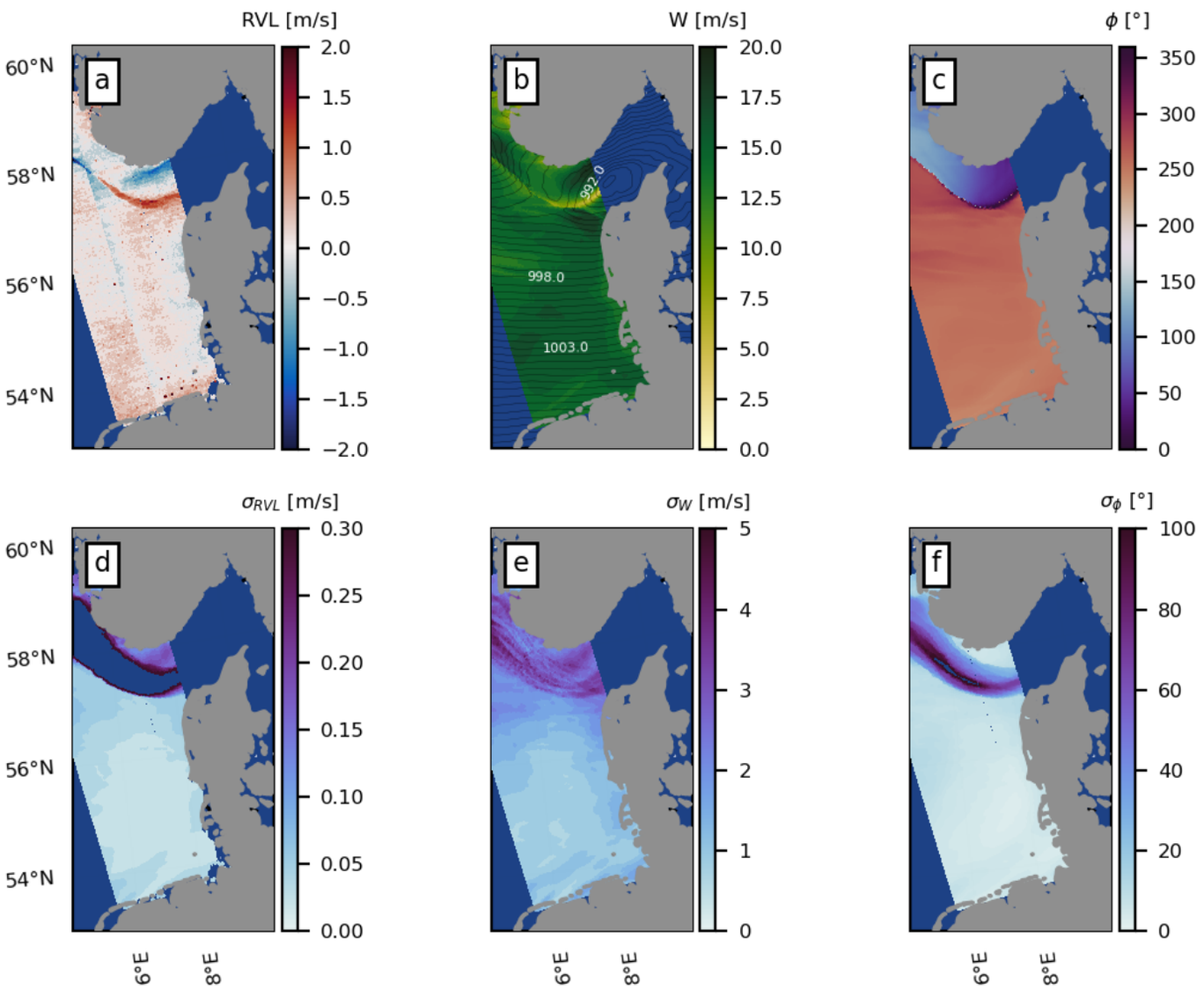

Figure 5. Same as Figure 3, but for a case with a passing front. Ascending pass 2023.01.04 at 17:18:55 UTC.

4.2 Wind speed and direction uncertainty analysis

The examples presented above depict different atmospheric scenarios where the RVL uncertainty levels exceed the 0.1 m/s threshold. We now assess the interplay between wind speed and direction uncertainties leading to these findings.

With wind speed uncertainty (σ_W) being bounded by the magnitude of the mean field ($\overline{W}$), we show in Figure 6 the 2D-histogram of the ratio $\overline{W}/\sigma_W$ and the standard deviation of wind direction with respect to the antenna-look direction (σ_ϕ, in degrees). The relationship between this ratio and the uncertainties in the wind-wave bias (b, $\sigma_{CDOP3SiX}$), and the retrieved RVL (c, σ_{RVL}) are also shown. The red line represent a fitted function of the type $\frac{\alpha}{x^\beta}$, where x is the ratio, and $\alpha = 2.57$, 5.54 and 0.26, and $\beta = 1.34$, 0.61 and 0.61 for Figure 6a, Figure 6b and Figure 6c, respectively. All data points in the 251 scenes were used.

Uncertainties in the wind direction (Figure 6a) steeply increase when the ratio is smaller than 5 (vertical line in the panels), i.e., $\sigma_W \geq 20\%$ of $\overline{W}$. We, therefore, expect high uncertainties in the wind-wave bias component and RVL estimation for such conditions, especially for low-wind speed cases (see Figure 2c, w). The figure also reveals that variations in σ_ϕ are smaller than 10° when $\sigma_W < 10\%$ of the mean wind speed. As the median of σ_W is around 0.9 m/s, wind direction uncertainties are then expected to be below 10° for mean wind fields above 8 m/s. The defined thresholds are displayed as horizontal lines in Figure 6b (2 Hz) and Figure 6c (0.1 m/s). The fitted function for both variables presents the same decay (rounded at the second decimal place), crossing the thresholds at $\overline{W}$ / $\sigma_W = 5$ (vertical line).

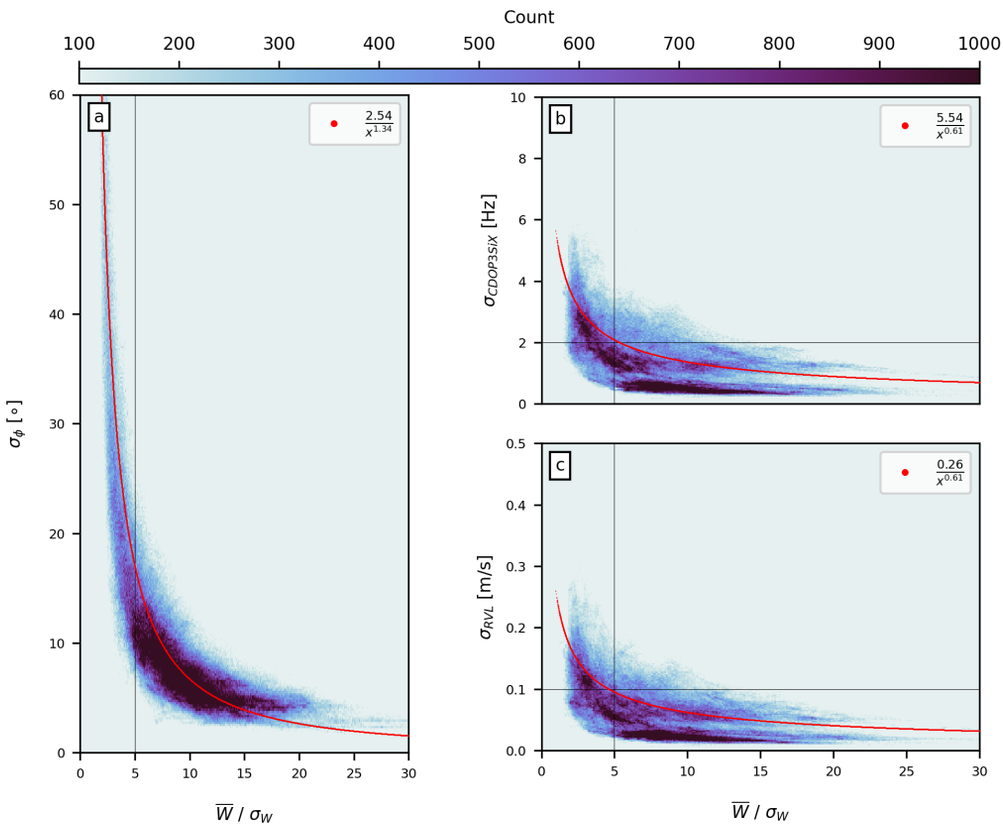

Figure 6. Heatmaps of $\overline{X}/\sigma_W$ and σ_ϕ [degree] (a); $\overline{X}/\sigma_W$ and $\sigma_{CDOP3SiX}[Hz]$ (b) and $\overline{X}/\sigma_W$ and σ_{RVL} [m/s] (c). Vertical lines represent $\overline{X}/\sigma_W = 5$ (see text for further explanation). Horizontal lines in b and c represent their respective thresholds: 2 Hz and 0.1 m/s, respectively. The red line represents the fitted exponential decay function of the type y = $\frac{\alpha}{x^\beta}$, where x is $\overline{X}/\sigma_W$.

Interesting points can be highlighted from these results. The relationships presented in Figure 6 indicate that the uncertainty of the wind direction is related to the relative magnitude of the mean wind field and its uncertainty. Specifically, for $\overline{X}/\sigma_W$ ¡ 5, the estimations tend to become less accurate when σ_W is greater than 20% of $\overline{W}$.

Figure 7 shows the ECDF for the ratio $\overline{W}/\sigma_W$ (a) and σ_ϕ (b, in degrees) for values satisfying (solid lines) and exceeding (dotted lines) the established uncertainty thresholds for $\sigma_{CDOP3SiX}$ (blue, 2 Hz) and σ_{RVL} (black, 0.1 m/s). Diamonds (circle) represent the 25, 50, 75, and 95 percentiles for when the threshold is exceeded, with red (gray) for $\sigma_{CDOP3SiX}$ (σ_{RVL}).

The ECDFs are equivalent to each other, i.e., uncertainties in the wind-wave bias removal are translated almost linearly to uncertainties in the RVL retrieval for the ranges considered here. The ECDF curves quickly diverge from each other and show reversed characteristics: while the $\overline{W}/\sigma_W$ ratio shows increasing probabilities towards high values (7 a), uncertainties in the wind direction are concentrated around 10° (7 b).

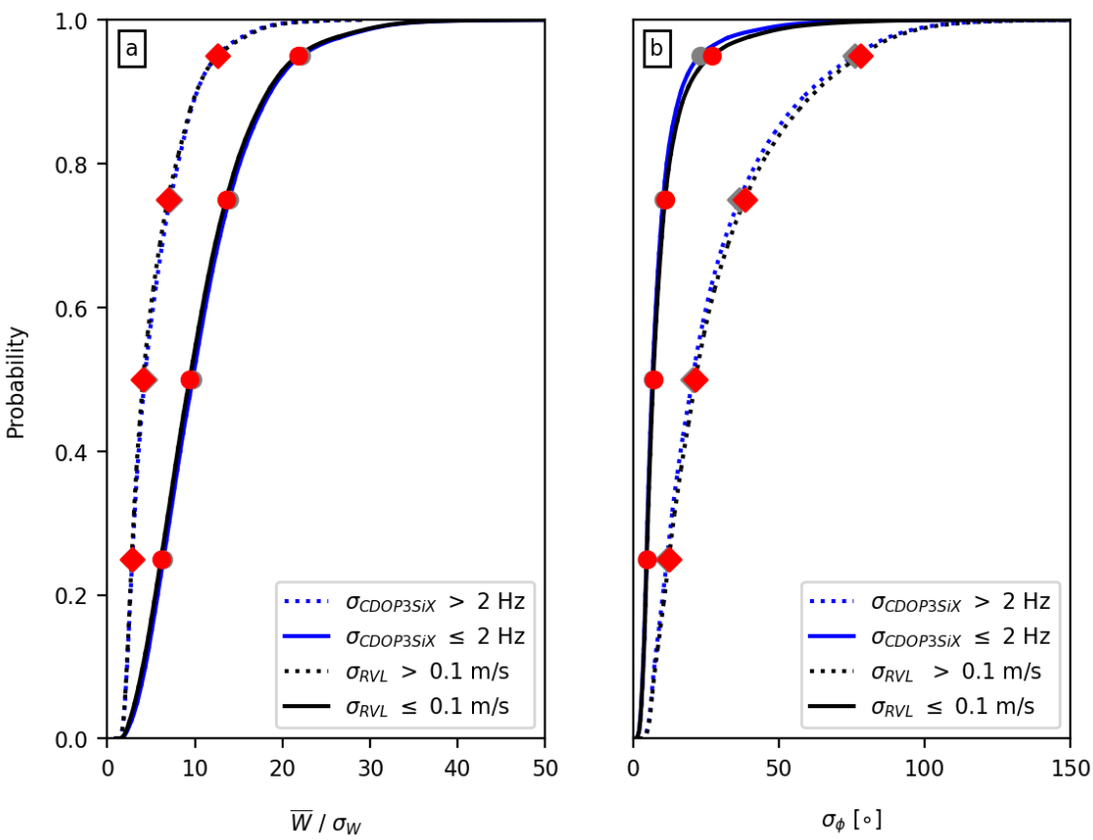

Figure 7. Empirical cumulative distribution function (ECDF) for $\overline{W}/\sigma_W$ (a) and σ_ϕ (b, in degrees). These variables are split into 2 groups based on the thresholds: $\sigma_{CDOP3SiX} \leq 2$ Hz (solid blue) and > 2 Hz (dotted blue); $\sigma_{RVL} \leq 0.1$ m/s (solid black) and > 0.1 m/s (dotted black). Diamond (circle) represents the 25, 50, 75, and 95 percentiles for when the threshold is exceeded (satisfied), with red (gray) for $\sigma_{CDOP3SiX}$ (σ_{RVL}).

4.3 Maximum Uncertainty Estimation

The previous section highlighted the interplay between modeled wind speed and direction on the uncertainty estimation of the wind-wave bias and RVL. We now proceed to quantify the maximum uncertainty in the input fields that satisfy the defined thresholds. Figure 8a and Figure 8b show the results obtained with the uncertainty propagation scheme (Subsection 3.2.1), whereas Figure 8c and Figure 8d were obtained through the ensemble ECDFs (see 3.2.2). Results for $\sigma_{CDOP3SiX}$ are not displayed to avoid further redundancy.

For low-wind speed ($\leq$ 4 - 5 m/s) and along-track angles (dashed lines), ΔW does not exceed 1 m/s. For the maximum uncertainty in the wind direction, one can see the correspondence between Figure 8b and Figure 2c. For the purple areas in the latter (variations above 2 Hz every 15°), $\Delta\phi$ ought to be below 10°. Two pitfalls of the simple model must nevertheless be stressed: it does not capture the asymmetric pattern (Figure 8a) and $\Delta\phi$ is weakly constrained for wind speed above 8 m/s (Figure 8b).

The maximum uncertainty estimated using the ensemble model through the evaluation of their ECDFs resembles the results obtained with the analytical uncertainty propagation. Overall, σ_W reaches up to 1.8 m/s for downwind angles (solid line), but are limited to approximately 1.2 m/s at upwind angles (Figure 8c). For mean wind speed values above 8 m/s, σ_ϕ does not exceed 25° and 15° for the corresponding angles. Similar values were found for the maximum uncertainty estimation in the wind-wave bias removal (not shown).

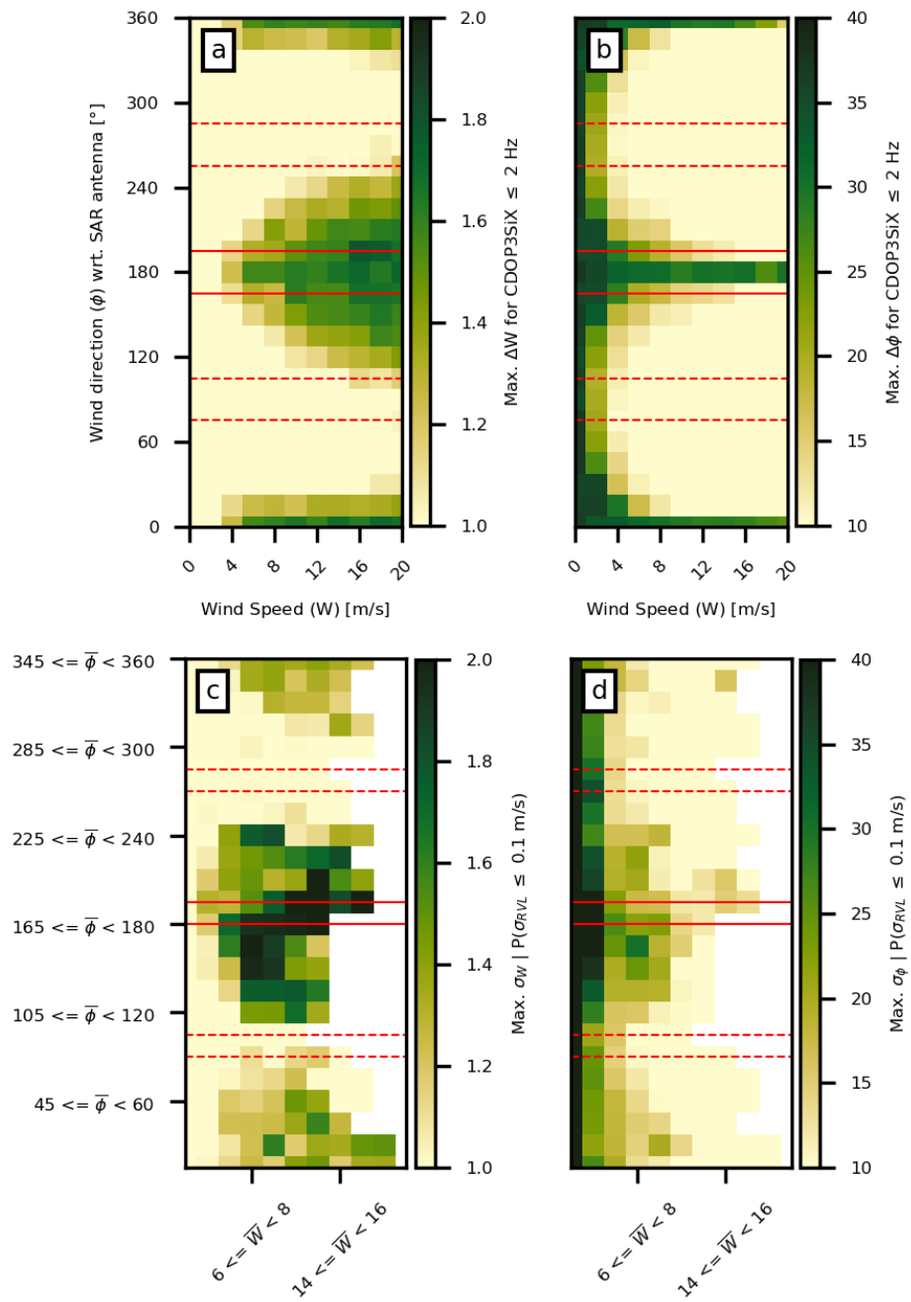

Figure 8. Wind speed (a and c) and direction (b and d) maximum uncertainty estimation. (a) and (b) show the wind-wave bias uncertainty estimation based on uncertainty propagation described in Subsection 3.2.1. (c) and (d) show RVL uncertainty estimation as described in Subsection 3.2.2. Red dashed (solid) lines represent along-track (downwind) wind directions with respect to the satellite antenna-look direction. See the text for further clarification.

5 Discussion and Concluding Remarks

Employing an operational atmospheric ensemble prediction system and 251 S1 scenes, we evaluated how uncertainties in the wind speed and direction affect wind-wave bias removal and RVL retrievals using the CDOP3SiX GMF. Focusing on the Skagerrak and the Kattegat, two major contributions were presented, namely the atmospheric conditions under which high uncertainty in the wind-wave bias removal and RVL retrieval is expected and the maximum uncertainty of wind speed and direction that suffices thresholds of 2 Hz and 0.1 m/s.

Three findings are reported here: (1) the ratio between the mean wind speed, $\overline{W}$, and wind speed ensemble spread, σ_W, ($\overline{W}/\sigma_W$) may serve as a proxy for RVL uncertainty estimation; (2) low wind speed and atmospheric fronts are the major cause of unreliable retrievals and (3) The maximum uncertainty depends on the SAR antenna-look direction. Ensemble predictions of wind speed, as presented here, show that the wind direction uncertainties grow exponentially when the wind speed uncertainty is equal to or greater than 20% of the mean wind speed (Figure 6a). Above this value, the wind-wave bias uncertainty solely due to the wind forcing can partially explain the residual variation of around 3.8 Hz after recalibration found by Moiseev et al. (2022). Considering MEPS median (mean) σ_W of 0.9 m/s (1 m/s), these findings indicate that the wind-wave bias and RVL uncertainties tend to exceed their respective thresholds in low-wind speed conditions ($<$ 4-5 m/s).

Similar results were described by Elyouncha et al. (2021), who stress that errors in the wind speed direction and retrieved ocean current direction increase at low-wind speed conditions. As the signal-to-noise ratio is reduced (Bao et al., 2017), a weaker Doppler information signal arrives at the antenna and hence increases the error in the DCA estimation. Additionally, atmospheric models exhibit large wind direction uncertainties at low wind speeds due to the relatively high wind speed and direction errors in scatterometers at that range (Pickett et al., 2003; Verspeek et al., 2010; Chou et al., 2013; Zhao et al., 2021). These two contributions, sensor and model-specific, hinder reliable RVL retrievals at low-wind speed ranges, and estimations should be treated with caution.

Previous research conducted by Johnsen (2016) and Elyouncha et al. (2019) showed that wind speed uncertainty lower than 2 m/s and 1 m/s is necessary to fulfill the established thresholds, respectively. The maximum wind speed and direction uncertainty estimations presented in this study (Fig. 8) extend the findings of Johnsen (2016) and Elyouncha et al. (2019) to a range of angles relative to the satellite antenna-look direction and wind direction. For upwind conditions at moderate to high wind speeds, σ_W rarely exceeds 1.5 m/s and σ_ϕ is stringent for $\overline{W} \geq 8$ m/s. For downwind conditions, values are about 0.5 m/s and 10° higher due to the lower sensitivity of CDOP3SiX at those angles. The estimation made by Elyouncha et al. (2019) holds for along-track winds.

Two remarks are further stressed: we found that the uncertainties alter the magnitude of the retrieval but not the directionality as much. In other words, if the NCC has a strong Doppler signal, the uncertainties in the wind field will impact its retrieved speed but not 'reverse' its direction. This ultimately expresses the need for accurate Doppler shift observations and removal of nongeophysical component removal (e.g., Moiseev et al., 2022; Martin et al., 2022). Secondly, reducing the atmospheric model spread to achieve the required criteria may lead to an exceeding over-confident ensemble system, of which misses extreme cases and is not able to reproduce the atmospheric variability. Rather than decreasing the atmospheric model spread, efforts should be put into investigations on the reliability and consistency of the ensemble estimations against independent observations.

Employing atmospheric ensemble models on the uncertainty quantification of wind-wave bias and RVL allows us to directly estimate errors that are due to wind uncertainty in a specific situation. Reliable uncertainty estimations of the NCC and JC are possible due to the satellite's orbit and the predominant wind direction in the region (west-northwest and east-northeast) for wind speeds between 6 and 12 m m/s. Although our results are model-dependent, further investigations using the global ECMWF ensemble product is expected to provide comparable results in other regions. The ECMWF-ENS product currently has the same spatial resolution as the HRES model, offering a new possibility for retraining GMFs in a Bayesian fashion and providing uncertainty estimations to RVL-product users. Future investigations should also include the impact of current-wave interactions on the wind-wave bias component based on wave ensemble prediction models.

References

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J.-F., Magne, R., Roland, A., ... Collard, F. (2010). Semiempirical dissipation source functions for ocean waves. part i: Definition, calibration, and validation. *Journal of Physical Oceanography*, *40*(9), 1917 - 1941. Retrieved from https://journals.ametsoc.org/view/journals/phoc/40/9/2010jpo4324.1.xml doi: 10.1175/2010JPO4324.1

Bao, Q., Lin, M., Zhang, Y., Dong, X., Lang, S., & Gong, P. (2017). Ocean surface current inversion method for a doppler scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(11), 6505-6516.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., ... Ødegaard Køltzow, M. (2017). The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Monthly Weather Review*, *145*(5), 1919 - 1935. Retrieved from https://journals.ametsoc.org/view/journals/mwre/145/5/mwr-d-16-0417.1.xml doi: 10.1175/MWR-D-16-0417.1

Breivik, Ø., & Arthur Allen, A. (2008). An operational search and rescue model for the norwegian sea and the north sea. *Journal of Marine Systems*, *69*(1), 99-113. Retrieved from https://www.sciencedirect.com/science/article/pii/S0924796307000383 (Maritime Rapid Environmental Assessment) doi: https://doi.org/10.1016/j.jmarsys.2007.02.010

Breivik, Ø., Carrasco, A., Haakenstad, H., Aarnes, O. J., Behrens, A., Bidlot, J.-R., ... Reistad, M. (2022). The impact of a reduced high-wind charnock parameter on wave growth with application to the north sea, the norwegian sea, and the arctic ocean. *Journal of Geophysical Research: Oceans*, *127*(3), e2021JC018196. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JC018196 (e2021JC018196 2021JC018196) doi: https://doi.org/10.1029/2021JC018196

Brekke, C., Espeseth, M., Dagestad, K.-F., Röhrs, J., Hole, L. R., & Reigber, A. (2021). Integrated analysis of multisensor datasets and oil drift simulations — A free-floating oil experiment in the open ocean. *Journal of Geophysical Research: Oceans*, *126*(1), e2020JC016499. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JC016499 doi: https://doi.org/10.1029/2020JC016499

Chapron, B., Collard, F., & Ardhuin, F. (2005). Direct measurements of ocean surface velocity from space: Interpretation and validation. *Journal of Geophysical Research: Oceans*, *110*(C7). doi: https://doi.org/10.1029/2004JC002809

Chou, K.-H., Wu, C.-C., & Lin, S.-Z. (2013). Assessment of the ASCAT wind error characteristics by global dropwind sonde observations. *Journal of Geophysical Research: Atmospheres*, *118*(16), 9011-9021. doi: https://doi.org/10.1002/jgrd.50724

Christensen, K., Sperrevik, A., & Broström, G. (2018). On the variability in the onset of the Norwegian Coastal Current. *Journal of Physical Oceanography*, *48*(3), 723 - 738. Retrieved from https://journals.ametsoc.org/view/journals/phoc/48/3/jpo-d-17-0117.1.xml doi: 10.1175/JPO-D-17-0117.1

Collard, F., Mouche, A., Chapron, B., Danilo, C., & Johannessen, J. (2008, 01). Routine high resolution observation of selected major surface currents from space. In *Proceedings of seasar, frascati.*

Copernicus Marine Service. (2023). *Global ocean physics reanalysis.* https://doi.org/10.48670/moi-00021. (Accessed: 2024-10-12)

Elyouncha, A., Eriksson, L., Broström, G., Axell, L., & Ulander, L. (2021). Joint retrieval of ocean surface wind and current vectors from satellite SAR data using a Bayesian inversion method. *Remote Sensing of Environment*, *260*, 112455. doi: https://doi.org/10.1016/j.rse.2021.112455

Elyouncha, A., Eriksson, L., Romeiser, R., & Ulander, L. (2019). Measurements

of sea surface currents in the Baltic Sea region using spaceborne along-track InSAR. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 8584-8599. doi: 10.1109/TGRS.2019.2921705

Engen, G., & Johnsen, H. (2015). *Sentinel-1 doppler and ocean radial velocity algorithm definition* (Tech. Rep.). Norut Report No. 19/2015, S1-TN-NRT-53-0658, version 1.4.

European Space Agency. (2024). *World ocean circulation: Understanding the ocean's role in climate.* https://www.worldoceancirculation.org/. (Accessed: 2024-10-12)

Frogner, I.-L., Andrae, U., Bojarova, J., Callado, A., Escribà, P., Feddersen, H., ... Vignes, O. (2019). HarmonEPS — the HARMONIE ensemble prediction system. *Weather and Forecasting*, *34*(6), 1909 - 1937. Retrieved from https://journals.ametsoc.org/view/journals/wefo/34/6/waf-d-19-0030_1.xml doi: 10.1175/WAF-D-19-0030.1

Frogner, I.-L., Andrae, U., Ollinaho, P., Hally, A., Hämäläinen, K., Kauhanen, J., ... Yazgi, K. (2022). Model uncertainty representation in a convection-permitting ensemble—SPP and SPPT in HarmonEPS. *Monthly Weather Review*, *150*(4), 775 - 795. Retrieved from https://journals.ametsoc.org/view/journals/mwre/150/4/MWR-D-21-0099.1.xml doi: 10.1175/MWR-D-21-0099.1

Gröger, M., Arneborg, L., Dieterich, A., C. Höglund, & Meier, H. (2019). Summer hydrographic changes in the baltic sea, kattegat and skagerrak projected in an ensemble of climate scenarios downscaled with a coupled regional ocean–sea ice–atmosphere model. *Climate Dynamics*, *53*, 5945–5966. doi: 10.1007/s00382-019-04908-9

Gusdal, Y., & Carrasco, A. (2012). *Validation of the Operational Wave Model WAM at met.no - Report 2011* (Tech. Rep.). Norwegian Meteorological Institute, Oslo. Retrieved from https://www.met.no/publikasjoner/met-report/met-report-2012/_/attachment/download/c7812738-6ff7-42cb-9304-16e8b71f66db:9258ba17832f5e235c0c7ea873592e3a2f4e21d7/MET-report-23-2012.pdf (Report No. 23/2012)

Hajduch, G., Bourbigot, M., Johnsen, H., & Piantanida, R. (2020). *Sentinel-1 Product Specification* (Tech. Rep.). European Space Agency. (S1-RS-MDA-52-7441, DI-MPC-PB, Ref: MPC-0240, Issue 3/7)

Hansen, M. W., Collard, F., Dagestad, K.-F., Johannessen, J. A., Fabry, P., & Chapron, B. (2011). Retrieval of sea surface range velocities from Envisat ASAR Doppler centroid measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(10), 3582-3592. doi: 10.1109/TGRS.2011.2153864

Hansen, M. W., Johannessen, J. A., Dagestad, K. F., Collard, F., & Chapron, B. (2011). Monitoring the surface inflow of Atlantic Water to the Norwegian Sea using Envisat ASAR. *Journal of Geophysical Research: Oceans*, *116*(C12). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JC007375 doi: https://doi.org/10.1029/2011JC007375

Johannessen, J., Chapron, B., Collard, F., Kudryavtsev, V., Mouche, A., Akimov, D., & Dagestad, K.-F. (2008). Direct ocean surface velocity measurements from space: Improved quantitative interpretation of Envisat ASAR observations. *Geophysical Research Letters*, *35*(22). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GL035709 doi: https://doi.org/10.1029/2008GL035709

Johnsen, H. (2016). *S1+CS feasibility study: Assessment of wind/wave/current retrieval methodology* (Tech. Rep.). NORUT. Retrieved from https://norceresearch.brage.unit.no/norceresearch-xmlui/handle/11250/2658242?show=full (Report No. 20/2016)

Johnsen, H., Nilsen, V., Engen, G., Mouche, A. A., & Collard, F. (2016). Ocean doppler anomaly and ocean surface current from Sentinel-1 TOPS mode. In *2016 IEEE International Geoscience and Remote Sensing Symposium*

(IGARSS) (p. 3993-3996). doi: 10.1109/IGARSS.2016.7730038

Leutbecher, M., & Palmer, T. (2008). Ensemble forecasting. *Journal of Computational Physics*, *227*(7), 3515-3539. Retrieved from https://www.sciencedirect.com/science/article/pii/S0021999107000812 (Predicting weather, climate and extreme events) doi: https://doi.org/10.1016/j.jcp.2007.02.014

Martin, A., Gommenginger, C. P., Jacob, B., & Staneva, J. (2022). First multi-year assessment of Sentinel-1 radial velocity products using HF radar currents in a coastal environment. *Remote Sensing of Environment*, *268*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0034425721004788 doi: https://doi.org/10.1016/j.rse.2021.112758

Melsom, A. (2005). Mesoscale activity in the North sea as seen in ensemble simulations. *Ocean Dynamics*, *55*(5), 338-350. doi: 10.1007/s10236-005-0016-3

Moiseev, A., Johannessen, J. A., & Johnsen, H. (2022). Towards retrieving reliable ocean surface currents in the coastal zone from the Sentinel-1 Doppler shift observations. *Journal of Geophysical Research: Oceans*, *127*(5), e2021JC018201. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JC018201 (e2021JC018201 2021JC018201) doi: https://doi.org/10.1029/2021JC018201

Moiseev, A., Johnsen, H., Hansen, M. W., & Johannessen, J. A. (2020). Evaluation of radial ocean surface currents derived from sentinel-1 iw doppler shift using coastal radar and lagrangian surface drifter observations. *Journal of Geophysical Research: Oceans*, *125*(4), e2019JC015743. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015743 (e2019JC015743 10.1029/2019JC015743) doi: https://doi.org/10.1029/2019JC015743

Moiseev, A., Johnsen, H., Johannessen, J. A., Collard, F., & Guitton, G. (2020). On removal of sea state contribution to Sentinel-1 Doppler shift for retrieving reliable ocean surface current. *Journal of Geophysical Research: Oceans*, *125*(9), e2020JC016288. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JC016288 (e2020JC016288 2020JC016288) doi: https://doi.org/10.1029/2020JC016288

Mouche, A. A., Collard, F., Chapron, B., Dagestad, K.-F., Guitton, G., Johannessen, J. A., ... Hansen, M. W. (2012). On the use of Doppler shift for sea surface wind retrieval from SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(7), 2901-2909. doi: 10.1109/TGRS.2011.2174998

NERSC. (2022). *Sentinel-1 IW ocean surface current radial velocity from ESA WOC project. ver. 2.0. obtained from CERSAT/Ifremer, Plouzane, France.* Retrieved 2024-01-30, from https://www.worldoceancirculation.org/Products#/metadata/eb10fc04-0d51-42b5-8788-f9b40130c2a1

Noer, G., Saetra, Ø., Lien, T., & Gusdal, Y. (2011). A climatological study of polar lows in the Nordic Seas. *Quarterly Journal of the Royal Meteorological Society*, *137*(660), 1762-1772. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.846 doi: https://doi.org/10.1002/qj.846

OceanDataLab Ltd. (2022). *Algorithm Theoretical Basis Document* (Tech. Rep.). (Doc. S1RVL-EXT-DEL05ATBD, Issue 1.0)

Pickett, M., Tang, W., Rosenfeld, L., & Wash, C. (2003). QuikSCAT satellite comparisons with nearshore buoy wind data off the U.S. West Coast. *Journal of Atmospheric and Oceanic Technology*, *20*(12), 1869 - 1879. doi: 10.1175/1520-0426(2003)020⟨1869:QSCWNB⟩2.0.CO;2

Romeiser, R., Runge, H., Suchandt, S., Kahle, R., Rossi, C., & Bell, P. S. (2014). Quality assessment of surface current fields from TerraSAR-X and TanDEM-X along-track interferometry and doppler centroid analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(5), 2759-2772. doi: 10.1109/TGRS.2013.2265659

Romeiser, R., Suchandt, S., Runge, H., Steinbrecher, U., & Grunler, S. (2010). First analysis of terrasar-x along-track insar-derived current fields. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(2), 820-829. doi: 10.1109/TGRS.2009.2030885

Sætre, R., Aure, J., & Ljøen, R. (1988). Wind effects on the lateral extension of the Norwegian Coastal Water. *Continental Shelf Research*, *8*(3), 239-253. Retrieved from `https://www.sciencedirect.com/science/article/pii/0278434388900313` doi: https://doi.org/10.1016/0278-4343(88)90031-3

Verspeek, J., Stoffelen, A., Portabella, M., Bonekamp, H., Anderson, C., & Saldana, J. F. (2010). Validation and calibration of ASCAT using CMOD5.n. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 386-395. doi: 10.1109/TGRS.2009.2027896

World Meteorological Organization. (2024). *Ocean surface currents vector.* `https://space.oscar.wmo.int/variables/view/ocean_surface_currents_vector`. (Accessed: 2024-10-12)

Zhao, K., Zhao, C., & Chen, G. (2021). Evaluation of chinese scatterometer ocean surface wind data: Preliminary analysis. *Earth and Space Science*, *8*(7), e2020EA001482. doi: https://doi.org/10.1029/2020EA001482

Open Research Section

The data set, as well as the collocated fields, are available on: `https://dataverse.no/privateurl.xhtml?token=ef4d8075-2aee-472d-8b95-3399eb8b5e0f`. NOTE: The folder will be made publicly available as supporting data for the published manuscript. MEPS and MyWaveWAM model outputs are freely available at: `https://thredds.met.no/`.

Acknowledgments

We acknowledge funding from the Research Council of Norway through grant no. 237906 (CIRFA). We also acknowledge ESA World Ocean Circulation project Contract No. 4000130730/20/I-NB for granting access to re-calibrated Sentinel-1 products and the CDOP3SiX.

/9

Conclusion & Future Work

9.1 Conclusions

This thesis focused on estimating the uncertainty of ocean surface currents and modeled trajectories using ocean and atmospheric ensemble prediction systems, oil slick observations, and ocean currents retrieved with SAR. This chapter provides the main conclusions and suggestions for future investigations. We begin this chapter by providing a quick overview of the need for constraints in uncertainty studies, moving to conclusions specifically related to the objectives defined in Chapter 1, and finalizing it with future work suggestions for different topics encompassing this thesis' scope.

9.1.1 Uncertainty under constraints

Uncertainties are intrinsic to both modeling and observational systems, necessitating constraints for effective quantification. The first two articles of this work leveraged low-uncertainty observations to evaluate model performance. In Paper I, one of the key challenges was integrating SAR observations with model outputs within a quantitative analytical framework, as the two-dimensional nature of the surface oil slicks required the use of different metrics to assess the position, extent, deviation angle, and overlap between modeled and observed slicks. Furthermore, evaluating the ensemble's performance using widely accepted metrics such as error-to-spread and rank histograms required adapting these methods to accommodate the two-dimensional context. Bridging the produced water time series and trajectory predictions through these metrics was fundamental to assessing the model's performance.

Unlike Papers I and II, where uncertainty arose from the modeling system and low-uncertainty observations were used to validate predictions, a different approach was needed to assess the uncertainty in the radial velocity estimates in Paper III. In this case, community-guided thresholds were employed to determine 'how uncertain' the estimations were. Employing these thresholds was essential to determine the impact of wind speed and direction uncertainties and their implications on the radial velocity estimations. Without observations or thresholds, the uncertainty quantification becomes an unconstrained problem, shifting the focus of uncertainty studies toward sensitivity analysis.

9.1.2 Objective-specific conclusions

Conclusions for each objective defined in Chapter 1 are presented here.

Objective 1: *Investigate the short-term forecasting capabilities and ensemble performance of an operational ocean ensemble prediction system using remotely*

sensed and in-situ data.

Accurately predicting the trajectory of an oil slick drift depends heavily on how well the driving forces are represented. Horizontal advection driven by ocean currents, winds, and Stokes drift are the primary contributors, and these factors were addressed in Paper I. We evaluated short-term drift simulations (6 hours) forced by Barents-2.5 EPS (ensemble model, coarser resolution) and NorKyst-800 (single output, higher resolution). Here, produced water slicks delineated in Sentinel-1 scenes were treated as ground truth, and the predictions were validated using several performance metrics (e.g. error-to-spread, rank histogram, and reliability diagram). Our results showed that the ensemble model can reproduce the uncertainty of the ocean surface velocities by promoting a spread and divergence of the possible drift pathways even at such short time scales. Additionally, over multiple cases, at least one ensemble member often equaled the performance of the high-resolution model. We argue that ensemble models should be used whenever possible, even when high-resolution models are available. A final key outcome of Paper I is that using only the ensemble wind field as the driving force for simulations led to highly underdispersive predictions, suggesting that the primary source of uncertainty in short-term simulations stems from ocean currents rather than the wind field.

Paper I was designed to demonstrate the Barents-2.5 EPS performance in an applicable situation, *i.e.*, oil slick drift. However, as snapshots, the information obtained with delineated slicks does not allow us to investigate how well the model represents the observed dynamics from the release point to the slick's final position. Higher-frequency observations were needed for this and drifter observations were then used in Paper II for this purpose. In the study, the ensemble prediction system was assessed in two areas of interest, namely the Barents Sea and the Fram Strait. The results reported in the paper indicate that Barents-2.5 EPS performs differently in the two regions. The model produced reliable predictions and demonstrated better calibration in the Barents Sea, where wind and tidal forces are the primary ocean surface current drivers. Additionally, the modeled rotary energy spectra closely matched observations, particularly in the low and inertial frequency ranges. However, the model's performance was less satisfactory in the Fram Strait where eddies and other small-scale phenomena play a larger role in influencing drifter trajectories. In Section 4.2.2, the results showed that the model error was the key factor affecting Barents-2.5 EPS' performance rather than its spread.

Objective 2: *Explore the applicability of ensemble prediction systems for oil slick drift modeling.*

An ensemble model, offering a range of possible outcomes, can provide valuable guidance to oil spill responders by not only delivering drift estimates but

also illustrating the spread of potential results, thereby indicating the confidence level of the predictions. Since uncertainties fluctuate across both time and space, forecasts may sometimes show significant divergence, suggesting that recovery resources should be distributed rather than concentrated in one location. Conversely, when simulations exhibit less variability, there is greater confidence that the oil is moving in the predicted direction. These scenarios were frequently observed in our study, highlighting the applicability of the Barents-2.5 EPS in oil slick drift modeling.

The Norwegian Coastal Administration and the Norwegian Clean Seas Association for Operating Companies utilize the OpenDrift particle tracking system (as demonstrated in Paper I and Paper II) alongside met-ocean models from the Norwegian Meteorological Institute as part of their emergency preparedness strategies for acute pollution incidents. We hope that the findings presented in this work will serve as a valuable knowledge base for operations in Norwegian waters and be further refined by both the scientific community and emergency responders.

Objective 3: *Investigate the uncertainty in ocean surface currents derived from Synthetic Aperture Radar imagery using an operational atmospheric ensemble prediction system.*

Papers I and II addressed ocean surface current uncertainties through drift trajectory simulations forced with an ensemble ocean model. Nevertheless, uncertainties are also present in observations and data processing. This was addressed in Paper III with radial velocities retrieved with SAR. In other words, we shifted the SAR data from a passive source of ground truth, a low-uncertainty validation tool, to an active source of observation, where uncertainties are no longer small.

In Paper III, we addressed the impact of wind speed and direction uncertainties on surface radial velocity retrieved using data from Sentinel-1. The uncertain wind fields were obtained from a regional operational atmospheric ensemble model and used with CDOP3SiX GMF to quantify the wind-wave bias component, a key step in the retrieval process. Our analysis revealed that uncertainties in wind speed and direction contributed over 5 Hz to the wind-wave bias correction and more than 0.3 m/s to the radial velocity estimation. Given that these magnitudes are comparable to the sensor's uncertainty after recalibration, it highlights that wind field uncertainties cannot be overlooked. We determined that retrievals under wind conditions below 5–6 m/s tend to be unreliable and should be treated with caution. Although the CDOP3SiX's sensitivity to wind direction relative to the satellite antenna look is known, our study was the first to quantify the maximum wind speed and direction that meet the scientific community-specified thresholds. In Skagerrak, where the prevailing wind di-

rections are west-northwest and east-northeast, ocean radial velocity retrievals generally show uncertainties below the desired threshold when wind speeds exceed the 5–6 m/s range. These results open a new possibility for downstream applications of radial velocity products, including data assimilation.

9.2 Future Work

The final component of this thesis provides a list of potential research topics to be investigated in the future.

- **Ensemble modeling improvement:**
  As observations were available for ensemble performance assessment in Papers I and II, we were able to quantitatively define the positive points and downsides of the Barents-2.5 EPS. Although the ensemble prediction system revealed satisfactory performance in Paper I, underdispersive and less skillful predictions are noticeable in Paper II. As discussed throughout this thesis, it is known that ensemble models, whether atmospheric or ocean, are not expected to provide the same performance everywhere. Improving the ensemble performance must be pursued by balancing error and spread. This can potentially be addressed for the former by assimilating sea level anomaly (SLA) from nadir-looking altimeters. Regarding spread, the impact of incorporating the SLA uncertainties in the ensemble system must be investigated. Other suggested approaches are:

  - *Model parameters perturbation:*
    One of the ensemble generation methods not discussed in this work is perturbed parameterizations. The method is based on perturbing parameters of key variables known beforehand to contribute to the ensemble spread. Although in use by ECMWF, defining these key variables and the perturbation scheme is not trivial.

  - *Multi-model atmospheric forcing:*
    Building on the Barents-2.5 EPS ensemble generation, which uses surface forcing from two atmospheric models, one could also explore how incorporating multiple atmospheric models from different centers influences the single-ensemble performance. As mentioned in Chapter 4, there is ongoing debate about whether an MME ensemble surpasses single-model atmospheric ensemble prediction systems. To date, investigations into how using MME as surface forcing in single-model ocean ensemble prediction systems impacts performance have not been undertaken. An ocean-dedicated comparison between MME and single-model ensemble in drift modeling would

also be of great value.

- **Ocean Radial Velocity**
  While we successfully demonstrated in Paper III how uncertainty in the wind field impacts the estimation of ocean radial velocity retrieved from SAR, we also highlight challenges and offer suggestions for future research.

  - *Wave-current interactions:*
    Surface gravity waves can be modulated by the underlying ocean current. As such, it is expected that wave models considering these interactions will provide different results from those not considering it, thus impacting the wave-wind bias estimation. Investigations on this topic are of great interest to provide a full overview of how the geophysical flow uncertainties impact radial velocity retrieval.

  - *Validation:*
    As mentioned throughout this thesis, observations play an important role in model validation. Although our uncertainty estimations are robust, validating the results with independent data has yet to be performed. Drifters can be employed for validation, but available in-situ data is often sparse. This challenge is not exclusive to this work, stressing the need for joint community efforts to collect new data, pursue new technologies, or conduct dedicated field campaigns.

- **Drift trajectory prediction improvement in the Fram Strait:**
  As mentioned in the previous section, the unsatisfactory ensemble performance of Barents-2.5 EPS in the Fram Strait stems from the model's error rather than spread. Nevertheless, we do not know which component, wind or currents (or both), that is the error source. Equipping drifters with anemometers, providing highly accurate wind velocity observations, can be helpful in determining further improvements in the model.

- **Data-driven ocean current forecast:**
  Both private organizations (e.g., Google, Huawei, and Nvidia) and state institutions (e.g., ECMWF and MET-Norway) have invested in weather forecasting using machine learning techniques. While still in the experimental phase, these methods have demonstrated performance comparable to state-of-the-art dynamic models. A notable advantage of data-driven simulations is their significantly reduced computation time once the models are trained. Expanding these concepts to ocean current forecasting represents an exciting avenue for future research. Notably, the ECMWF has already announced plans for data-driven ensemble modeling, paving the way for innovative applications in ocean forecasting as

well.

- **Further improvements in oil slick drift modeling:**
  Our main goal in this thesis was to investigate uncertainties in ocean surface currents, but the drift of oil slicks depends on other factors that were not considered here. It is desirable that further investigations also consider different oil types, vertical mixing, and weathering processes. Together with the ensemble ocean and wind fields, one can obtain a full overview of the uncertainties present in oil slick drift prediction and, more importantly, those that actually contribute to the spread of possible outcomes.

- **Decision making under uncertainty:**
  A final suggestion for future work is on the visualization of uncertainties. Our work has focused mainly on the validation of ensemble models, but we are aware that this information also needs to be understood by those using the products. Displaying the uncertainty estimates in trajectory modeling, especially when decisions are at stake, must be informative and clear. This step, ideally, should be developed in cooperation with responders. Confidence ellipses were used in Paper I to communicate uncertainty, but other types of illustrations may be preferable for longer-period simulations. Further research could focus, for instance, on employing contours of probability bounds. Addressing uncertainty communication is one of the frontiers in operational ocean ensemble modeling.

Despite the need for further research into the different topics covered in this work, we conclude that ocean ensemble modeling and the correct representation of surface current uncertainties are of great usefulness and applicable to different branches of operational oceanography. This ultimately reinforces the relevance of advancing our expertise, developing independent ensemble generation methods, and moving towards a new era of ocean prediction systems.

Bibliography

1 United Nations Conference on Trade and Development (UNCTAD). (2021). *Advancing the potential of sustainable ocean-based economies: Trade trends, market drivers, and market access - A first assessment*. United Nations. Geneva, Switzerland. https://unctad.org/system/files/official-document/ditctedinf2021d2_en.pdf

2 Sokkeldirektoratet. (2022). *Ressursrapport 2022* (tech. rep.). Sokkeldirektoratet. https://www.sodir.no/globalassets/1-sodir/publikasjoner/rapporter/ressursrapporter/2022/no/ressursrapport-2022.pdf

3 Robertsen, R., Iversen, A., Nyrud, T., Erraia, J., & Blomgren, A. (2022). *Ringvirknings og verdiskapingsanalyser i norsksjømatnæring 2020–2022* (tech. rep.). Nofima. https://hdl.handle.net/11250/3046520

4 Palmer, T. (2019). The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 12–24. https://doi.org/https://doi.org/10.1002/qj.3383

5 Kalnay, E. (2002). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

6 Jung, T., Klinker, E., & Uppala, S. (2005). Reanalysis and reforecast of three major european storms of the twentieth century using the ECMWF forecasting system. part ii: Ensemble forecasts. *Meteorological Applications*, *12*(2), 111–122. https://doi.org/https://doi.org/10.1017/S1350482705001623

7 Evensen, G., & van Leeuwen, P. (1996). Assimilation of geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, *124*(1), 85–96. https://doi.org/10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO;2

8 Spaulding, M. (2017). State of the art review and future directions in oil spill modeling. *Marine Pollution Bulletin*, *115*(1), 7–19. https://doi.org/https://doi.org/10.1016/j.marpolbul.2017.01.001

9 Barke, C., Kourafalou, V., Beegle-Krause, C., Boufadel, M., Bourassa, M., Buschang, S., Androulidakis, Y., Chassignet, E., Dagestad, K.-F., Danmeier, D., et al. (2020). Progress in operational modeling in support of oil spill response. *Journal of Marine Science and Engineering*, *8*(9). https://doi.org/https://doi.org/10.3390/jmse8090668

10 Keramea, P., Spanoudaki, K., Zodiatis, G., Gikas, G., & Sylaios, G. (2021). Oil spill modeling: A critical review on current trends, perspectives, and challenges. *Journal of Marine Science and Engineering*, *9*(2). https://doi.org/10.3390/jmse9020181

11 Röhrs, J., Sutherland, G., Jeans, G., Bedington, M., Sperrevik, A., Dagestad, K.-F., Gusdal, Y., Mauritzen, C., Dale, A., & LaCasce, J. (2023). Surface currents in operational oceanography: Key applications, mechanisms, and methods. *Journal of Operational Oceanography*, *16*(1), 60–88. https://doi.org/10.1080/1755876X.2021.1903221

12 Helland-Hansen, B., & Nansen, F. (1909). *The Norwegian Sea; its Physical Oceanography based upon the Norwegian researches 1900-1904* (tech. rep. No. 2). Report on Norwegian Fishery and Marine Investigations.

13 Balwada, D., Xie, J.-H., Marino, R., & Feraco, F. (2022). Direct observational evidence of an oceanic dual kinetic energy cascade and its seasonality. *Science Advances*, *8*(41), eabq2566. https://doi.org/10.1126/sciadv.abq2566

14 Chelton, D., Gaube, P., Schlax, M., et al. (2011). The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science*, *334*(6054), 328–332. https://doi.org/10.1126/science.1208897

15 Stommel, H. (1948). The westward intensification of wind-driven ocean currents. *Eos, Transactions American Geophysical Union*, *29*(2), 202–206. https://doi.org/https://doi.org/10.1029/TR029i002p00202

16 Munk, W. (1950). On the wind-driven ocean circulation. *Journal of Atmospheric Sciences*, *7*(2), 80–93. https://doi.org/10.1175/1520-0469(1950)007<0080:OTWDOC>2.0.CO;2

17 Ekman, V. (1905). On the influence of the Earth's rotations on ocean-currents. *Quarterly Journal of the Royal Meteorological Society*, *2*(11), 1–53.

18 Rio, M.-H., & Hernandez, F. (2003). High-frequency response of wind-driven currents measured by drifting buoys and altimetry over the world ocean. *Journal of Geophysical Research: Oceans*, *108*(C8). https://doi.org/https://doi.org/10.1029/2002JC001655

19 Cushman-Roisin, B., & Beckers, J.-M. (2011a). Chapter 8 - the Ekman layer. In B. Cushman-Roisin & J. Beckers (Eds.), *Introduction to geophysical fluid dynamics* (pp. 239–270, Vol. 101). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-088759-0.00008-0

20 Scott, R., & Wang, F. (2005). Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *Journal of Physical Oceanography*, *35*(9), 1650–1666. https://doi.org/10.1175/JPO2771.1

21 Ajayi, A., Le Sommer, J., Chassignet, E., Molines, J.-M., Xu, X., Albert, A., & Dewar, W. (2021). Diagnosing cross-scale kinetic energy exchanges from two submesoscale permitting ocean models [e2019MS001923 2019MS001923]. *Journal of Advances in Modeling Earth Systems*, *13*(6), e2019MS001923. https://doi.org/https://doi.org/10.1029/2019MS001923

22 Korotenko, K. (2018). Effects of mesoscale eddies on behavior of an oil spill resulting from an accidental deepwater blowout in the Black Sea: An assessment of the environmental impacts. *PeerJ*, *6*. https://doi.org/10.7717/peerj.5448

23 Walker, N., Pilley, C., Raghunathan, V., & other. (2011). Impacts of Loop Current frontal cyclonic eddies and wind forcing on the 2010 Gulf of Mexico Oil Spill. *Washington DC American Geophysical Union Geophysical Monograph Series*, *195*, 103–116. https://doi.org/10.1029/2011GM001120

24 Kourafalou, V., & Androulidakis, Y. (2013). Influence of Mississippi river induced circulation on the Deepwater Horizon oil spill transport. *Journal of Geophysical Research: Oceans*, *118*(8), 3823–3842. https://doi.org/https://doi.org/10.1002/jgrc.20272

25 Androulidakis, Y., Kourafalou, V., Le Hénaff, M., et al. (2021). The role of mesoscale dynamics over Northwestern Cuba in the Loop Current evolution in 2010, during the Deepwater Horizon incident. *Journal of Marine Science and Engineering*, *9*(2). https://doi.org/10.3390/jmse9020188

26 Haza, A., Özgökmen, T., & Hogan, P. (2016). Impact of submesoscales on surface material distribution in a gulf of Mexico mesoscale eddy. *Ocean Modelling*, *107*, 28–47. https://doi.org/https://doi.org/10.1016/j.ocemod.2016.10.002

27 McWilliams, J. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *472*(2189), 20160117. https://doi.org/10.1098/rspa.2016.0117

28 McWilliams, J. C. (2019). A survey of submesoscale currents. *Geoscience Letters*, *6*(1), 3. https://doi.org/10.1186/s40562-019-0133-3

29 Cushman-Roisin, B., & Beckers, J.-M. (2011b). Chapter 5 - diffusive processes. In B. Cushman-Roisin & J. Beckers (Eds.), *Introduction to geophysical fluid dynamics* (pp. 131–161, Vol. 101). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-088759-0.00005-5

30 Thorpe, S. A. (2007). Turbulence, heat and waves. In *An introduction to ocean turbulence* (pp. 1–36). Cambridge University Press.

31 Jones, C., Dagestad, K.-F., Breivik, Ø., Holt, B., Röhrs, J., Christensen, K., Espeseth, M., Brekke, C., & Skrunes, S. (2016). Measurement and modeling of oil slick transport. *Journal of Geophysical Research: Oceans*, *121*(10), 7759–7775. https://doi.org/https://doi.org/10.1002/2016JC012113

32 Röhrs, J., Dagestad, K.-F., Asbjørnsen, H., Nordam, T., Skancke, J., Jones, C. E., & Brekke, C. (2018). The effect of vertical mixing on the horizontal drift of oil spills. *Ocean Science*, *14*(6), 1581–1601. https://doi.org/10.5194/os-14-1581-2018

33 Davidson, P. (2015, June). *Turbulence: An Introduction for Scientists and Engineers*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198722588.001.0001

34 Koszalka, I., LaCasce, J., Andersson, M., et al. (2011). Surface circulation in the nordic seas from clustered drifters. *Deep Sea Research Part I:*

Oceanographic Research Papers, *58*(4), 468–485. https://doi.org/https://doi.org/10.1016/j.dsr.2011.01.007

35 Rühs, S., Zhurbas, V., Koszalka, I., et al. (2018). Eddy diffusivity estimates from Lagrangian trajectories simulated with ocean models and surface drifter data — a case study for the Greater Agulhas System. *Journal of Physical Oceanography*, *48*(1), 175–196. https://doi.org/10.1175/JPO-D-17-0048.1

36 Christensen, K., Sperrevik, A., & Broström, G. (2018). On the variability in the onset of the Norwegian Coastal Current. *Journal of Physical Oceanography*, *48*(3), 723–738. https://doi.org/10.1175/JPO-D-17-0117.1

37 Melsom, A. (2005). Mesoscale activity in the North Sea as seen in ensemble simulations. *Ocean Dynamics*, *55*, 338–350. https://doi.org/https://doi.org/10.1007/s10236-005-0016-3

38 Albretsen, J., & Røed, L. (2010). Decadal long simulations of mesoscale structures in the northern North Sea/Skagerrak using two ocean models. *Ocean Dynamics*, *60*(4), 933–955. https://doi.org/10.1007/s10236-010-0296-0

39 Trodahl, M., Isachsen, P., Lilly, J., et al. (2020). The regeneration of the Lofoten vortex through vertical alignment. *Journal of Physical Oceanography*, *50*(9), 2689–2711. https://doi.org/10.1175/JPO-D-20-0029.1

40 Årthun, M., Eldevik, T., Smedsrud, T., Skagseth, Ø., et al. (2012). Quantifying the influence of Atlantic heat on Barents Sea ice variability and retreat. *Journal of Climate*, *25*(13), 4736–4743. https://doi.org/10.1175/JCLI-D-11-00466.1

41 Lundesgaard, Ø., Sundfjord, A., Lind, S., Nilsen, F., & Renner, A. H. H. (2022). Import of atlantic water and sea ice controls the ocean environment in the northern Barents Sea. *Ocean Science*, *18*(5), 1389–1418. https://doi.org/10.5194/os-18-1389-2022

42 Rieke, O., Årthun, M., & Dörr, J. S. (2023). Rapid sea ice changes in the future Barents Sea. *The Cryosphere*, *17*(4), 1445–1456. https://doi.org/10.5194/tc-17-1445-2023

43 von Appen, W.-J., Baumann, T., Janout, M., et al. (2022). Eddies and the distribution of eddy kinetic energy in the Arctic Ocean. *Oceanography*, *35*. https://doi.org/10.5670/oceanog.2022.122

44 Wekerle, C., Hattermann, T., Wang, Q., Crews, L., von Appen, W.-J., & Danilov, S. (2020). Properties and dynamics of mesoscale eddies in Fram Strait from a comparison between two high-resolution ocean–sea ice models. *Ocean Science*, *16*(5), 1225–1246. https://doi.org/10.5194/os-16-1225-2020

45 Wang, Q., Koldunov, N., Danilov, S., et al. (2020). Eddy kinetic energy in the Arctic Ocean from a global simulation with a 1-km Arctic [e2020GL088550 10.1029/2020GL088550]. *Geophysical Research Letters*, *47*(14), e2020GL088550. https://doi.org/https://doi.org/10.1029/2020GL088550

46 Abbe, C. (1901). The physical basis of long-range weather forecasts. *Monthly Weather Review*, *29*(12), 551–561. https://doi.org/10.1175/1520-0493(1901)29[551c:TPBOLW]2.0.CO;2

47 Bjerknes, V. (2009). Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik (the problem of weather prediction, considered from the viewpoints of mechanics and physics) [(translated and edited by Volken E. and S. Brönnimann)]. *Meteorologische Zeitschrift - METEOROL Z*, *18*, 663–667. https://doi.org/10.1127/0941-2948/2009/416

48 Richardson, L. (2007). *Weather prediction by numerical process* (2nd ed.). Cambridge University Press.

49 Charney, J., Fjörtoft, R., & Von Neumann, J. (1950). Numerical integration of the barotropic vorticity equation. *Tellus*, *2*(4), 237–254. https://doi.org/https://doi.org/10.1111/j.2153-3490.1950.tb00336.x

50 Woodring, J., Petersen, M., Schmeißer, A., Patchett, J., Ahrens, J., & Hagen, H. (2016). In situ eddy analysis in a high-resolution ocean climate model. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 857–866. https://doi.org/10.1109/TVCG.2015.2467411

51 Soufflet, Y., Marchesiello, P., Lemarié, F., Jouanno, X., J. andCapet, Debreu, L., & Benshila, R. (2016). On effective resolution in ocean models. *Ocean Modelling*, *98*, 36–50. https://doi.org/https://doi.org/10.1016/j.ocemod.2015.12.004

52 Stull, R. (1988). Turbulence closure techniques. In R. Stull (Ed.), *An introduction to boundary layer meteorology* (pp. 197–250). Springer Netherlands. https://doi.org/10.1007/978-94-009-3027-8_6

53 Lévy, M., Resplandy, L., Klein, P., et al. (2012). Grid degradation of submesoscale resolving ocean models: Benefits for offline passive tracer transport. *Ocean Modelling*, *48*, 1–9. https://doi.org/https://doi.org/10.1016/j.ocemod.2012.02.004

54 Matsuzaki, Y., & Fujita, I. (2017). In situ estimates of horizontal turbulent diffusivity at the sea surface for oil transport simulation. *Marine Pollution Bulletin*, *117*(1), 34–40. https://doi.org/https://doi.org/10.1016/j.marpolbul.2016.10.026

55 Nummelin, A., Busecke, J., Haine, T., et al. (2021). Diagnosing the scale- and space-dependent horizontal eddy diffusivity at the global surface ocean. *Journal of Physical Oceanography*, *51*(2), 279–297. https://doi.org/10.1175/JPO-D-19-0256.1

56 Shchepetkin, A., & McWilliams, J. (2005). The regional oceanic modeling system (roms): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, *9*(4), 347–404. https://doi.org/https://doi.org/10.1016/j.ocemod.2004.08.002

57 Umlauf, L., & Burchard, H. (2005). Second-order turbulence closure models for geophysical boundary layers. A review of recent work. *Continental Shelf Research*, *25*(7–8), 795–827. https://doi.org/10.1016/j.csr.2004.08.004

58 Kundu, P., Cohen, I., & Dowling, D. (2016). Chapter 3 - kinematics. In P. Kundu, I. Cohen, & D. Dowling (Eds.), *Fluid mechanics (sixth edition)* (Sixth Edition, pp. 77–108). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-405935-1.00003-4

59 Hughes, P. (1956). A determination of the relation between wind and sea-surface drift. *Quarterly Journal of the Royal Meteorological Society*, *82*(354), 494–502. https://doi.org/https://doi.org/10.1002/qj.49708235412

60 Till, J. W., Eisenman, I., Ceroli, A., & Constantinou, N. (2022). How winds and ocean currents influence the drift of floating objects. *Journal of Physical Oceanography*, *52*(5), 907–916. https://doi.org/10.1175/JPO-D-20-0275.1

61 Brekke, C., Espeseth, M., Dagestad, K.-F., Röhrs, J., Hole, L., & Reigber, A. (2021). Integrated analysis of multisensor datasets and oil drift simulations — a free-floating oil experiment in the open ocean [e2020JC016499 2020JC016499]. *Journal of Geophysical Research: Oceans*, *126*(1). https://doi.org/https://doi.org/10.1029/2020JC016499

62 van Sebille, E., Griffies, S., Abernathey, R., et al. (2018). Lagrangian ocean analysis: Fundamentals and practices. *Ocean Modelling*, *121*, 49–75. https://doi.org/https://doi.org/10.1016/j.ocemod.2017.11.008

63 Boyce, W., & DiPrima, R. (2017). *Elementary differential equations and boundary value problems* (10th). Wiley.

64 Nordam, T., & Duran, R. (2020). Numerical integrators for Lagrangian oceanography. *Geoscientific Model Development*, *13*(12), 5935–5957. https://doi.org/10.5194/gmd-13-5935-2020

65 Dagestad, K.-F., Röhrs, J., Breivik, Ø., & Ådlandsvik, B. (2018). Opendrift v1.0: A generic framework for trajectory modelling. *Geoscientific Model Development*, *11*(4), 1405–1420. https://doi.org/10.5194/gmd-11-1405-2018

66 NOAA Office of Response and Restoration (ORR). (2023). Adios oil database [GitHub repository]. https://github.com/NOAA-ORR-ERD/adios_oil_database

67 Lorenz, E. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, *20*(2), 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

68 Epstein, E. (1969a). The role of initial uncertainties in predicion. *Journal of Applied Meteorology and Climatology*, *8*(2), 190–198. https://doi.org/10.1175/1520-0450(1969)008<0190:TROIUI>2.0.CO;2

69 Stammer, D. (2014). Ocean modeling and data assimilation. In E. Njoku (Ed.), *Encyclopedia of remote sensing* (pp. 446–454). Springer New York. https://doi.org/10.1007/978-0-387-36699-9_119

70 Liu, Y., & Weisberg, R. (2011). Evaluation of trajectory modeling in different dynamic regions using normalized cumulative lagrangian separation. *Journal of Geophysical Research: Oceans*, *116*(C9). https://doi.org/https://doi.org/10.1029/2010JC006837

71 De Dominicis, M., Pinardi, N., Zodiatis, G., & Archetti, R. (2013). MEDSLIK-II, a Lagrangian marine surface oil spill model for short-term forecasting – part 2: Numerical simulations and validations. *Geoscientific Model Development*, *6*(6), 1871–1888. https://doi.org/10.5194/gmd-6-1871-2013

72 Dagestad, K.-F., & Röhrs, J. (2019). Prediction of ocean surface trajectories using satellite derived vs. modeled ocean currents. *Remote Sensing of Environment*, *223*, 130–142. https://doi.org/https://doi.org/10.1016/j.rse.2019.01.001

73 Garcia-Pineda, O., Androulidakis, Y., Le Hénaff, M., Kourafalou, V., Hole, L., Kang, H., Staples, G., Ramirez, E., & DiPinto, L. (2020). Measuring oil residence time with gps-drifters, satellites, and Unmanned Aerial Systems (UAS). *Marine Pollution Bulletin*, *150*, 110644. https://doi.org/https://doi.org/10.1016/j.marpolbul.2019.110644

74 Androulidakis, Y., Kourafalou, V., Özgökmen, T., Garcia-Pineda, O., Lund, B., Le Hénaff, M., Hu, C., Haus, B. K., Novelli, G., Guigand, C., Kang, H., Hole, L., & Horstmann, J. (2018). Influence of river-induced fronts on hydrocarbon transport: A multiplatform observational study. *Journal of Geophysical Research: Oceans*, *123*(5), 3259–3285. https://doi.org/https://doi.org/10.1029/2017JC013514

75 Lumpkin, R., Özgökmen, T., & Centurioni, L. (2017). Advances in the application of surface drifters. *Annual Review of Marine Science*, *9*(Volume 9, 2017), 59–81. https://doi.org/https://doi.org/10.1146/annurev-marine-010816-060641

76 Robinson, I. (2014). Ocean, measurements and applications. In E. Njoku (Ed.), *Encyclopedia of remote sensing* (pp. 469–480). Springer New York. https://doi.org/10.1007/978-0-387-36699-9_121

77 Isern-Fontanet, J., Ballabrera-Poy, J., Turiel, A., & García-Ladona, E. (2017). Remote sensing of ocean surface currents: A review of what is being observed and what is being assimilated. *Nonlinear Processes in Geophysics*, *24*(4), 613–643. https://doi.org/10.5194/npg-24-613-2017

78 Villas Bôas, A., Ardhuin, F., Ayet, A., Bourassa, M., Brandt, P., Chapron, B., Cornuelle, B. D., Farrar, J., Fewings, M., Fox-Kemper, B., Gille, S., Gommenginger, C., Heimbach, P., Hell, M., Li, Q., Mazloff, M., Merrifield, S., Mouche, A., Rio, M., . . . van Sebille, E. (2019). Integrated observations of global surface winds, currents, and waves: Requirements and challenges for the next decade. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00425

79 Hauser, D., Abdalla, S., Ardhuin, F., Bidlot, J.-R., Bourassa, M., Cotton, D., Gommenginger, C., Evers-King, H., Johnsen, H., Knaff, J., Lavender, S., Mouche, A., Reul, N., Sampson, C., Steele, E., & Stoffelen, A. (2023). Satellite remote sensing of surface winds, waves, and currents: Where are we now? *Surveys in Geophysics*, *44*(5), 1357–1446. https://doi.org/10.1007/s10712-023-09771-2

80 Chapron, B., Johannessen, J., & Collard, F. (2014). Ocean surface velocity. In E. Njoku (Ed.), *Encyclopedia of remote sensing* (pp. 461–469). Springer New York. https://doi.org/10.1007/978-0-387-36699-9_164

81 Amores, A., Jordà, G., Arsouze, T., & Le Sommer, J. (2018). Up to what extent can we characterize ocean eddies using present-day gridded altimetric products? *Journal of Geophysical Research: Oceans*, *123*(10), 7220–7236. https://doi.org/https://doi.org/10.1029/2018JC014140

82 Amores, A., Jordà, G., & Monserrat, S. (2019). Ocean eddies in the Mediterranean Sea from satellite altimetry: Sensitivity to satellite track location. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00703

83 Poulain, P.-M., Menna, M., & Mauri, M. (2012). Surface geostrophic circulation of the Mediterranean Sea derived from drifter and satellite altimeter data. *Journal of Physical Oceanography*, *42*(6), 973–990. https://doi.org/10.1175/JPO-D-11-0159.1

84 Coquereau, A., & Foukal, N. P. (2023). Evaluating altimetry-derived surface currents on the south Greenland shelf with surface drifters. *Ocean Science*, *19*(5), 1393–1411. https://doi.org/10.5194/os-19-1393-2023

85 European Space Agency. (n.d.[a]). Satellite frequency bands [Last accessed: 2024-10-13]. https://www.esa.int/Applications/Connectivity_and_Secure_Communications/Satellite_frequency_bands

86 Emery, W., & Camps, A. (2017). Chapter 5 - radar. In W. Emery & A. Camps (Eds.), *Introduction to satellite remote sensing* (pp. 291–453). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-809254-5.00005-1

87 Lohse, J. (2021). *On Automated Classification of Sea Ice Types in SAR Imagery* [Doctoral dissertation, UiT The Arctic University of Norway]. https://munin.uit.no/handle/10037/20606

88 European Space Agency. (n.d.[b]). https://sentiwiki.copernicus.eu/web/s1-mission#S1-Mission-Wave

89 Farr, T. (1993). Radar interactions with geologic surfaces. In *Guide to Magellan image interpretation* (pp. 45–56, Vol. 93). National Aeronautics and Space Administration, Jet Propulsion Laboratory.

90 Valenzuela, G. (1978). Theories for the interaction of electromagnetic and oceanic waves — A review. *Boundary-Layer Meteorology*, *13*(1), 61–85. https://doi.org/10.1007/BF00913863

91 Wright, J. (1968). A new model for sea clutter. *IEEE Transactions on Antennas and Propagation*, *16*(2), 217–223. https://doi.org/10.1109/TAP.1968.1139147

92 Li, F.-k., Held, D. N., Curlander, J. C., & Wu, C. (1985). Doppler parameter estimation for spaceborne Synthetic-Aperture Radars. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-23*(1), 47–56. https://doi.org/10.1109/TGRS.1985.289499

93 Madsen, S. (1989). Estimating the doppler centroid of SAR data. *IEEE Transactions on Aerospace and Electronic Systems*, *25*(2), 134–140. https://doi.org/10.1109/7.18675

94 European Space Agency (ESA). (2013). *Sentinel-1 Level-1 Detailed Algorithm Definition* [Last accessed: 2024-11-14]. European Space Agency. https://sentinel.esa.int/documents/247904/1877131/Sentinel-1-Level-1-Detailed-Algorithm-Definition

95 Miranda, N., Rosich, B., Santella, C., & Grion, M. (2005, April). Review of the impact of ERS-2 piloting modes on the SAR doppler stability.

96 Chapron, B., Collard, F., & Ardhuin, F. (2005). Direct measurements of ocean surface velocity from space: Interpretation and validation. *Journal of Geophysical Research: Oceans*, *110*(C7). https://doi.org/https://doi.org/10.1029/2004JC002809

97 Shuchman, R., Lyzenga, D., & Klooster, A. (1981). *Exploitation of SAR data for measurement of ocean currents and wave velocities* (tech. rep.) (Prepared for the National Earth Satellite Service, NOAA, Washington, DC. Contract No. MO-A01-78-00-4322). Environmental Research Institute of Michigan. Ann Arbor, MI. https://ntrs.nasa.gov/api/citations/19830005504/downloads/19830005504.pdf

98 Moiseev, A., Johnsen, H., Johannessen, J. A., Collard, F., & Guitton, G. (2020). On removal of sea state contribution to Sentinel-1 Doppler shift for retrieving reliable ocean surface current [e2020JC016288 2020JC016288]. *Journal of Geophysical Research: Oceans*, *125*(9), e2020JC016288. https://doi.org/https://doi.org/10.1029/2020JC016288

99 Elyouncha, A., Eriksson, L., Romeiser, R., & Ulander, L. (2019). Measurements of sea surface currents in the Baltic Sea region using spaceborne along-track InSAR. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 8584–8599. https://doi.org/10.1109/TGRS.2019.2921705

100 Moiseev, A., Johannessen, J. A., & Johnsen, H. (2022). Towards retrieving reliable ocean surface currents in the coastal zone from the Sentinel-1 Doppler shift observations [e2021JC018201 2021JC018201]. *Journal of Geophysical Research: Oceans*, *127*(5), e2021JC018201. https://doi.org/https://doi.org/10.1029/2021JC018201

101 Martin, A., Gommenginger, C. P., Jacob, B., & Staneva, J. (2022). First multi-year assessment of Sentinel-1 radial velocity products using HF radar currents in a coastal environment. *Remote Sensing of Environment*, *268*. https://doi.org/https://doi.org/10.1016/j.rse.2021.112758

102 Fan, S., Zhang, B., Moiseev, A., Kudryavtsev, V., Johannessen, J., & Chapron, B. (2023). On the use of dual co-polarized radar data to derive a sea surface doppler model—part 2: Simulation and validation. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–9. https://doi.org/10.1109/TGRS.2023.3246771

103 Carpenter, A. (2019). Oil pollution in the north sea: The impact of governance measures on oil pollution over several decades. *Hydrobiologia*, *845*, 109–127. https://doi.org/https://doi.org/10.1007/s10750-018-3559-2

104 OSPAR Commission. (2021). Offshore Installations — Webpage. OSPAR Commission [Last accessed: 2024-10-10]. https://oap.ospar.org/en/ospar-

assessments/quality-status-reports/qsr-2023/thematic-assessments/offshore-industry/activities/

105 OSPAR Commission. (2017). OSPAR report on discharges, spills and emissions from offshore oil and gas installations in 2017 [Last accessed: 2024-10-10]. https://oap.ospar.org/en/ospar-assessments/committee-assessments/offshore-industry/discharges-spills-emissions/annual-reports/discharges-spills-and-emissions-offshore-oil-and-gas-2017/

106 Transportation Research Board and National Research Council. (2003). *Oil in the sea iii: Inputs, fates, and effects*. The National Academies Press. https://doi.org/10.17226/10388

107 Fingas, M. (2012). Chapter 3 - Oil spills: Why do they happen and how often? In *The basics of oil spill cleanup* (pp. 33–42). Taylor & Francis Group.

108 Alpers, W., Holt, B., & Zeng, K. (2017). Oil spill detection by imaging radars: Challenges and pitfalls. *Remote Sensing of Environment*, *201*, 133–147. https://doi.org/https://doi.org/10.1016/j.rse.2017.09.002

109 Espeseth, M., Brekke, C., Jones, C., Holt, B., & Freeman, A. (2020). The impact of system noise in polarimetric SAR imagery on oil spill observations. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(6), 4194–4214. https://doi.org/10.1109/TGRS.2019.2961684

110 Jones, C., & Holt, B. (2018). Experimental l-band airborne sar for oil spill response at sea and in coastal waters. *Sensors*, *18*(2). https://doi.org/10.3390/s18020641

111 Jones, C. (2023). An automated algorithm for calculating the ocean contrast in support of oil spill response. *Marine Pollution Bulletin*, *191*, 114952. https://doi.org/https://doi.org/10.1016/j.marpolbul.2023.114952

112 Quigley, C., Johansson, M., & Jones, C. (2023). An investigation on the damping ratio of marine oil slicks in synthetic aperture radar imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 5488–5501. https://doi.org/10.1109/JSTARS.2023.3285145

113 Skrunes, S., Brekke, C., & Eltoft, T. (2014). Characterization of marine surface slicks by Radarsat-2 multipolarization features. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(9), 5302–5319. https://doi.org/10.1109/TGRS.2013.2287916

114 European Maritime Safety Agency. (n.d.). CleanSeaNet - Detections and Feedback data (2015-2023) [Last accessed: 2024-10-13]. https://www.emsa.europa.eu/csn-menu/csn-service/items.html?cid=122&id=4645

115 Daniel, P., Josse, P., Dandin, P., Gouriou, V., Marchand, M., & Tiercelin, C. (2001). Forecasting the Erika Oil Spills. *International Oil Spill Conference Proceedings*, *2001*(1), 649–655. https://doi.org/10.7901/2169-3358-2001-1-649

116 Xu, W., Li, X., Wei, Y., Tang, Z., Cheng, Y., & Pichel, W. (2013). Satellite observations and modeling of oil spill trajectories in the Bohai Sea. *Marine Pollution Bulletin*, *71*(1), 107–116. https://doi.org/https://doi.org/10.1016/j.marpolbul.2013.03.028

117 Zodiatis, G., Coppini, G., Perivoliotis, L., Lardner, R., Alves, T., Pinardi, N., Liubartseva, S., De Dominicis, M., Bourma, E., & Sepp Neves, A. (2018). Numerical modeling of oil pollution in the Eastern Mediterranean Sea. In A. Carpenter & A. Kostianoy (Eds.), *Oil pollution in the Mediterranean Sea: Part I: The International Context* (pp. 215–254). Springer International Publishing. https://doi.org/10.1007/698_2017_131

118 Skrunes, S., Johansson, A., & Brekke, C. (2019). Synthetic aperture radar remote sensing of operational platform produced water releases. *Remote Sensing*, *11*(23). https://doi.org/10.3390/rs11232882

119 Johansson, A., Skrunes, S., Brekke, C., & Isaksen, H. (2021). Multi-mission remote sensing of low concentration produced water slicks. *EUSAR 2021; 13th European Conference on Synthetic Aperture Radar*, 1–6.

120 Johannessen, J. A., Kudryavtsev, V., Akimov, D., Eldevik, T., Winther, N., & Chapron, B. (2005). On radar imaging of current features: 2. mesoscale eddy and current front detection. *Journal of Geophysical Research: Oceans*, *110*(C7). https://doi.org/https://doi.org/10.1029/2004JC002802

121 Dagestad, K.-F., Johannessen, J., Kerbaol, V., Collard, F., Kudryavtsev, V., Akimov, D., Chapron, B., Wang, H.-J., Wang, Z., & He, M.-X. (2006). Marine monitoring of the South- and East China seas based on Envisat ASAR. In E. P. D. C. Estec (Ed.), *Proceedings of dragon programme mid-term results, santorini, greece 27 june-] july 2005 (ESA sp-611, january 2006). h. lacoste (ed.) isbn no: 92-9092-922-7. pp.27-36?* (pp. 27–+, Vol. 611). https://archimer.ifremer.fr/doc/00706/81828/

122 Mouche, A., Collard, F., Chapron, B., Dagestad, K.-F., Guitton, G., Johannessen, J., Kerbaol, V., & Hansen, M. (2012). On the use of Doppler shift for sea surface wind retrieval from SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(7), 2901–2909. https://doi.org/10.1109/TGRS.2011.2174998

123 Tollinger, M., Graversen, R., & Johnsen, H. (2021). High-resolution polar low winds obtained from unsupervised SAR wind retrieval. *Remote Sensing*, *13*(22). https://doi.org/10.3390/rs13224655

124 Hansen, M. W., Collard, F., Dagestad, K.-F., Johannessen, J. A., Fabry, P., & Chapron, B. (2011). Retrieval of sea surface range velocities from envisat asar doppler centroid measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(10), 3582–3592. https://doi.org/10.1109/TGRS.2011.2153864

125 Bojinski, S., Verstraete, M., Peterson, T., et al. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, *95*(9), 1431–1443. https://doi.org/10.1175/BAMS-D-13-00047.1

126 Romeiser, R., Suchandt, S., Runge, H., Steinbrecher, U., & Grunler, S. (2010). First analysis of terrasar-x along-track insar-derived current fields. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(2), 820–829. https://doi.org/10.1109/TGRS.2009.2030885

127 Johannessen, J. A., Chapron, B., Collard, F., Kudryavtsev, V., Mouche, A., Akimov, D., & Dagestad, K.-F. (2008). Direct ocean surface velocity measurements from space: Improved quantitative interpretation of Envisat ASAR observations. *Geophysical Research Letters*, *35*(22). https://doi.org/https://doi.org/10.1029/2008GL035709

128 Collard, F., Mouche, A., Chapron, B., Danilo, C., & Johannessen, J. (2008). Routine high resolution observation of selected major surface currents from space. *Proceedings of SEASAR, Frascati*.

129 Martin, A., Gommenginger, C., Marquez, J., Doody, S., Navarro, V., & Buck, C. (2016). Wind-wave-induced velocity in ATI SAR ocean surface currents: First experimental evidence from an airborne campaign. *Journal of Geophysical Research: Oceans*, *121*(3), 1640–1653. https://doi.org/https://doi.org/10.1002/2015JC011459

130 Elyouncha, A., Eriksson, L., Broström, G., Axell, L., & Ulander, L. (2021). Joint retrieval of ocean surface wind and current vectors from satellite SAR data using a Bayesian inversion method. *Remote Sensing of Environment*, *260*, 112455. https://doi.org/https://doi.org/10.1016/j.rse.2021.112455

131 Frogner, I.-L., Andrae, U., Bojarova, J., Callado, A., Escribà, P., Feddersen, H., Hally, A., Kauhanen, J., Randriamampianina, R., Singleton, A., Smet, G., van der Veen, S., & Vignes, O. (2019). HarmonEPS — the HARMONIE ensemble prediction system. *Weather and Forecasting*, *34*(6), 1909–1937. https://doi.org/10.1175/WAF-D-19-0030.1

132 Johnsen, H. (2016). *S1+CS feasibility study: Assessment of wind/wave/current retrieval methodology* (tech. rep.) (Report No. 20/2016). NORUT. https://norceresearch.brage.unit.no/norceresearch-xmlui/handle/11250/2658242?show=full

133 World Meteorological Organization. (2024). Ocean surface currents vector [Accessed: 2024-10-12].

134 Reeves, R. (2014). Edward Lorenz revisiting the limits of predictability and their implications: An interview from 2007. *Bulletin of the American Meteorological Society*, *95*(5), 681–687. https://doi.org/10.1175/BAMS-D-13-00096.1

135 Saltzman, B. (1962). Finite amplitude free convection as an initial value problem—i. *Journal of Atmospheric Sciences*, *19*(4), 329–341. https://doi.org/10.1175/1520-0469(1962)019<0329:FAFCAA>2.0.CO;2

136 University of Waterloo, Faculty of Mathematics. (2012). Chaos Theory and the Lorenz Equation: History, Analysis, and Application [Last accessed: 2024-10-10]. https://links.uwaterloo.ca/pmath370w14/PMATH370/lorenz%20Latex.pdf

137 Lorenz, E. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, *21*(3), 289–307. https://doi.org/https://doi.org/10.1111/j.2153-3490.1969.tb00444.x

138 Leutbecher, M., & Palmer, T. (2008). Ensemble forecasting [Predicting weather, climate and extreme events]. *Journal of Computational Physics*, *227*(7), 3515–3539. https://doi.org/https://doi.org/10.1016/j.jcp.2007.02.014

139 Epstein, E. (1969b). Stochastic dynamic prediction. *Tellus*, *21*(6), 739–759. https://doi.org/10.3402/tellusa.v21i6.10143

140 Buizza, R. (2019). Introduction to the special issue on "25 years of ensemble forecasting". *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 1–11. https://doi.org/https://doi.org/10.1002/qj.3370

141 Toth, Z., & Kalnay, E. (1997). Ensemble forecasting at NCEP and the Breeding Method. *Monthly Weather Review*, *125*(12), 3297–3319. https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2

142 Hoffman, R., & Kalnay, E. (1983). Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, *35A*(2), 100–118. https://doi.org/https://doi.org/10.1111/j.1600-0870.1983.tb00189.x

143 de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., Schepers, D., et al. (2022). Coupled data assimilation at ecmwf: Current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, *148*(747), 2672–2702. https://doi.org/https://doi.org/10.1002/qj.4330

144 Fortin, V., Abaza, M., Anctil, F., & Turcotte, T. (2014). Why should ensemble spread match the RMSE of the ensemble mean? *Journal of Hydrometeorology*, *15*(4), 1708–1713. https://doi.org/10.1175/JHM-D-14-0008.1

145 Eckel, F., & Mass, C. (2005). Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting. *Weather and Forecasting*, *20*(3), 328–350. https://doi.org/10.1175/WAF843.1

146 Hamill, T. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review*, *129*(3), 550–560. https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2

147 Carton, J. (1987). How predictable are the geostrophic currents in the recirculation zone of the north atlantic? *Journal of Physical Oceanography*, *17*(6), 751–762. https://doi.org/10.1175/1520-0485(1987)017<0751:HPATGC>2.0.CO;2

148 Adamec, D. (1989). Predictability of quasi-geostrophic ocean flow: Sensitivity to varying model vertical resolution. *Journal of Physical Oceanography*, *19*(11), 1753–1764. https://doi.org/10.1175/1520-0485(1989)019<1753:POQGOF>2.0.CO;2

149 Brasseur, P., Blayo, E., & Verron, J. (1996). Predictability experiments in the north atlantic ocean: Outcome of a quasi-geostrophic model with assimilation of topex/poseidon altimeter data. *Journal of Geophysical Research: Oceans*, *101*(C6), 14161–14173. https://doi.org/https://doi.org/10.1029/96JC00665

150 Liu, Y., Weisberg, H., Robert, Hu, C., & Zheng, L. (2011). Trajectory forecast as a rapid response to the Deepwater Horizon oil spill. In *Monitoring and*

modeling the Deepwater Horizon oil spill: A record-breaking enterprise (pp. 153–165). American Geophysical Union (AGU). https://doi.org/https://doi.org/10.1029/2011GM001121

151 Macfadyen, A., Watabayashi, G., Barker, C., & Beegle-Krause, C. J. (2011). Tactical modeling of surface oil transport during the Deepwater Horizon spill response. In *Monitoring and modeling the Deepwater Horizon oil spill: A record-breaking enterprise* (pp. 167–178). American Geophysical Union (AGU). https://doi.org/https://doi.org/10.1029/2011GM001128

152 Zodiatis, G., De Dominicis, M., Perivoliotis, L., Radhakrishnan, H., Georgoudis, E., Sotillo, M., Lardner, R., Krokos, G., Bruciaferri, D., Clementi, E., Guarnieri, A., Ribotti, A., Drago, A., Bourma, E., Padorno, E., Daniel, P., Gonzalez, G., Chazot, C., Gouriou, V., ... Mancini, M. (2016). The Mediterranean Decision Support System for Marine Safety dedicated to oil slicks predictions [Physical, chemical and biological observations and modelling of oil spills in the Mediterranean Sea]. *Deep Sea Research Part II: Topical Studies in Oceanography*, *133*, 4–20. https://doi.org/https://doi.org/10.1016/j.dsr2.2016.07.014

153 De Dominicis, M., Bruciaferri, D., Gerin, R., Pinardi, N., Poulain, P., Garreau, P., Zodiatis, G., Perivoliotis, L., Fazioli, L., Sorgente, R., & Manganiello, C. (2016). A multi-model assessment of the impact of currents, waves and wind in modelling surface drifters and oil spill. *Deep Sea Research Part II: Topical Studies in Oceanography*, *133*, 21–38. https://doi.org/https://doi.org/10.1016/j.dsr2.2016.04.002

154 French-McCay, D., Tajalli-Bakhsh, T., Jayko, K., Spaulding, M., & Li, Z. (2018). Validation of oil spill transport and fate modeling in Arctic ice. *Arctic Science*, *4*(1), 71–97. https://doi.org/10.1139/as-2017-0027

155 Clemen, R., & Murphy, A. (1986). Objective and subjective precipitation probability forecasts: Some methods for improving forecast quality. *Weather and Forecasting*, *1*(3), 213–218. https://doi.org/10.1175/1520-0434(1986)001<0213:OASPPF>2.0.CO;2

156 Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2008). Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, *134*(630), 241–260. https://doi.org/https://doi.org/10.1002/qj.210

157 Coelho, E., Hogan, P., Jacobs, G., Thoppil, P., Huntley, H., et al. (2015). Ocean current estimation using a multi-model Ensemble Kalman Filter during the Grand Lagrangian Deployment experiment (GLAD). *Ocean Modelling*, *87*, 86–106. https://doi.org/https://doi.org/10.1016/j.ocemod.2014.11.001

158 Ziehmann, C. (2000). Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus A: Dynamic Meteorology and Oceanography*, *52*(3), 280–299. https://doi.org/10.3402/tellusa.v52i3.12266

159 Hagedorn, R., Doblas-Reyes, F., & Palmer, T. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting — I. Basic concept. *Tellus A: Dynamic Meteorology and Oceanography*, *57*(3), 219–233. https://doi.org/10.3402/tellusa.v57i3.14657

160 Hagedorn, R., Buizza, R., Hamill, T., Leutbecher, M., & Palmer, T. (2012, January). Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. https://doi.org/10.21957/9nfamplly

161 Titley, H., Bowyer, R., & Cloke, H. (2020). A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 531–545. https://doi.org/https://doi.org/10.1002/qj.3712

162 Krishnamurti, T., Kishtawal, C. M., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C., Gadgil, S., & Surendran, S. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, *285*(5433), 1548–1550. https://doi.org/10.1126/science.285.5433.1548

163 Rixen, M., & Ferreira-Coelho, E. (2007). Operational surface drift prediction using linear and non-linear hyper-ensemble statistics on atmospheric and ocean models [Marine Environmental Monitoring and Prediction]. *Journal of Marine Systems*, *65*(1), 105–121. https://doi.org/https://doi.org/10.1016/j.jmarsys.2004.12.005

164 Rixen, M., Ferreira-Coelho, E., & Signell, R. (2008). Surface drift prediction in the adriatic sea using hyper-ensemble statistics on atmospheric, ocean and wave models: Uncertainties and probability distribution areas [Maritime Rapid Environmental Assessment]. *Journal of Marine Systems*, *69*(1), 86–98. https://doi.org/https://doi.org/10.1016/j.jmarsys.2007.02.015

165 van den Berge, L., Selten, F. M., Wiegerinck, W., & Duane, G. S. (2011). A multi-model ensemble method that combines imperfect models through learning. *Earth System Dynamics*, *2*(1), 161–177. https://doi.org/10.5194/esd-2-161-2011

166 Schevenhoven, F., Keenlyside, N., Counillon, F., Carrassi, A., Chapman, W., et al. (2023). Supermodeling: Improving predictions with an ensemble of interacting models. *Bulletin of the American Meteorological Society*, *104*(9), E1670–E1686. https://doi.org/10.1175/BAMS-D-22-0070.1

167 Khade, V., Kurian, J., Chang, P., Szunyogh, I., Thyng, K., & Montuoro, R. (2017). Oceanic ensemble forecasting in the Gulf of Mexico: An application to the case of the Deep Water Horizon oil spill. *Ocean Modelling*, *113*, 171–184. https://doi.org/https://doi.org/10.1016/j.ocemod.2017.04.004

168 Li, G., Iskandarani, M., Le Hénaff, M., Winokur, J., Le Maître, O., & Knio, O. (2016). Quantifying initial and wind forcing uncertainties in the Gulf of Mexico. *Comput. Geosci.*, *20*, 1133–1153. https://doi.org/https://doi.org/10.1007/s10596-016-9581-4

169 Counillon, F., & Bertino, L. (2009). High-resolution ensemble forecasting for the Gulf of Mexico eddies and fronts. *Ocean Dynamics*, *59*, 83–95. https://doi.org/https://doi.org/10.1007/s10236-008-0167-0

170 Wei, M., Jacobs, G., Rowley, C., Barron, C., Hogan, P., Spence, P., Smedstad, O., Martin, P., Muscarella, P., & Coelho, E. (2016). The performance of the US Navy's RELO ensemble, NCOM, HYCOM during the period of glad at-sea experiment in the Gulf of Mexico [The Gulf of Mexico Ecosystem - before, during and after the Macondo Blowout]. *Deep Sea Research Part II: Topical Studies in Oceanography*, *129*, 374–393. https://doi.org/https://doi.org/10.1016/j.dsr2.2013.09.002

171 Jacobs, G., D'Addezio, J., Ngodock, H., & Souopgui, I. (2021). Observation and model resolution implications to ocean prediction. *Ocean Modelling*, *159*, 101760. https://doi.org/https://doi.org/10.1016/j.ocemod.2021.101760

172 Pinardi, N., Bonazzi, A., Dobricic, S., Milliff, R., Wikle, C., & Berliner, L. M. (2011). Ocean ensemble forecasting. part II: Mediterranean forecast system response. *Quarterly Journal of the Royal Meteorological Society*, *137*(657), 879–893. https://doi.org/https://doi.org/10.1002/qj.816

173 Leroux, S., Brankart, J.-M., Albert, A., Brodeau, L., Molines, J.-M., Jamet, Q., Le Sommer, J., Penduff, T., & Brasseur, P. (2022). Ensemble quantification of short-term predictability of the ocean dynamics at a kilometric-scale resolution: A Western Mediterranean test case. *Ocean Science*, *18*(6), 1619–1644. https://doi.org/10.5194/os-18-1619-2022

174 Melsom, A., Counillon, F., LaCasce, J., et al. (2012). Forecasting search areas using ensemble ocean circulation modeling. *Ocean Dynamics*, *62*, 1245–1257. https://doi.org/https://doi.org/10.1007/s10236-012-0561-5

175 Holm, H. H., Sætra, M. L., & Brodtkorb, A. R. (2020). Data assimilation for ocean drift trajectories using massive ensembles and GPUs. In R. Klöfkorn, E. Keilegavlen, F. Radu, & J. Fuhrmann (Eds.), *Finite Volumes for Complex Applications IX - methods, theoretical aspects, examples* (pp. 715–723). Springer International Publishing.

176 Beiser, F., Holm, H., Sætra, M., Melsom, N., & Christensen, K. (2024). Combining barotropic and baroclinic simplified models for drift trajectory predictions. *Journal of Operational Oceanography*, *0*(0), 1–20. https://doi.org/10.1080/1755876X.2024.2364975

177 Jorda, G., Comerma, E., Bolaños, R., & Espino, M. (2007). Impact of forcing errors in the CAMCAT oil spill forecasting system. A sensitivity study. *Journal of Marine Systems*, *65*(1), 134–157. https://doi.org/https://doi.org/10.1016/j.jmarsys.2005.11.016

178 Kampouris, K., Vervatis, V., Karagiorgos, J., & Sofianos, S. (2021). Oil spill model uncertainty quantification using an atmospheric ensemble. *Ocean Science*, *17*(4), 919–934. https://doi.org/10.5194/os-17-919-2021

179 Norwegian Petroleum Directorate. (2022). *Resource report 2022: Norwegian resources* (tech. rep.) (Last accessed: 2024-10-12). Norwegian

Petroleum Directorate. https://www.sodir.no/globalassets/1-sodir/publikasjoner/rapporter/ressursrapporter/2022/en/resource-report-2022.pdf

180 Cervantes-Hernández, P., Celis-Hernández, P., Ahumada-Sempoal, M., et al. (2024). Combined use of SAR images and numerical simulations to identify the source and trajectories of oil spills in coastal environments. *Marine Pollution Bulletin*, *199*, 115981. https://doi.org/https://doi.org/10.1016/j.marpolbul.2023.115981

181 Varona, H., Noriega, C., Calzada, A., et al. (2024). Effects of meteo-oceanographic conditions on the weathering processes of oil spills in northeastern Brazil. *Marine Pollution Bulletin*, *198*, 115828. https://doi.org/https://doi.org/10.1016/j.marpolbul.2023.115828

ISBN: