NATIONAL & KAPODISTRIAN UNIVERSITY OF ATHENS

MASTER THESIS

Quantification of ocean model uncertainties in an ensemble of high-resolution Bay of Biscay simulations

Author:
Ioannis KARAGIORGOS

Supervisor:
Assist. Prof. Sarantis SOFIANOS

A thesis submitted for the degree of Master of Science

in the

Oceanography and Management of the Marine Environment
Specialization: Physical Oceanography

Athens, Greece
July 2019

QUANTIFICATION OF OCEAN MODEL UNCERTAINTIES IN AN ENSEMBLE OF HIGH-RESOLUTION BAY OF BISCAY SIMULATIONS

Inter-Institutional Postgraduate Course

Oceanography and Management of the Marine Environment

NATIONAL & KAPODISTRIAN UNIVERSITY OF ATHENS

Dissertation Committee:

Assist. Prof. **Sarantis Sofianos** (Supervisor)
University of Athens
Department of Physics

Prof. **Maria Tombrou-Tzella**
University of Athens
Department of Physics

Dr. **George Triantafyllou**
Research Director of HCMR
Institute of Oceanography

Athens, Greece
July 2019

Thesis supervisor:
Sarantis SOFIANOS

Author:
Ioannis KARAGIORGOS

QUANTIFICATION OF OCEAN MODEL UNCERTAINTIES IN AN ENSEMBLE OF HIGH-RESOLUTION BAY OF BISCAY SIMULATIONS

Abstract

Quantifying model uncertainty and error bounds is a key outstanding challenge in ocean state estimation and prediction. In this study, we investigate the impact of different ensemble generation methods in terms of forecast uncertainty representation within the context of a regional/coastal model. As a case study for open-ocean and coastal shelf dynamics, we use a coupled ocean–biogeochemical configuration of the ocean model NEMO (Nucleus for European Modelling of the Ocean) for the Bay of Biscay at $1/36°$ resolution. An ocean ensemble forced by the atmospheric model ensemble product of the European Centre for Medium-Range Weather Forecasts (ECMWF-EPS) is compared with an ensemble that generated through applying stochastic perturbations to an unperturbed (control) run. The stochastic perturbations are chosen with the aim to represent potential model forecast uncertainties from assumptions subject to erroneous atmospheric forcing, improper ocean model parameterizations, and ecosystem state uncertainties. The comparison is made through ensemble statistics focusing on the spatio-temporal variability of model spread for the coupled system. In summary, we find that the ECMWF-EPS forcing does not produce a large wind spread compared to the stochastic modelling approach. As a result of this, the stochastic method can be considered as the primary approach to generate upper-ocean model uncertainties. This study is concluded with a qualitative evaluation of the model forecasts with respect to the observations.

Keywords: Ocean modelling; Ensemble forecast; Model uncertainties; Stochastic modelling; NEMO; Bay of Biscay

ΠΟΣΟΤΙΚΟΠΟΙΗΣΗ ΤΩΝ ΑΒΕΒΑΙΟΤΗΤΩΝ ΑΡΙΘΜΗΤΙΚΟΥ ΜΟΝΤΕΛΟΥ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ ΓΙΑ ΤΟ ΒΙΣΚΑΪΚΟ ΚΟΛΠΟ ΜΕ ΜΕΘΟΔΟΥΣ ΤΥΠΟΥ "ΑΝΣΑΜΠΛ"

Περίληψη

Η ποσοτικοποίηση της αβεβαιότητας των αριθμητικών μοντέλων αποτελεί σημαντική περιοχή μελέτης για την εκτίμηση και την πρόβλεψη της κατάστασης των ωκεανών. Στην παρούσα εργασία, διερευνάται η επίδραση διαφορετικών μεθόδων παραγωγής προσομοιώσεων τύπου "άνσάμπλ" στην αναπαράσταση των αβεβαιοτήτων ενός περιοχικού αριθμητικού μοντέλου. Για τη μελέτη αυτή χρησιμοποιήσαμε ένα συζευγμένο φυσικό-βιογεωχημικό μοντέλο υψηλής ανάλυσης (1/36°) για το Βισκαϊκό κόλπο, βασισμένο στην πλατφόρμα ωκεανογραφικών μοντέλων NEMO. Συγκρίναμε δύο κύριες μέθοδους παραγωγής ανσάμπλ προσομοιώσεων για το μοντέλο του ωκεανού. Η πρώτη μέθοδος ενσωματώνει πολυμεταβλητά ατμοσφαιρικά ανσάμπλ του Ευρωπαϊκό Κέντρο Μεσο-πρόθεσμων Μετεωρολογικών Προγνώσεων (ECMWF-EPS), ενώ η δεύτερη διαταράσσει το πείραμα αναφοράς με τη χρήση στοχαστικών μεθόδων. Οι στοχαστικές διαταραχές επιλέγονται με στόχο να αντιπροσωπεύσουν πιθανές αβεβαιότητες στην πρόγνωση του μοντέλου, που οφείλονται σε παραδοχές υποκείμενες σε εσφαλμένη ατμοσφαιρική δράση, εσφαλμένης παραμετροποίησης μοντέλου και αβεβαιότητας της κατάστασης του οικοσυστήματος. Η σύγκριση γίνεται μέσω στατιστικής ανάλυσης των αποτελεσμάτων, εστιάζοντας κυρίως στην χωρο-χρονική εξέλιξη της διασποράς του συζευγμένου συστήματος. Συνοπτικά, διαπιστώνουμε ότι οι προβλέψεις του ECMWF-EPS δεν παρουσιάζουν μεγάλη διασπορά ανέμου στο περιοχικό μας μοντέλο σε σύγκριση με την στοχαστική μέθοδο μοντελοποίησης. Ως αποτέλεσμα αυτού, η στοχαστική μέθοδος μπορεί να θεωρηθεί ως καταλληλότερη για την δημιουργία αβεβαιοτήτων στο ωκεάνιο μοντέλο. Τέλος, γίνεται μια ποιοτική αξιολόγηση των αποτελεσμάτων σε σχέση με τις παρατηρήσεις.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and motivation

Many factors contribute to errors in regional/coastal ocean model forecasts. These may be related to different sources of model uncertainties, such as physical parameterizations, atmospheric forcing, lateral boundary and initial conditions, bathymetry, river runoff, and so forth. In general, these many sources of uncertainty are difficult to assess. However, the sensitivity of ocean simulations to the different sources of model uncertainties is necessary to assess the forecast skill, which is a key asset for ocean data assimilation, and to produce reliable probabilistic hindcasts and forecasts both in coupled and un-coupled systems.

The quantification and representation of model uncertainties (often formally expressed as model error covariance) both physically and statistically are required in various data assimilation techniques such as the Ensemble Kalman (EnKF) filter (Evensen, 2003), the Singular Evolutive Extended Kalman (SEEK) filter (Pham and Jacques Roubaud, 1998) and the Ensemble Optimal Interpolation (EnOI; Oke et al., 2002; Evensen, 2003). Thus, the success of data assimilation and forecast accuracy depends crucially on the realism of model uncertainties, which in turn depends on the realism of the representation method.

Ensemble methods are a useful way to represent the uncertainties associated with complex systems. The idea of an ensemble forecast is simple enough. Instead of running just one forecast with an unknown error, a set of slightly different forecasts runs describing the evolution of the probability density function (pdf) of the model results. The difficulty lies in constructing the ensemble in such a way that it correctly reflects uncertainties in our knowledge of ocean initial state and modelling of the relevant processes. There are standard methods to represent uncertainty in the ocean for ensemble forecasts, namely: perturbed initial state, perturbed physics (parameters and parametrizations) and multi-physics (such as multi-models).

The main goal of this study is to investigate the impact of different ensemble generation methods in terms of model errors, focusing on upper-ocean processes. As a case study, we use the Bay of Biscay, which covers a broad spectrum of open-ocean and coastal physical processes. The simulations are performed using a coupled ocean physics and biogeochemistry configuration of the ocean model NEMO (Nucleus for European Modeling of the Ocean), with a high horizontal analysis (1/36°) in the experimental region. The aim is to identify the spatial and temporal distribution of model uncertainty for several physical-biogeochemical variables and to evaluate the statistical consistency (reliability) of the model forecasts with respect to the observations.

We use two main approaches to generate an ensemble of simulations in order to represent model uncertainties associated with unresolved or poorly-resolved processes in the coupled physical-biogeochemical system. In the first method, we produce an ocean ensemble using the atmospheric ensemble product of the European Centre for Medium-range Weather Forecasting system (ECMWF-EPS). This ensemble includes several ocean simulations in which a different atmospheric member of the ECMWF-EPS system used as forcing. The second approach is based on stochastic perturbation of several variables/parameters of a control run, generating an ocean ensemble. We also examine a third approach that merges the two above methods using both the ECMWF-EPS and the ocean model stochastic perturbations.

Two schemes of stochastic parameterization are considered: a stochastically perturbed parameterization tendency (SPPT) scheme which has proven particularly effective in the atmosphere (e.g. Palmer et al., 2009), and a stochastic parameterization of unresolved fluctuations (SPUF) scheme similar to Brankart et al. (2015). These schemes can help to improve the representation of small scale, high frequency, and unpredictable processes in a statistical sense, rather than trying to actually resolve these processes. While they are already widely used in weather forecasting, the application of these schemes in the ocean component is fairly recent.

The thesis structures as follows. In chapter 2, we describe the model implementation and ensemble-forecast-generating methods. In chapter 3, the model uncertainties are discussed on a temporal and spatial scale, focusing on the ensemble spread of upper-ocean properties. We also identify the principal mechanisms associated with the ocean response to the different ensemble-generation protocols, and we present a qualitative evaluation with respect to observational products. Finally, we summarize and discuss the results in chapter 4.

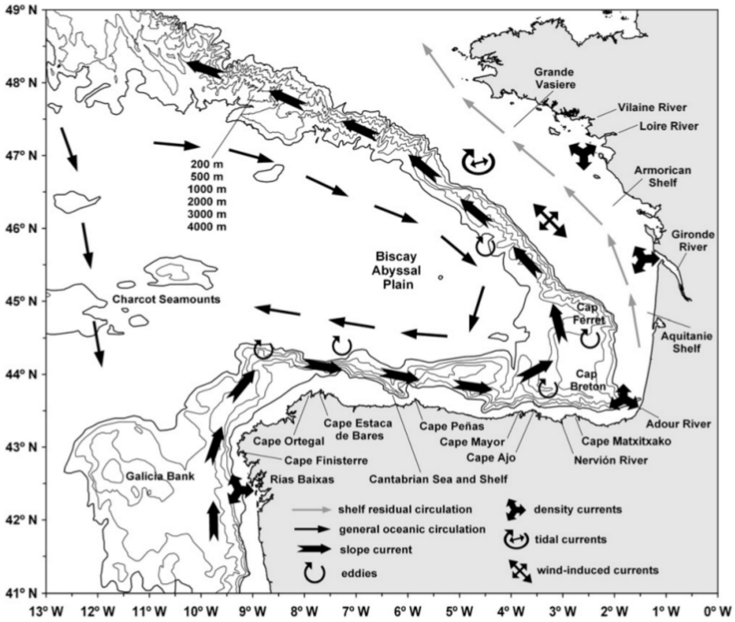
1.2 The study region

Due to the rich nature of its dynamics, the Bay of Biscay, located in the North-eastern Atlantic (Fig. 1.1), is an ideal laboratory for this study. In this region, the circulation constrained by a complex topography (e.g. varying width of the continental shelf, canyons), river runoffs, strong tidal currents, eddies, and large scale circulation. Fig. 1.1A provides a schematic view of the major features of the water circulation occurring in the Bay.

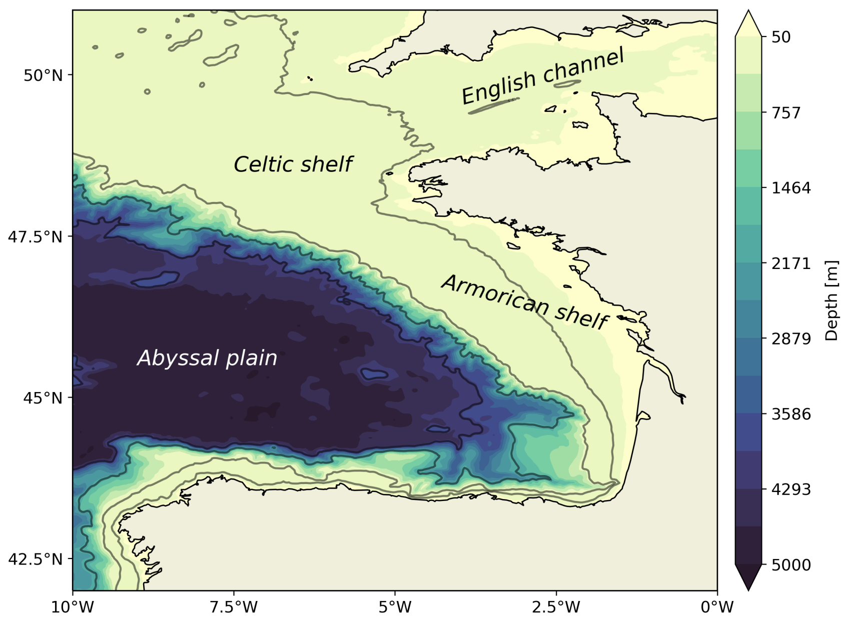
In the central part of the Bay of Biscay, the Abyssal plain, with depths around 4000 m reaching 5500 m, the surface circulation is relatively weak (Charria et al., 2013); it is mainly cyclonic in winter and anticyclonic in summer. The continental margin current instabilities yield long-lived anticyclonic eddies, called Slope Water Oceanic eDDIES (SWODDIES) by Pingree and Le Cann (1992).

Along the slope, which indicates the transition region above the shelf break and the open-ocean part, the circulation exhibits seasonal variations and a vertical baroclinic structure (Pingree and Le Cann, 1990). A typical recurrent dynamic feature of the slope circulation is the warm extension of the Iberian Poleward Current (IPC) along the northern Iberian coast during autumn/winter (Le Cann and Serpette, 2009; Garcia-Soto et al., 2002), which have been referred as the "Navidad" event (Pingree and Le Cann, 1992).

The continental shelf is wide in the northern part of the Bay of Biscay (up to 180 km) but narrows to the south (up to 30 km). The circulation here may be explained as a combination of wind-driven, tidally induced, and density-driven flows (Pingree and Le Cann, 1989). At the shelf break, the interactions between the barotropic tidal currents and the steep topography generate internal tides that propagate both on- and off-shelf from about $5°$ to $9° W$ (Pingree et al., 1986). They appear to be responsible for significant mixing and upwelling of nutrients over the shelf break.



(A)



(B)

FIGURE 1.1: (A) Features of the general water circulation in the Bay of Biscay (modified from Koutsikopoulos and Le Cann, 1996). (B) The Bay of Biscay model domain and bathymetry in meters. The black lines represent the 100, 200, 2000, 4000 m isobaths.

Chapter 2

Methodology and experimental design

2.1 Model configuration

In this study, all simulations have been performed based on a high-resolution regional configuration of NEMO (Nucleus for European Modelling of the ocean; Madec, 2012) modelling framework, targeting coastal to open-ocean applications. The model domain covers the Bay of Biscay and the western part of the English Channel between 11–0.26°W and 41–51.5°N, which is a subgrid of the Iberian-Biscay-Irish (IBI) model (more details and validation procedure given in Maraldi et al., 2013). This domain includes the Bay of Biscay abyssal plain, the shelf break and a variety of wide and narrow expanses of the continental shelf (Fig. 1.1B).

2.1.1 The coupled physical-biogeochemical NEMO/PISCES model

The primitive model equation NEMO platform in its stable version 3.6 is used, which realistically represents the ocean dynamics and its interactions with the biogeochemical component. The main part of the NEMO system is the 3D hydrostatic OPA (Ocean PArallelised; Madec et al., 1998) module which computes the oceanic circulation. The biogeochemical model PISCES-v2 (Aumont et al., 2015) is coupled on-line (one-way coupling and at the same resolution) to OPA via the TOP2 component that manages the advection-diffusion equations of passive tracers and biogeochemical source and sinks terms. The coupling is done every two time-steps (i.e., 150 s for physics and 300 s for biogeochemistry), and the biogeochemistry has no feedback on the physics. This on-line high-frequency coupling is optimal compared to an off-line approach, in terms of the conservation of tracers Gutknecht et al. (2016).

The primitive equations are discretized on a curvilinear Arakawa C-grid which is a refined subset at 1/36°(≈2–3 km) of the so-called "ORCA" tripolar grid (Bernard et al., 2006). The model grid is an exact 3:1 refinement of the PSY2V4R2 parent grid providing initial and boundary conditions. The chosen resolution characterizes the model as eddy-resolving in the abyssal plain and eddy-permitting on the shelves. The vertical grid has 50 geopotential levels as in the parent grid model, with a resolution decreasing from $\sim$ 1 m in the upper 10 m to more than 400 m in the deep ocean. The vertical mixing is parameterized using a k-epsilon scheme (Umlauf and Burchard, 2003), and the 3^{rd} order QUICKEST scheme is used (Leonard, 1979) for tracer advection. For the biogeochemical processes, an Euler integration scheme is used which differs from the scheme used in physics, i.e., leap-frog scheme. The advection scheme is the same as in physics, i.e., QUICKEST. More details for the physical component of the numerical set-up provided by Maraldi et al. (2013) and Vervatis et al. (2016).

The NEMO/PISCES set-up is the same as in Vervatis et al. (2019). The PISCES architecture includes 24 prognostic variables simulating the biogeochemical cycles of oxygen, carbon, and the main nutrients controlling phytoplankton growth, i.e., nitrate, ammonium, phosphate, silicic acid, and iron. The model distinguishes four plankton functional types based on size, including two phytoplankton compartments (nano-phytoplankton and diatoms) and two zooplankton classes (micro-zooplankton and meso-zooplankton). The distinction of the two phytoplankton size classes, along with the description of multiple nutrient co-limitations allows the model to represent ocean productivity, across different biogeographic ocean provinces (Longhurst, 1998). The phytoplankton prognostic variables are the total biomass in C, Fe, Si (only for diatoms) and chlorophyll and hence the Fe/C, Si/C, Chl/C ratios are variable. This allows a more accurate conversion of phytoplankton into chlorophyll concentrations, which is of great importance for comparisons with proxy ocean colour satellite data. PISCES also distinguishes three non-living pools for organic carbon: small Particulate Organic Matters (sPOM), big Particulate Organic Matters (bPOM; different settling velocities with sPOM) and semi-labile Dissolved Organic Carbon (DOC).

In the coupled NEMO/PISCES configuration, the biogeochemical variables are considered as tracers with their evolution determined by the advective-diffusive Eq. 2.1.1:

$$\frac{\partial C}{\partial t} = -\overbrace{\nabla(uC)}^{\text{A}} - \overbrace{K_h \nabla_h^2 C}^{D_h} + \overbrace{\frac{\partial}{\partial z}(K_z \frac{\partial C}{\partial z})}^{D_v} + SMS(C), \tag{2.1}$$

where C is the model state vector including the 24 PISCES prognostic variables. The terms A, D_h and D_v represent respectively the advection, the horizontal diffusion and the vertical diffusion computed by the physical model. SMS is the "Source Minus Sink" term of the PISCES model budget due to the biogeochemical sources and sinks.

2.1.2 Forcing and boundary conditions

Meteorological fields are provided by the European Center for Medium-Range Weather Forecasts (ECMWF, https://www.ecmwf.int/) with a resolution of three hours in time and 9 km or 18 km in the horizontal depending from the forecast product, as we will see below. Wind stress and heat surface fluxes are computed, within the model, with bulk formulae (Large and Yeager, 2004) using a set of atmospheric variables which is summarized in Table 2.1. Freshwater fluxes are also considered in the model implementation.

TABLE 2.1: Atmospheric variables from the ECMWF used to calculate the atmospheric forcing both and on the ensemble and the control run simulations.

10 m U wind component
10 m V wind component
2 m air temperature
2 m relative humidity
2 m specific humidity
Surface downward solar radiation
Surface downward thermal radiation
Total precipitation
Mean sea level pressure

Initial and open boundary conditions (temperature, salinity, current velocities, and sea surface height) are provided by the Mercator Ocean operational system over the North Atlantic area at 1/12°, named PSY2v3. A monthly climatology provides the river's discharge contribution at the sites of the three major rivers Loire (2.19 °W 47.27°N), Gironde (0.73°E 45.26°N) and Adour (1.52°E 43.53°N) acts as a forcing mechanism for coastal non-linear processes in the Armorican shelf.

Tidal forcing is implemented as the sum of 11 tidal harmonics (M2, S2, K2, N2, K1, O1, P1, Q1, M4, Mf, and Mm) provided by the TPXO 7.1 global tide model (Egbert et al., 1994). Finally, an Inverse Barometer (IB) signal is added to the sea level at the boundaries, since the parent system does not include any atmospheric pressure forcing. See Maraldi et al. (2013) for a detailed description.

The open boundaries of the biogeochemical model are forced by the global system BIOMER4V1R1 (resolution: 1/2°; http://www.mercator-ocean.fr/), providing 3D global weekly mean analysis of dissolved iron, nitrate, phosphate, silicate, oxygen, chlorophyll, phytoplankton concentrations, and primary production parameters.

2.2 Ocean ensemble generation methods

In this section, we will describe the ensemble forecast methods that we use to generate ocean-biogeochemical ensemble forecasts over the Bay of Biscay domain (Fig. 1.1B).

2.2.1 The ECMWF-EPS system

The first method studied here uses forecast products from the ECMWF global ensemble prediction system (ECMWF-EPS). The ECMWF-EPS is one of the most successful prediction systems and proved to be extremely useful in a wide range of applications (Buizza, 2006). However, the capability of this product in generating an ensemble of regional ocean forecast is tested here. The atmospheric forcing variables of ECMWF-EPS product used in this study are summarized in Table 2.1. In Fig. 2.1, we also illustrate an example of differences, in forcing variables for two members, over the Bay of Biscay domain.

The key feature of the ECMWF-EPS system is the usage of singular vectors to perturb the atmospheric forecast initial conditions (Lacarra and Talagrand, 1988; Farrell, 1990). The forecast products consist of one control forecast starting from the best guess initial conditions, and 50 members starting from slightly perturbed initial conditions at a horizontal resolution of approximately 18 km.

2.2.2 Stochastic approaches

We use the SPPT (Stochastically Perturbed Parametrisation Tendencies) scheme for the ocean in a similar way to weather forecasting systems (Buizza et al., 1999). The SPPT scheme addresses model uncertainty due to the physics parameterization schemes by perturbing the physics tendencies or parameters using multiplicative

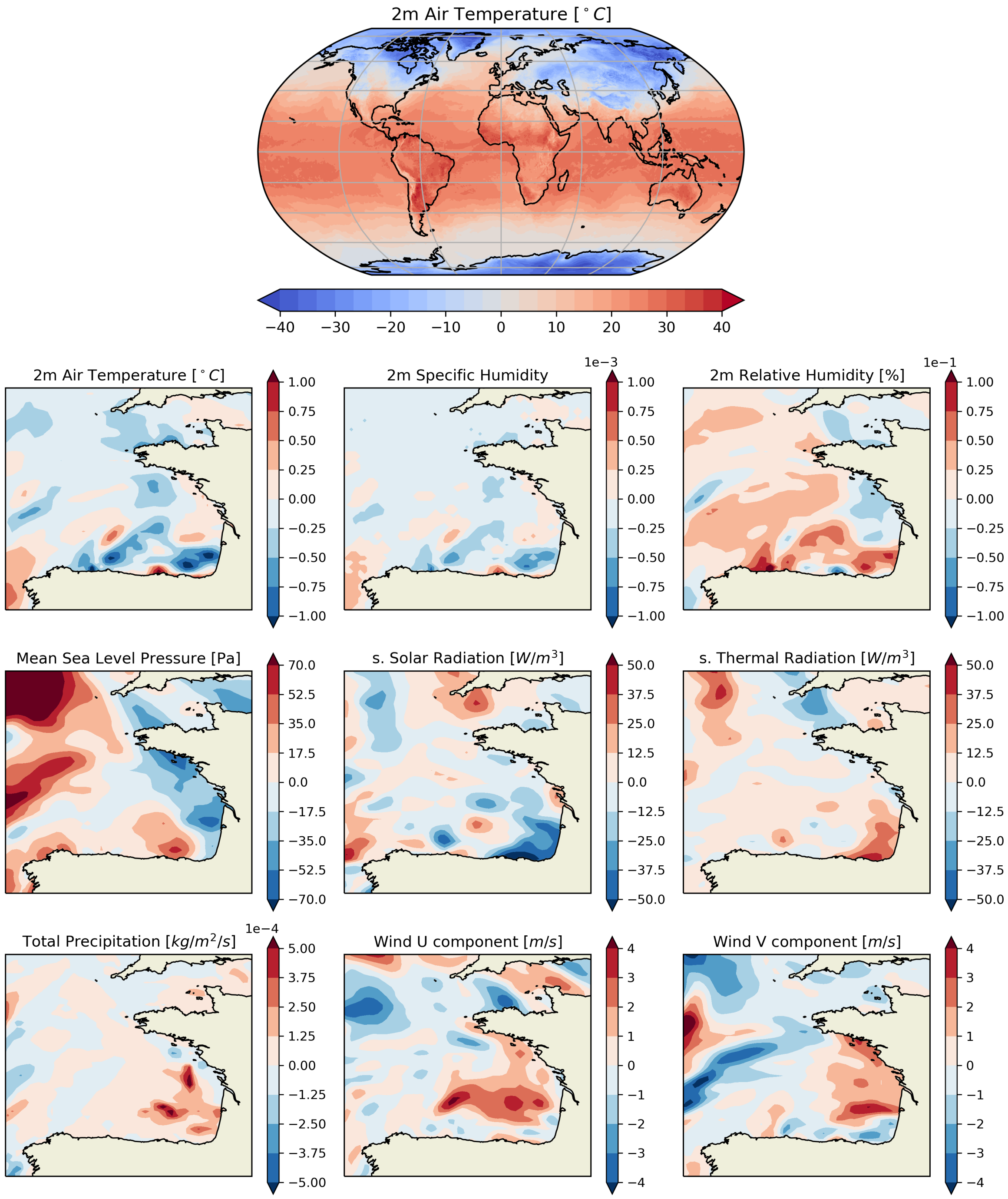


FIGURE 2.1: The ECMWF-EPS atmospheric ensemble product. Row one: 2m air temperature in the global; rows two-five: the difference of members 1 and 2 for the atmospheric forcing variables in the Bay of Biscay domain. Date: 30-Jan-2017 15:00:00.

noise; the word "tendency" refers to the change in a variable over time. Let $\mathcal{X}$ denote the net parametrized physics tendency of any variable. For an unperturbed tendency $\mathcal{X}_c$, the perturbed tendency $\mathcal{X}_p$ is computed as:

$$\mathcal{X}_p = (1+\xi)\mathcal{X}_c, \tag{2.2}$$

where ξ is the zero-mean random perturbation. The multiplication with the random number ξ takes place at every time step and every grid node.

There are a few properties of the random number that needs to be defined, namely temporal and spatial correlations, and the shape of the distribution. The temporal correlation is given by a Gaussian auto-regressive process of first-order (AR(1)) (Brankart et al., 2015):

$$\xi_{k+1} = e^{-1/\tau}\xi_k + (\sigma + \sqrt{1-e^{-2/\tau}})w + \mu(1-e^{-1/\tau}), \tag{2.3}$$

where k is the model time-step, w is a Gaussian white noise, μ, σ and τ the mean, the standard deviation (uncertainty amplitude) and the correlation timescale, respectively.

The spatial correlation is introduced to the system by replacing the white noise of w in Eq. 2.2.2. For more realistic spatial scales, instead of implementing white noise, we introduce anisotropic 2D correlated spatial scales based on long-range spatial correlations modelled through an anisotropic Gaussian recursive filter (Riishøjgaard, 1998; Purser et al., 2003). More in specific, we generate 2D equal probability density contours, using a bivariate normal distribution function (Gaussian) over a few random grid-points in the model domain. The centers of the distributions are called modes. Then, the stochastic patterns are produced by the linear combination of these unimodal probability distributions. Since downscaling projection methods (i.e for the atmospheric forcing) are subject to many sources of model uncertainties, this approach is of vital importance in high-resolution regional configurations.

For the generation of the stochastic ensembles, we use the SPPT scheme focused on uncertainties from:

- Surface boundary conditions perturbing the atmospheric forcing
  - U and V wind component ($\mathcal{U}$)
  - Air temperature (T_{air})
  - Sea level pressure (SLP)
- Improper parametrizations perturbing ocean model coefficients

- Wind drag (c_d)
- Turbulent fluxes (c_e, c_h)
- Bottom drag (c_b)

- State variables perturbing the "Sources Minus Sink" (SMS) term of the ecosystem in Eq. 2.1.1

The spatio-temporal correlation parameters for applying the auto-regressive processes are presented in the Table 2.2.

TABLE 2.2: Statistical parameters defining the first-order auto-regressive processes in the context of SPPT scheme.

Perturbed variables	Uncertainty amplitude (σ)	Correlation timescales (τ)	Spatial scales (σ_r)	Distribution
$\mathcal{U}$	0.4	3 days	1 °	Gaussian
T_{air}	0.1	15 days	2 °	Gaussian
SLP	0.01	5 days	3 °	Gaussian
C_d, C_e, C_h	0.1	3 days	0.5 °	Gaussian
C_b	0.2	30 days	0.2 °	Gaussian
SMS	0.8	10 days	0.5 °	Log-normal

We also perturb the nonlinear equation of state $\rho(T, S)$, using the stochastic parameterization of unresolved fluctuations scheme (SPUF; Brankart et al., 2015). The SPUF scheme is based on random walks sampling gradients from the state vector (T/S), in order to simulate the effect of unresolved temperature and salinity fluctuations (ΔT and ΔS) (Brankart, 2013). The random walks are parameterized with independent first-order auto-regressive processes. The statistical parameters defining the random walks (Table 2.3) were selected in an ad hoc way to inflate the ensemble spread. No special tuning has been preceded.

TABLE 2.3: Statistical parameters defining the random walks in the context of SPUF scheme.

Number of random walks	p=1
Horizontal standard deviation	1 grid-point
Vertical standard deviation	0.5 grid-point
Correlation time-scale	τ = 5 days

2.2.3 Experiments

Results from four simulations will be considered in the next chapters:

1. A control run (hereinafter CR), forced by the unperturbed atmospheric fields (ECMWF-HRES product; 9km resolution) and without other stochastic parametrization, has been first carried out for the ocean-biogeochemical model to develop coherent structures starting from PSY2V4R2 analysis and identify the main physical processes of the Bay of Biscay. The CR starts on January 2015 and includes the period in which the ensemble experiment was performed in a second step (December 2016 to June 2017). Since all ensembles were initialized by using the ocean and the biogeochemical states of the CR (without perturbing the initial conditions), the deterministic CR serves as a reference for the ensemble experiments.

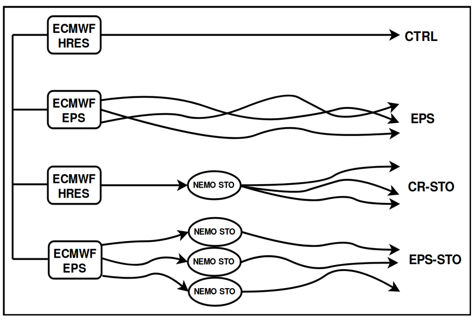
2. An ocean ensemble that incorporates atmospheric ensembles (i.e., using the ECMWF-EPS system). We call this experiment the EPS ocean ensemble forecast. The EPS is constructed, forcing several ocean runs with different members of the ECMWF-EPS atmospheric ensemble.

3. An ocean ensemble based on stochastic approaches (i.e., SPPT and SPUF schemes). We call this experiment the CR-STO ocean ensemble forecast. The CR-STO ensemble generated by perturbing variables/parameters (Table 2.2) of the CR.

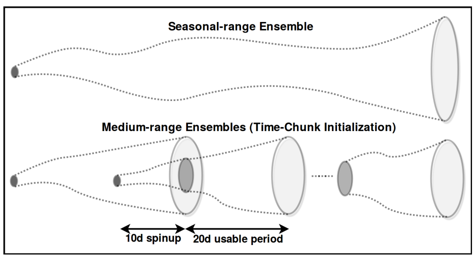
   Several ocean-atmosphere and biogeochemistry sources of uncertainty are considered by perturbing the relevant variables/parameters of the CR, and an ensemble of simulations run with the perturbed fields.

4. An ocean ensemble runs identical to EPS, but in this case, a perturbation applied on each member with the same stochastic protocol used in the CR-STO experiment. We call this merged experiment the EPS-STO ocean ensemble forecast.

The flowchart of the ensemble-based forecasts generation is presented in Fig. 2.2A. In addition, the Table 2.4 briefly summarize all the performed simulations.



(A)



(B)

FIGURE 2.2: The ensemble generation strategy.

In order to be able to properly compare different periods of the year with identical forecast lead time, we perform a short- to medium-range ensembles (we call this method time-chunk initialization or hereinafter "time-chunked" ensembles) over repeated periods (see Fig. 2.2B). Time-chunk initialization is a common practice for operational forecasting systems and permits to take into account the ensemble spin-up period and give access to aging errors within a given forecast lead time, which does not occur in the seasonal range ensemble. The length of the ensemble time-chunks is adjusted between 10 days and up to a month. The short spin-up of 10 days is meant to bring the ensemble up to empirical consistency, and the Usable Period (UP) of "ensemble forecasts" spanning the rest of the time-chunk, is meant to assess model uncertainties and probabilities over lead times within the length of the UP, i.e., 20 days. In general, time-chunked ensembles could be useful to access probabilistic skill scores with respect to forecast lead time mimicking repeated forecasting cycles, but this application is out of the scope of this work.

TABLE 2.4: Table of simulations.

Experiment	Contents	Members	Stochastic perturbations
CR	Control run incorporating the atmospheric ECMWF-HRES product	-	no
EPS	Ocean ensemble incorporating the atmospheric ECMWF-EPS ensemble product in the form of a time-chunked ensemble	10	no
CR-STO	Stochastic time-chunked ocean ensemble based on SPPT and SPUF methods incorporating the ECMWF-HRES product	20	yes
EPS-STO	Combining the EPS and CR-STO ensembles, incorporating the atmospheric ECMWF-EPS product, in the form of a time-chunked ensemble	20	yes
EPS-SR	Seasonal-range ensemble incorporating the atmospheric ECMWF-EPS product	20	no

All the simulations were performed at ECMWF's High-Performance Computing Facilities (HPCF) using resources of ECMWF's special project SCRUM2. In the context of ensemble experiments, the present configuration uses the enhanced MPI strategy of NEMO for double parallelization in the spatial domain and the ensemble dimensions (Bessières et al., 2017). The latter means that ensemble simulations are carried out by just one call to the executable of the coupled NEMO-PISCES system. The configuration scales-out using 96 processors of domain decomposition per ensemble member, excluding land processors. Thus, due to the limited computational resources, only 20 members for CR-STO, EPS-STO and EPS-SR and 10 for EPS were carried out instead of 50 members which are the overall target (i.e., the same number of members as in the ECMWF-EPS perturbed forecasts).

2.3 Diagnostics–Consistency analysis

The ocean model response to the various experimental setups (Table 2.4) is investigated using ensemble statistics, mainly the ensemble mean and spread. The ensemble mean is the average of all ensemble members. The ensemble spread, computed as the standard deviation about the ensemble mean, provides a measure of the forecast uncertainty on the model results. On average, small spread indicates high theoretical forecast accuracy; large spread indicates low theoretical forecast accuracy.

For the model evaluation with respect to observational products, we use the data-space consistency statistics tool of SDAP (Sequoia Data Assimilation Platform), whose functions are interfaced with the NEMO platform and its biogeochemical component PISCES. This tool calculates first- and second-order statistics about the distribution of observational samples, ensemble samples, and innovation (observation minus Ensemble – OmE) samples. For this task, observations are perturbed with a Gaussian random number, in order to generate data distributions with the proper error standard deviation for each network. The observations are considered independent (no cross-correlations), and their Error Covariance Matrix (ECM) is diagonal. In particular, the following consistency statistics are calculated:

- *SPREAD*: mean spread (st.dev.) of the ensemble
- *OmE*: center of the mean OmE distribution (should be close to 0)
- *DOmE*: st.dev. of the debiased mean OmE distribution (should be larger than SPREAD)

Besides, we calculate ensemble quantiles both in the model and data space to assess the ensemble median Q2(50%), the mid-spread Q1(25%)-Q3(75%) and the ensemble envelope Q0(1%)-Q4(99%).

Chapter 3

Results

In this chapter, we want to investigate how different ensemble generation strategies lead to different properties of the ensemble spread both on the open ocean and the shelf regions for different space and time scales. We aim at discussing two distinct areas in the Bay of Biscay, namely the Armorican shelf and the Abyssal plain, both governed by different physical-biogeochemical processes.

3.1 Comparison between ECMWF-EPS and stochastic perturbed wind distributions

The uncertainties in the wind forcing fields are expected to have a significant impact on different space and time scales on the induced ocean response. For example, in the Bay of Biscay, the upper-ocean processes include Ekman currents, mesoscale/sub-mesoscale shelf break exchanges, slope currents, vertical mixing, upwelling/downwelling and density-driven currents reshaped by the wind field.

For this reason, we examine the wind speed standard deviation generated by the different ensemble forecasts, as described in chapter 2. Fig. 3.1b shows the time evolution of domain-averaged wind spread for the EPS and stochastic CR-STO and EPS-STO ensembles, respectively. The EPS wind spread is smaller compared to those from the stochastic ensembles by an average of about 0.4 m/s. Besides, for the stochastic methods, the temporal variability of wind spread follows the mean wind speed variability of the CR (Fig. 3.1a). The EPS-STO method has slightly higher spread values than the CR-STO method, especially in cases where the spread is minimum. Thus, the spread is not added linearly when we use ECMWF-EPS in combination with the stochastic implementation.

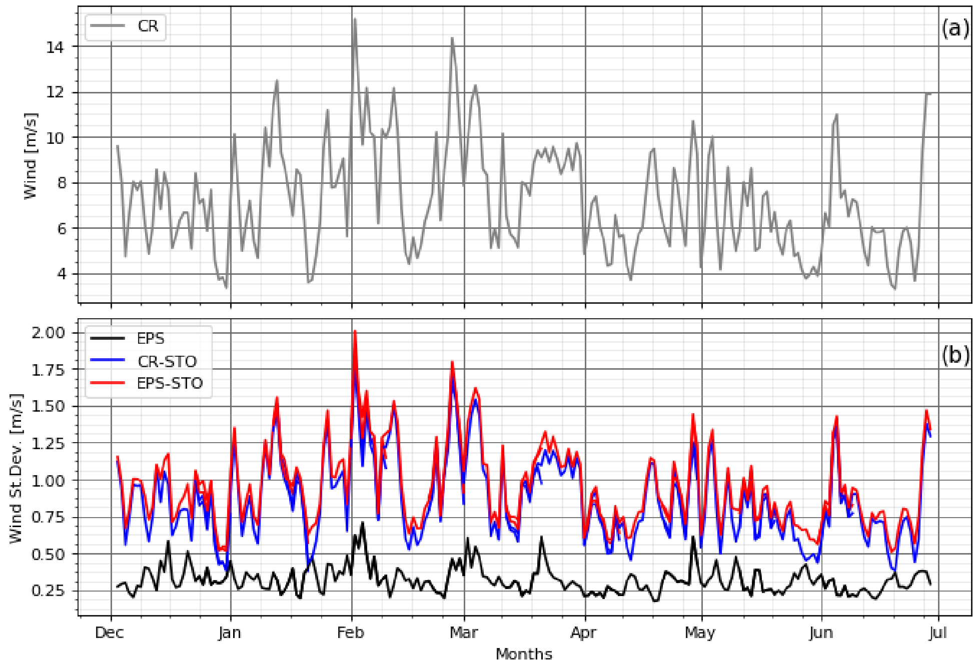


FIGURE 3.1: The temporal evolution of domain-averaged (a) wind speed derived from CR, (b) ensemble standard deviation of wind speed generated by the different ensemble experiments.

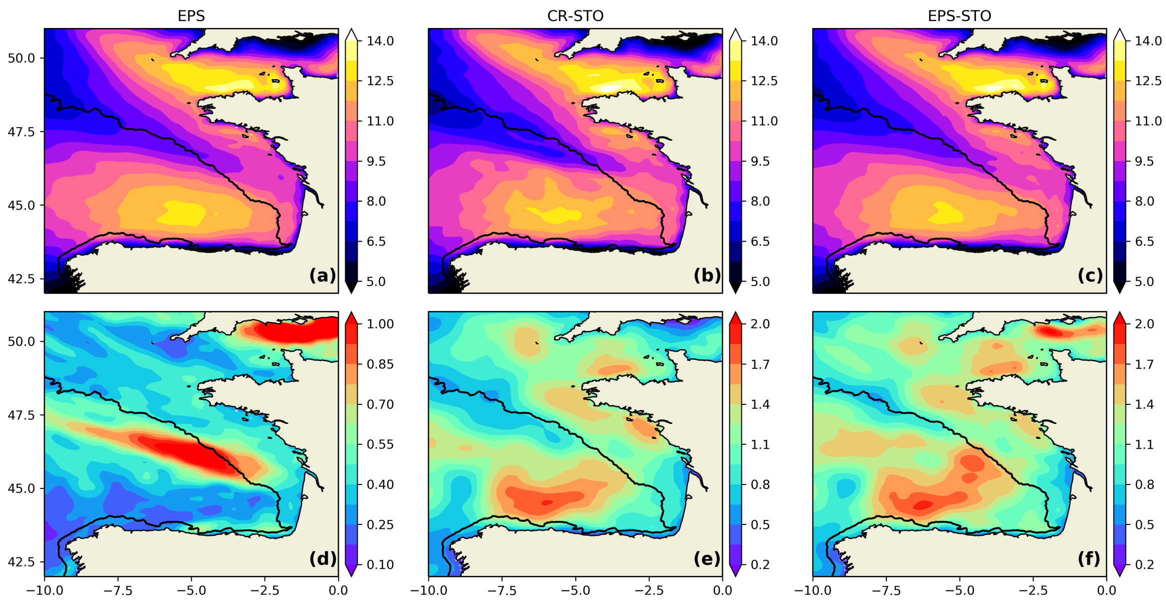


FIGURE 3.2: Snapshots of the ensemble mean (a-c) and spread (d-f) for the wind speed on 01 May 2017. Results correspond to the EPS, CR-STO, and EPS-STO ensembles.

As shown in Fig. 3.2d for a given date, the ECMWF-EPS forcing are not capable of giving significant spread to wind patterns that cover broad-domain scales (Fig. 3.2a). In contrast, it presents high spread values in areas with a strong wind speed gradient (Fig. 3.2a) associated with colocated ensemble members. On the other hand, the stochastic perturbations can produce high values of spread with spatial scales consistent with the wind ensemble mean (see Fig. 3.2b,c, and Fig. 3.2e,f respectively).

In the sections below, we will show that higher uncertainties in the wind field with eventually smaller scales are important in generating an ensemble response at the ocean mesoscale.

3.2 Ensemble spread of upper-ocean properties

3.2.1 Temporal evolution

In Fig. 3.3 and 3.5, we present the impact of the different ensemble-generation-methods on the temporal evolution of the model spread (i.e., 1σ) focusing on surface ocean variables such as SST, surface CHL, and SSH. Smaller model spread occurs for the time-chunked EPS ensemble mainly attributed to the smaller wind spread compared to the stochastic ensembles CR-STO and EPS-STO respectively. The ensemble envelope also presented for the EPS-STO ensemble. We discuss results over the Abyssal plain and the Armorican shelf in model space for 20 members, over the same period from 03/12/2016 to 30/06/2017.

As shown in Fig. 3.3b,d, the EPS ensemble which incorporates the ECMWF-EPS forcing (no stochastic formulation for the ocean state), has a smaller spread in the SST compared with the stochastic ensembles CR-STO and EPS-STO, though their differences are small during the transition period in March-April. The stochastic ensembles seem identical in terms of spread during winter, but differences are found in the spread peaks during the summer, coming from the different atmospheric forcing of ECMWF (CR-STO: 9km; EPS-STO: 18km). The model uncertainties are more extensive in the Armorican shelf than in the Abyssal plain, especially in the winter, as shown by the larger ensemble envelope and spread (Fig. 3.3c,d).

The SST spread evolution follows the seasonal cycle of mixed-layer depth, showing similar behavior both on the shelf and the open ocean. The deepening of the mixed-layer in the winter period (February-March) reduces the impact of the atmospheric forcing on the SST spread modulation. On the other hand, the shoaling of mixed-layer depth in summer (June), increases the impact of wind forcing uncertainties on the SST model errors (Fig. 3.3a,c), therefore enhancing SST spread. More in specific, the SST spread is small during winter at about 0.1°C for the Abyssal plain and 0.2°C for the Armorican shelf, and it is increased during summer at about 0.5°C and 0.6°C respectively. The ensemble envelope is also broader in the summer season than in the winter.

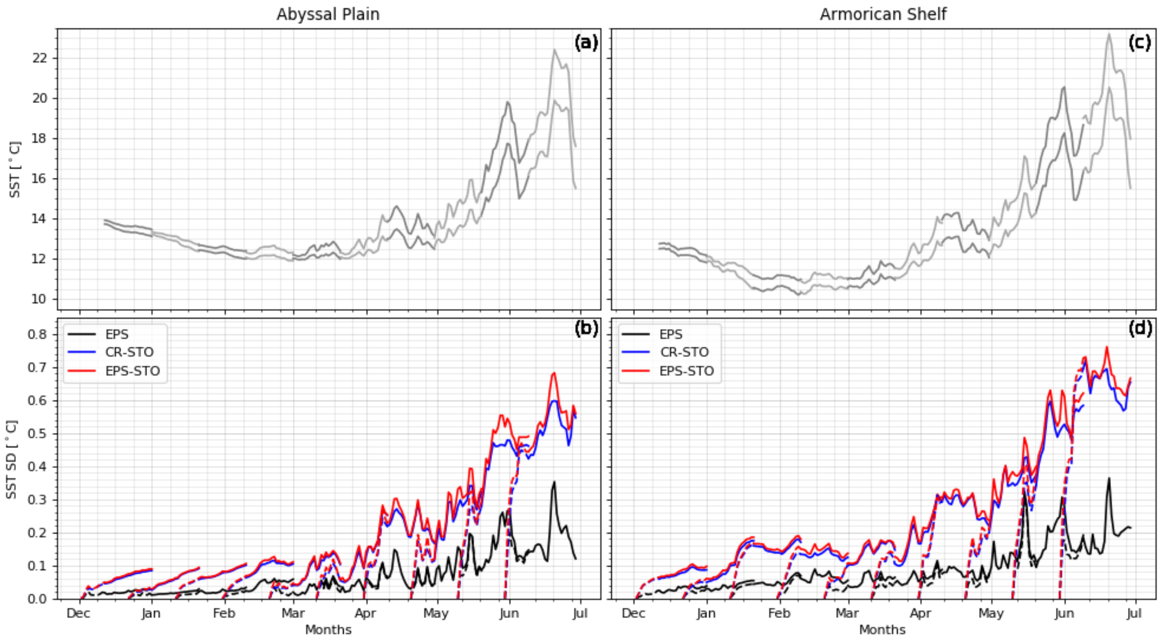


FIGURE 3.3: Temporal evolution of the ensemble statistics for the SST in model space. Upper panels present EPS-STO envelope (quantiles: Q0(cold member); Q4(warm member)). Lower panels show ensemble spread for EPS, CR-STO and EPS-STO ensembles. A spin-up period of 10 days (dashed line; overlapped in each time-chunk) and a Usable Period (continuous line) with the forecast lead time of 20 days is presented. (a), (c) correspond to spatially-averaged values for the Abyssal Plain, and (b), (d) to Armorican shelf respectively.

In Fig. 3.4 we show that the ECMWF-EPS forcing is also not able to augment significant ensemble spread for the SSH. The stochastic ensembles CR-STO and EPS-STO are identical in terms of spread both in the Armorican shelf and in the Abyssal plain, and they have significantly higher spread than the EPS ensemble throughout the running period. The SSH fields are de-tided, filtered over 25 h and daily-averaged. Since we focus on the baroclinic impact of IB and not on the barotropic part, the Inverse Barometer (IB) component also subtracted from the SSH. As clearly seen, the SSH presents highly variable model errors over the Armorican shelf (Fig. 3.4b) compared to the Abyssal plain (Fig. 3.6a). This result is associated with uncertainties induced from high-frequency shelf processes, especially during winter, where the wind uncertainties modulate the shelf dynamics.

Over the Abyssal plain (Fig. 3.4a), the SSH spread appears to be less variable in time. The spread evolution for the time-chunked stochastic ensembles shows a maximum ($\sim 0.7cm$) during the end of winter (February-March). This fact may be associated with the intense mesoscale eddy activity during the winter, which has the potential to increase the spread in SSH. An interesting remark in this area is the different behavior of the seasonal range EPS-SR ensemble, where after about a month of spin-up the spread is progressively increased until the end of the running period. This fact is related to the decorrelation of mesoscale activity, namely the increasing differences with time on the representation of eddies between members (Lucas et al., 2008). This behavior also appears on the shelf towards the end of the run but to a lesser extent.

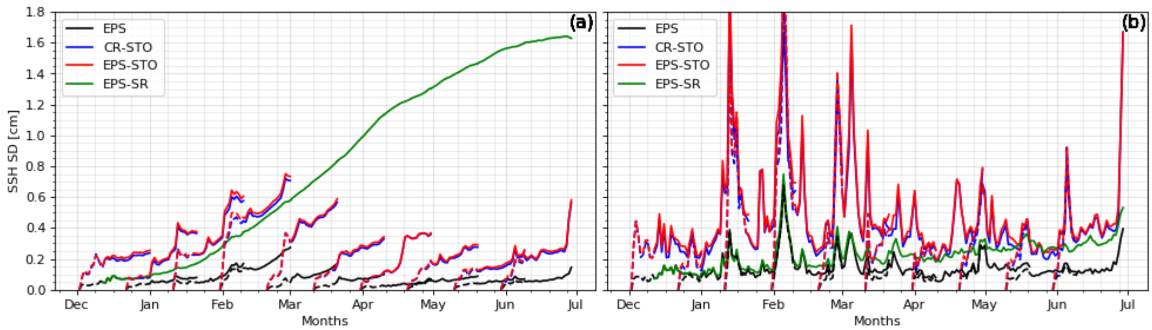


FIGURE 3.4: Temporal evolution of the spatially-averaged SSH ensemble spread (in cm) for the EPS, CR-STO, EPS-STO, and EPS-SR experiments, respectively. A spin-up period of 10 days (dashed line; overlapped in each time-chunk) and a Usable Period (continuous line) with the forecast lead time of 20 days is presented. The inverse barometer has been removed. (a) correspond to the Abyssal plain and (b) to the Armorican shelf.

In the case of surface chlorophyll, very different behavior presented between the shelf and the open ocean. High chlorophyll dispersion is observed over the Armorican shelf (Fig. 3.5c) throughout the run (particularly after the late winter). In contrast, the Abyssal plain (Fig. 3.5a) generally shows low chlorophyll dispersion with a peak during the spring bloom (March-early April).

In Fig. 3.5b,d is highlighted the higher impact of stochastic methods on CHL spread compared to the EPS method, indicating that biogeochemical uncertainties arise from uncertainties in ocean physics (i.e., SST, SSH). The onset of the spring bloom drives a CHL spread maximum over the Abyssal plain ($\sim 0.04 mg/m^3$ for the stochastic ensembles) in March-early April (Fig. 3.5b). This maximum is related to the strong biogeochemical dynamics when seasonal re-stratification occurs, and consequently, the mixed layer becomes shallower. From May onwards, the chlorophyll spread drops sharply to lowest values in summer ($\sim 0.01 mg/m^3$), when water column stratification prevents nutrient supply to euphotic layers. Over the shelf, coastal runoff and river plumes provide continual input of nutrients. In combination with physical processes, such as coastal upwelling, internal waves, and tidal fronts, these nutrients modulate intense uncertainty in chlorophyll throughout the run (above $0.02 mg/m^3$ on average).

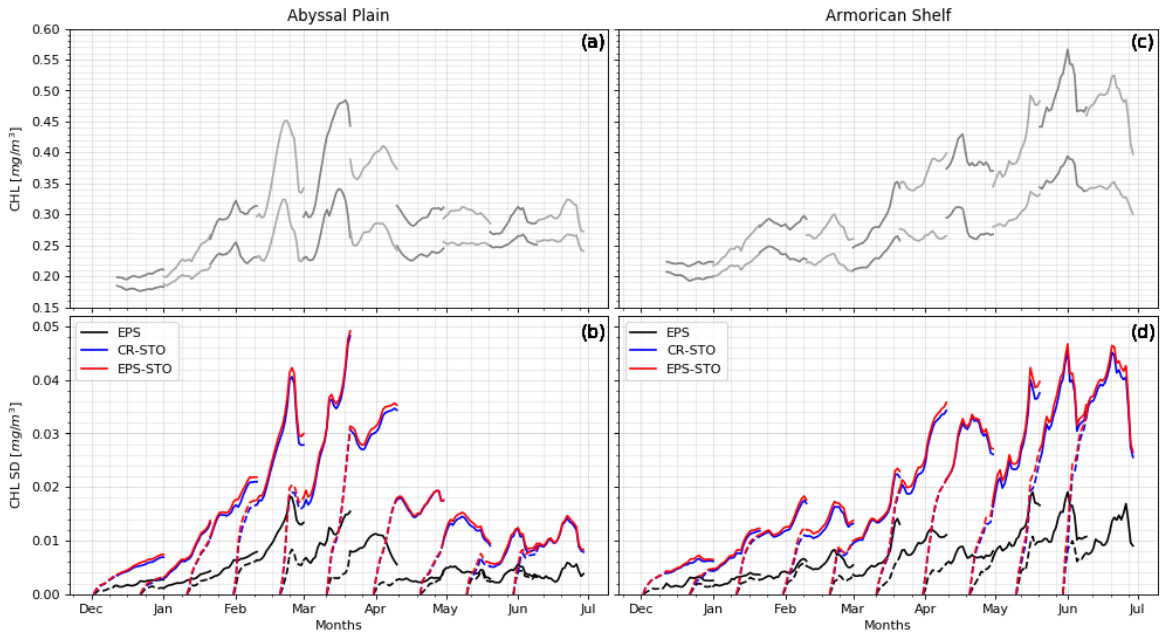


FIGURE 3.5: Identical to Fig. 3.3 for the surface chlorophyll.

3.2.2 Spatial patterns

In this subsection, we examine the spatial patterns of the model spread for the different ensemble-forecast-generating methods.

In a first step, we investigate the skill of the ensemble mean forecasts to produce the main dynamical processes of the Bay of Biscay. Fig. 3.6 and 3.7 shows maps of the ensemble mean (top panels) and spread (bottom panels) for the SST in two contrast days (winter and summer). Model uncertainties appear to linked with processes being different between the open ocean and the shelves, and between seasons respectively. Even if the ensemble experiments have identical ensemble mean, they present differences in spread patterns, especially in the summer period. The ensemble spread localized in a few regions where ocean response to wind forcing is significant. Also, a background ocean mesoscale signal is more widespread, with specific regional maxima that coincide with variable regions of the Bay of Biscay general circulation. As we expected, the stochastic approach presents a higher spread intensity compared to the EPS method.

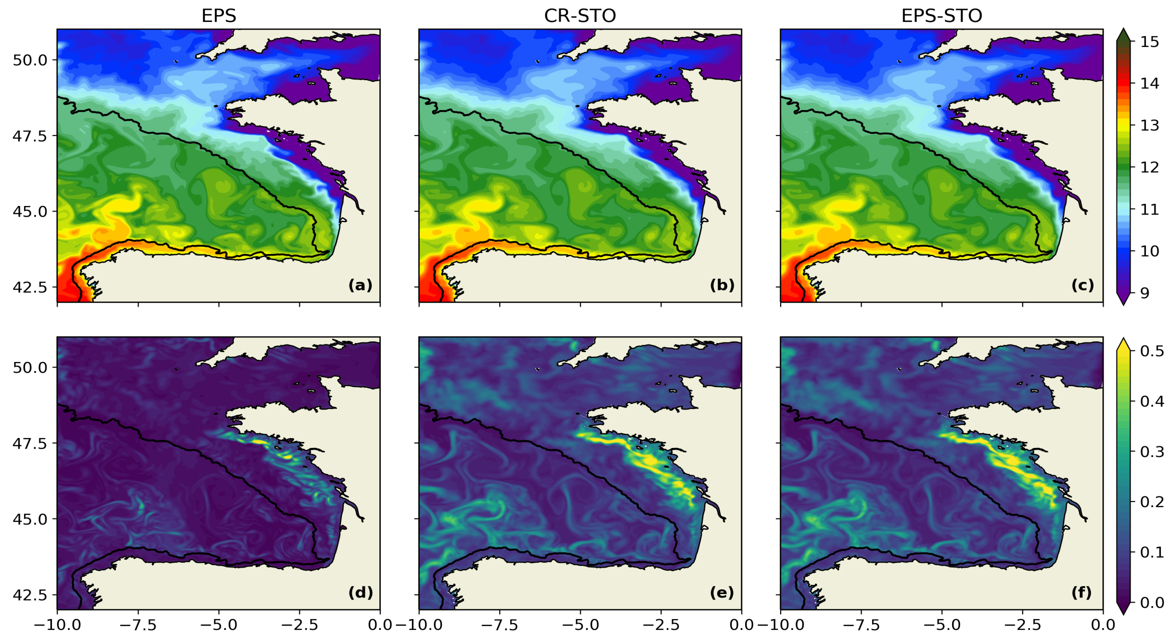


FIGURE 3.6: Snapshots of the ensemble mean (a-c) and spread (d-f) for sea surface temperature (in °C) on 08 February 2017. Results correspond to the EPS, CR-STO, and EPS-STO ensembles.

In winter, the persistence of a warm surface water mass in the southern part of the Bay, the main signature of the so-called Navidad event, is highlighted in Fig. 3.6a-c. The Navidad event generates a surface temperature gradient contrasting the colder temperatures of the northern side of the Bay. In summer (Fig. 3.7(a)-(c)), the simulation's ensemble mean re-produces the "warm pool" in the southeastern part of the Bay of Biscay with temperature exceeding 21 °C (Lazure et al., 2009).

In the Abyssal plain, the SST spread is relatively weak in winter (<0.3 °C), with filament-like patterns around the mesoscale features (Fig. 3.6d-f). Higher values are associated with the active dynamical processes of the southern continental slope, due to the evolving SST, and subsequent generation and evolution of SWODDIES, in turn linked to the Navidad event. The SST spread is higher in summer (Fig. 3.7d-f) compared to winter and can be mainly associated with thermal fluxes due to less mesoscale activity. The maximum SST spread ($\sim$ 1°C) occurs in the southeastern corner of the Bay of Biscay due to the mixed layer depth minimum (not shown).

Over the Armorican shelf, a thermal front divides colder and fresher coastal waters related to river plumes, from the relatively warmer open-ocean waters (Fig. 3.6a-c). Two major rivers, the Loire and the Gironde, are responsible for large outflows during the winter, contributing to establish a strong stratification over the shelf. In summer (Fig. 3.7a-c), the interactions of tidal currents with bottom topography are also responsible for the formation of seasonal thermal fronts, such as Ushant

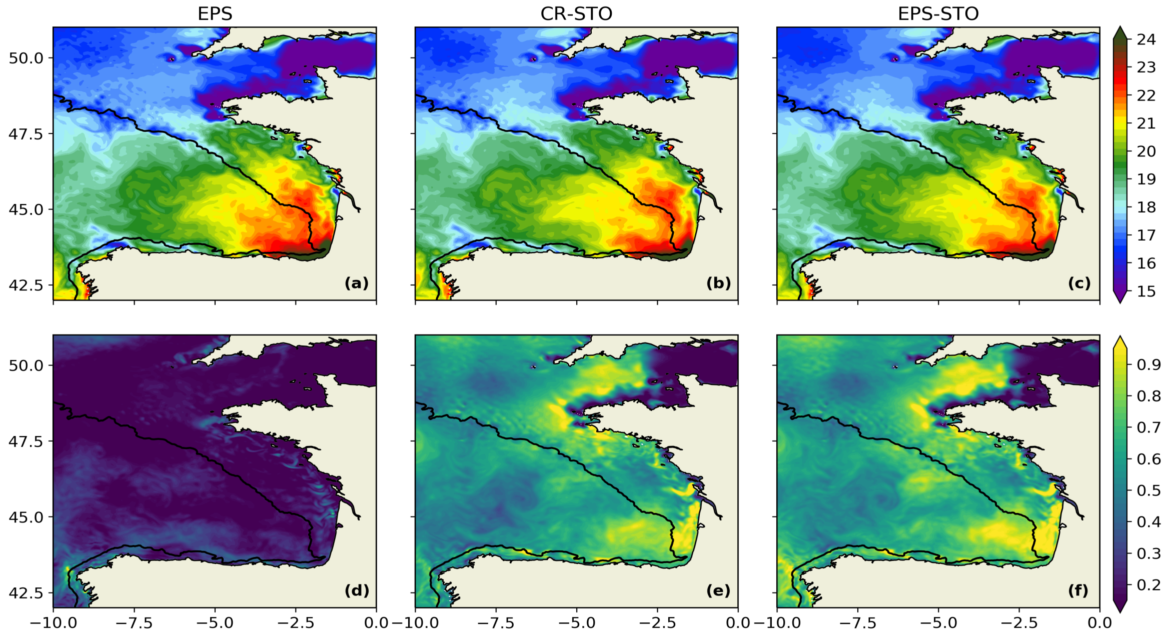


FIGURE 3.7: Identical to Fig. 3.6 for 19 June 2017.

front off western Brittany (48.2 °N, 5.6 °W), with cold waters in the vicinity of the coast and warmer water outside the front (Renaudie et al., 2011; Pasquet et al., 2012; Boyer et al., 2013). The internal tides also induce intense vertical mixing in the shelf, which prevents the formation of seasonal stratification and results in a cold SST tongue along with the shelf break from 11 °W to 5 °W. In addition, the coastal areas of the southern continental shelf and the French coast (vicinity of islands; Loire and Gironde estuaries) principally marked by the presence of local wind-induced upwelling events during summer (Pingree, 1984; Botas et al., 1990; Puillat et al., 2004).

The winter SST spread was found to be larger over the shelf (>0.4 °C for the stochastic ensembles) compared to the Abyssal plain (Fig. 3.6d-f). The maximum values associated with the position of the thermal front whose variability may arise by the combined effect of the wind uncertainties and the river outflow. In summer (Fig. 3.7d-f), the SST spread pattern presents a remarkable difference between stochastic and EPS ensembles in contrast with the winter case. More in specific, the main difference is located on the northern part of the Armorican shelf at the entrance of the English Channel. In stochastic ensembles, we have higher model SST spread which may be induced by the perturbation of the bottom drag coefficient (3.7g-h) since the above area dynamically controlled by tides and their interaction with the seabed. The imprint of the Ushant thermal front is also revealed at the south-western edges of the English Channel and dominates the SST pattern with small SST spread.

In Fig. 3.8a, the modeled surface relative vorticity for the CR at a given date is presented. Eddies and filament-like structures can be seen through large relative vorticity values over the Abyssal plain. The spread of SSH is associated with these structures, mostly in the stochastic CR-STO than in the EPS ensemble (Fig. 3.8b-c). An interesting remark for the open ocean (i.e., the Abyssal plain) is that the pattern of SSH spread is colocated with the pattern of SST spread (see Fig. 3.6d-f). Over the Armorican shelf, the SSH spread is right at the coast and may correspond to waves caused by the wind.

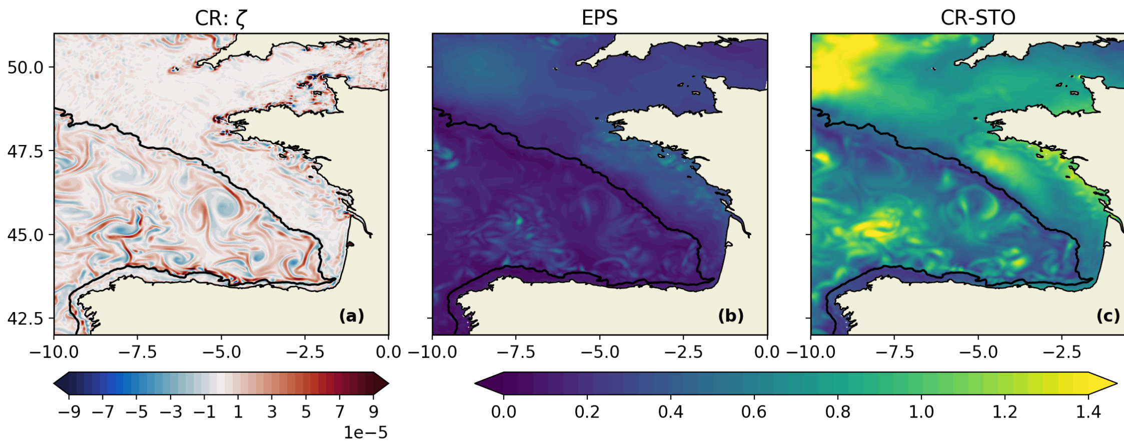


FIGURE 3.8: (a) Snapshot of surface relative vorticity ζ (in s^{-1}) for the CR on 08 February 2017. (b-c) correspond to sea surface height spread (in cm) for EPS and CR-STO experiments, respectively. The inverse barometer has been removed.

Fig. 3.9b-c presents snapshots of surface CHL spread for the EPS and CR-STO ensembles during the spring bloom period (22/03/2017). As clearly seen, the CHL spread follows the eddy mesoscale activity (Fig. 3.9a) over the Abyssal plain, indicating the strong dependence of the biogeochemical field on physical processes. The spread pattern of CHL is different compared with the pattern in SST. The latter presents a filament-like structure, while the former increasing above the center of the anticyclonic eddies (Fig, 3.9a; especially at 46 $^{\circ}N$, 5 $^{\circ}W$), where the surface chlorophyll exhibits high values (Caballero et al., 2016). Over the Armorican shelf, the progressively diffuse processes of river plumes lead to high CHL spread.

In Fig. 3.10, we focus on upwelling events along the Iberian and Spanish coast during summer. The imprint of these processes on the CHL concentration and the relevant spread is presented in Fig. 3.10a-b. The uncertainties in upwelling-favorable winds induce variability in the upwelled nutrients, which in turn have a significant impact on primary production in coastal areas. An interesting feature of upwelled waters is the westward off-shore advection along the 200 m isobaths, in the form of filament structure. This feature is visible in the CHL ensemble mean (Fig. 3.10a; $\sim$ $3^{\circ}W, 44^{\circ}N$) and indicates the transportation of biological material towards to

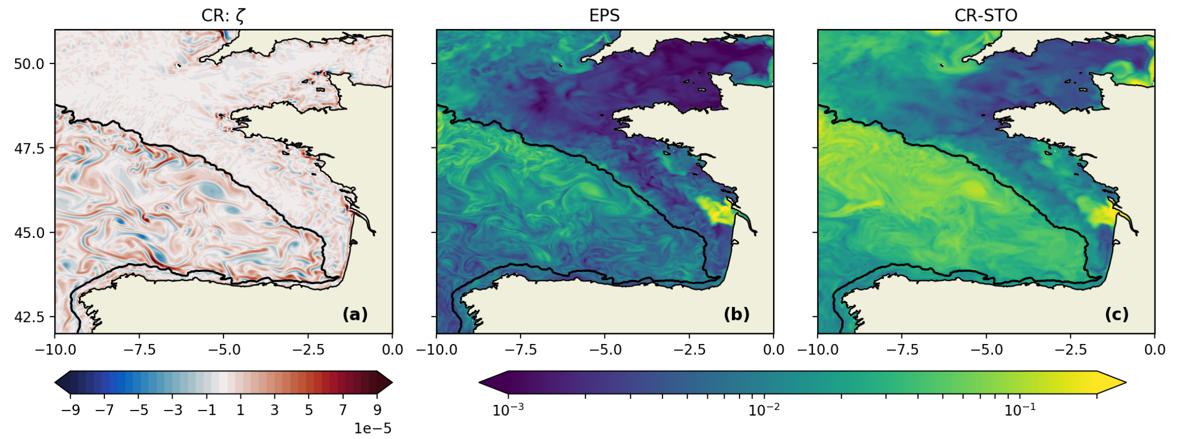


FIGURE 3.9: Identical to Fig. 3.8 for surface chlorophyll concentration (in mg/m^3) on 22 March 2017.

oceanic waters. The "4 $^\circ W$ eddy" (Fig. 3.10c) described by Caballero et al. (2014) have a strong influence on chlorophyll distribution leading to high spread values (Fig. 3.10a-b).

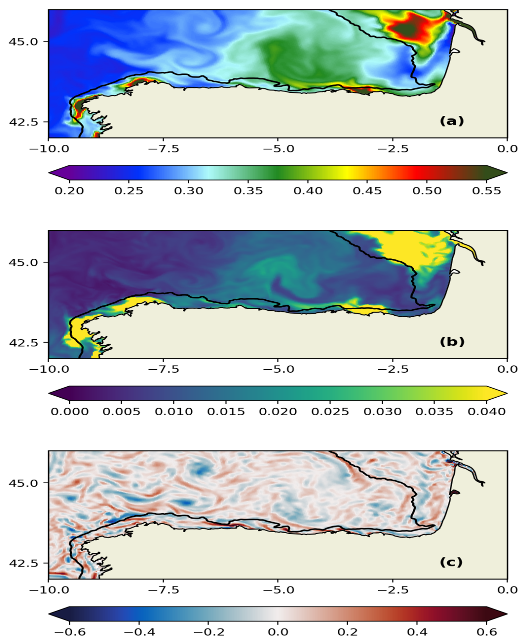


FIGURE 3.10: Snapshots of the ensemble mean (a) and spread (b) of surface chlorophyll concentration (in mg/m^3) for the CR-STO ensemble on 19 June 2017. (c) correspond to the surface relative vorticity ζ (in $s^{-1}(\times 10^{-4})$) for the CR on the same date.

3.3 Ensemble-based innovation statistics

In this section, the time-chunked stochastic ensemble EPS-STO undergo verification with respect to CMEMS (Copernicus Marine Environment Monitoring Service;

http://marine.copernicus.eu/) observational products. In particular, the model-data misfits (hereafter called "innovation" for simplicity) are examined, focusing on upper-ocean variables. Fig. 3.11 and 3.12, shows examples of consistency metrics for SST and sea level anomaly (SLA), based on innovation samples for different datasets and periods. The calculations were carried out with the SDAP platform using only the Usable Period (UP) of the EPS-STO time-chunked ensemble, both for the Abyssal plain and Armorican shelf.

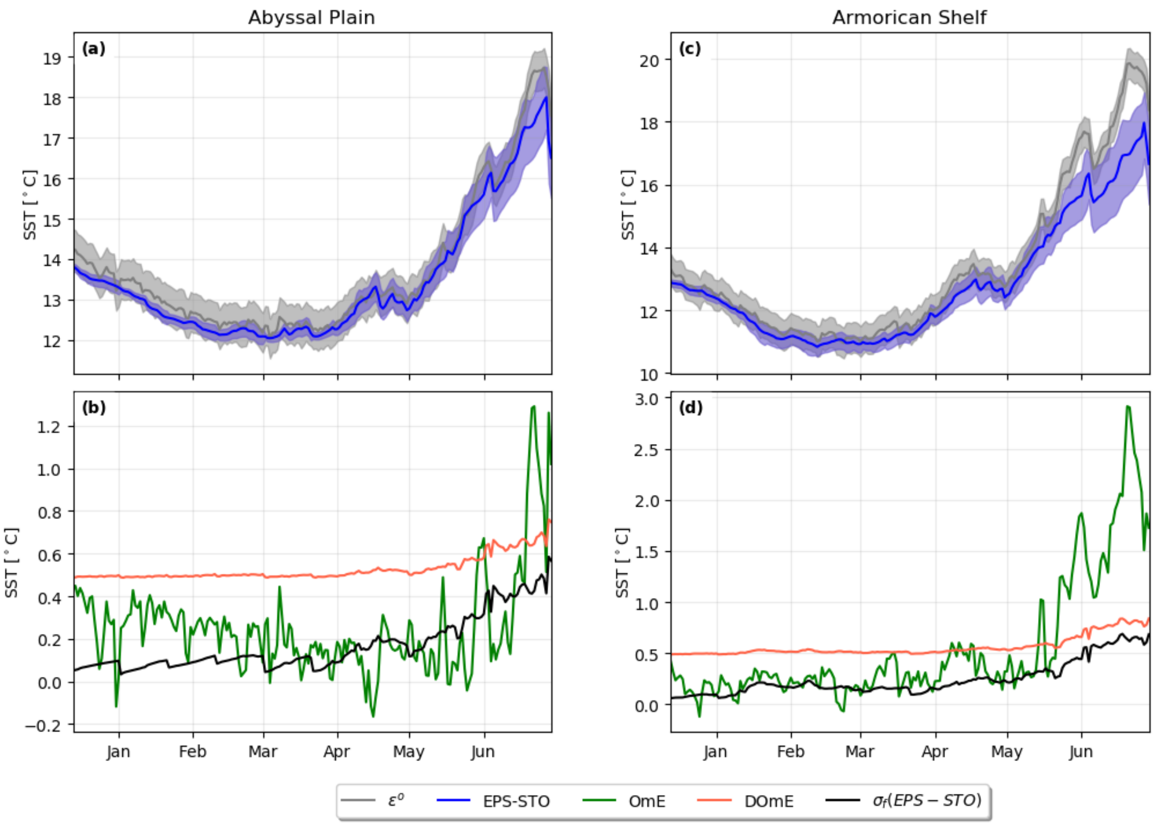


FIGURE 3.11: Ensemble envelope of EPS-STO (blue) in OSTIA-SST (obs. error in grey) (upper panels), and EPS-STO innovation statistics/spread (lower panels). (a-b) correspond to the Abyssal plain and (c-d) to Armorican shelf respectively.

For the SST evaluation, we use a high-resolution (0.05° × 0.05°) OSTIA SST L4 gap-free gridded dataset. OSTIA is a near-real-time (NRT) daily-mean product of foundation SST free of diurnal variability. In Fig. 3.11, the EPS-STO ensemble is compared to a data ensemble generated by the random perturbation of the OSTIA SST product using a Gaussian law and assuming an observational error with st.dev. of $0.5°C$. The model and data ensemble envelope (Fig. 3.11a,c) appear to be compatible with each other since vicinities overlapped most of the time though the EPS-STO is cold-biased over the Armorican shelf during summer (June-July) (Fig. 3.11c). The statistical properties showing that the ensemble is consistent, are the model error estimate σ_f being lower than the innovation spread $DOmE$, and the ensemble mean

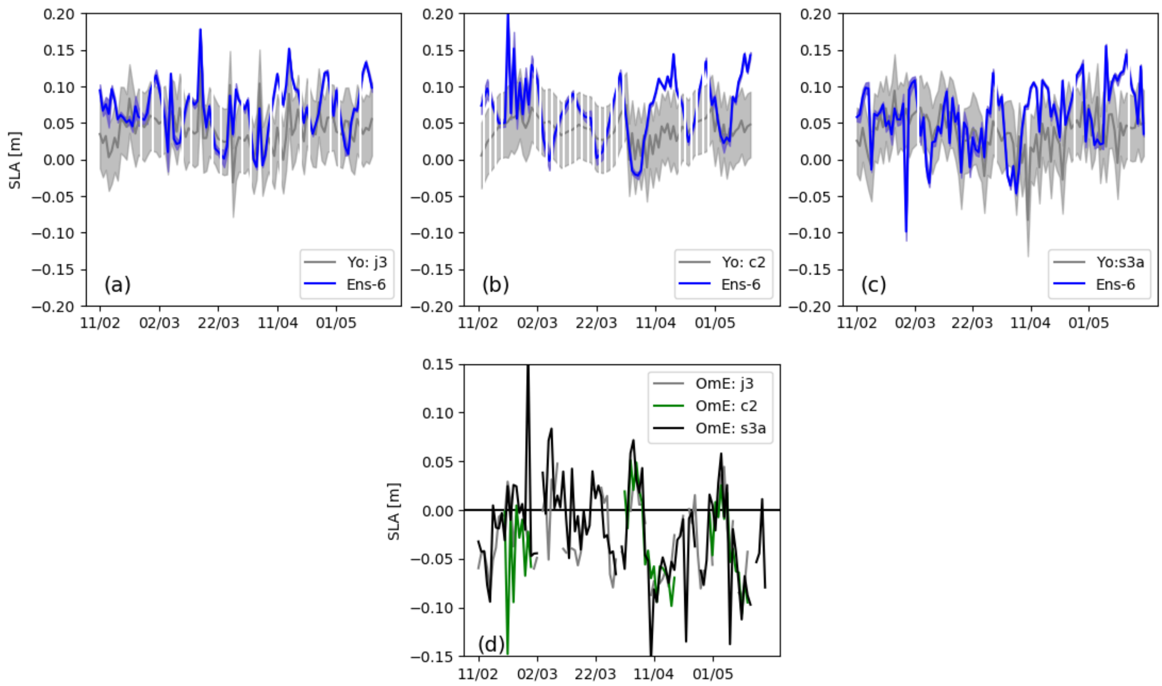
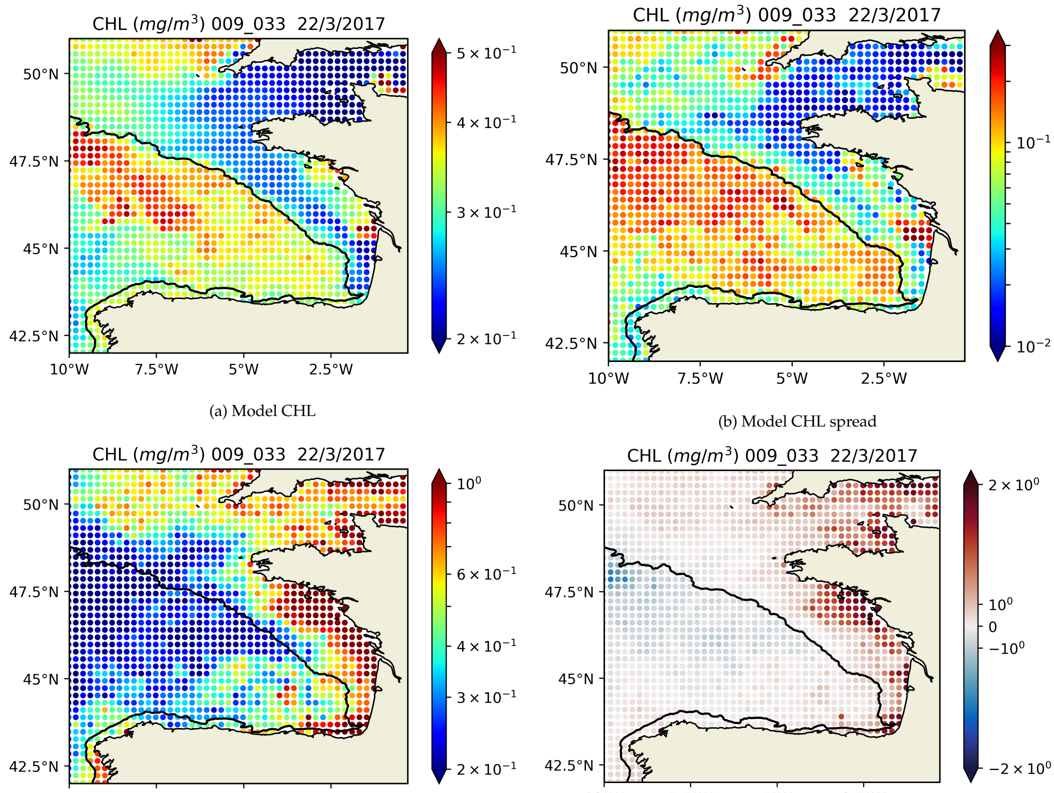


FIGURE 3.12: (a-c) EPS-STO model equivalent SLA envelope (blue) in along-track SLA data (obs. error in grey). (d) innovation metric OmE (in meters), in view of three different altimetry missions in data space (from left to right (a-c)): Jason-3, Cryosat-2 and Sentinel-3A, for the period 01 Feb to 25 Mar 2017.

innovation vector OmE contained within the observational error interval $\pm 0.5°C$, which are desired conditions in both cases (Fig. 3.11b,d).

To evaluate the model SSH, we use L3 along-track seal-level anomaly (SLA) products from different altimetry missions (Jason-3, Cryosat-2, and Sentinel-3A; 14 km resolution). For the SLA model equivalent calculations, we subtracted from the modeled SSH the mean SSH of the control run for the period 2015-2018. The model also includes pressure forcing, and therefore, an inverted barometer (IB) correction is applied to the model and observations. Also, both data and model include tides. In Fig. 3.12, an ensemble generated by the perturbation of the SLA data (assuming an observational Gaussian error with st.dev. of 0.05 m) is compared to EPS-STO ensemble. In Fig. 3.12a-c, we could see the partial overlapping between the model and along-track data distributions. The EPS-STO ensemble appears to have more energy on weekly timescales, resulting in several disjoint uncertainty vicinities. However, we find that OmE stays in a fair number of cases within the observational error interval of 0.05 m (Fig. 3.12d).

For the chlorophyll evaluation, we use a gridded NRT product provided through CMEMS which based on a multi-sensors/algorithms approach. The OC L4 product is reconstructed from L3 reprocessed daily composites applying space-time optimal



(a) Model CHL

(b) Model CHL spread

(c) S3A-OC CHL

(d) OmE CHL

FIGURE 3.13: EPS-STO mean (a) and spread (b) in data space for 22 March 2017. (c) correspond to the satellite observational data and (d) presents the observations minus model metric (OmE) for the same date.

interpolation to fill in missing data (4 km resolution). The innovation statistics are calculated in log space by applying an anamorphosis function to transform ecosystem log-normal distributions into Gaussian distributions. An ensemble of OC (L4) CHL data (generated with random perturbation of CHL data assuming an observational error with st.dev. of 0.3 mg/m^3) is also compared with the EPS-STO ensemble. Fig. 3.13 shows ensemble statistics and observations in data space for CHL during the spring bloom (22/032017). The strong phytoplankton bloom dynamics leads to a high ensemble mean and spread values throughout the Abyssal plain (3.13a-b). Since the OmE metric is close to zero (Fig. 3.13d), the EPS-STO model-data samples appear to be compatible with each other in this region. Over the Armorican shelf, the EPS-STO ensemble is underestimated compared to the observations (3.13d).

Chapter 4

Summary and discussion

This study aimed to explore the response of a coupled ocean-biogeochemical model to different ensemble generation methods. A regional high-resolution NEMO model for the Bay of Biscay was used to perform ensemble simulations. Three ensemble generation strategies were compared. In the first approach, an atmospheric model ensemble (the ECMWF-EPS product) was used to force our ocean model configuration. In the second approach, the ECMWF-HRES deterministic atmospheric forcing was used, and ocean model ensembles were generated by perturbing stochastically several variables in the ocean model. Finally, in the third approach, the two above methods were merged using both the ECMWF-EPS and the ocean model stochastic perturbations. In order to mimic the repeated data assimilation cycles used in most forecasting systems, the simulations were performed with a specific forecast lead time ranging from a few days to weeks (we call this method time-chunked initialization or hereinafter time-chunked ensembles). Those time-chunked ensembles were also compared with a seasonal range ensemble which spanning a period of seven months.

Our stochastic implementation was based on a stochastically perturbed parameterization tendency (SPPT) scheme, which has proven to be an important element of an ensemble forecast system in numerical weather prediction. The SPPT scheme was applied to several sources of model uncertainties in the coupled system such as erroneous atmospheric forcing, improper ocean model parameterizations, and ecosystem model state uncertainties (Brankart et al., 2015). The relatively simple SPPT scheme injects noise into the prognostic equations with an amplitude proportional to the deterministically parameterized tendencies. In addition to the SPPT scheme, the impact of unresolved variability in salinity and temperature on the equation of state was simulated with the stochastic parameterization of unresolved fluctuations (SPUF) scheme, as proposed by Brankart (2013) and Brankart et al. (2015).

In this study, we focused on the model response in upper-ocean variables (SST, SSH, and CHL). Our analysis was mainly based on the ensemble spread, its time evolution, and spatial distribution. Four ensembles were carried out over a period of seven months for the years 2016-2017. We also investigated the consistency of our ensembles against CMEMS observational products using the SEQUOIA Data Assimilation Platform (SDAP).

Our results show that the ocean stochastic implementation was able to produce larger spread in the wind field compared to the approach using the ECMWF-EPS forcing. Since the wind forcing has a major impact on upper-ocean properties, the ocean stochastic method enabled the generation of a larger spread for SST, SSH, and surface CHL. For our domain of application, the combination of the two methods slightly augments the ensemble spread for the wind and the upper-ocean variables.

The most uncertain component of the ocean model response was found to be the mesoscale eddy activity in the Abyssal plain and the high-frequency processes on the shelf. We showed that over the Abyssal plain, eddies and meanders form the spatio-temporal evolution of the ensemble spread. Higher spread values that were observed over the shelves are associated with coastal river runoff processes and upwelling events. The forecasts spread with a lead time of a few days and up to two/three weeks appear to be comparable in magnitude with the expected data errors for most observational networks. Since there is no need to run extended period experiments to obtain large ensemble spread, the similar expected data errors lay the basis for an ensemble-based data assimilation coastal ocean forecasting system.

The quantitative evaluation of model forecast skill with respect to the SST observational product showed that the time-chunked ensembles based on ocean stochastic physics are fairly consistent with the data distribution. The joint probability associated with both model and data ensemble distributions appeared to be always non-zero, showing the potential use of data assimilation. The consistency analysis of the stochastic ensembles with respect to the along-track SLA data showed a partial overlapping between the model and data distributions. Regarding chlorophyll, the stochastic ensemble appeared to be skillful mostly above the Abyssal plain. Statistical consistency was not verified for chlorophyll in coastal regions, like for instance over the Armorican shelf, most likely indicating model ensemble overconfidence (i.e., small ensemble spread).

This study is a step towards the estimation of model uncertainties using ensemble approaches in the context of coastal/regional high-resolution ocean modelling. Future steps include: a) the use of our model ensembles in a physical-biogeochemical multivariate data assimilation scheme and b) the investigation of additional sources of model uncertainty to inflate the ensemble spread, such as uncertainties in the initial and/or boundary conditions.

Bibliography

O. Aumont, C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, 8(8):2465–2513, 2015. doi: 10.5194/gmd-8-2465-2015. URL https://www.geosci-model-dev.net/8/2465/2015/.

B. Bernard, G. Madec, T. Penduff, J.-M. Molines, A.-M. Treguier, J. Le Sommer, A. Beckmann, A. Biastoch, C. Böning, J. Dengg, C. Derval, E. Durand, S. Gulev, E. Remy, C. Talandier, S. Theetten, M. Maltrud, J. McClean, and B. De Cuevas. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dynamics*, 56(5-6):543–567, 2006. ISSN 1616-7341. doi: 10.1007/s10236-006-0082-1. URL http://link.springer.com/10.1007/s10236-006-0082-1.

L. Bessières, S. Leroux, J.-M. Brankart, J. M. Molines, M. P. Moine, P. A. Bouttier, T. Penduff, L. Terray, B. Barnier, and G. Sérazin. Development of a probabilistic ocean modelling system based on NEMO 3.5: Application at eddying resolution. *Geoscientific Model Development*, 10(3):1091–1106, 2017. ISSN 19919603. doi: 10.5194/gmd-10-1091-2017.

J. A. Botas, E. Fernandez, A. Bode, and R. Anadon. A persistent upwelling off the Central Cantabrian Coast (Bay of Biscay). *Estuarine, Coastal and Shelf Science*, 30(2):185–199, 1990. ISSN 0272-7714. doi: https://doi.org/10.1016/0272-7714(90)90063-W. URL http://www.sciencedirect.com/science/article/pii/027277149090063W.

A. Boyer, G. Charria, B. Le Cann, P. Lazure, and L. Marié. Circulation on the shelf and the upper slope of the Bay of Biscay. *Continental Shelf Research*, 55:97–107, 2013. ISSN 0278-4343. doi: https://doi.org/10.1016/j.csr.2013.01.006. URL http://www.sciencedirect.com/science/article/pii/S0278434313000162.

J.-M. Brankart. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. *Ocean Modelling*, 66:64–76, 2013. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2013.02.004. URL http://www.sciencedirect.com/science/article/pii/S1463500313000309.

J.-M. Brankart, G. Candille, F. Garnier, C. Calone, A. Melet, P. A. Bouttier, P. Brasseur, and J. Verron. A generic approach to explicit simulation of uncertainty in the NEMO ocean model. *Geoscientific Model Development*, 8(5):1285–1297, 2015. ISSN 19919603. doi: 10.5194/gmd-8-1285-2015.

R. Buizza. *The ECMWF Ensemble Prediction System*, page 459–488. Cambridge University Press, 2006. doi: 10.1017/CBO9780511617652.018.

R. Buizza, M. Milleer, and T. N. Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560):2887–2908, 1999. doi: 10.1002/qj.49712556006. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712556006.

A. Caballero, L. Ferrer, A. Rubio, G. Charria, B. H. Taylor, and N. Grima. Monitoring of a quasi-stationary eddy in the Bay of Biscay by means of satellite, in situ and model results. *Deep Sea Research Part II: Topical Studies in Oceanography*, 106:23–37, 2014. ISSN 0967-0645. doi: https://doi.org/10.1016/j.dsr2.2013.09.029. URL http://www.sciencedirect.com/science/article/pii/S0967064513003706.

A. Caballero, A. Rubio, S. Ruiz, B. Le Cann, P. Testor, J. Mader, and C. Hernández. South-Eastern Bay of Biscay eddy-induced anomalies and their effect on chlorophyll distribution. *Journal of Marine Systems*, 162:57–72, 2016. ISSN 0924-7963. doi: https://doi.org/10.1016/j.jmarsys.2016.04.001. URL http://www.sciencedirect.com/science/article/pii/S0924796316300355.

G. Charria, B. Lazure, P. Le Cann, A. Serpette, G. Reverdin, S. Louazel, F. Batifoulier, F. Dumas, A. Pichon, and Y. Morel. Surface layer circulation derived from Lagrangian drifters in the Bay of Biscay. *Journal of Marine Systems*, 109-110 (SUPPL.):S60–S76, 2013. ISSN 09247963. doi: 10.1016/j.jmarsys.2011.09.015. URL http://dx.doi.org/10.1016/j.jmarsys.2011.09.015.

G. Egbert, A. Bennett, and M. Foreman. TOPEX/POSEIDON tides estimated using a global inverse model. *Journal of Geophysical Research: Oceans*, 99(C12):24821–24852, 1994. doi: 10.1029/94JC01894. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC01894.

G. Evensen. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. ISSN 1616-7228. doi: 10.1007/s10236-003-0036-9. URL https://doi.org/10.1007/s10236-003-0036-9.

B. F. Farrell. Small Error Dynamics and the Predictability of Atmospheric Flows. *Journal of Atmospheric Sciences*, 47:2409–2416, Oct. 1990. doi: 10.1175/1520-0469(1990)047<2409:SEDATP>2.0.CO;2.

C. Garcia-Soto, R. D. Pingree, and L. Vald es. Navidad development in the southern Bay of Biscay: Climate change and swoddy structure from remote sensing and in situ measurements. *Journal of Geophysical Research*, 107(C8):3118, 2002. ISSN 0148-0227. doi: 10.1029/2001JC001012. URL http://doi.wiley.com/10.1029/2001JC001012.

E. Gutknecht, G. Reffray, M. Gehlen, D. Triyulianti, I. Berlianty, and P. Gaspar. Evaluation of an operational ocean model configuration at 1/12° spatial resolution for the Indonesian seas (NEMO2.3/INDO12) – Part 2: Biogeochemistry. *Geoscientific Model Development*, 9(4):1523–1543, 2016. doi: 10.5194/gmd-9-1523-2016. URL https://www.geosci-model-dev.net/9/1523/2016/.

C. Koutsikopoulos and B. Le Cann. Physical processes and hydrological structures related to the Bay of Biscay anchovy. *Scientia Marina*, 60:9–19, 1996.

J.-F. Lacarra and O. Talagrand. Short-range evolution of small perturbations in a barotropic model. *Tellus A*, 40A(2):81–95, 1988. doi: 10.1111/j.1600-0870.1988.tb00408.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.1988.tb00408.x.

G. Large and S. Yeager. Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies. 2004.

P. Lazure, V. Garnier, F. Dumas, C. Herry, and M. Chifflet. Development of a hydrodynamic model of the Bay of Biscay. Validation of hydrology. *Continental Shelf Research*, 29(8):985–997, 2009. ISSN 02784343. doi: 10.1016/j.csr.2008.12.017.

B. Le Cann and A. Serpette. Intense warm and saline upper ocean inflow in the southern Bay of Biscay in autumn-winter 2006-2007. *Continental Shelf Research*, 29 (8):1014–1025, 2009. ISSN 02784343. doi: 10.1016/j.csr.2008.11.015.

B. P. Leonard. A stable and accurate convective modelling procedure based on quadratic upstream interpolation. *Computer Methods in Applied Mechanics and Engineering*, 19(1):59–98, 1979. ISSN 0045-7825. doi: https://doi.org/10.1016/0045-7825(79)90034-3. URL http://www.sciencedirect.com/science/article/pii/0045782579900343.

M. Lucas, N. Ayoub, B. Barnier, T. Penduff, and P. De Mey. Stochastic study of the temperature response of the upper ocean to uncertainties in the atmospheric forcing in an Atlantic OGCM. *Ocean Modelling*, 20(1):90–113, 2008. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2007.07.006. URL http://www.sciencedirect.com/science/article/pii/S1463500307001011.

G. Madec. NEMO ocean engine. *Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL)*, (27):357pp., 2012. ISSN 1288-1619. URL http://eprints.soton.ac.uk/64324/.

G. Madec, P. Delecluse, M. Imbard, and L. C. Opa 8 ocean general circulation model - reference manual. Technical report, LODYC/IPSL Note 11, 1998.

C. Maraldi, J. Chanut, B. Levier, N. Ayoub, P. De Mey, G. Reffray, F. Lyard, S. Cailleau, M. Drévillon, E. A. Fanjul, M. G. Sotillo, P. Marsaleix, and the Mercator Research and Development Team. NEMO on the shelf: assessment of the Iberia–Biscay–Ireland configuration. *Ocean Sci*, 9:745–771, 2013. doi: 10.5194/os-9-745-2013. URL www.ocean-sci.net/9/745/2013/.

P. Oke, J. S. Allen, R. N. Miller, G. D. Egbert, and P. M. Kosro. Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research: Oceans*, 107(C9):5–25, 2002. doi: 10.1029/2000JC000511. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JC000511.

T. N. Palmer, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer. Stochastic parametrization and model uncertainty. *ECMWF Technical Memoranda*, (598):42, 2009. doi: 10.21957/ps8gbwbdv. URL https://www.ecmwf.int/node/11577.

A. Pasquet, T. Szekely, and Y. Morel. Production and dispersion of mixed waters in stratified coastal areas. *Continental Shelf Research*, 39-40:49–77, 2012. ISSN 02784343. doi: 10.1016/j.csr.2012.04.001. URL http://dx.doi.org/10.1016/j.csr.2012.04.001.

D. T. V. Pham and M.-C. Jacques Roubaud. A singular evolutive extended kalman filter for data assimilation in oceanography, 1998.

R. Pingree, G. T. Mardell, and A. L. New. Propagation of internal tides from the upper slopes of the Bay of Biscay. *Nature*, 321(6066):154–158, 1986. ISSN 0028-0836. doi: 10.1038/321154a0. URL http://www.nature.com/doifinder/10.1038/321154a0.

R. D. Pingree. Some Applications of Remote Sensing to Studies in the Bay of Biscay, Celtic Sea and English Channel. In J. C. J. Nihoul, editor, *Remote Sensing of Shelf Sea Hydrodynamics*, volume 38 of *Elsevier Oceanography Series*, pages 287–315. Elsevier, 1984. doi: https://doi.org/10.1016/S0422-9894(08)70617-2. URL http://www.sciencedirect.com/science/article/pii/S0422989408706172.

R. D. Pingree and B. Le Cann. Celtic and Armorican slope and shelf residual currents. *Progress in Oceanography*, 23(4):303–338, 1989. ISSN 0079-6611. doi: https://doi.org/10.1016/0079-6611(89)90003-7. URL http://www.sciencedirect.com/science/article/pii/0079661189900037.

R. D. Pingree and B. Le Cann. Structure, strength and seasonality of the slope currents in the bay of biscay region. *Journal of the Marine Biological Association of the United Kingdom*, 70(4):857–885, 1990. doi: 10.1017/S0025315400059117.

R. D. Pingree and B. Le Cann. Three anticyclonic slope water oceanic eDDIES (SWODDIES) in the Southern Bay of Biscay in 1990. *Deep Sea Research Part A. Oceanographic Research Papers*, 39(7):1147–1175, 1992. ISSN 0198-0149. doi: https://doi.org/10.1016/0198-0149(92)90062-X. URL http://www.sciencedirect.com/science/article/pii/019801499290062X.

I. Puillat, P. Lazure, A.-M. Jegou, L. Lampert, and P. Miller. Hydrographical variability on the French continental shelf in the Bay of Biscay, during the 1990s, 2004. URL https://archimer.ifremer.fr/doc/00000/779/.

R. J. Purser, W.-S. Wu, D. F. Parrish, and N. M. Roberts. Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances. *Monthly Weather Review*, 131(8):1536–1548, 2003. doi: 10.1175//2543.1. URL https://doi.org/10.1175//2543.1.

C. Renaudie, Y. Morel, G. Hello, H. Giordani, and R. Baraille. Observation and analysis of mixing in a tidal and wind-mixed coastal region. *Ocean Modelling*, 37(3):65–84, 2011. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2011.01.004. URL http://www.sciencedirect.com/science/article/pii/S1463500311000072.

L. P. Riishøjgaard. A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus A*, 50(1):42–57, 1998. doi: 10.1034/j.1600-0870.1998.00004.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0870.1998.00004.x.

L. Umlauf and H. Burchard. A generic length-scale equation for geophysical turbulence models. *Journal of Marine Research*, (2):235–265, 2003.

V. Vervatis, C. E. Testut, P. D. Mey, N. Ayoub, J. Chanut, and G. Quattrocchi. Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4DEnOI based on stochastic modeling of the wind forcing. *Ocean Modelling*, 100:1–19,

2016. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2016.01.003. URL http://www.sciencedirect.com/science/article/pii/S1463500316000044.

V. D. Vervatis, P. De Mey-Frémaux, N. Ayoub, S. Sofianos, C.-E. Testut, M. Kailas, J. Karagiorgos, and M. Ghantous. Physical-biogeochemical regional ocean model uncertainties stemming from stochastic parameterizations and potential impact on data assimilation. *Geoscientific Model Development Discussions*, 2019:1–38, 2019. doi: 10.5194/gmd-2019-31. URL https://www.geosci-model-dev-discuss.net/gmd-2019-31/.

Acknowledgements

First of all, I would like to acknowledge my supervisor, Sarantis Sofianos, who gave me the opportunity to work on such an interesting subject. Many thanks to Vasileios Vervatis for his precious, daily supervision during the last year. Our discussions and cooperation were crucial for this project. Last but not least, I want to thank my friends for their motivation and support during my studies.