HAL open science

Assessment of a regional physical–biogeochemical stochastic ocean model. Part 1: Ensemble generation

Vassilios D Vervatis, Pierre de Mey-Frémaux, Nadia Ayoub, John Karagiorgos, Malek Ghantous, Marios Kailas, Charles-Emmanuel Testut, Sarantis Sofianos

To cite this version:

Vassilios D Vervatis, Pierre de Mey-Frémaux, Nadia Ayoub, John Karagiorgos, Malek Ghantous, et al.. Assessment of a regional physical–biogeochemical stochastic ocean model. Part 1: Ensemble generation. Ocean Modelling, 2021, 160, pp.101781. 10.1016/j.ocemod.2021.101781 . hal-03441625

HAL Id: hal-03441625

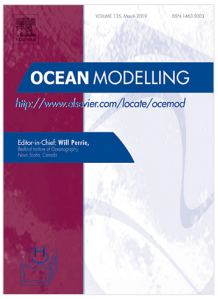
https://hal.science/hal-03441625v1

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proof



Assessment of a regional physical-biogeochemical stochastic ocean model. Part 1: Ensemble generation

Vassilios D. Vervatis, Pierre De Mey-Frémaux, Nadia Ayoub, John Karagiorgos, Malek Ghantous, Marios Kailas, Charles-Emmanuel Testut, Sarantis Sofianos

PII: S1463-5003(21)00031-7
DOI: https://doi.org/10.1016/j.ocemod.2021.101781
Reference: OCEMOD 101781

To appear in: *Ocean Modelling*

Received date : 18 May 2020
Revised date : 19 January 2021
Accepted date : 20 February 2021

Please cite this article as: V.D. Vervatis, P. De Mey-Frémaux, N. Ayoub et al., Assessment of a regional physical-biogeochemical stochastic ocean model. Part 1: Ensemble generation. *Ocean Modelling* (2021), doi: https://doi.org/10.1016/j.ocemod.2021.101781.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Elsevier Ltd. All rights reserved.

Journal Pre-proof

Assessment of a regional physical-biogeochemical stochastic ocean model. Part 1: ensemble generation

Vassilios D. Vervatis (1), Pierre De Mey-Frémaux (2), Nadia Ayoub (2), John Karagiorgos (1), Malek Ghantous (3), Marios Kailas (1), Charles-Emmanuel Testut (4), and Sarantis Sofianos (1)

(1) University of Athens, Department of Physics, Athens, Greece. (2) LEGOS/CNRS, Toulouse, France. (3) Météo France, Toulouse, France. (4) Mercator Ocean International, Ramonville St. Agne, France.

Correspondence to: Vassilios D. Vervatis (vervatis@oc.phys.uoa.gr)

Abstract. In this article, Part 1 of a two-part series, we run and evaluate the skill of a regional physical-biogeochemical stochastic ocean model based on NEMO. The domain covers the Bay of Biscay at 1/36° resolution, as a case study for open-ocean and coastal shelf dynamics. We generate ensembles of uncertainties from assumptions related to errors in the atmospheric forcing, the ocean model parameterizations and in the sources and sinks of the biogeochemical variables. The resulting errors are found to be mainly driven by the wind forcing uncertainties, with the rest of the perturbed forcing and parameters locally influencing the ensemble spread. Biogeochemical uncertainties arise from intrinsic ecosystem model errors and from errors in the physical state. Uncertainties in physical forcing and parameterization are found to have a larger impact on chlorophyll spread than uncertainties in ecosystem sources and sinks. The ensembles undergo quantitative verification with respect to observations, focusing on upper-ocean properties. Despite a tendency for ensembles to be generally under-dispersive, they appear to be reasonably consistent with respect to sea surface temperature data. The largest statistical sea-level biases are observed in coastal regions. These biases hint at the presence of high-frequency error sources currently unaccounted for, and suggest that the ensemble-based uncertainties are unfit to model error covariances for assimilation. Model ensembles for chlorophyll appear to be consistent with ocean colour data only at times. The stochastic model is qualitatively evaluated by analysing its ability at generating consistent multivariate incremental model corrections. Corrections to physical properties are associated with large-scale biases between model and data, with diverse characteristics in the open-ocean and the shelves. Mesoscale features imprint their signature on temperature and sea-level corrections, as well as on chlorophyll corrections due to the vertical velocities associated with vortices. Small scale local corrections are visible over the shelves. Chlorophyll information has measurable impact on physical variables.

Keywords: ensemble modelling, model uncertainties, stochastic physics-biogeochemistry, ocean colour, data assimilation, Bay of Biscay

1 Introduction

For a while now, the oceanographic community has been aware that regional and coastal ocean models have some quite specific requirements for both the methods and the observational data used for an effective data assimilation scheme (see extensive reviews e.g. in *De Mey-Frémaux,* 2000; *Edwards et al.*, 2015; *Pinardi et al.,* 2017; *Fujii et al.*, 2019). But the success of data assimilation in a particular case will depend not only on the power of the methodology and data themselves, but on how realistic our estimates of observational errors and model errors are (*Oke and Sakov*, 2008). Error processes in a model constitute a subspace of the complete space spanned by all of its variables as the model is integrated. Important sources of model errors in coastal domains are the presence of the coast and shallower water, strong bathymetry gradients, inputs from rivers, and are forced by (among other things) pressure and current fields from ocean-scale mass balances, circulation, tides and eddies, winds and air pressure variations, and non-uniform density (*Kourafalou et al.*, 2015a; 2015b).

In addition to the effects of local geography mentioned above, the dynamics of coastal (nested) models are largely controlled by the open boundary conditions (*Ghantous et al.*, 2020). *Thacker et al.* (2012) presented how model uncertainties in the open boundaries manifest as model sea-level uncertainties in the Gulf of Mexico. The Bay of Biscay's dynamics are influenced by the North East Atlantic circulation, especially along the southern slope, with the seasonal reversal of the Iberian poleward current. At depth, the entrance of Mediterranean water masses from the south has been shown to influence the Bay of Biscay's hydrology between 600m and 1500m depth (*Koutsikopoulos and Le Cann*, 1996), and

potentially the circulation through interactions between eddies and deep, salty lenses (see for instance *Carton et al*., 2013).

One common approach to realistically represent error processes and their time evolution is stochastic modelling (e.g. *Adhler et al.*, 1996; and recent studies for regional configurations in *Melsom et al.*, 2012; *Sakov et al.*, 2012; *Quattrocchi et al.,* 2014; *Vandenbulcke and Barth,* 2015, 2019; *Vervatis et al.,* 2016). *Lucas et al.* (2008) generated ensembles using NEMO (Nucleus for European Modelling of the Ocean; http://www.nemo-ocean.eu/; *Madec*, 2012), based on a configuration for the North Atlantic at 1/4° resolution and multivariate EOF modes. The ensemble capabilities of NEMO have been discussed in the literature following the SANGOMA (http://www.data-assimilation.net/) and OCCIPUT projects (*Penduff et al.*, 2014), focusing on global and regional academic configurations spanning from 2° to 1/4° resolution and from seasonal to decadal time scales (*Brankart*, 2013; *Brankart et al.*, 2015; *Candille et al.*, 2015; *Garnier et al.*, 2016; *Bessières et al.*, 2017).

In this paper, we configure, run and assess a stochastic ocean model in the Bay of Biscay, consisting of physical-biogeochemical ensemble simulations with the NEMO-PISCESv2 model (Pelagic Interactions Scheme for Carbon and Ecosystem Studies volume 2; *Aumont et al.*, 2015), as a step towards coastal and regional data assimilation. The work is based on recent advances in NEMO explicitly simulating the effects of model uncertainties using an "ensemble generator" (e.g. the same as a scalar time-stepping numerical model can be seen as a "state generator"), modified here specifically for stochastic parameterizations in high-resolution configurations. Our "classic" stochastic modelling approach is based on introducing stochastic degrees of freedom of forcings and model parameterizations (the model response to those stochastic degrees of freedom being the so-called "system errors" whose second-order moments are usually modelled as the Q matrix in a data assimilation context). This stochastic model undergoes validation in Parts 1 and 2 of this study.

A generic perturbation approach based on first-order autoregressive processes - AR(1) - is proposed for the coupled physical and biogeochemical models. Autoregressive processes are based on statistical models operating under the hypothesis that the past state has an effect on the present state. First-order denotes that the current value is based on the immediately preceding value. A comprehensive analysis for the stochastic formulation of NEMO is given in *Brankart* (2013) and *Brankart et al.* (2015). Some theoretical background for probabilistic ocean modelling, with technical details on implementation strategies based on NEMO (e.g. online ensemble diagnostics, connection with observation operators and data assimilation systems) is provided by *Bessières et al.* (2017).

Brankart et al. (2015) introduced two approaches for ensemble ocean modelling: the first is the stochastic perturbed parameterized tendencies (SPPT; *Buizza et al.,* 1999) and the second is the stochastic parameterization of unresolved fluctuations (SPUF). The SPPT implementation, which we use in this study, aims to generate perturbations on the models' parameterized tendencies (referred to here as the models' time derivative) and implements Monte-Carlo techniques to obtain a probability density function (*pdf*) of these tendencies. The stochastically derived parameterized tendencies are added to the models' non-parameterized tendencies (the latter assumed free of uncertainties). The SPUF implementation (not used in this study) is based on random walks sampling gradients from the state vector and adding them to the models' solution. In another study, *Ollinaho et al.* (2017) proposed a stochastic perturbed parameters (SPP) scheme, perturbing a number of model parameters whose values are presumed to contain errors. As in their work, model errors of physical parameters are introduced in this study, by applying spatiotemporally varying perturbations.

Marine biogeochemical data assimilation is increasingly being used in operational platforms as a tool to improve ocean forecasting systems. However, the subject is still immature with several challenges remaining, as for example using multivariate increments and validating non-assimilated variables in biogeochemical models. The Ensemble Kalman Filter (EnKF; *Evensen*, 2003) was first used with a simple 1D ecosystem model by *Eknes and Evensen* (2002), and later, by *Allen et al*. (2003), to control the evolution of zooplankton and nutrients by assimilating chlorophyll. *Simon and Bertino* (2009) extended the EnKF to include a Gaussian anamorphosis transformation, accounting for non-Gaussian biogeochemical distributions. *Ciavatta et al*. (2014), instead of assimilating chlorophyll, adapted the

EnKF to assimilate the SeaWiFS (https://oceancolor.gsfc.nasa.gov/SeaWiFS/) light attenuation coefficient incorporating a bio-optical model. Biogeochemical data assimilation mostly relies on satellite chlorophyll *a*. Recent studies have shed light on the assimilation of ocean colour plankton functional types, further improving marine ecosystem simulations (*Ciavatta et al*., 2016, 2018).

In light of these advances, the ocean modelling and data assimilation communities have put significant effort in improving the predictive skill of biogeochemical models (*Song et al.*, 2016; *Mattern et al.*, 2017; *Mattern et al.*, 2018; *Kaufman et al.*, 2018; *Yu et al.*, 2018; *Ford*, 2019; *Goodliff et al.*, 2019). Meanwhile, there is an increasing interest in biogeochemical model optimization (*Wang et al.*, 2020), as well as for coupled probabilistic systems and methods for their evaluation (*Candille et al.*, 2015; *Garnier et al.*, 2016; *Santana-Falcón et al.*, 2020). Errors in biogeochemical models stem from parameterizations and unresolved scales, and can be investigated by uncertainties in initial and atmospheric conditions (*Verdy and Mazloff*, 2017; *Fransner et al.*, 2020), in biogeochemical tracers and their sources and sinks (*Brankart et al.*, 2015), and model parameters (e.g. nutrient limitations, growth and mortality rates, grazing etc.; *Garnier et al.*, 2016; *Gharamti et al.*, 2017). Moreover, biogeochemical model performance is strongly dependent on ocean dynamics and on choices made in the assimilation scheme. The use of (pre-) operational biogeochemistry data assimilation is of vital importance to assess these systems and their products, and advance on the management of marine ecosystems (*Gehlen et al.*, 2015).

The most useful statistical properties of ensembles include their mean, spread, and - especially for data assimilation - their covariances. Higher order moments can also sometimes prove useful (e.g. *Quattrocchi et al.,* 2014). Where observations are available the ensemble mean can reveal biases, while the ensemble spread can measure model sensitivity to perturbed quantities used to generate the ensemble. When data are available and their uncertainties reasonably well known, consistency analyses can be performed between ensemble-based uncertainties and observational errors (*Edwards et al.*, 2015), using differences between modelled and observed variable values (the innovation vectors).

We use the above techniques to validate our ensembles, and hence the underlying stochastic model. We describe and implement consistency analysis methods to produce a quantitative assessment of the stochastic modelling of coupled physical and biogeochemical processes. Two other methods to assess the empirical consistency of ensembles will be discussed in a companion article (*Vervatis et al*., 2021). We complement this quantitative assessment with a qualitative approach, in which the stochastic model is evaluated by exploring its ability to generate consistent multivariate incremental model corrections. We illustrate this by means of multivariate representers and "stochastic" EnKF incremental analyses for a few members.

The scientific objectives of this paper focus on the generation of ensembles in high-resolution coastal and regional models. The study aims at guiding future ensemble-based modelling strategies, in support of data assimilation and probabilistic forecasting approaches. On the basis of prior knowledge in the literature for stochastic approaches and on what is feasible in terms of computational resources, we investigate physical-biogeochemical model uncertainties and what the impacts of the choice of different sets of perturbations and observational networks are on the analysed ocean state. In Section 2, we begin by describing the coupled physical-biogeochemical model and the methods to evaluate the stochastic model with respect to satellite observations. The ensembles and the model-data consistency assessment are discussed in the results section (Section 3). A discussion and concluding remarks are presented in Section 4.

2 Methods and data

2.1 The coupled physical-biogeochemical deterministic ocean model

We use the release 3.6 of the NEMO ocean model in a regional configuration encompassing the Bay of Biscay and the western part of the English Channel (Fig. 1). The configuration, named BISCAY36, is a subgrid of the operational Copernicus Marine Environment Monitoring Service (CMEMS) Iberia-Biscay-Ireland (IBI) configuration (*Sotillo et al.*, 2015) previously applied and validated (*Quattrocchi et al.,* 2014; *Vervatis et al.,* 2016). The horizontal resolution is 1/36° (about 3 km and 2 km in the meridional and zonal directions respectively). For a complete description of the original BISCAY36 configuration, and for validation details, the reader is referred to *Maraldi et al.* (2013).

In our BISCAY36 implementation, the NEMO ocean engine OPA (Océan Parallélisé) is coupled on-line with the passive tracer package TOP2 and the biogeochemical model PISCESv2. The model solves for 24 prognostic variables simulating the biogeochemical cycles of oxygen, carbon and the main nutrients controlling phytoplankton growth: nitrate, ammonium, phosphate, silicic acid and iron. The model distinguishes four plankton functional types based on size, including two phytoplankton compartments (nanophytoplankton and diatoms) and two zooplankton classes (microzooplankton and mesozooplankton). 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 biogeochemical model tracks phytoplankton biomass in units of carbon, iron, silicon (the latter only for diatoms) and chlorophyll. The ratios between carbon and the latter three variables can change, which allows for a more accurate conversion from phytoplankton to chlorophyll concentrations, which is of great importance for comparisons with ocean colour satellite data. The ratio between carbon, nitrogen and phosphorus is kept constant at the Redfield values of 122:16:1 following *Takahashi et al*. (1985). PISCESv2 also distinguishes three non-living pools for organic carbon: small Particulate Organic Matter (sPOM), big Particulate Organic Matter (bPOM; different settling velocities from sPOM) and semi-labile Dissolved Organic Carbon (DOC).

The physical and biogeochemical models are coupled on-line, using a high coupling frequency for the conservation of tracers of once every two time-steps, i.e. 150 s for physics and 300 s for biogeochemistry. The coupling is one-way, so that the ocean forcing is applied to the biogeochemical model but there is no feedback. The primitive equations and the tracer transport model are discretized on the same 1/36° curvilinear Arakawa C-grid. The TOP2 package controls the advection-diffusion equations of the passive and biogeochemical tracers. The numerical scheme for the biogeochemical processes is forward in time (Euler) and differs from the classical leap-frog scheme used for the physics. The advection scheme is the same as for the physics, i.e. QUICKEST (*Leonard*, 1979), but using the limiter of *Zalesak* (1979). These options have been tested by *Gutknecht et al.* (2016) and are now used in the IBI-MFC operational system (http://marine.copernicus.eu/).

The meteorological fields are provided by the ECMWF (European Center for Medium-Range Weather Forecasts). The initial state and open boundary conditions are acquired from the daily archives of the CMEMS infrastructure for physics, and the weekly archives for biogeochemistry. Details of the physical model set-up are described in *Vervatis et al.* (2016). The open boundaries of the biogeochemical model are forced by the global system BIOMER4V1R1 (resolution: 1/2°; http://www.mercator-ocean.fr/), which provides (via the CMEMS archives) a 3D global weekly mean analysis for dissolved iron, nitrate, phosphate, silicate, oxygen, chlorophyll and phytoplankton concentrations expressed as their carbon contents.

2.2 The stochastic model

The main issues to be faced when building a stochastic model is first to identify and select the relevant quantities to perturb (depending on the dynamics and on the objectives of the stochastic model) and secondly to generate the perturbations. This is what we propose to investigate in this study, the general context being an operational system assimilating satellite data in coastal and regional domains.

The ensembles of simulations are generated by perturbing some quantities that are considered as major sources of errors for the model. These quantities consist of forcing fields at the boundaries, parameters and state variables; these are referred to as "parameterized tendencies" (see *Buizza et al.*, 1999), in the sense that the time evolution of the model state variables can be written as follows: $\partial_t M(\boldsymbol{x},\boldsymbol{u},\boldsymbol{p},t) = D(\boldsymbol{x},t) + P(\boldsymbol{x},\boldsymbol{u},\boldsymbol{p},t)$, where $\partial_t M(\boldsymbol{x},\boldsymbol{u},\boldsymbol{p},t)$ is the time derivative of the model $M(\boldsymbol{x},\boldsymbol{u},\boldsymbol{p},t)$, as a function of the model state vector $\boldsymbol{x}$ (e.g. temperature, salinity or currents), the forcing $\boldsymbol{u}$ and the vector of the model parameters $\boldsymbol{p}$. $D(\boldsymbol{x},t)$ is the dynamics tendency of the ocean state for well-resolved, non-parameterized processes, equal to the advection term $-\nabla\cdot(\boldsymbol{xU})$, where $\boldsymbol{U}$ the vector of the Eulerian velocity, and $P(\boldsymbol{x},\boldsymbol{u},\boldsymbol{p},t)$ is the tendency of parameterized processes. The latter is perturbed based on assumptions on the amplitude and space-time structure of atmospheric forcing uncertainties, model errors in physical parameterizations and sources-minus-sinks (SMS) biogeochemical model errors (state

variables are not explicitly perturbed in our experiments); the scientific basis of these assumptions is discussed in the next sections.

In this study, the perturbations, hereafter noted ξ, are computed using the stochastic perturbed parameterized tendencies (SPPT; *Buizza et al.*, 1999) and the stochastic perturbed parameters (SPP; *Ollinaho et al.*, 2017) schemes. Assuming that the past value of the uncertainties affects the present one, the ξ are modelled as first-order auto-regressive processes. In practice, at every time-step, Gaussian AR(1) processes are generated following the expression defining:

$$\xi_{k+1} = e^{-\frac{1}{\tau}} \cdot \xi_k + \left(\sigma \cdot \sqrt{1 - e^{-\frac{2}{\tau}}}\right) \cdot w + \mu \cdot \left(1 - e^{-\frac{1}{\tau}}\right) \quad (1)$$

where k is the model time-step, w is white Gaussian noise, μ, σ and τ the mean, standard deviation (uncertainty amplitude) and correlation timescale, respectively. Here, we assume the errors are unbiased, so μ is set to zero. The perturbation fields are different for each perturbed tendency or parameter, and vary under the assumption of spatial and temporal correlated scales. Also, the fields are bounded $\xi \in [-1 \quad 1]$ to retain the sign of the tendency (*Palmer et al.*, 2009).

Brankart et al. (2015) generated independent Gaussian autoregressive processes at every model grid point and introduced spatial dependence in the ξ 2D or 3D fields, by applying a Laplacian filtering operator $\mathcal{F}[\,]$ to the white noise w in Eq. (1) as $\hat{\xi} \propto \mathcal{F}[w]$. The Laplacian filter implemented in NEMO is equivalent to the five-point smoothing operator described by *Shapiro* (1970). This is a low-pass, first-order Shapiro filter used to remove small scale grid noise. When applied to white noise on a coarse-resolution domain (typically of a few tens of kilometres), a few Laplacian passes are sufficient to introduce correlation at scales of a few hundreds of kilometres. In this context, the Laplacian-Shapiro filter acts to "blur" the white noise at every model grid point and is computationally efficient (*Brankart et al.*, 2015; *Garnier et al.*, 2016). By contrast, in high-resolution configurations of a few kilometres (e.g. BISCAY36), the Laplacian-Shapiro filter is not an optimal approach to attenuate noise and produce long-range correlations (cf. Appendix A; Fig. A1); therefore, another approach is required.

Here, long-range spatial correlations (on the order of 10 to 100 km) are explicitly calculated by applying a 2D Gaussian function to estimate iso-correlation contours with a *pdf* given by the mathematical expression:

$$p(\boldsymbol{X}|\overline{\boldsymbol{X}}, \boldsymbol{\Sigma}) \propto exp(-\tfrac{1}{2}(\boldsymbol{X} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \overline{\boldsymbol{X}})) \quad (2)$$

where $\boldsymbol{X} = (x, y)$ represents the whole model domain, $\overline{\boldsymbol{X}} = (\bar{x}, \bar{y})$ is the central grid point of the distribution (the mode) and $\boldsymbol{\Sigma} = diag[\sigma_x^2, \sigma_y^2]$ is the bivariate diagonal covariance kernel controlling the length scales of the perturbation. The spatial covariance kernel has anisotropic (important in coastal domains) e-folding length scale variances in the meridional σ_x^2 and zonal σ_y^2 directions, varying randomly around a typical correlation length per variable and across the ensemble members. In order to generate a multimodal pattern (with two or more local maxima; $m \geq 2$), we recursively implement the Gaussian function at m random central grid points $\overline{\boldsymbol{X}}^{(i)}$ in the model domain, resulting in a finite sum of Gaussian distributions $p(\boldsymbol{X}|\overline{\boldsymbol{X}}^{(m)}, \Sigma^{(m)}) \propto p\left(\boldsymbol{X}\middle|\overline{\boldsymbol{X}}^{(m)}, \boldsymbol{\Sigma}^{(m)}\right) + \sum_{i=1}^{m-1} p\left(\boldsymbol{X}\middle|\overline{\boldsymbol{X}}^{(i)}, \boldsymbol{\Sigma}^{(i)}\right)$. The recursive implementation of the function further increases the perturbation's anisotropy. The number of recursive iterations $i = 1, \ldots, m$ (and therefore the number of modes) depends on the correlation length within a given model domain. For instance, a correlation length half the size of the model domain yields approximately a bimodal distribution. The Gaussian function in Eq. (2), is used as a squared exponential (apodization) covariance function, so instead of the normalization factor $1/2\pi|\Sigma|^{1/2}$ we used the perturbation uncertainty amplitude defined by σ in Eq. (1). Stochastic perturbations are generated by substituting the Laplacian convolutions in the white noise w in Eq. (1), with a multimodal Gaussian pattern spanning the whole model domain i.e. $\mathcal{F}[w] \rightarrow p(\boldsymbol{X}|\overline{\boldsymbol{X}}^{(m)}, \boldsymbol{\Sigma}^{(m)})$. AR(1) processes estimated by Eq. (1), are performed at every model grid point shifting and altering the Gaussian modes in space and time. The implementation is compatible with the NEMO MPI double parallelization environment in the

spatial domain and across an ensemble (*Bessières et al.,* 2017). The algorithm is available from the open-source software platform Zenodo with doi:10.5281/zenodo.2556530.

2.2.1 Atmospheric forcing uncertainties

The atmospheric forcing in coastal and regional applications constitutes a major source of ocean model uncertainty and is likely to lead to large-scale biases (*Vervatis et al.*, 2016). We investigate here uncertainties in the wind velocities $U_{air} = (u_{air}, v_{air})$, the sea level pressure (SLP) and the air temperature (T_{air}), i.e. $u \in \{U_{air}, SLP, T_{air}\}$. The ECMWF atmospheric fields u are multiplied by AR(1) stochastic processes ξ (Figs. 2a-f) following the SPPT scheme:

$$u_k \rightarrow (1 + \alpha \cdot \xi_k) \cdot u_k \tag{3}$$

at every time-step k, where α is an optional tapering value in the interval [0 1] (*Buizza et al*., 1999). To calculate the AR(1) stochastic parameterization ξ using Eq. (1), we select representative values for the uncertainty amplitude σ, the average spatial correlation length $\sigma_{x,y}$ used in Eq. (2) and the correlation timescale τ for each tendency.

Output diagnostics from a one-year control run are analysed to estimate the distribution of *pdf*s (not shown) and tune the stochastic parameterizations for all atmospheric variables of interest. Wind velocity U_{air} and SLP are distributed normally, whereas the distribution T_{air} is bimodal due to the seasonal cycle. We assume that uncertainties are related to synoptic timescales, e.g. atmospheric phenomena such as storms, and we set a temporal correlation length of a few days for all atmospheric variables. The synoptic timescales are also verified by time-lagged autocorrelation methods applied to the control run. The spatial scales of the atmospheric fluctuations are determined over synoptic timescales. The signal-to-noise ratio is assigned according to the statistical laws of the *pdf*s. These stochastic parameterizations for Eq. (1) are in agreement with other studies in the literature (e.g. *Palmer et al*., 2009).

2.2.2 Model uncertainties in physical parameterizations

Fluxes of momentum, heat and mass are the key quantities linking the air and sea. The physical processes related to them are parameterized in terms of bulk coefficients, which are deduced from empirical laws, incorporating wind speed dependent coefficients and feedback from the sea state. In this study, we assume that model errors are in part due to limitations of these parameterizations of air-sea interaction. Stochastic perturbations varying spatiotemporally are imposed on the models' momentum drag c_d, latent heat c_e and sensible heat c_h coefficients. The AR(1) distribution and temporal scales are the same as those we prescribed for the wind; the spatial scales on the other hand, are assumed to be that of the ocean state and are set to a few Rossby radii of deformation. The same stochastic pattern is applied to all coefficients in the CORE bulk formulae (*Large and Yeager*, 2004). The positiveness of the coefficients is verified by tapering methods and for different stability conditions and wind speed regimes in the bulk formulae. Like Eq. (3), the SPP scheme introduces spatiotemporal perturbations following a normal distribution of AR(1) processes (other options may include simpler perturbations of white noise uncorrelated in space and time), expressed as:

$$p_k \rightarrow (1 + \alpha \cdot \xi_k) \cdot p_k \tag{4}$$

In order to represent model uncertainties for tidal mixing over the shelves, stochastic fluctuations based on flux boundary conditions are also imposed at the bottom layer. The bottom drag parameterizations c_b are based on model assumptions for the vertical shear, the mixing scheme and the nature of the seabed (rocky, sandy or muddy), when known, which modify the bottom boundary layer. Because in many ocean models the bottom drag is approximated as a permanent feature (e.g. constant minimum values in the abyssal plain as in *Maraldi et al*., 2013), large temporal scales up to one month are imposed in Eq. (1). The true nature of the seabed is unknown and consequently so are the dominant scales in the bottom layer, so we simply apply white noise and Laplacian filtering to introduce AR(1) spatial scales. The formulation of the bottom drag follows a quadratic log law, with minimum positive values clamped at 2.5×10^{-3} in the abyssal plain and maximum values observed in the shallow areas of the English Channel.

2.2.3 Biogeochemical sources-minus-sinks model uncertainties

Marine biogeochemical models encompass many sources of uncertainty stemming from unresolved model processes and poorly constrained parameterizations. These uncertainties fall into two broad categories: unresolved biodiversity and unresolved scales. The first category refers to the biodiversity restriction of only a few tens of species resolved by the model in an effort to reduce state variables (*Le Quéré et al*., 2005). This category also includes errors in the parameterization or missing biogeochemical processes, those that control the feedback between sub-systems (ecosystem, chemistry, oxygen and carbonate models; *Gehlen et al.*, 2011). These errors emerge from having a limited number of compartments, which often leads to a crude parameterization of their processes. The second category - errors due to unresolved scales - includes errors related to the model resolution and the species included in the model (e.g. lack of diurnal migration in the model), and the method for handling them is similar to that for small scale unresolved physical processes.

The SPUF scheme appears to be the most natural method to simulate uncertainties for both categories. However, one must consider that performing random walks with a large state vector could be computationally expensive. Hence, as with the other sources of error, unresolved biodiversity is here explored via the SPPT AR(1) scheme (*Brankart et al*., 2015; *Garnier et al*., 2016), introducing uncertainties in the sources-minus-sinks terms for each of the 24 biogeochemical tracers C, using the stochastic field ξ described in Eq. (1) at every time step k, as

$$SMS_k(C) \rightarrow SMS_k(C) \cdot e^{\left(\xi_k - \frac{\sigma^2}{2}\right)}. \tag{5}$$

As an example of this term, phytoplankton sources and sinks are photosynthesis, respiration, death and grazing. The biogeochemical model parameters themselves are not perturbed.

Garnier et al. (2016) used SeaWIFS at 1/12° resolution to assess the unresolved scales of a coarser 1/4° model and tune the AR(1) stochastic parameterizations. In this study, we are unable to assess the unresolved scales of the high-resolution BISCAY36 model, with respect to a dense observational network. Therefore, stochastic parameterizations are deduced from a one-year control run, mainly focusing on the statistical properties of chlorophyll. The sources-minus-sinks perturbations are assumed to have a lognormal distribution and large uncertainty amplitude σ of up to 60%. The bias correction term $\sigma^2/2$ (*Simon and Bertino*, 2009) is part of the model tuning, implying that the ensemble mean is equal to the unperturbed SMS term. The spatial scales are representative of a few Rossby radii of deformation and the correlation time-scale is set to10 days, typical for the underlying mesoscale dynamics and associated upwelling patterns. We perturb the state of the sources-minus-sinks terms in all three dimensions and across all vertical levels with the same stochastic pattern. Sensitivity experiments showed that this approach yielded uncertainty regimes with an ensemble spread increasing in time. We did test alternatives by applying different stochastic patterns per level or per tracer but these degraded the impact of the method.

2.3 Experimental design

In order to test which error sources are most significant for the biogeochemical uncertainties, and to better understand the couplings within the stochastic model, we designed three ensemble experiments: perturbing only the physics (EnsP), only the biogeochemistry (EnsB), and perturbing both simultaneously (EnsPB). In the coupled simulation, the evolution of the biogeochemical tracers is described by the advection-diffusion equation:

$$\partial_t C = \overbrace{\overbrace{-\underbrace{\nabla(u \cdot C)}_{advection} - \underbrace{K_h \nabla_h^2 C}_{\substack{horizontal\\diffusion}} + \underbrace{\partial_z(K_z\, \partial_z C)}_{\substack{vertical\\diffusion}}}^{EnsP} + \overbrace{\underbrace{SMS(C)}_{biology}}^{EnsB}}^{EnsPB} \tag{6}$$

where on the right hand of Eq. (6) the first term represents the advective transport of tracers along isopycnals, and the second and third terms the 3D parameterized diffusion processes. The last term denotes all biological processes affecting the concentration of tracers C, due to the sources-minus-sinks terms $SMS(C)$. The three experiments defined above are indicated schematically in Eq. (6).

In Table 1, we summarize our settings for the stochastic parameterizations in the sensitivity experiments perturbing individual sources of model uncertainty and the seasonal-range ensembles EnsP, EnsB, EnsPB. We also include a reference ensemble (Ens0) and a control run (CR). A detailed presentation of the Ens0 ensemble is given by *Vervatis et al*. (2016), but it suffices to say here that it differs fundamentally from the EnsP, EnsB, EnsPB ensembles in that it was generated by performing stochastic modelling of the wind forcing based on empirical orthogonal functions (EOF); there is no other variable perturbed in Ens0.

A major challenge in ensemble forecasting is to identify the most important sources of model error and properly configure the stochastic parameterizations. In order to tackle this, we adopted the following two-step procedure: firstly, we identified which error sources the model was most sensitive to in the Bay of Biscay, by running medium-range ensembles of 20 members for one month during the spring bloom, namely April, 2012 (S1-5, SP, SB, SPB experiments in Table 1). Several stochastic parameterizations, deduced from the output diagnostics of a one-year control run, were tested on a case-by-case basis for each sensitivity experiment. A final selection of perturbations was established, based on their impact on the spread of upper-ocean variables as well as on their computational efficiency (e.g. the requirement that no members blow up during the ensemble run). Based on these experiments, we decided not to perturb variables and parameters that the model in this configuration is insensitive to, such as, for example the photosynthetically active radiation coefficient k_{PAR} for the penetrative solar radiation (sensitivity experiments not shown). Results from the sensitivity experiments are discussed in Section 3.1. For the second step, we used these perturbations to run seasonal-range ensembles of 40 members each (EnsP, EnsB, EnsPB experiments in Table 1). The stochastic model skill of the seasonal-range ensembles is evaluated with respect to observations in Sections 3.2 and 3.3.

Figure 2 shows examples of AR(1) stochastic patterns drawn for several variables of the coupled system, in order to introduce model errors. We generate wind uncertainties with smaller spatial scales, varying around an average of $\sigma_{x,y}\sim 1°$ (Figs. 2a-d), compared to the work presented by *Vervatis et al.* (2016), who used in their study broader patterns based on dominant EOF modes (cf. Appendix A; Fig. A2). Figures 2a-c depict the spatial variations of the wind stochastic pattern for member-001, over a temporal correlation period of three days. In the same simulation stochastic patterns with larger spatial scales can be applied to other atmospheric variables: for example, scales of $\sigma_{x,y}\sim 2°$ and $\sigma_{x,y}\sim 3°$ have been adopted for the air temperature and sea level pressure respectively (Figs. 2e-f). Figure 2h illustrates the effect of the Laplacian filtering after 100 iterations in BISCAY36. It shows noisy spatial patterns, not representative of most oceanic processes.

In the same vein as *Vervatis et al.* (2016), a free run (i.e. the CR in Table 1) is carried out with no stochastic parameterizations. Starting from the PSY2V4R2 analysis (1/12° resolution; *Lellouche et al., 2013*; http://www.mercator-ocean.fr/) the free run is used to generate a five-month spin-up, from July to November 2011, to allow the ocean model to develop coherent structures and represent the main physical processes in the Bay of Biscay. The same period is used for the spin-up of the on-line coupled biogeochemical model. The post spin-up solution is used as the initial conditions for all other experiments. The free run is then extended from December 2011 to June 2012 and serves as a reference for the above-mentioned ensemble experiments; hence, we refer to it as the one-year control run. In Fig. 3, we present the ocean state of the control run on April 30, 2012, for a number of surface variables: sea surface height (SSH), sea surface temperature (SST), sea surface salinity (SSS), total chlorophyll, and nanophytoplankton and diatom concentrations. In supplementary material, we verify the model's skill to simulate subsurface nutrient conditions compared to world ocean atlas climatology (*Garcia et al.*, 2018).

The seasonal-range ensembles were run from December 2011 to June 2012. The computational resources for the ensemble experiments are presented in Appendix B. All sensitivity (S1-5, SP, SB, SPB) and seasonal-range (EnsP, EnsB, EnsPB) ensembles were initialized by using the ocean and the biogeochemical states of the control run, with the exact same initial conditions. No data was assimilated and the perturbation mechanism remains at work throughout the whole simulation period. A total number of 40 ensemble restarts was archived, where pseudo-random seed numbers (with different random sequences) were saved for all stochastic parameterizations. These ensemble restarts were used to ensure the reproducibility of the stochastic patterns upon initialization of the ensembles. This capability allowed

us to investigate the growth rate of model uncertainties, with the same forecast lead time, under different atmospheric and ocean states.

2.4 Datasets

We focus on upper-ocean properties for sea level anomaly (SLA), SST and surface total chlorophyll. The observational networks were accessed via the CMEMS infrastructure (http://marine.copernicus.eu/) and are summarized in Table 2. The observational errors include both measurement and representativity errors, and are usually unknown. Here, the observational errors were taken from those reported at CMEMS and averaged to create a single representative value for each product (Table 2).

High-resolution SST observations are necessary when it comes to validating ensembles at eddy-resolving scales, such as in BISCAY36 at 1/36° resolution. We use the OSTIA SST L4 gap-free gridded dataset which is a near-real-time daily-mean product of foundation SST free of diurnal variability. The data set is provided with an estimate of their uncertainty which is 0.5°C on average. The model proxy for the foundation SST has been chosen to be the temperature interpolated at 10 metres depth.

For the sea-level, we use the CMEMS L3 along-track product based on several altimetry missions and which provides SLA at a resolution of 14 km with uncertainties of 0.05 m. Both data and model include tides and both observed and simulated SLA are detided. The model includes pressure forcing, and therefore an inverted barometer correction is applied to the model and observations (a dynamic atmospheric correction is applied to the sea-level observations). In order to calculate the SLA model equivalent, we use the mean dynamic topography of the parent model IBI36 (Benkiran, pers. comm.).

Of central importance to this work is the consistency analysis of biogeochemical ensembles, where physics and biogeochemistry are both perturbed, against ocean colour products. The ocean colour method exploits different radiation wavelengths and reflectances emitted from the sea surface, affected by phytoplankton and corresponding to different water types. In this study, we use surface total chlorophyll produced for the Global Ocean in the framework of the ESA ocean colour climate change initiative (OC-CCI), made available through CMEMS. This is a merged data records product collected from multiple sensors and ocean satellite passages provided in gridded format at 4 km resolution. The ocean colour L4 product is reconstructed from L3 reprocessed daily composites applying 8-day temporal averaging to fill in missing data, though some gaps remain.

The satellite chlorophyll *a* data are provided with a scaled (%) observational error in comparison to the signal and therefore, the error has spatial distribution. We decided not to use this information because it would be harder to interpret pattern-consistency and because here we mostly focus on the investigation of model errors. Instead, we estimated a static error at 0.3 mg/m^3, which is moderately larger (smaller) in the abyssal plain (shelf areas) compared with the scaled error. The ocean colour chlorophyll proxy in models is often taken as an average of chlorophyll over the top 10% of the euphotic layer (*Prunet et al.*, 1996). In our case, estimating a euphotic layer of approximately 50 m, we decided to tune the observation operator to return a model proxy as the mean value of the first 5 m of the water column. This is in agreement also with calculations based on $1/k_{PAR}$.

2.5 Consistency analysis framework

2.5.1 Ensemble-based innovation metrics

In Section 3.2, we evaluate the stochastic model, by quantifying the ability of the ensembles at representing the model's errors with respect to observations. Our method assesses the statistical consistency of the following distributions: (a) ensemble-based uncertainties, and (b) empirically-based uncertainties for upper-ocean properties. As the system is dynamically heterogeneous, the consistency is checked locally.

The empirically-based uncertainties are represented by an ensemble of observations for each dataset. The observations are perturbed with a Gaussian random number to generate data distributions with the proper error standard deviation for each network. All observational errors are considered independent (no cross-correlations) for each dataset so that their error covariance matrix is diagonal (a useful, but not entirely true assumption). For the consistency analysis of EnsP perturbing only the physics, we used the SST and

SLA observational networks. The chlorophyll *a* product is used for the consistency check of EnsB and EnsPB biogeochemical model uncertainties. The statistical consistency between the simulated and empirical ensembles is evaluated by comparing the shape of the ensemble envelopes locally and as a function of time. For each dataset, we also compute the innovation vector, that is the model-data misfits, and compare the second-order moments: the innovation variance across the ensemble is the result of uncertainties in both the model and observations. Under the assumption that errors in the model and in observations are uncorrelated, the innovation variance should be close to the sum of the model uncertainty variance (estimated from the ensemble variance) and the observational error variance. The diagnosis proposed by *Desroziers et al.* (2005) for data assimilation is based on a similar assumption.

Denoting $y_{j,n}^{o}$ the measured value of observation n for ensemble member j, and $y_{j,n}^{f}$ the model forecast equivalent for observation n for member j, the following ensemble-based consistency statistics in data space may be calculated:

$$OmE_{bias} = \overline{y_{j,n}^{o} - y_{j,n}^{f}}^{j,n} \tag{7}$$

$$OmE = \sqrt{\overline{\left(y_{j,n}^{o} - y_{j,n}^{f}\right)^{2}}^{j,n}} \tag{8}$$

$$DOmE = \sqrt{\overline{\overline{\left(y_{j,n}^{o} - y_{j,n}^{f}\right)^{2}}^{j} - \left(\overline{y_{j,n}^{o} - y_{j,n}^{f}}^{j}\right)^{2}}^{n}} \tag{9}$$

where the j −overbar averages are taken over the ensemble members $j = 1, \dots, N$ and the n −overbar averages are taken over all observations $n = 1, \dots, p$ of a given type within a given region, and the combined j, n −overbar denotes a double average taken first over the ensemble and second over all observations of a given type within a given region. OmE stands for *Observation minus Ensemble* metric (it is an ensemble-wide extension of the so-called *Norm of the innovation vector* used in data assimilation). $DOmE$ stands for *Debiased* OmE metric defining the *mean spread of innovation*, which should be larger than the ensemble spread, and can be used even if the OmE_{bias} is not small. For an unbiased system, i.e. no biases between observations $y_{j,n}^{o}$ and model $y_{j,n}^{f}$, $\forall n$, we get $OmE_{bias} = 0$ and $DOmE = OmE$. In Appendix C, we provide a detailed description of the notation for the ensemble-based innovation metrics.

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

Let us also define the support of a probability density function *(pdf)* as the smallest closed set outside of which the *pdf* vanishes. For a *pdf* defined in $\mathbb{R}$, the *pdf* envelope is defined as the range between the minimum and maximum values of the support.

In order to illustrate the dependency of assessment results on geographic region, we focus on two distinct areas in the Bay of Biscay, namely the abyssal plain and the Armorican shelf. Both areas exhibit strong spatiotemporal variability, governed by diverse physical-biogeochemical open-ocean and coastal shelf processes, respectively.

2.5.2 Multivariate incremental analysis

The stochastic model skill is qualitatively evaluated by means of multivariate representers (defined below) and EnKF-type incremental analyses (*Evensen*, 2003). The incremental analysis corresponds to the correction step in the sequential data assimilation scheme; in other words, at a given date, we compute the correction to the controls using an estimate of the model and data errors and assuming a given value for the innovations, but then we do not restart the model from the corrected state. This is what makes it different from a data assimilation experiment. A by-product will be to illustrate the potential impact of

our ensemble-modelled uncertainties on data assimilation at analysis time. In all cases, we investigate the impact of observations on unobserved variables, such as other data types or subsurface variables. The novelty in this exercise is to showcase how stochastic modelling of different components in the coupled physical-biogeochemical system contributes to incremental analyses. We emphasize the fact that biogeochemical model errors in Eq. (6) have been generated in simulation experiments of increasing complexity, perturbing separately the physics and biogeochemistry. The calculations are conducted with SDAP (Sequoia Data Assimilation Platform, https://sourceforge.net/projects/sequoia-dap/), whose functions were expanded to interface with the NEMO platform and its biogeochemical component PISCESv2.

Let us define $\boldsymbol{H}$ to be the observational operator and $\boldsymbol{P}^f$ the error covariance matrix of the prior controls. In this study, the controls are the following state variables: SSH, T, S, Chl. In an ensemble-based context, the prior error covariance matrix is approximated by N samples (i.e. members) in a decomposed form as $\boldsymbol{P}^f = \boldsymbol{S}^f \boldsymbol{S}^{f^T}$, where

$$\boldsymbol{S}^f = \frac{1}{\sqrt{N-1}} \prod_{j=1}^{N} \left(\boldsymbol{x}_j^f - \overline{\boldsymbol{x}^f} \right) \tag{10}$$

is the square root matrix of the error-subspace, with ensemble mean

$$\overline{\boldsymbol{x}^f} = \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{x}_j^f , \tag{11}$$

$\boldsymbol{x}^f$ being the control vector.

The representers are the influence functions of single observations: the correction added to the forecast during the analysis step can be written as a linear combination of representers weighted by the innovation (*Bennett et al.*, 1996). The matrix of representers is $\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T$. In Section 3.3, we analyze zero-lag representers, i.e. where observations and controls are at the same time (operator $\boldsymbol{H}$ has no time component).

The corrections are computed from the Kalman gain matrix

$$\boldsymbol{K} = (\varrho \circ \boldsymbol{P}^f)\, \boldsymbol{H}^T \, [\, \boldsymbol{H}\, (\varrho \circ \boldsymbol{P}^f)\, \boldsymbol{H}^T + \boldsymbol{R}\,]^{-1} \tag{12}$$

multiplied by the member-dependent innovation vector $\boldsymbol{d} = \boldsymbol{y}^o - H(\boldsymbol{x}^f)$, where $\boldsymbol{y}^o$ denotes the (member-independent) observation vector. The symbol $\circ$ denotes the Schur (element-wise) product of two matrices and ϱ is the localization function (*Gaspari and Cohn*, 1999). Here, unlike the analysis step of a "stochastic" EnKF (and from the previous Section 2.5.1), we chose not to perturb the observations (i.e. as in a "deterministic" EnKF, cf. *Sakov and Oke*, 2008; *Yu et al.*, 2018). There is a trade-off - using unperturbed observations does lead to the loss of statistical consistency for the second-order moments and forbids calculating innovation-based diagnostics, but properly calculated EnKF increments for individual members would be harder to interpret physically because of the observation noise. Convolution with a localization function in Eq. (12) constrains spurious long-distance correlations resulting from the small ensemble size. We used a localization Gaussian function with cut-off radius 3° and e-folding scale 0.2°. Vertical localization is not applied to make results as "raw" and easy to interpret as possible.

In Section 3.3, we compute and analyse the multivariate EnKF-type incremental analyses $\boldsymbol{K} \cdot \boldsymbol{d}$ for the first two members (001/002) of the three long-range ensembles (EnsP, EnsB and EnsPB), as well as for different subsets and seasons, using multiple observations together. Thinning techniques are applied to high density SST and ocean colour L4 observational networks (*Zuo et al.*, 2019), in order to reduce the computational time and optimize data assimilation performance.

We define specific criteria to assess the skill of the multivariate representers and the incremental analyses, in the context of different perturbation methods (i.e. Ens0 and EnsP, EnsB, EnsPB). First, we examine whether the increment correction is large constraining the model ensemble. A subsequent criterion is whether the increment (regardless of its magnitude) has a physical-biogeochemical meaning attributed to specific dynamical processes in the area, showing also potential to constrain the ensemble. Additional criteria are used to investigate the consistency of the increments (a) when calculated from a smaller number of members (testing whether we have convergence of covariances), (b) for different periods (e.g.

flow dependent errors relevant to different seasons), (c) when assimilating different datasets (e.g. SST and chlorophyll *a*) and (d) their impact on unobserved variables (e.g. SSS).

3 Results

3.1 Ensemble sensitivity experiments

To assess error patterns prior to the seasonal-range ensemble simulations, we ran a series of medium-range ensemble simulations. These small ensembles are useful for tuning the stochastic parameterizations and to verify that their spreads are moderate compared to the seasonal-range ensembles with longer spin-up period.

In Fig. 4, we present maps for April 30, 2012, of model uncertainties for all medium-range ensembles, S1-5, SP, SB and SPB, after a one-month spin-up. If we compare the effect on SST of wind perturbations alone (Fig. 4a) to all perturbations together (Fig. 4f) we see little difference; likewise, for the SSH (Figs. 4g, l). This implies that error regimes for SST and SSH are mainly driven by the wind forcing. Wind uncertainties have a large impact on upper-ocean uncertainties for both the geostrophic and the Ekman components, defined by Sverdrup dynamics. Imposing the same perturbation field for both u and v wind velocities, i.e. not changing the wind azimuth, results in similar uncertainties for the vorticity and Ekman pumping, further enhancing model errors.

The rest of the perturbed variables locally augment the ensemble spread in filament-like patterns in the periphery of eddies, near river plumes and over the shelf slope due to energy trapping (cf. Fig. 4 with respect to Fig. 3). After the wind, air temperature uncertainties have the next greatest impact on SST (Fig. 4b) and chlorophyll (Fig. 5b). Uncertainties in the wind drag coefficient have moderate impact on the wind stress and consequently on the spread of surface variables. Compared to the spread associated with uncertainties in the wind forcing they are approximately an order of magnitude smaller (Figs. 4d, j). This is because the expression for the wind drag coefficient c_d has a correction dependency based on the different wind speed regimes (i.e. discrete c_d values that the perturbation may be ineffective to change the class of these values). The ocean response to SLP forcing has two components: the first-order Inverse Barometer (IB), which is isostatic and dominant at large scales, and the second-order non-IB, which depends mostly on the geographic region. The IB response to SLP perturbations has spatial scales that are equal to or larger than the external Rossby radius and the IB pumping on the abyssal plain (*Tai,* 1993). Subtracting the IB response from the SSH leaves us with the non-IB, which in our case has a limited impact on SST and SSH (Figs. 4c, i). Uncertainties in the bottom drag coefficient amplify error regimes, mostly for SST and less for SSH, along the shelf break and on the shelves, and especially on the macrotidal area of the English Channel dynamically controlled by strong tidal currents and fronts (Figs. 4e, k).

Biogeochemical uncertainties arise from the ecosystem model's intrinsic errors and from errors in the physical state variables (Fig. 5). All sensitivity experiments perturbing the physics leave an imprint on chlorophyll uncertainties, which are sometimes significant compared to the values of the control run and to other model variables, like for instance the SSH and SST. When perturbing only the biogeochemical model, we implement an identical stochastic pattern across all variables and vertical levels in order to increase the impact of the method. Since the physics are not perturbed, the biogeochemical uncertainties are passively advected via the ocean circulation. In all cases, physical model errors are found to have a larger impact on chlorophyll spread than those of the biogeochemical perturbations. It is straightforward to see why: if the ocean physics are identical for all members (as is the case for the SB ensemble) it behaves as a strong dynamical attractor for the system; vertical nutrient fluxes are of course sensitive to the velocity field and in particular its divergence, the latter scaling over the continental shelf break (*Karagiorgos et al.*, 2020), a process which might dominate other sources of uncertainty, such as those of the SMS of the biogeochemical model. In general, the spread is largest in SPB where both physics and biogeochemical perturbations are applied simultaneously. Biogeochemical uncertainties resulting from errors in the coupled system depend on the geographic location as well. An example is the chlorophyll uncertainties in the English Channel compared to the mesoscale field in the abyssal plain.

There is a stark geographic difference in spread between nanophytoplankton chlorophyll and diatom chlorophyll (Figs. 5i-n). This is expected from the different model parameters of the phytoplankton

classes and the additional requirements in nutrient supply of Si for the primary production of diatoms. Larger uncertainties are observed for the nanophytoplankton, especially in the open-ocean, whilst diatoms have larger uncertainties over the shelves and in the English Channel, where errors are dynamically controlled by tides. It is evident that the uncertainty regimes for each class follow the different size class chlorophyll abundance patterns in the region (cf. Fig. 5 with respect to Fig. 3), especially during the spring bloom.

Interestingly, even though model uncertainties are in general larger when both physical and biogeochemical models are perturbed, a small decrease in chlorophyll spread may be observed over the abyssal plain in the presence of coherent eddies. This effect is attributed to two facts: (i) the ensembles are not constrained by data assimilation and therefore model errors with long forecast lead time from different periods, may partially overlap or even cancel each other with "flow-dependent" errors (also called "errors of the day"); (ii) sub-mesoscale physical and biogeochemical processes with length scales smaller than 10 km can be resolved in the abyssal plain (*Levy et al.*, 2012a; 2012b; *Charria et al.*, 2017), therefore not contributing to the mesoscale error patterns being different in EnsB and EnsPB. This does not change the fact that the dominant spatial scales in the abyssal plain are characteristic of the underlying quasigeostrophic mesoscale features, due to the mesoscale vertical velocity field.

3.2 Quantitative evaluation of stochastic model skill

3.2.1 SST L4 gap-free gridded observations

Figure 6 shows examples of consistency metrics, as defined in Section 2.5.1, based on daily innovation samples. An ensemble of OSTIA SST is compared with EnsP (Figs. 6a, c). The EnsPB ensemble is statistically identical, since there is no feedback from the biogeochemical model into the circulation model. The calculations are carried out through the whole period of the seasonal-range ensemble. For the SP ensemble in Figs. 6b, d, calculations are performed for April 2012.

The model and data ensembles in both regions are consistent with each other since the supports of the *pdfs* overlap (Figs. 6a, c). Assimilating these observations, with these error estimates, with an ensemble or Bayesian filter would probably be well-posed, at least in data space (additional criteria to the *pdf* supports may need to be verified). In both regions, the model error estimate is lower than the innovation spread $DOmE$, and the ensemble mean innovation vector OmE_{bias} is mostly contained within the observational error interval ± 0.5 °C (Figs. 6b, d), showing that the ensemble is consistent. The spread indicator slowly evolves in time, mainly increasing, consistently with the fact that the perturbation mechanism remains active throughout the period. However, there are occasions where the EnsP spread is reduced and the ensemble envelope is under-dispersive and sometimes biased with respect to the observational *pdf*. This occurs over a short period during the spring shoaling of the thermocline and mainly over the Armorican continental margin. This model overconfidence could be associated with missing re-stratification processes in our stochastic protocol, such as the vertical subgrid scale physics like the vertical eddy viscosity and diffusivity coefficients calculated in the turbulent closure scheme, which are not perturbed.

An interesting feature when comparing the seasonal-range EnsP and medium-range SP ensembles, is that the SST model errors over the abyssal plain increase with similar rates given a forecast lead time of a few days (Fig. 6d), despite the fact that they have different numbers of members and initial conditions (i.e. December 2011 for EnsP and April 2012 for SP). The stochastic restarts in these two experiments are identical and therefore the stochastic fields applied in the perturbed variables and parameters are alike. However, after this initial period of a few days' forecast lead time the model errors appear to increase with different rates, pertaining to the different ocean-atmosphere states during December for EnsP and April for SP. The latter is more apparent in the Armorican margins area (Fig. 6b).

In Figs. 6b, d, we compare the EnsP SST spread with the reference ensemble Ens0 (*Vervatis et al.*, 2016). Both stochastic approaches in Ens0 and EnsP ensembles can be considered as variants of the perturbed tendency scheme. In the case of the AR(1) processes we have the option to perturb several variables in EnsP, compared to the stochastic modelling of only the wind forcing in Ens0. Therefore, the former yields a larger SST spread during winter and at the end of the run, with their differences ranging between about 0.05 °C to 0.1 °C, and with larger differences observed in the open-ocean. At the end of the run, both

ensembles become biased with respect to the observations (more so over the shelves; cf. also *Vervatis et al.,* 2016 for Ens0 and Figs. 6b, d for EnsP), since they are not constrained by data assimilation. During spring, in the presence of a strong thermocline, the spread is reduced in both ensembles and regions (though more over the shelf), suggesting that there are missing error processes in both stochastic protocols (i.e. in addition to the wind for Ens0 and all variables perturbed in EnsP).

3.2.2 SLA L3 along-track observations

The data space consistency analysis for EnsP SSH is carried out with the sea-level L3 along-track product in Table 2. Figure 7a shows the distributions corresponding to averages of all SLA along track observations crossing the BISCAY36 domain, as well as the EnsP and Ens0 supports of *pdfs* in data space. Both model ensembles appear to have more energy at the weekly timescale, occasionally at shorter timescales, and overall at the seasonal timescale compared to the observations. The minimum levels of spread in both model and data are reached in mid-February, 2012; for the rest of the simulation the SLA model equivalent shows larger variability than the data. There is a notable difference in spread between the EnsP and Ens0 ensembles, with the former's being greater possibly because of the SLA response to SLP perturbations, as well as the different stochastic approaches in the wind perturbations. EnsP model uncertainties are comparable with those of the observational error at 0.05 m, reaching a few cm in magnitude.

We also find that in about one-third of cases the OmE_{bias} stays within the observational error interval (Fig. 7b). The most consistent configuration is over the abyssal plain, where the spread is increased due to mesoscale decorrelation of eddies after spin-up. In contrast, the largest statistical biases are observed in coastal regions and in the English Channel. This is also confirmed by the box-and-whisker plots that are plotted for several regions (Figs. 7c-e). These serve as a means to visualize both the distributions and their consistency. There are many cases where the joint uncertainty associated with both sources of information is clearly nonzero. However, several instances of strong bias are evident in two of the regions, sometimes to the point that the supports of the *pdf*s become disjoint, e.g. during the week between 06 to 12 March, 2012 (Figs. 7c-e). If those *pdf*s are used in an Ensemble filter, disjoint supports will lead to ill-posed assimilation problems, meaning that a solution will be obtained since all analysis schemes are convex, but the result will be meaningless.

When inconsistencies occur, it is generally because other error processes are active in the model in addition to the ones generated by the range of EnsP perturbations. This is more apparent in Ens0, since in this particular ensemble there are no perturbations in SLP. Both ensembles appear to be occasionally under-dispersive. In the English Channel, there are missing error processes such as the residual tidal error, which is enhanced by the presence of local tidal fronts and occasional Kelvin waves propagating along the coasts, and the non-isostatic response to atmospheric pressure (e.g. the non-IB response). The missing error sources are also particularly evident in the English Channel, hinting at the presence of high-frequency errors currently unaccounted for, although we cannot currently identify them with the tools at hand. Another possible reason for the observed statistical inconsistency is the low coverage of altimeter observations. This is exacerbated by the larger errors which altimetric data are subject to in coastal areas, which leads to less available data there. Future wide-swath altimetry products (e.g. SWOT, https://swot.jpl.nasa.gov/) are likely to provide better coverage in coastal regions and stronger constraints on models overall.

3.2.3 Ocean colour L4 gridded observations

Figure 8 shows results of biogeochemical ensembles and innovation statistics in the data space assuming an observational error of 0.3 mg/m^3. Innovation statistics are calculated in log space applying an anamorphosis function to transform chlorophyll lognormal distributions into Gaussian distributions. The lognormal assumption for chlorophyll *a* has shown advantages in data assimilation systems (*Song et al.*, 2012). Among the seasonal-range ensembles the EnsPB exhibits the largest chlorophyll spread, with EnsB being the least dispersive, as seen in the medium-range ensembles (Fig. 5). The model ensemble in the Armorican shelf shows a strong bias with respect to the ocean colour data, with disjoint *pdf*s supports that may lead to an ill-posed assimilation problem if those error estimates were used in a data assimilation system. On the other hand, in Fig. 8c, the EnsPB model-data samples appear to be marginally compatible

with each other over the abyssal plain (also for EnsP and EnsB, not shown). There are also periods with disjoint *pdf*s supports with the most prominent in late March. During this period, observations show a strong phytoplankton bloom in the open-ocean, weakly present in all ensemble members (but with the correct phase), resulting in disjoint *pdf*s. Subsequent, less intense blooms in the open-ocean just before and after this event are also found in both model and data samples, and in those cases the *pdf*s overlap partially. Just as with the previously discussed SST and SLA networks, assimilating these ocean colour observations with these error estimates in the open-ocean would probably be marginally well-posed, in the sense that the joint probability associated with both sources of information appears to be nonzero only at times. The most consistent configuration for EnsPB after the spin-up period appears to be May-June and only for the abyssal plain.

In Figs. 8b, d, we illustrate the innovation metrics defined in Section 2.5.1. All biogeochemical model ensembles are under-dispersive during the first three months of the simulation, associated also with the fact that primary production is low in winter. After this period dispersion slowly increases over time. Towards the end of the run chlorophyll uncertainties in the Armorican shelf exceed 0.15 mg/m^3 (Fig. 8b). These model errors are small compared with the chosen observational error and with ocean colour data measured by satellite, often yielding a model-data misfit at about ~1 mg/m^3. However, considering that chlorophyll has a lognormal distribution with near-zero positive values, they are not small with respect to the model's overall chlorophyll concentration of about 0.3 mg/m^3. In the abyssal plain, the model-data ensemble mean misfit expressed by the OmE_{bias} metric, is well defined within the observational error interval for most of the period, except during the phytoplankton bloom observed in spring (Fig. 8d). In general, the control run underestimates chlorophyll abundance compared with ocean colour data and subsequently leads to under-dispersive biogeochemical model ensembles. Inconsistency is mostly attributed to the low bias. An examination of the temporal evolution of the spread resulting from different stochastic protocols, reveals that chlorophyll uncertainty variations are mainly controlled by physical processes and their errors. Biogeochemical processes and their uncertainties have only a moderate impact on model errors, one reason could be the unaccounted errors in nutrient initial conditions (cf. supplementary material using the world ocean atlas nutrient climatology; *Garcia et al.*, 2018), except during periods of phytoplankton blooms where both components of the coupled system are important. In line with this, biogeochemical model errors are higher during periods of high biological productivity, than in periods of low biological productivity (e.g. comparison of EnsPB and SPB in Figs. 8b, d).

Examining the mean model spread for all ensembles against the *mean spread of the innovation* $DOmE$, the former is always contained within the latter (Figs. 8b, d). However, this is not sufficient to verify consistency. This is because there is a constant bias of greater amplitude than the model dispersion, especially in coastal regions.

3.3 Qualitative assessment of stochastic model skill through its impact on multivariate incremental analysis

3.3.1 Ensemble-based single observation representers

In Figs. 9a-f, we illustrate examples of zero-lag representers for three individual observations in different locations on May 07, 2012. In order to address the qualitative criteria in Section 2.5.2, the representers are calculated from 40 members of EnsP (Figs. 9a-d), also from a 20-member subset of EnsP (Fig. 9e) and finally from 40 members of Ens0 (Fig. 9f), and are shown as correlations between SST and other surface variables of the control vector. The correlation structures reveal differences between the abyssal plain and coastal areas, as well as between variables. In the English Channel and over the south Armorican shelf, the filament-shaped structures for SST, SSS and surface chlorophyll are linked to near-shore features, such as river discharges (e.g. Loire river plume), mid-shelf thermal fronts and tidal fronts. The SSS pattern is dipolar, which could be explained by meridional plume displacement (Fig. 9b). Spring bloom is suggested by the negative correlation between SST and chlorophyll over the shelf (Fig. 9d). As the surface layers heat up during spring plankton depletion follows the bloom.

SST and SSH are largely decorrelated in the domain, probably because of the large-scale atmospheric forcings directly influencing SST in the spring, as well as low-frequency mesoscale variability (Fig. 9c). Due to the mixed conditions over the inner shelf at that time of year, the SSH response is relatively large-

scale as it is associated with barotropic processes at the scale of the external Rossby radius. In the case of smaller ensembles, i.e. 20 members (Fig. 9e) or under-dispersive ensembles, i.e. SSH in Ens0 (Fig. 9f), correlations due to partially-converged statistics increasingly contaminate the pattern of representers. These patterns are however not very different from those of larger or more dispersive ensembles, at least in their general shapes and signs, and mostly the correlations or negative correlations are amplified. Similar results of rather broad and symmetrical structures are also found in the case of SSH single observation representers (not shown). Model errors associated with SST and SSH, at that time of the year and for those ensembles, may have a limited impact on the ensemble model skill.

Zero-lag representers of single-observation ocean colour data are calculated from EnsPB. The most important finding is that correlations behave differently in the open-ocean and over the shelves (Figs. 9g-i). Chlorophyll autocorrelation structures appear broad and symmetrical (with respect to the single-observation location) in the open-ocean, with scales dictated by the vertical velocity field of the underlying mesoscale features. This is in agreement with the biogeochemical model error regimes for the medium-range ensembles depicted in Fig. 5. Over the shelves, chlorophyll correlations are more dipolar in nature, when calculated for instance with respect to the SST and SSS fields, representing smaller-scale local conditions. Filament-shaped structures of negative correlations between chlorophyll and SST are seen again with the EnsPB stochastic protocol, indicating that model errors in primary production are mainly controlled by model uncertainties in physical processes, implying once again the possibility to improve ecosystem model skill performance by perturbing the physics alone.

3.3.2 Incremental analysis using SST L4

In this section, we qualitatively assess the data assimilation impact of the stochastic model in correcting the surface variables at analysis time. In Figs. 10a, b, we illustrate the SST correction of the first two members 001 and 002 on May 07, 2012, using the prior error covariance matrix $\boldsymbol{P}^f$ (in a decomposed form of the square root matrix $\boldsymbol{S}^f$ as in Eq.(10)) from all 40 members of EnsP and assimilating the OSTIA SST dataset. SST corrections are consistent between members, suggesting that they are meaningful and that they could enhance model skill. The increments reveal a large-scale north-south pattern, which is consistent with the dynamics of the region and the model's cold bias with respect to the observations for this period over the Irish shelf and the English Channel. They also correct mesoscale processes over the abyssal plain and shelf processes near the river plumes. An interesting feature is the correction of the Bay of Biscay sub-gyre located between 4°-6°W and 44°-46°N, confirmed by consistent increments in SSH, SSS and especially in surface total chlorophyll, this last hinting at sub-gyre scale changes in the vertical velocity (Figs. 10c-e). Increments of opposite signs for SST and chlorophyll indicate that physical processes in the Bay of Biscay, such as tidal mixing, slope currents, river plumes and open-ocean mesoscale activity, play an important role in the biological productivity of the area. Nonzero chlorophyll corrections, calculated by cross-covariances of the coupled system, offer potential for improving ecosystem model skill, even if only physical data are assimilated.

In order to qualitatively evaluate the impact of EnsP on the 3D temperature, salinity and chlorophyll model update, we examine the increment profiles for the first member 001 at two specific locations over the abyssal plain and the Armorican shelf (Fig. 11). The increments are able to capture the region's spatiotemporal variability over two different seasons and in areas with different characteristics, suggesting that model skill performance depends on flow-dependent errors in the physics and biogeochemistry. The correction profiles reflect a fairly deep mixed layer in winter and a shallower mixed layer in spring, with chlorophyll changes at sub-surface layers hinting at ongoing bloom-related changes. The increments are associated with the shallow thermocline depth during spring (~10-30 m) and with the depth of the euphotic layer (~40-50 m) influencing the sub-surface vertical corrections of chlorophyll. Over the abyssal plain, at depths greater than 1000 m the vertical T/S corrections are possibly linked to the upper-ocean low-frequency mesoscale circulation, which affects the deep vortex dynamics. This is an intriguing result showing that deep model errors can be controlled by ensemble-based data assimilation methods, in which model ensembles are generated by perturbing surface variables and assimilating data of near-surface ocean properties.

3.3.3 Incremental analysis using ocean colour L4

In Fig. 12, we see that most of the total chlorophyll correction on May 07, 2012, as seen from the increments of the first member, arises from uncertainties in the physics (i.e. EnsP and EnsPB), though biogeochemical model uncertainties in most areas in the domain are still significant. We focus on that specific date, because of the good spatial data coverage (ensembles are projected in data space in Figs. 12d-f), in order the analysis to be meaningful. When the physics are not perturbed (i.e. EnsB) one can see coherent mesoscale features in the correction (e.g. over the abyssal plain), since the physics and hence the dynamics are the same for all members. At 15 m depth, a signature of the ocean bottom relief can be seen in the correction fields, perpendicularly to the seabed features (resulting in a parallel crest signal) over the Celtic shelf near 7°-8°W and 48°-49°N. Biogeochemical processes and consequently ecosystem model performance are sensitive to bottom Ekman flow transport due to nutrient upwelling, especially in shallow shelf regions, and possibly to internal tides and waves induced by barotropic tides and winds.

High positive increments over the shelves for all three ensembles suggest that chlorophyll abundance is underestimated in those areas during subsequent spring blooms (Figs. 12a-c). This holds true over the continental shelf break near 3°-4°W and 46°-47°N, an area dynamically controlled by both barotropic and baroclinic tidal processes, which contribute to vertical mixing and enhance primary production. It is worth noting that in some coastal areas where the coupled system appears to underestimate chlorophyll abundance, a very small correction is applied instead, due to model overconfidence (i.e. under-dispersive ensembles lacking model skill) with respect to observational errors. In order to qualitatively assess the ensemble model skill, we may compare the chlorophyll increments contributed by different observational networks. Figure 10e and Fig. 12a present different correction patterns when different observations are assimilated, hinting at dissimilar processes captured by the prior error covariance matrices in combination with each network, in these cases by SST or ocean colour respectively. Here the chlorophyll correction is rather small when ocean colour is not assimilated, possibly because of weak cross-covariances between ocean physics and biogeochemical properties. Interestingly, we observe different signs in the increments, e.g. over the Celtic shelf, as calculated from the different model ensembles of the coupled system (Figs. 12a-c), and in particular between EnsP perturbing only physics and EnsB perturbing only biogeochemistry sources and sinks. In the present configuration, we conclude that the assimilation of ocean colour is important, and the specific stochastic method less so, when we seek to enhance model performance in marine ecosystems.

3.3.4 Incremental analysis using both SST L4 and ocean colour L4

In this section, we qualitatively assess the multivariate impact of the stochastic model in both temperature and total chlorophyll, from two different observational networks assimilated simultaneously, namely the OSTIA SST L4 and the ocean colour L4 (cf. Table 2). The changes that the extra observation network brings to the analyses vary from moderate to locally large for all variables. In Figs. 13a-d, we depict the correction fields for May 07, 2012 of the state vector surface variables for the first member, based on ensemble covariances of EnsPB with 40 members.

The north-south SST correction pattern is not as distinct as that when assimilating only OSTIA SST, especially in the English Channel area (Fig. 13a vs. Fig. 10a). In addition, the increments are greater, especially over the Irish shelf. This last aspect is also true for other surface variables, such as the SSH and SSS (Figs. 13b-c vs. Figs. 10c-d). We attribute this effect to the presence of ocean colour data in a situation where error *pdf*s might be disjoint, even when the data assimilation problem is convex, which suggests weak ensemble model skill. In these situations, the analysis scheme possibly impacts biogeochemical properties more than the ocean physics do. Similarly, chlorophyll correction values are moderately decreased when both observational networks are assimilated, compared to the analysis when assimilating only ocean colour (Fig. 13d vs. Fig. 12c). Compared to the correction of physical properties, the chlorophyll increments seem to be less sensitive to changes when multivariate observing networks are brought together. Thus, the ensemble ecosystem model skill appears to be less sensitive compared to the ocean model skill when ocean colour is assimilated, along with other more traditional networks like SST.

We investigated the convergence of covariances and its impact on the incremental analysis to assess model performance, incorporating different ensembles and ensemble sizes (i.e. how effective are the

ensembles EnsP vs. EnsPB and 10 vs. 40 members at extracting reliable covariances; Figs. 13a, d, e-h). Differences in the SST analyses between EnsPB and EnsP with 40 members in a multivariate context are minor (Fig. 13a vs. Fig. 13e). If we use fewer ensemble members, the analyses for both ocean physics and biogeochemistry properties resemble the correction patterns of larger ensembles. However, the increments are noticeably less smooth because covariances are calculated from partially converged statistics (Figs. 13e-f vs. Figs. 13g-h). As a final remark, we note that when chlorophyll is assimilated in conjunction with OSTIA SST, the analysis scheme moderately affects small scale processes for all variables, augmenting the increment values around coherent ocean dynamical features; this is suggestive of good stochastic model skill.

4 Discussion and conclusions

In this paper, Part 1 of a two-part series, we evaluated the skill of a NEMO-based physical-biogeochemical stochastic model. To build this stochastic model, existing NEMO stochastic modules were complemented by integrating a 2D Gaussian function to introduce spatial correlations in the stochastic perturbations. *Brankart et al.* (2015) generated correlations by applying a spatial filter in global and regional coarse domains. Here, the correlations have variable and anisotropic spatial scales, which is of particular importance in high-resolution coastal configurations. Another option would have been to use a tensor form that would make those patterns coast-aware (*Barth et al*., 2009). We decided not to follow that route here, since this would have added significant complexity while, we felt, not being critical for the specific BISCAY36 configuration (e.g. there are not many islands and islets) and for the variables we chose to perturb (e.g. atmospheric forcing). Recursive filters and Gaussian functions have been used in the literature to calculate correlation lengths, but most of them focused on large spatial scales in coarse global domains and few of them have been used recently for biogeochemistry (*Storto et al.,* 2014; *Kuhn et al*., 2019).

Our stochastic implementation was based on the SPPT and SPP schemes, and first-order autoregressive processes (*Brankart et al.,* 2015; *Ollinaho et al.,* 2017), applied to several sources of model uncertainties in the coupled system. These uncertainties emerge mainly from atmospheric forcing uncertainties, model errors in physical parameterizations and biogeochemical model sources and sink term uncertainties. Wind uncertainties were found to dominate all other atmosphere-ocean sources of model errors. Biogeochemical model uncertainties resulting from perturbations in physics were greater than those perturbing the SMS concentration of the biogeochemical tracers.

Sensitivity studies of the stochastic model, e.g. with respect to (1) initial and boundary conditions (biases in this stochastic context), (2) parameter lognormal perturbations (*Ollinaho et al.,* 2017), and (3) perturbations of the unresolved scales, would be an ulterior step and are not included in this study.

The quantitative assessment of ensembles with respect to the gridded gap-free OSTIA SST L4, assuming an observational Gaussian error of 0.5 °C standard deviation, suggests that the seasonal-range ensembles where only the physics were perturbed are fairly consistent with the data distribution. Analysing the consistency of the ensembles with respect to the along-track SLA L3 CMEMS product (observational Gaussian error of 0.05 m standard deviation), we could see the presence of strong biases between the model and along-track data distributions. All ensembles were under-dispersive for sea-level, especially in coastal regions. The sea-level model-data misfits were found to be associated with strong SSH spatial gradients, in particular in the shelf regions such as the English Channel and Celtic Sea. High-frequency error processes are currently unaccounted for in the English Channel (e.g. open boundary conditions in the English Channel, which do not deal properly with high-frequency processes). Probabilistic attribution approaches (see e.g. *Hannart et al.,* 2016), in addition to the quantitative assessment presented in this study, may provide more insight.

Our consistency analysis in log-space showed a statistical spin-up time for chlorophyll on the order of 3 months. During this spin-up period the biogeochemical model ensembles EnsP, EnsB and EnsPB were under-dispersive and biased with respect to gridded ocean colour L4 data to which a log-transformed, observational Gaussian error of 0.3 mg/m^3 standard deviation was applied. Statistical consistency was not always maintained for chlorophyll as it was for SST and, to a lesser extent, for SSH. It is difficult to attribute those error patterns to specific physical or biogeochemical processes, without a further

probabilistic attribution analysis. The investigation of uncertainty regimes for other state variables of PISCESv2 is beyond the scope of this study, however we mention in passing that the observed error patterns in upper-ocean ecosystem properties, for instance for zooplankton and nutrients, are similar to those of chlorophyll (not shown).

The low chlorophyll bias may be a sign that the biogeochemical model's initial conditions for our ensembles (after the 5-month control run spin-up) are incorrect (cf. supplementary material using the world ocean atlas nutrient climatology; *Garcia et al.*, 2018). Perturbing the physical forcing and the biogeochemical SMS terms cannot properly account for uncertainties in the initial biogeochemical state. These initial uncertainties could lead to unrealistic adjustments in the properties that are being perturbed.

Finally, we used the stochastic model output to generate multivariate increments, by means of representers and EnKF-type analyses with the aim to provide a qualitative assessment step towards developing an assimilation scheme. One objective was to assess the impact of observations on unobserved variables, such as other data types or subsurface variables. Specific qualitative criteria were used to assess the stochastic model skill at analysis time.

Based on these qualitative criteria, we summarize here the most important findings in the incremental analysis, arranging them into three broad categories. (i) Concerning the physics alone: corrections to physical properties are associated with large-scale biases between the open-ocean and the shelves; small scale local corrections are mainly visible over the shelves in near-shore coastal areas, which may be explained by meridional river plume migration, mid-shelf thermal fronts, barotropic tides and possibly baroclinic tides. (ii) Concerning the vertical profiles: incremental analysis on the water column structure reveals vertical changes linked to seasonal variability of the thermocline, such as for example the extended mixed layer depth in winter and the spring shoaling of the thermocline; increments are also observed at depths near the euphotic layer (~40-50 m) controlling the sub-surface vertical corrections of chlorophyll. (iii) Concerning the interplay between the physics and the biogeochemical model: the scales of the correction patterns in the abyssal plain are often characteristic of the underlying quasigeostrophic mesoscale features associated with vortices; most of the chlorophyll correction arises from uncertainties in the physics, but biogeochemical model errors tend to enhance this effect; as expected, assimilating chlorophyll has a measurable impact on physical variables, e.g. when assimilating both ocean colour and SST datasets, the increments are larger than when assimilating SST only.

Results from the two assessment approaches in Sections 3.2 and 3.3 are analysed together in an attempt to evaluate the model skill in physics and biogeochemistry. The skill of the perturbation method at properly estimating model errors is in general improved for physics, though data assimilation performance to correct these errors depends strongly on the multivariate analysis. The biogeochemical model is not as sensitive as the physical model to the multivariate analysis and its skill is mainly defined by the assimilation of chlorophyll. The biogeochemical model errors appear to be complementary to those of the dynamical model alone and they contribute moderately to the incremental analysis. However, both increments derived from stochastic physics and biogeochemistry have marked differences in their spatial patterns. In some specific areas, the stochastic perturbations in physics and biogeochemistry lead to an under-dispersive spread. This is a weakness of the stochastic model which results in a limited impact of data assimilation in these areas, especially on the biogeochemical variables.

As our ensembles were sometimes found to be under-dispersive, for future efforts additional approaches should be envisaged to augment ensemble spread, in particular for biogeochemical variables; to list just a few: (a) inflation techniques (*Hamill et al.*, 2001; *Anderson*, 2009) in the initial conditions may open some degrees of freedom in the first few time-steps of the model run, (b) perturbing the biogeochemical parameters (see also *Garnier et al*., 2016) in addition to SMS concentrations, (c) activation of the feedback of biology onto the physics in the NEMO-PISCESv2 coupled system, and (d) incorporating atmospheric ensembles, such as the ECMWF Ensemble Prediction System (https://apps.ecmwf.int/archive-catalogue/).

Uncertainties in the open boundary conditions, either in the numerical scheme or in the prescribed values for active boundaries, are expected to contribute significantly to the model error budget (see also *Kim et al*., 2011). So, it is natural to consider perturbing the open boundary conditions. This would very likely

lead to an increase in the ensemble spread from the surface to at least 1500 m. However, perturbing the open boundaries is not straightforward: first because of the need to ensure physical consistency between the perturbed variables, and second because the errors in the prescribed fields at the open boundaries are usually unknown. A favorable solution is when an ensemble of nesting (or parent) solutions is available and provides an estimate of open boundary uncertainties to the child model (*Ghantous et al.*, 2020). However, in the absence of a larger-scale ensemble to force the boundaries, a boundary error would appear to the stochastic model as a systematic error (bias) across all members, which was beyond the scope of this paper, but will be important to address in the future.

The success of physical-biogeochemical data assimilation will also depend on different observational networks brought together, e.g. remote sensing in synergy with in-situ observations, and their multivariate nature (*Song et al.*, 2016; *Verdy and Mazloff*, 2017; *Mattern et al.*, 2017; *Mattern et al.*, 2018; *Kaufman et al.*, 2018; *Ford*, 2019; *Goodliff et al.*, 2019). Regarding in-situ observations, the increasing availability of high-resolution profiles provided by Biogeochemical-Argo floats (http://biogeochemical-argo.org), enables us to investigate the ocean interior across a wide range of spatiotemporal scales. The development of the Biogeochemical-Argo network is expected to substantially improve the quality of biogeochemical models, by incorporating vigorous validation, improved parameterizations (*Wang et al.*, 2020), and novel data assimilation techniques, including ensemble-based methods (*Yu et al.*, 2018).

Acknowledgments. This work was carried out as part of the Copernicus Marine Environment Monitoring Service (CMEMS) "Stochastic Coastal/Regional Uncertainty Modelling (SCRUM)" Service Evolution project. CMEMS is implemented by Mercator Ocean International in the framework of a delegation agreement with the European Union. Part of this research was also made possible through the IKY Scholarships Programme and co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the action entitled "Reinforcement of Postdoctoral Researchers", in the framework of the Operational Programme "Human Resources Development Programme, Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) 2014-2020. The contribution of P. De Mey-Frémaux and N. Ayoub is supported by the Centre National de la Recherche Scientifique (CNRS). We acknowledge the use of the ECMWF's computing and archive facilities in this research. This work was also supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility – ARIS – under project ID PA002007. We thank four anonymous reviewers for their constructive comments.

Appendix A: Spatial correlation length scales

The response function of the first-order Shapiro filter (*Shapiro*, 1970) measures the skill of the filter to attenuate noise and is given by the equation $R^m = cos^{2\cdot m}(\pi \cdot dx/L)$, where dx the mesh grid resolution and L the correlation length scale produced after repeated iterations m. Figure A1 shows the performance of the Laplacian-Shapiro filter to attenuate noise for two different mesh grid resolutions. Figure A1a refers to a coarse-resolution domain at $1/4°$ (e.g. ORCA025, dx~25 km) and Fig. A1b refers to a high-resolution domain at $1/36°$ (e.g. BISCAY36, dx~2.5 km). For the coarse-resolution domain the Shapiro filter is able to attenuate noise efficiently (e.g. in most cases more than 50%) and produce long-range correlations of a few hundreds of kilometres. For the high-resolution domain the Shapiro filter may introduce length scales of a few tens of kilometres, but the noise is not attenuated for the large scales (cf. also Fig. 2h).

Figure A2 shows the first and second order statistical moments of two different methods for performing stochastic modelling of the wind forcing. In this study, we introduced wind uncertainties by means of SPPT AR(1) processes and a 2D Gaussian function with length scales $\sigma_{x,y}$~1° (Fig. A2a). In a previous study, we used EOF modes to generate wind forcing uncertainties (*Vervatis et al.,* 2016). The most important attribute of both methods is that they have similar ensemble means (and their ensemble means are also similar to the unperturbed wind; not shown), whilst there are marked differences in their ensemble spreads.

Journal Pre-proof

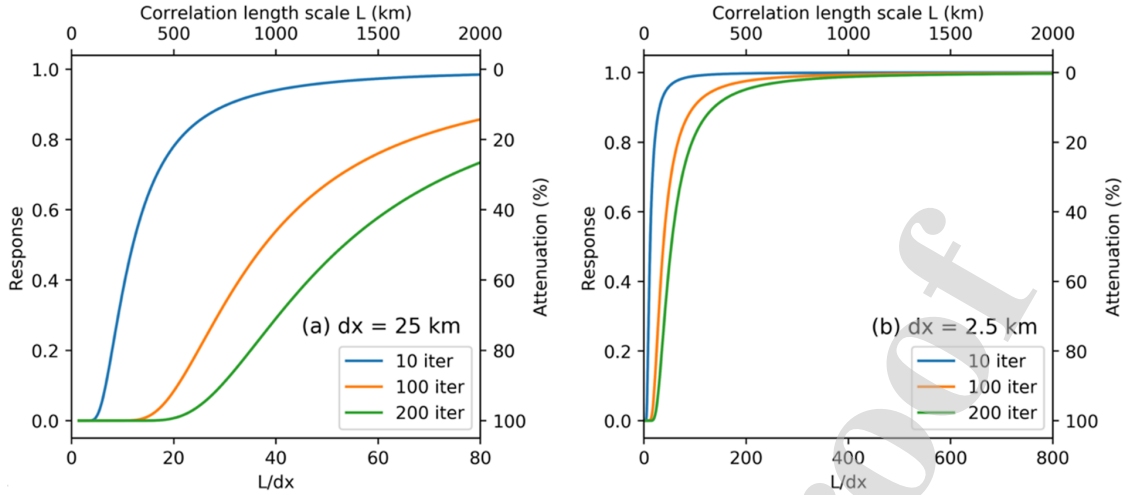


Figure A1 Response function R^m (no units) and noise attenuation (%) of the Laplacian-Shapiro filter for different number of iterations m, as a function of the correlation length scale L (km) and for different model resolutions: (a) dx~25 km, (b) dx~2.5 km.

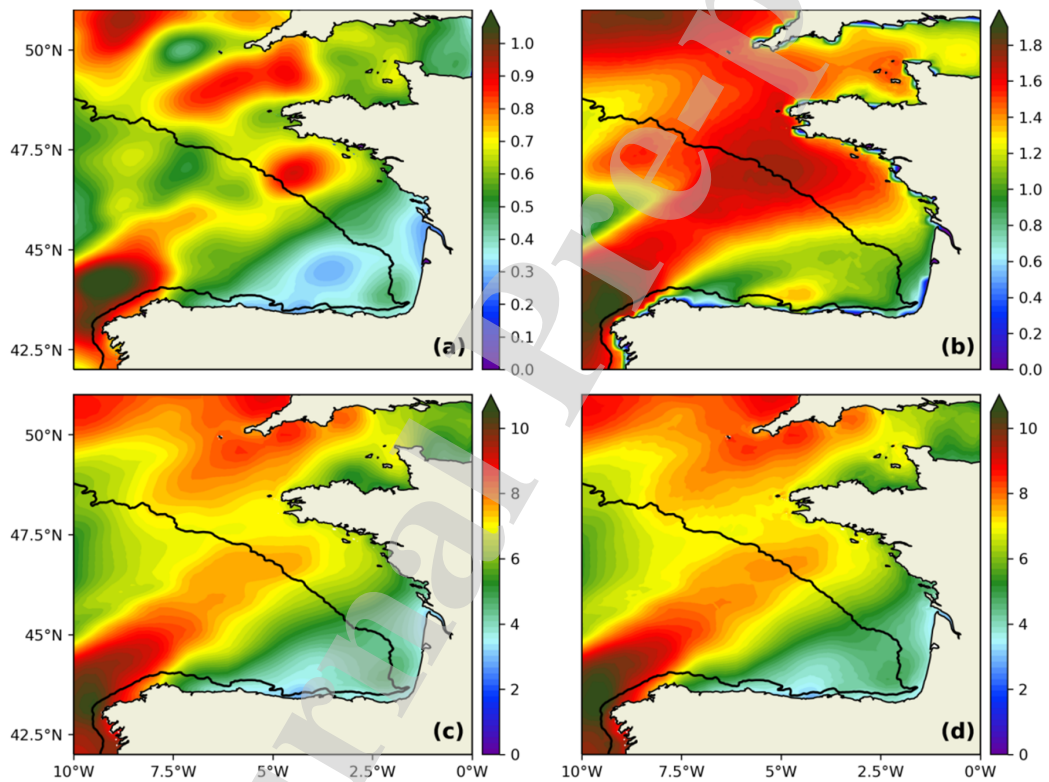


Figure A2 (a, b) EnsP and Ens0 wind ensemble spread, and (c, d) wind ensemble mean on May 07, 2012. Stochastic modelling in EnsP is by means of SPPT AR(1) using a 2D Gaussian function with length scales $\sigma_{x,y}$~1° and in Ens0 by using EOF modes (*Vervatis et al.,* 2016). Colorbar units in m/s.

Appendix B: Computational resources and performance

We briefly present here the most recent setup compiling and running the code at ECMWF HPCF. The model ensembles were generated on CCA and CCB clusters, which are Cray XC40 systems integrating Intel Broadwell nodes, with 36 cores per node and 128 GB (2400 MHz DDR4) memory per node. The code was compiled under the Intel Broadwell software environment using the Cray Development Toolkit (CDT) cdt/17.03, with intel/17.0.3.053 compiler, and the following libraries: cray-netcdf-hdf5parallel/4.4.1.1 and cray-hdf5-parallel/1.10.0.1. The same environment was used for the compilation of XIOS version 2.0. We used -O3 optimization in the FCFLAGS of the compilation architecture file.

The model output consists of daily files of the ocean state vector and the two classes of chlorophyll, as well as three-day averages of 14 3D-biogeochemical variables.

We made used of NEMO's enhanced MPI strategy whose features allow for parallelization in both the spatial domain and across ensemble members (*Bessières et al.*, 2017). Ensemble simulations were then carried out by just one call of the executable of the coupled NEMO-PISCESv2 system. BISCAY36 scales out using 96 processor cores of domain decomposition per ensemble member, excluding land processors. The configuration used the NEMO I/O and was connected to an external server (i.e. XIOS controlled by an XML file), thus increasing the total number of processors for the ensemble simulations, including those handling the I/O specifications (i.e. model variables, domains, grid, output frequencies etc.). Our ensemble experiments were designed to fit the scalability limits encountered in current operational systems. *Vervatis et al*. (2016) showed that with increasing ensemble size, results for BISCAY36 converge in the range of 20 to 40 members. We used the same range of members resulting in a scalability problem of the order of 10^3 cores.

Taking into account the ECMWF's hardware and software specifications, we tested the following resource geometries: (a) for 10 members, we used 960 NEMO processors and 48 XIOS servers filling a total of 28 nodes, and (b) for 20 members, we used 1920 NEMO processors and 24 XIOS servers filling a total of 54 nodes. The ensemble simulations were submitted as batch jobs for a 30-day run. For these examples, the ECMWF's job epilogue during production indicated a runtime average of about 489 minutes, with runtime standard deviation of approximately 29 minutes, including the first and last reading and writing time-steps.

Appendix C: Notation for ensemble-based innovation metrics

Let us note $Y_n = \overline{y_{j,n}^o - y_{j,n}^f}^{\,j}$ the ensemble mean of the innovation for observation n as:

$$Y_n = \frac{1}{N}\sum_{j=1}^{N}\left(y_{j,n}^o - y_{j,n}^f\right) \tag{C1}$$

where $j = 1, \dots, N$ the size of the ensemble and $n = 1, \dots, p$ the number of observations. With this notation, OmE_{bias} in Eq. (7) writes:

$$OmE_{bias} = \frac{1}{p}\sum_{n=1}^{p} Y_n \tag{C2}$$

as the sum over p observations of the ensemble mean of the innovation. From Eq. (8) the square of OmE writes:

$$OmE^2 = \frac{1}{p}\sum_{n=1}^{p}\left[\frac{1}{N}\sum_{j=1}^{N}(y_{j,n}^o - y_{j,n}^f)^2\right] \tag{C3}$$

We now take the square of $DOmE$ from Eq. (9) using the Y_n notation in (C1) as:

$$DOmE^2 = \frac{1}{p}\sum_{n=1}^{p}\left[\frac{1}{N}\sum_{j=1}^{N}(y_{j,n}^o - y_{j,n}^f)^2 - Y_n^2\right] \tag{C4}$$

This can be written:

$$DOmE^2 = \frac{1}{p}\sum_{n=1}^{p}\left[\frac{1}{N}\sum_{j=1}^{N}(y_{j,n}^o - y_{j,n}^f)^2\right] - \frac{1}{p}\sum_{n=1}^{p} Y_n^2 \tag{C5}$$

and using Eq. (C3) we get:

$$DOmE^2 = OmE^2 - \frac{1}{p}\sum_{n=1}^{p} Y_n^2 \tag{C6}$$

If $OmE_{bias} = 0$ then from Eq. (C2) we get $\frac{1}{p}\sum_{n=1}^{p} Y_n = 0$. In general, this yields $\frac{1}{p}\sum_{n=1}^{p} {Y_n}^2 \neq 0$ and subsequently $DOmE \neq OmE$ as seen from Eq. (C6), unless we have an unbiased system, i.e. no biases between model and observations with $Y_n = 0, \forall n$.

References

Allen, J. I., Eknes, M., and Evensen, G., 2003. An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. Annales Geophysicae, 21, 399-411.

Anderson, J.L., 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus A, 61: 72-83. doi:10.1111/j.1600-0870.2008.00361.x.

Adler, R., P. Müller and B.L. Rozovskii, 1996. Stochastic Modelling in Physical Oceanography. Progress in Probability series, vol. 39. Birkhaüser, 466pp.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M., 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. Geosci. Model Dev. 8, 2465-2513.

Barth, A., Alvera-Azcárate, A., Beckers, J.-M., Weisberg, R. H., Vandenbulcke, L., Lenartz, F., and Rixen, M., 2009. Dynamically constrained ensemble perturbations-application to tides on the West Florida Shelf, Ocean Sci., 5, 259-270, https://doi.org/10.5194/os-5-259-2009.

Bennett, A.F., Chua B.S., and Leslie, L.M., 1996. Generalized inversion of a global numerical weather prediction model. Meteor. Atmos. Physics, 60, 165-178.

Bessières, L., Leroux, S., Brankart, J.M., Molines, J.M., Moine, M.P., Bouttier, P.A., Penduff, T., Terray, L., Barnier, B., and Sérazin, G., 2017. Development of a probabilistic ocean modelling system based on NEMO 3.5: Application at eddying resolution. Geosci. Model Dev., 10, 1091–1106.

Brankart, J.-M., 2013. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. Ocean Modell., 66, 64–76, http://dx.doi.org/10.1016/j.ocemod.2013.02.004.

Brankart, J.-M., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P.-A., Brasseur, P., and Verron, J., 2015. A generic approach to explicit simulation of uncertainty in the NEMO ocean model. Geosci. Model Dev. 8, 1285-1297.

Buizza, R., Miller, M., and Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quart. J. Roy. Meteor. Soc., 125, 2887-2908, doi:10.1002/qj.49712556006.

Candille, G., Brankart, J.-M., and Brasseur, P., 2015. Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation. Ocean Science 11, 425–438.

Carton, X., Le Cann, B., Serpette, A., and Dubert, J., 2013. Interactions of surface and deep anticyclonic eddies in the Bay of Biscay, J. of Mar. Sys., 109-110, doi:10.1016/j.jmarsys. 2011.09.014.

Charria, G., Theetten, S., Vandermeirsch, F., Yelekçi, Ö., and Audiffren, N., 2017. Interannual evolution of (sub)mesoscale dynamics in the Bay of Biscay, Ocean Sci., 13, 777–797, https://doi.org/10.5194/os-13-777-2017.

Ciavatta, S., Brewin, R. J. W., Sk_akala, J., Polimene, L., de Mora, L., Artioli, Y., and Allen, J.I., 2018. Assimilation of oceancolor plankton functional types to improve marine ecosystem simulations. Journal of Geophysical Research: Oceans, 123, 834–854. https://doi.org/10.1002/2017JC013490.

Ciavatta, S., S. Kay, S. Saux-Picart, M. Butenschön, and Allen, J.I., 2016. Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem, J. Geophys. Res. Oceans, 121, 1824–1845, doi:10.1002/2015JC011496.

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., and Allen, J.I., 2014. Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. Progress in Oceanography, 127, 74-95.

De Mey, P., 2000. Scientific Feasibility of Data Assimilation in MERCATOR. MERCATOR report, reference MOO-ST-431-8-MER, Sept. 2000.

Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc. 131, 3385–3396. https://doi.org/10.1256/qj.05.108.

Edwards, C. A., A. M. Moore, I. Hoteit, and Cornuelle, B. D., 2015. Regional Ocean Data Assimilation. Annual Review of Marine Science, 7:1, 21-42.

Eknes, M., and Evensen, G., 2002. An ensemble Kalman filter with a 1-D marine ecosystem model. Journal of Marine Systems, 36(1), 75-100.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343-367.

Ford, D., 2019. Assessing the role and consistency of satellite observation products in global physical-biogeochemical ocean reanalysis. Ocean Sci. Discuss. 1–25. https://doi.org/10.5194/os-2019-118.

Fransner, F., Counillon, F., Bethke, I., Tjiputra, J., Samuelsen, A., Nummelin, A., Olsen, A., 2020. Ocean Biogeochemical Predictions—Initialization and Limits of Predictability. Front. Mar. Sci. 7. https://doi.org/10.3389/fmars.2020.00386.

Fujii Y, Rémy E, Zuo H, Oke P, Halliwell G, Gasparin F, Benkiran M, Loose N, Cummings J, Xie J, Xue Y, Masuda S, Smith GC, Balmaseda M, Germineaud C, Lea DJ, Larnicol G, Bertino L, Bonaduce A, Brasseur P, Donlon C, Heimbach P, Kim Y, Kourafalou V, Le Traon P-Y, Martin M, Paturi S, Tranchant B and Usui N, 2019. Observing System Evaluation Based on Ocean Data Assimilation and Prediction Systems: On-Going Challenges and a Future Vision for Designing and Supporting Ocean Observational Networks. Front. Mar. Sci. 6:417. doi: 10.3389/fmars.2019.00417.

Garcia, H. E., K. Weathers, C. R. Paver, I. Smolyar, T. P. Boyer, R. A. Locarnini, M. M. Zweng, A. V. Mishonov, O. K. Baranova, D. Seidov, and J. R. Reagan, 2018. World Ocean Atlas 2018, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). A. Mishonov Technical Ed.; NOAA Atlas NESDIS 84, 35 pp.

Garnier, F., Brankart, J.M., Brasseur, P., and Cosme, E., 2016. Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data. Journal of Marine Systems, 155, 59–72.

Gaspari, G., and S. E. Cohn, 1999. Construction of correlation functions in two and three dimensions. Quart. J. Roy. Meteor. Soc., 125, 723–757.

Gehlen, M., Gruber, N., Gangstø, R., Bopp, L., and Oschlies, A., 2011. Biogeochemical consequences of ocean acidification and feedbacks to the earth system, Ocean Acidification, 1, 230–248.

Ghantous, M., Ayoub, N., De Mey-Frémaux, P., Vervatis, V., and Marsaleix, P., 2020. Ensemble downscaling of a regional ocean model. Ocean Modell., 145, https://doi.org/10.1016/j.ocemod.2019.101511.

Gharamti, M. E., J. Tjiputra, I. Bethke, A. Samuelsen, I. Skjelvan, M. Bentsen, and L. Bertino, 2017. Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites, Ocean Modelling, 112, 65-89, https://doi.org/10.1016/j.ocemod.2017.02.006.

Goodliff, M., Bruening, T., Schwichtenberg, F., Li, X., Lindenthal, A., Lorkowski, I., Nerger, L., 2019. Temperature assimilation into a coastal ocean-biogeochemical model: assessment of weakly and strongly coupled data assimilation. Ocean Dyn. https://doi.org/10.1007/s10236-019-01299-7.

Gutknecht, E., Reffray, G., Gehlen, M., Triyulianti, I., Berlianty, D., and Gaspar, P., 2016. Evaluation of an operational ocean model configuration at 1/12 spatial resolution for the Indonesian seas (NEMO2.3/INDO12)–Part 2: Biogeochemistry, Geosci. Model Dev., 9, 1523–1543, doi:10.5194/gmdd-9-1523-2016.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001. Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. Mon. Wea. Rev., 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Hannart, A., Carrassi, A., Bocquet, M., Ghil, M., Naveau, P., Pulido, M., et al., 2016. DADA: data assimilation for the detection and attribution of weather and climate-related events. Climatic Change, Springer Verlag, 136 (2), pp.155-174.

Karagiorgos, J., V. Vervatis, and S. Sofianos, 2020. The Impact of Tides on the Bay of Biscay Dynamics. J. Mar. Sci. Eng., 8, 617.

Kaufman, D. E., Friedrichs, M. A. M., Hemmings, J. C. P., and Smith Jr., W. O., 2018. Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea, Biogeosciences, 15, 73–90, https://doi.org/10.5194/bg-15-73-2018.

Kim, S., Samelson, R.M., and Snyder, C., 2011. Toward an uncertainty budget for a coastal ocean model, Mon. Wea. Rev., 139, doi: 10.1175/2010MWR3352.1.

Kourafalou, V.H., P. De Mey, M. Le Hénaff, G. Charria, C.A. Edwards, R. He, M. Herzfeld, A. Pascual, E.V. Stanev, J. Tintoré, N. Usui, A.J. van der Westhuysen, J. Wilkin and X. Zhu, 2015. Coastal Ocean Forecasting: system integration and evaluation, Journal of Operational Oceanography, 8:sup1,s127-s146, DOI: 10.1080/1755876X.2015.1022336.

Kourafalou, V.H., P. De Mey, J. Staneva, N. Ayoub, A. Barth, Y. Chao, M. Cirano, J. Fiechter, M. Herzfeld, A. Kurapov, A.M. Moore, P. Oddo, J. Pullen, A. van der Westhuysen & R.H. Weisberg, 2015. Coastal Ocean Forecasting: science foundation and user benefits, Journal of Operational Oceanography, 8:sup1, s147-s167, DOI: 10.1080/1755876X.2015.1022348.

Koutsikopoulos, C. and Le Cann, B., 1996. Physical processes and hydrological structures related to the Bay of Biscay anchovy, Scientia Marina., 60 (Supl.2), 9-19.

Kuhn, A. M., Dutkiewicz, S., Jahn, O., Clayton, S., Rynearson, T. A., Mazloff, M. R., and Barton, A. D. (2019). Temporal and spatial scales of correlation in marine phytoplankton communities. Journal of Geophysical Research: Oceans, 124, 9417–9438, https://doi.org/10.1029/2019JC015331.

Large, W. and Yeager, S., 2004. Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies, CGD Division of the National Center for Atmospheric Research, NCAR Technical Note: NCAR/TN-460+STR.

Lévy, M., D. Iovino, L. Resplandy, P. Klein, G. Madec, A.-M. Tréguier, S. Masson, and K. Takahashi, 2012a. Large-scale impacts of submesoscale dynamics on phytoplankton: Local and remote effects, Ocean Modelling, Vol. 43-44, pp. 77-93, https://doi.org/10.1016/j.ocemod.2011.12.003.

Lévy, M., L. Resplandy, P. Klein, X. Capet, D. Iovino, and C. Ethé, 2012b. Grid degradation of submesoscale resolving ocean models: Benefits for offline passive tracer transport, Ocean Modelling, Vol. 48, pp. 1-9, https://doi.org/10.1016/j.ocemod.2012.02.004.

Lellouche, J.-M., Le Galloudec, O., Drévillon, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.-E., Bricaud, C., Bourdallé-Badie, R., Tranchant, B., Benkiran, M., Drillet, Y., Daudin, A., and De Nicola, C., 2013. Evaluation of global monitoring and forecasting systems at Mercator Océan, Ocean Sci., 9, 57-81, https://doi.org/10.5194/os-9-57-2013.

Leonard, B.P., 1979. A stable and accurate convective modelling procedure based on quadratic upstream interpolation. Computer Methods in Applied Mechanics and Engineering, 19, 59-98.

Le Quéré, C., Harrison, S.P., Prentice, I.C., Buitenhuis, E.T., Aumont, O., Bopp, L., Claustre, H., Da Cunha, L.C., Geider, R., Giraud, X., Klaas, C., Kohfeld, K.E., Legendre, L., Manizza, M., Platt, T., Rivkin, R.B., Sathyendranath, S., Uitz, J., Watson, A.J., and Wolf-Gladrow, D., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Glob. Change Biol., 11, 2016–2040.

Longhurst, A.R., 1998. Ecological Geography of the Sea, 398 pp. Academic, San Diego, Calif.

Lucas M., N. Ayoub, B. Barnier, T. Penduff and De Mey, P., 2008. Stochastic study of the temperature response of the upper ocean to uncertainties in the atmospheric forcing in an Atlantic OGCM, Ocean Modell., Vol.20, 1, 90-113, doi: 10.1016/j.ocemod.2007.07.006.

Madec, G., 2012. Nemo ocean engine, Tech. rep., NEMO team.

Maraldi, C., B. Levier, N. Ayoub, J. Chanut, P. De Mey, F. Lyard, G. Reffray, S. Cailleau, M. Drévillon, E. A. Fanjul, M. Garcia Sotillo, P. Marsaleix, and the Mercator Team, 2013. NEMO on the shelf: Assessment of the Iberia–Biscay–Ireland configuration. Ocean Science 2013/9, 745-771.

Mattern, J.P., Edwards, C.A., and Moore, A.M., 2018. Improving Variational Data Assimilation through Background and Observation Error Adjustments. Mon. Wea. Rev., 146, 485–501, https://doi.org/10.1175/MWR-D-17-0263.1.

Mattern, J.P., Song, H., Edwards, C.A., Moore, A.M., Fiechter, J., 2017. Data assimilation of physical and chlorophyll a observations in the California Current System using two biogeochemical models. Ocean Model. 109, 55–71. https://doi.org/10.1016/j.ocemod.2016.12.002.

Melsom, A., Counillon, F., LaCasce, J. H. and Bertino, L., 2012. Forecasting search areas using ensemble ocean circulation modeling, Ocean Dyn., 62(8), 1245–1257, doi:10.1007/s10236-012-0561-5.

Oke, P. R., and P. Sakov, 2008. Representation Error of Oceanic Observations for Data Assimilation. J. Atmos. Oceanic Technol., 25, 1004–1017, https://doi.org/10.1175/2007JTECHO558.1.

Ollinaho, P., S. Lock, M. Leutbecher, P. Bechtold, A. Beljaars, A. Bozzo, R. M. Forbes, T. Haiden, R. J. Hogan and Sandu, I., 2017. Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. Q.J.R. Meteorol. Soc., 143:408-422, doi:10.1002/qj.2931.

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer, and Weisheimer, A., 2009. Stochastic Parameterization and Model Uncertainty, ECMWF, pp. 1-42 Technical Memorandum No. 598.

Penduff, T., B. Barnier, L. Terray, G. Sérazin, S. Gregorio, J.-M. Brankart, M.-P. Moine, J.-M. Molines, and Brasseur, P., 2014. Ensembles of eddying ocean simulations for climate, CLIVAR Exchanges No. 65, Vol. 19, No. 2.

Pinardi, N., L. Cavaleri, G. Coppini, P. De Mey, C. Fratianni, J. Huthnance, P.F.J. Lermusiaux, A. Navarra, R. Preller, and Tibaldi, S., 2017. From weather to ocean predictions: an historical viewpoint. Journal of Marine Research, 75, 103-159.

Prunet, P., Minster, J.-F., Ruiz-Pino, D., and Dadou, I., 1996. Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean: 1. Method and preliminary results, Global Biogeochem. Cycles, 10(1), 111– 138, doi:10.1029/95GB03436.

Quattrocchi, G., P. De Mey, N. Ayoub, V. Vervatis, C.-E. Testut, G. Reffray, J. Chanut and Drillet, Y., 2014. Characterisation of errors of a regional model of the Bay of Biscay in response to wind uncertainties: a first step toward a data assimilation system suitable for coastal sea domains. Journal of Operational Oceanography, Volume 7, Number 2, August 2014, pp. 25-34(10).

Sakov, P., and Oke, P. R., 2008. A deterministic formulation of the ensemble Kalman Filter: an alternative to ensemble square root filters. Tellus A 60, 361–371, http://dx.doi.org/10.1111/j.1600-0870.2007.00299.x.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R. and Korablev, A., 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Sci., 8(4), 633–656, doi:10.5194/os-8-633-2012.

Santana-Falcón, Y., Brasseur, P., Brankart, J. M., and Garnier, F., 2020. Assimilation of chlorophyll data into a stochastic ensemble simulation for the North Atlantic ocean, Ocean Sci. Discuss., https://doi.org/10.5194/os-2020-6, accepted.

Shapiro, R., 1970. Smoothing, filtering, and boundary effects, Rev. Geophys., 8(2), 359-387, doi:10.1029/RG008i002p00359.

Simon, E., and Bertino, L., 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. Ocean Sci. 5, 495-510.

Song, H., Edwards, C.A., Moore, A.M., Fiechter, J., 2012. Incremental four-dimensional variational data assimilation of positive-definite oceanic variables using a logarithm transformation, Ocean Modell., 54-55, 1-17, https://doi.org/10.1016/j.ocemod.2012.06.001.

Song, H., Edwards, C.A., Moore, A.M., Fiechter, J., 2016. Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations. Ocean Model. 106, 159–172. https://doi.org/10.1016/j.ocemod.2016.06.005.

Sotillo, M. G., S. Cailleau, P. Lorente, B. Levier, R. Aznar, G. Reffray, A. Amo-Baladrón, J. Chanut, M. Benkiran, and Alvarez-Fanjul, E., 2015. The MyOcean IBI Ocean Forecast and Reanalysis Systems: operational products and roadmap to the future Copernicus Service, Journal of Operational Oceanography, 8:1, 63-79, DOI: 10.1080/1755876X.2015.1014663.

Storto, A., S. Masina, and S. Dobricic, 2014. Estimation and impact of nonuniform horizontal correlation length scales for global ocean physical analyses. Journal of Atmospheric and Oceanic Technology, Volume 31, pp. 2330-2349, doi: 10.1175/JTECH-D-14-00042.1.

Takahashi, T., Broecker, W. S., and Langer, S., 1985. Redfield ratio based on chemical data from isopycnal surfaces, J. Geophys. Res., 90, 6907–6924.

Tai, C.K., 1993. On the Quasigeostrophic oceanic response to atmospheric pressure forcing: the Inverted Barometer pumping. NOAA Techinacal Memorandum NOS OES 005.

Thacker, W.C., A. Srinivasan, M. Iskandarani, O.M. Knio, and Le Hénaff, M., 2012. Propagating boundary uncertainties using polynomial expansions, Ocean Modelling, Vol. 43–44, pp. 52-63, https://doi.org/10.1016/j.ocemod.2011.11.011.

Vandenbulcke, L., and Barth, A., 2015. A stochastic operational forecasting system of the Black Sea: Technique and validation. Ocean Modell., 93, 7-21.

Vandenbulcke, L. and Barth, A., 2019. Upscaling of a local models into a larger-scale models, Ocean Sci., 15, 291-305, https://doi.org/10.5194/os-15-291-2019.

Verdy, A. and M. R. Mazloff, 2017. A data assimilating model for estimating Southern Ocean biogeochemistry, J. Geophys. Res. Oceans, 122, 6968-6988, doi:10.1002/ 2016JC012650.

Vervatis, V. D., C.E. Testut, P. De Mey, N. Ayoub, J. Chanut, and Quattrocchi, G., 2016. Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4D EnOI based on stochastic modelling of the wind forcing. Ocean Modell., 100, 1-19, http://dx.doi.org/10.1016/j.ocemod.2016.01.003.

Vervatis, V. D., P. De Mey-Frémaux, N. Ayoub, J. Karagiorgos, S. Ciavatta, R. Brewin, and Sofianos. S., 2021. Assessment of a regional physical-biogeochemical stochastic ocean model. Part 2: empirical consistency. Ocean Modell., under review.

Wang, B., Fennel, K., Yu, L., and Gordon, C., 2020. Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. Biogeosciences, 17, 4059-4074, https://doi.org/10.5194/bg-17-4059-2020.

Yu, L., Fennel, K., Bertino, L., Gharamti, M.E., Thompson, K.R., 2018. Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling. Ocean Model. 126, 13–28. https://doi.org/10.1016/j.ocemod.2018.04.005.

Zalesak, S.T., 1979. Fully multidimensional flux corrected transport algorithms for fluids. J. Comput. Phys., 31.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M., 2019. The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Sci., 15, 779–808, https://doi.org/10.5194/os-15-779-2019.

Table 1. Table of simulations: control run (CR), ensemble medium-range sensitivity experiments (S1-5, SP, SB, SPB) and seasonal-range ensembles (EnsP, EnsB, EnsPB). The table shows the stochastic parameterizations of first-order autoregressive processes AR(1), deduced from the statistical properties of the one-year CR. Spatial correlations are explicitly calculated by applying a 2D Gaussian function with variable anisotropic length scales (1std: $\sigma_{x,y}$ in degrees), temporal correlations (τ in days) and uncertainty

amplitude (1std: σ no units) of the 2D Gaussian distributions (prior to the anamorphosis function if applied). The spatial correlation value for the bottom drag coefficient c_b is an approximation after 100 passes of the Laplacian operator. Ens0 is a seasonal-range ensemble performing stochastic modelling of the wind forcing based on EOF modes (*Vervatis et al*., 2016).

experiment	perturbed variables	uncertainty amplitude (σ)	correlation timescales (τ)	spatial scales ($\sigma_{x,y}$)	distribution
CR	one-year unperturbed free run (July 2011-June 2012)				
Ens0	wind perturbations based on EOF modes (*Vervatis et al.,* 2016)				
	atmospheric forcing				
S1	U_{air}	0.3	3 days	1°	Gaussian
S2	T_{air}	0.1	15 days	2°	Gaussian
S3	SLP	0.01	5 days	3°	Gaussian
	ocean model parameterizations				
S4	c_d, c_e, c_h	0.1	3 days	0.5°	Gaussian
S5	c_b	0.2	30 days	0.2°	Laplace flt*
	synthesis of ocean-atmosphere model uncertainties				
SP	S1-5	medium range ens (April 2012; 20 mem)			
EnsP	SP	seasonal range ens (December 2011–June 2012; 40 mem)			
	biogeochemical state				
	ocean-atmosphere state identical to the CR for all members				
		0.6	10 days	0.5°	Lognormal**
SB	$SMS(C)$	medium range ens (April 2012; 20 mem)			
EnsB	SB	seasonal range ens (December 2011–June 2012; 40 mem)			
	synthesis of coupled ocean-biogeochemical model uncertainties				
SPB	SP+SB	medium range ens (April 2012; 20 mem)			
EnsPB	EnsP+EnsB	seasonal range ens (December 2011–June 2012; 40 mem)			

abbreviations: flt-filter; ens-ensemble; mem-members; U_{air}-u, v wind velocities; T_{air}-air temperature; SLP-sea level pressure; c_d, c_e, c_h-wind drag and turbulent coefficients; c_b-bottom drag; $SMS(C)$-sources minus sinks of biogeochemical tracers C. *100 passes of the Laplacian filter in a Gaussian distribution per model grid point (spatial scales estimated by the response function of the first-order Shapiro filter). **A lognormal anamorphosis function is applied to the $SMS(C)$ of the 24 PISCESv2 prognostic variables.

Journal Pre-proof

Journal Pre-proof

Table 2. CMEMS observation product identifiers (http://marine.copernicus.eu/).

daily frequency		Error
SST gridded 0.05°	SST_GLO_SST_L4_NRT_OBS_010_001	0.5 °C
SLA along track 14 km	SEALEVEL_GLO_PHY_L3_REP_OBS_008_062	0.05 m
8-day frequency		
Chl *a* gridded 4 km	OCEANCOLOUR_GLO_CHL_L4_REP_OBS_009_093	0.3 mg/m^3

Journal Pre-proof

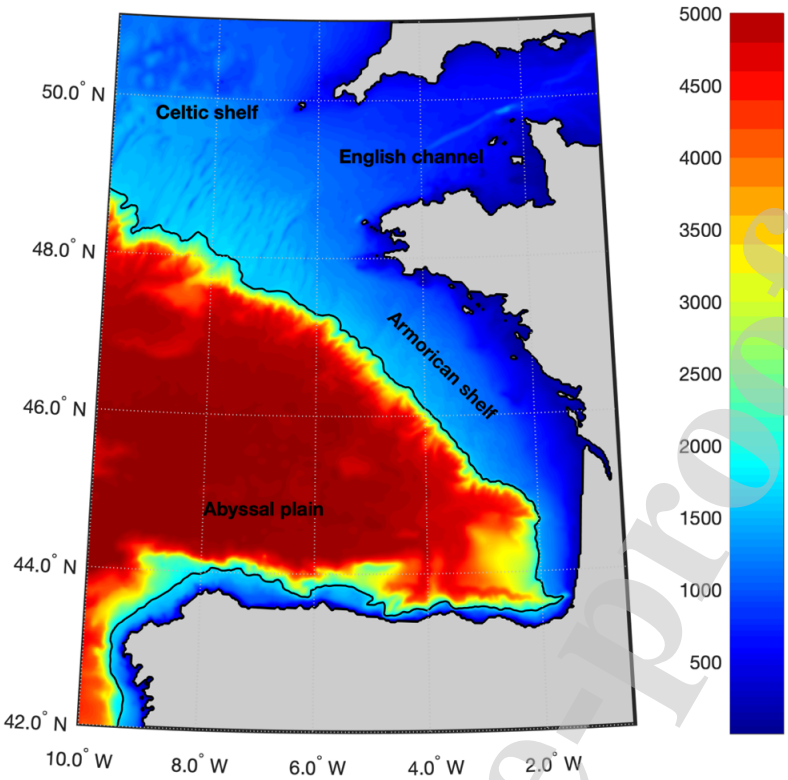


Figure 1 BISCAY36 model domain and bathymetry (m). Black line denotes the 200 m isobath.

Journal Pre-proof

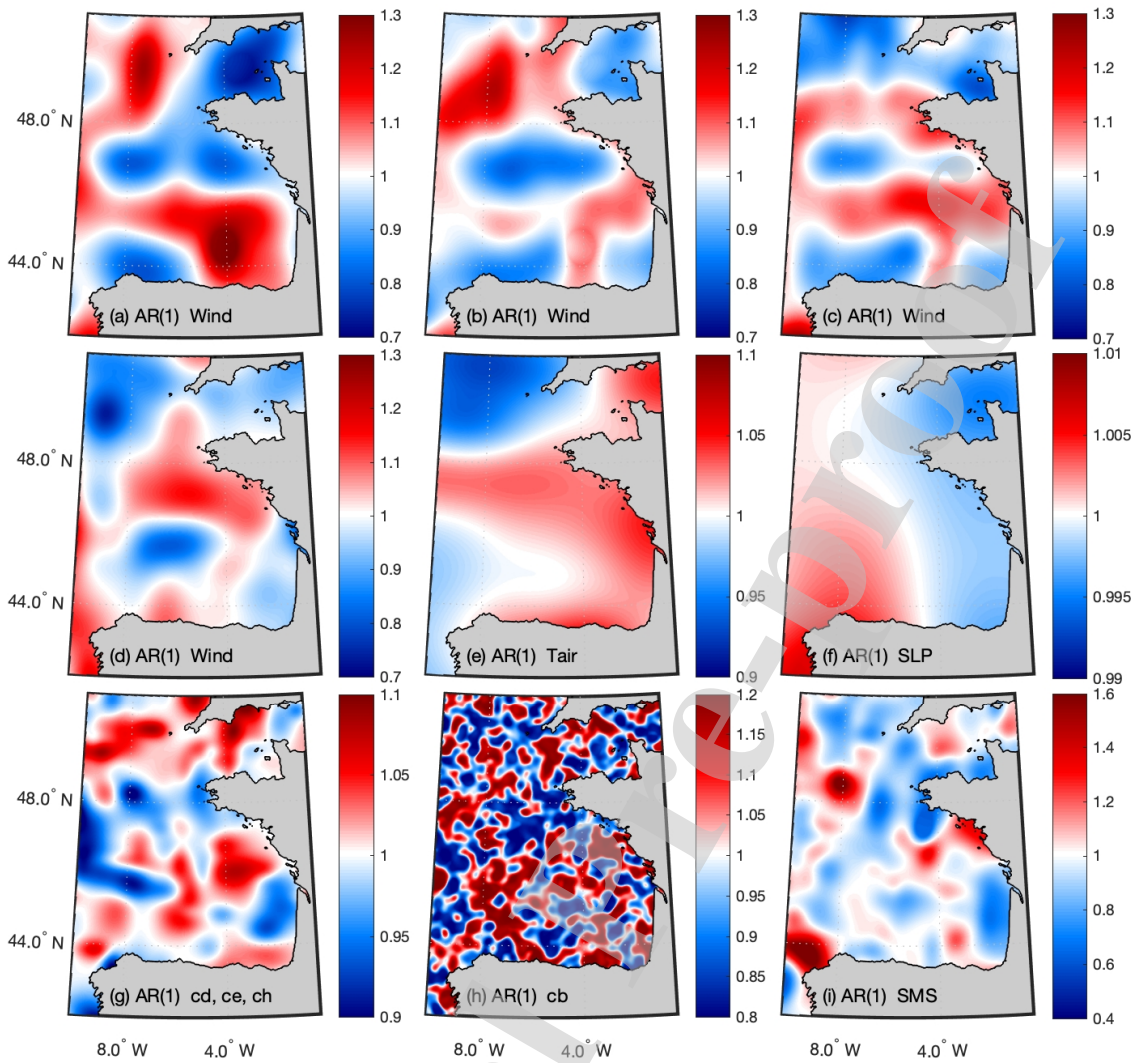


Figure 2 Stochastic uncertainty patterns (no units) explicitly calculated by applying a 2D Gaussian function in first-order autoregressive processes: (a) initial stochastic wind pattern $(1 + \alpha \cdot \xi_k)$ as in Eq. (3) of member-001; all other subplots as in (a) for: (b-c) second and third day respectively, (d) member 002, (e) air temperature, (f) sea level pressure, (g) stochastic pattern $(1 + \alpha \cdot \xi_k)$ as in Eq. (4) of wind drag and turbulent coefficients, (h) bottom drag coefficient by applying a Laplacian filter, (i) stochastic pattern $e^{\left(\xi_k - \frac{\sigma^2}{2}\right)}$ as in Eq. (5) of sources-minus-sinks biogeochemical tracers.

Journal Pre-proof

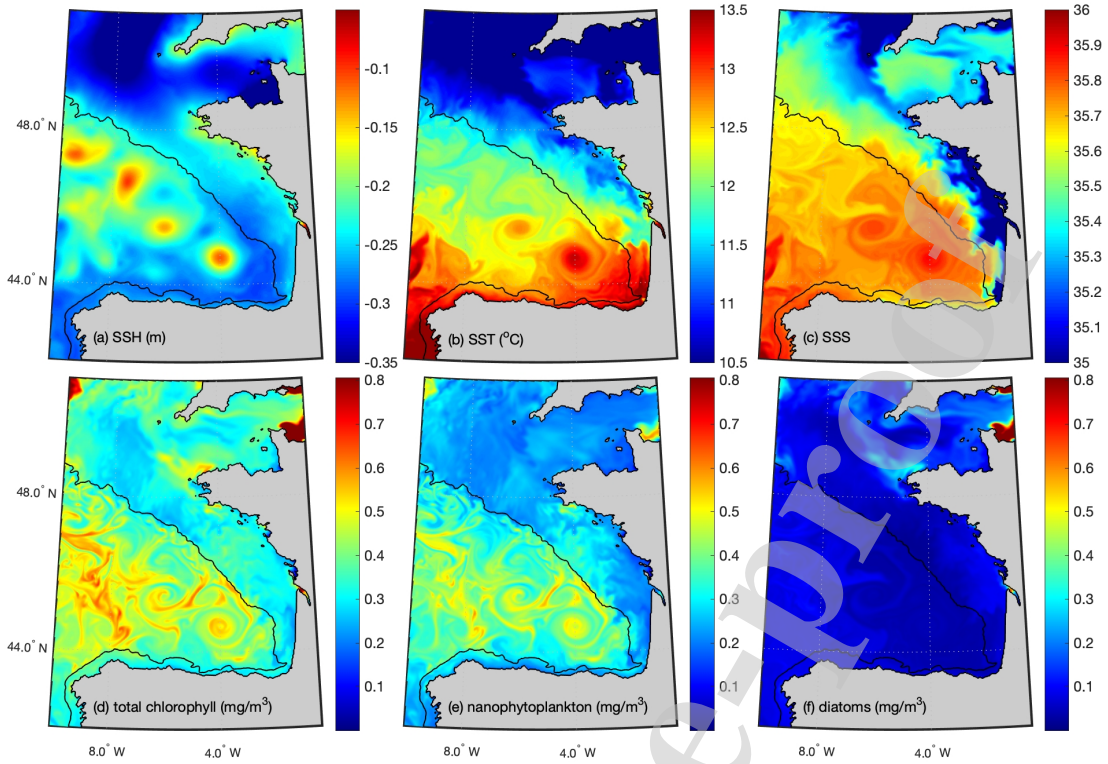


Figure 3 Control run ocean model surface variables on April 30, 2012: (a) SSH (m), (b) SST (°C), (c) SSS, (d-f) left to right: total surface chlorophyll, nanophytoplankton and diatoms (mg/m^3).

Journal Pre-proof

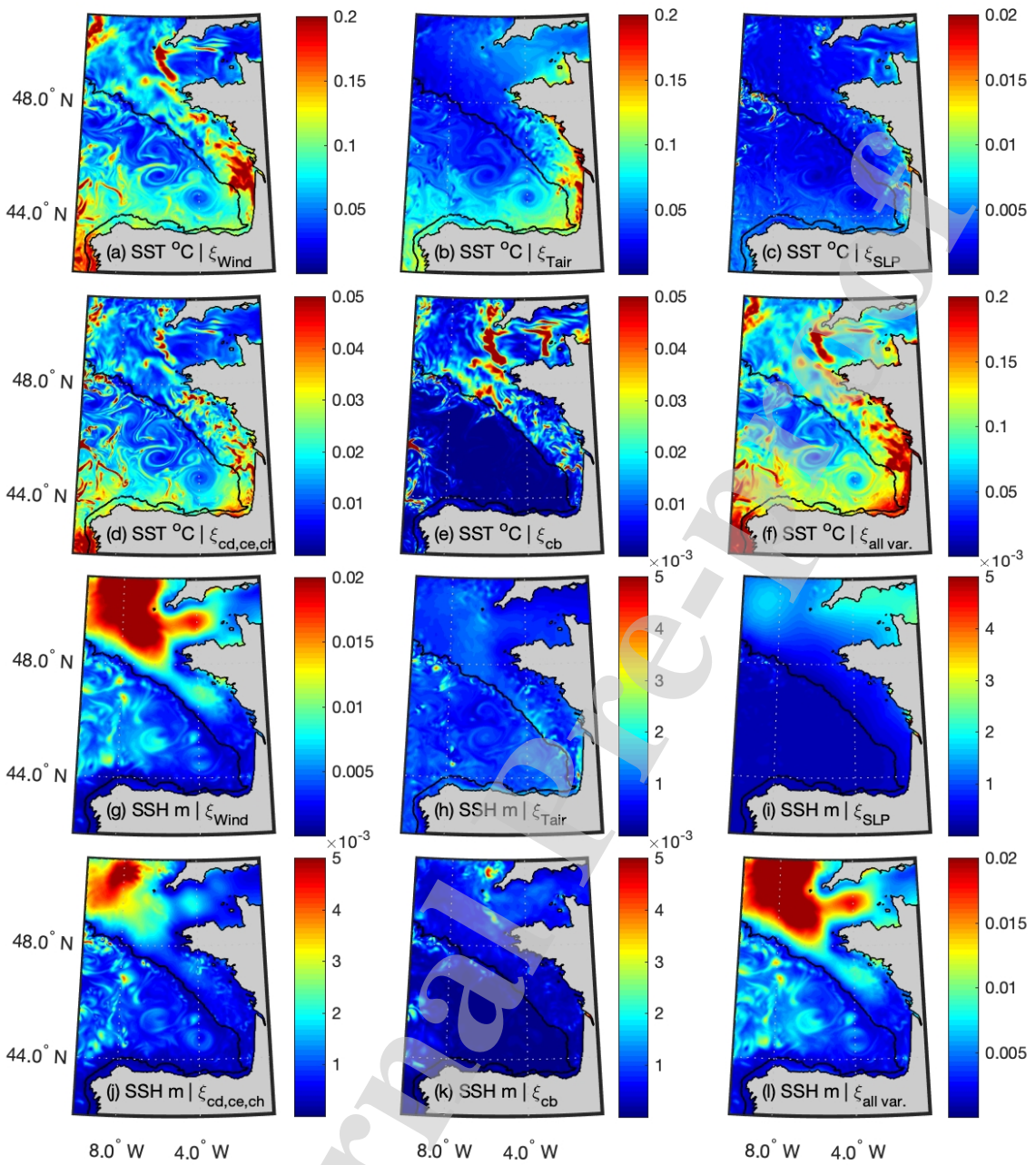


Figure 4 Model uncertainties (i.e. 1std ensemble spread) of medium-range ensembles S1-5 and SP on April 30, 2012: (a-f) SST spread (°C) and (g-l) SSH spread (m) perturbing the (a, g) wind, (b, h) air temperature, (c, i) sea level pressure, (d, j) wind drag and turbulent coefficients, (e, k) bottom drag coefficient, (f, l) all variables together. Note the different colorbars with units varying up to 0.2 °C in SST and up to 0.02 m in SSH.

Journal Pre-proof

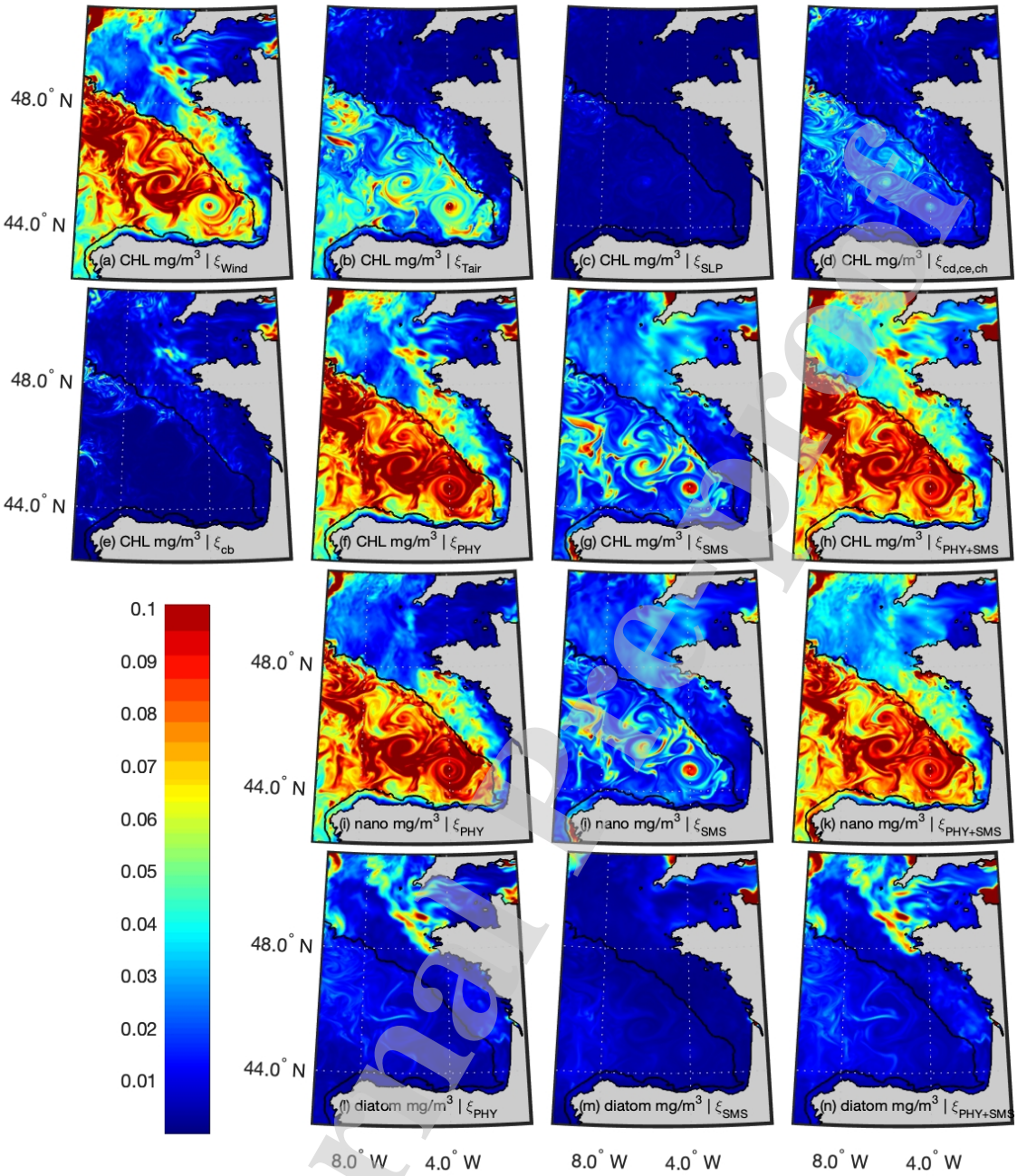


Figure 5 (a-f) S1-5 and SP model uncertainties (i.e. 1std ensemble spread) of total surface chlorophyll (mg/m^3) perturbing physics (PHY), (g) SB chlorophyll spread perturbing $SMS(C)$, (h) SPB chlorophyll spread perturbing all physical-biogeochemical variables. (i-k) and (l-n) are the same as (f-h) but for nanophytoplankton and diatom spread.

Journal Pre-proof

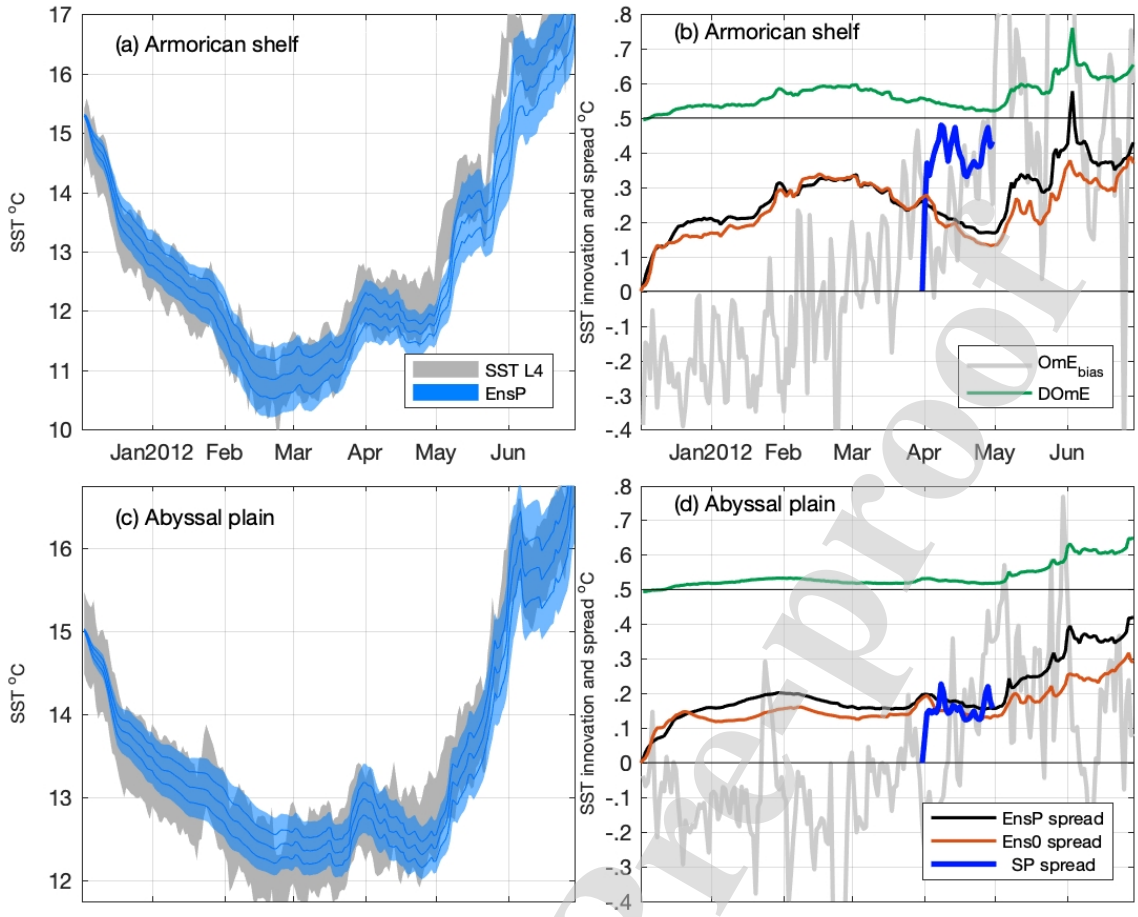


Figure 6 (a, c) OSTIA SST L4 (°C) observation distribution and EnsP ensemble inter-quantile ranges in data space averaged over the Armorican shelf and the abyssal plain, (b, d) innovation statistics and spread for the two regions, respectively; the thin horizontal line denotes the observational error 0.5 °C (the legend is split across two subplots).

Journal Pre-proof

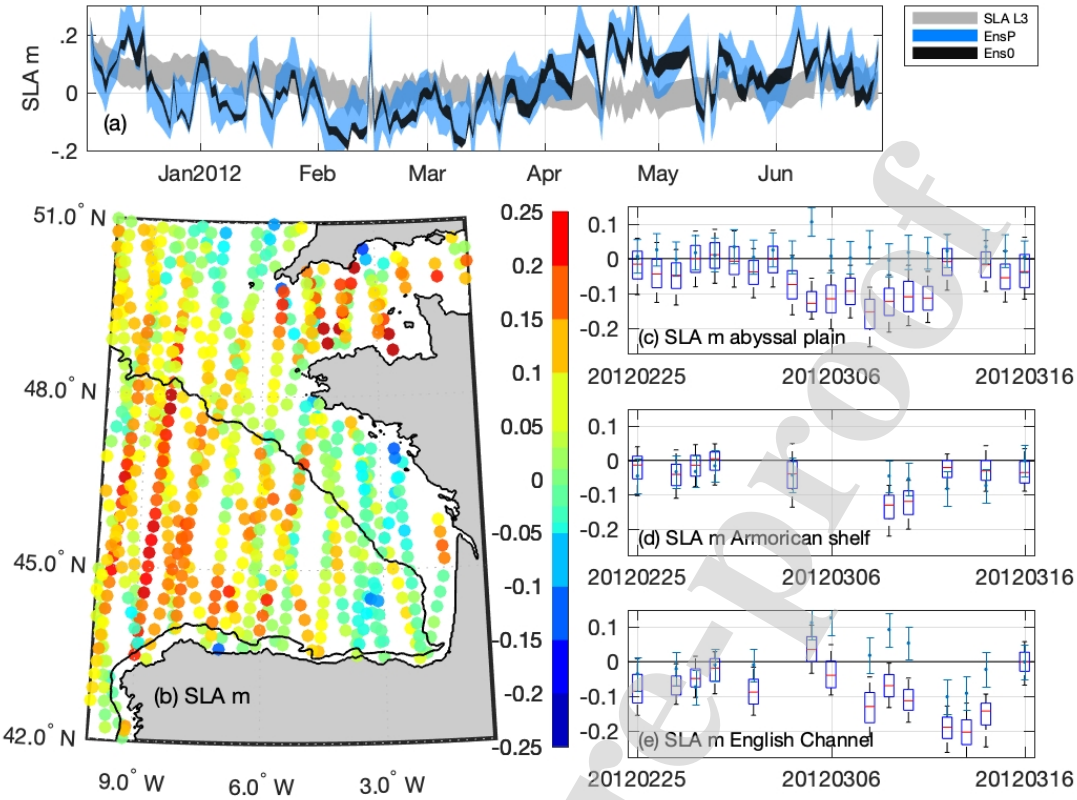


Figure 7 (a) SLA along track L3 observation distribution (m) and EnsP and Ens0 ensembles in data-space, (b) OmE_{bias} map using EnsP for the period starting on February 25, 2012 and for three consecutive weeks, (c) EnsP box-whisker plots and observation error bars averaged over the abyssal plain, (d) the Armorican shelf and (e) the English Channel.

Journal Pre-proof

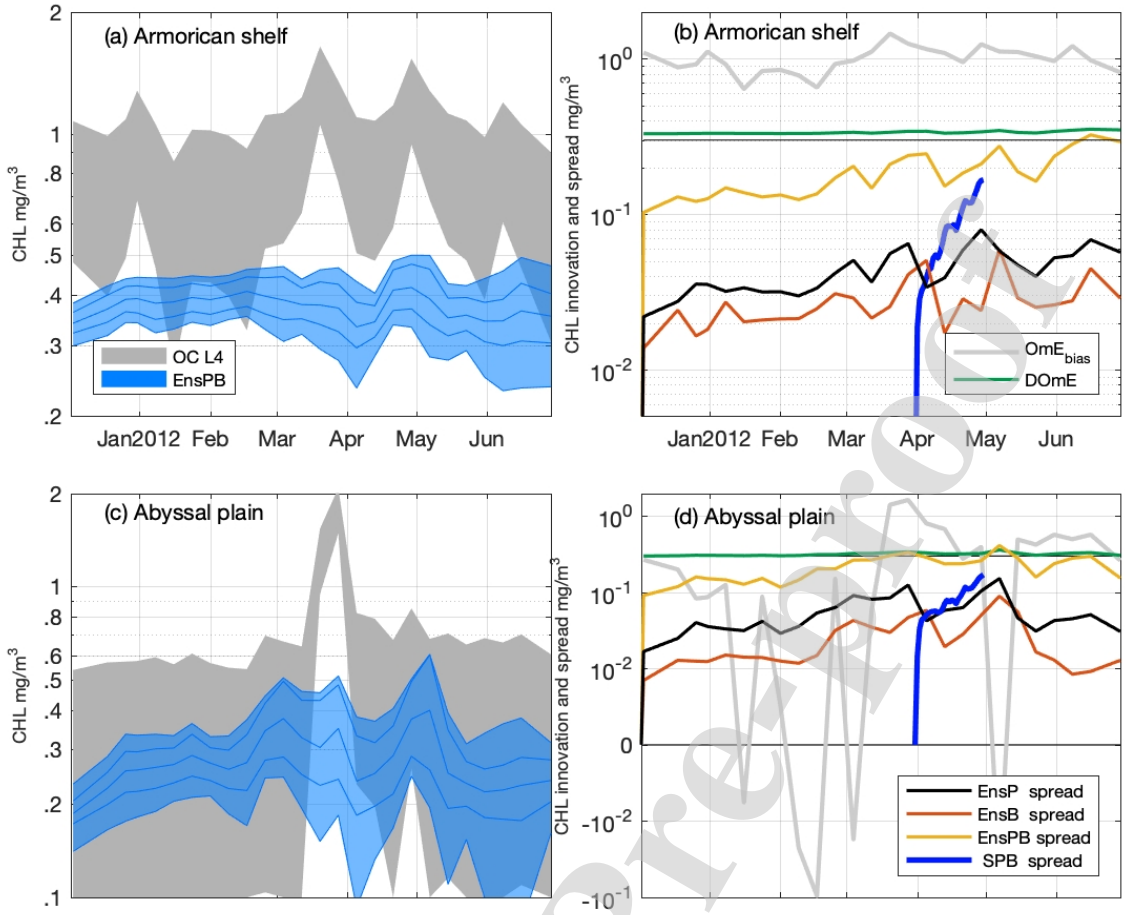


Figure 8 Same as Fig. 6 for ocean colour L4 total surface chlorophyll observations and biogeochemical model ensembles, with innovation statistics calculated in log space. Observational error: 0.3 mg/m^3.

Journal Pre-proof

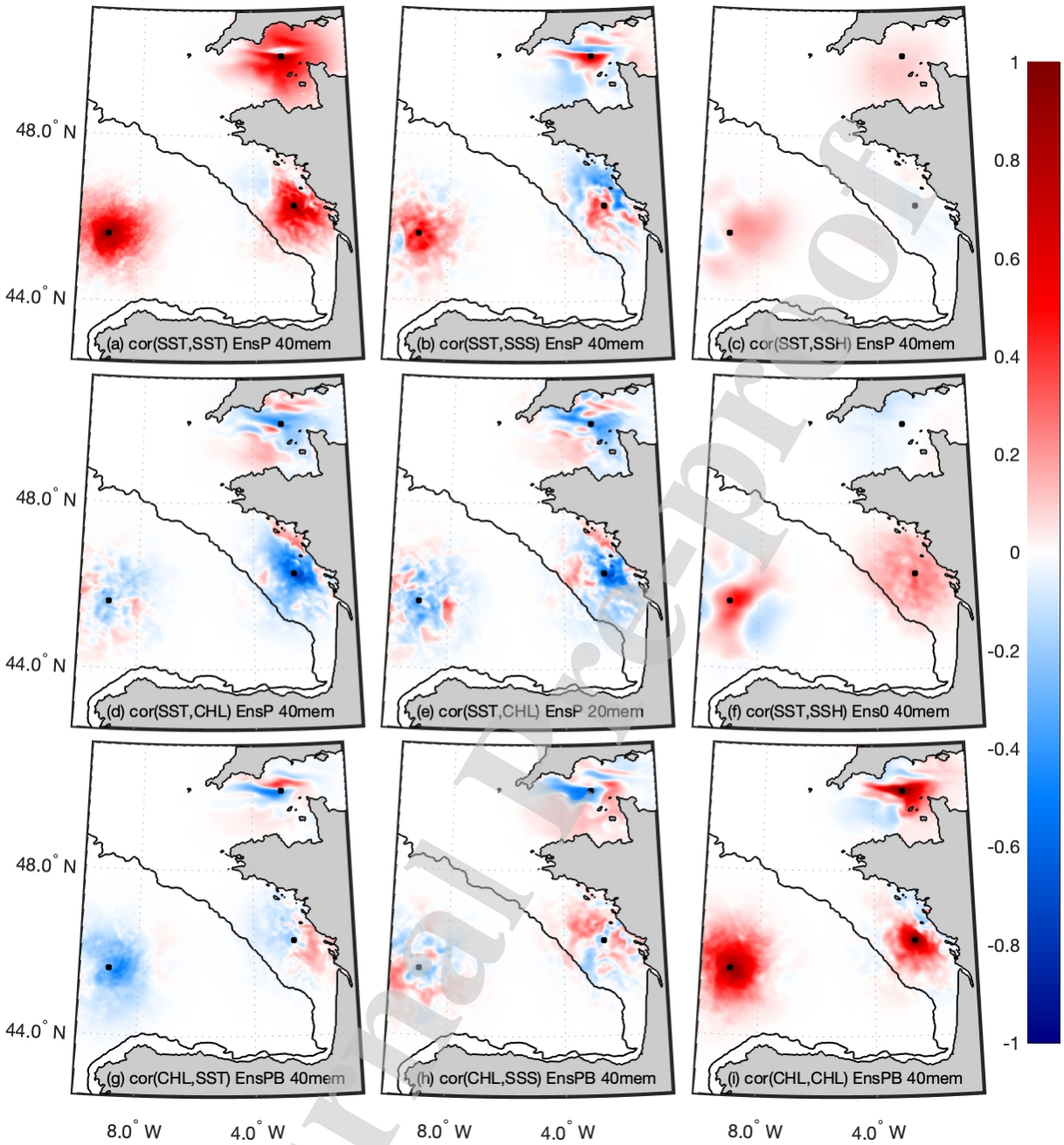


Figure 9 Zero-lag single observation representers for three different surface locations (black markers) in the abyssal plain [9°W 45.6°N], the Armorican shelf [2.6°W 46.3°N] and the English Channel [2.9°W 49.9°N], calculated as correlations between OSTIA SST L4 and all surface variables in the control vector, derived from 40 members of EnsP ensemble on May 07, 2012: (a) cor(SST,SST), (b) cor(SST,SSS), (c) cor(SST,SSH), (d) cor(SST,CHL), (e) same as (d) for 20 members, (f) same as (c) for Ens0. Correlations between ocean colour L4 chlorophyll observations and surface variables in EnsPB control vector: (g) cor(CHL,SST), (h) cor(CHL,SSS) and (i) cor(CHL,CHL). A localization Gaussian function is applied with cut-off radius 3° and e-folding scale 0.2°.

Journal Pre-proof

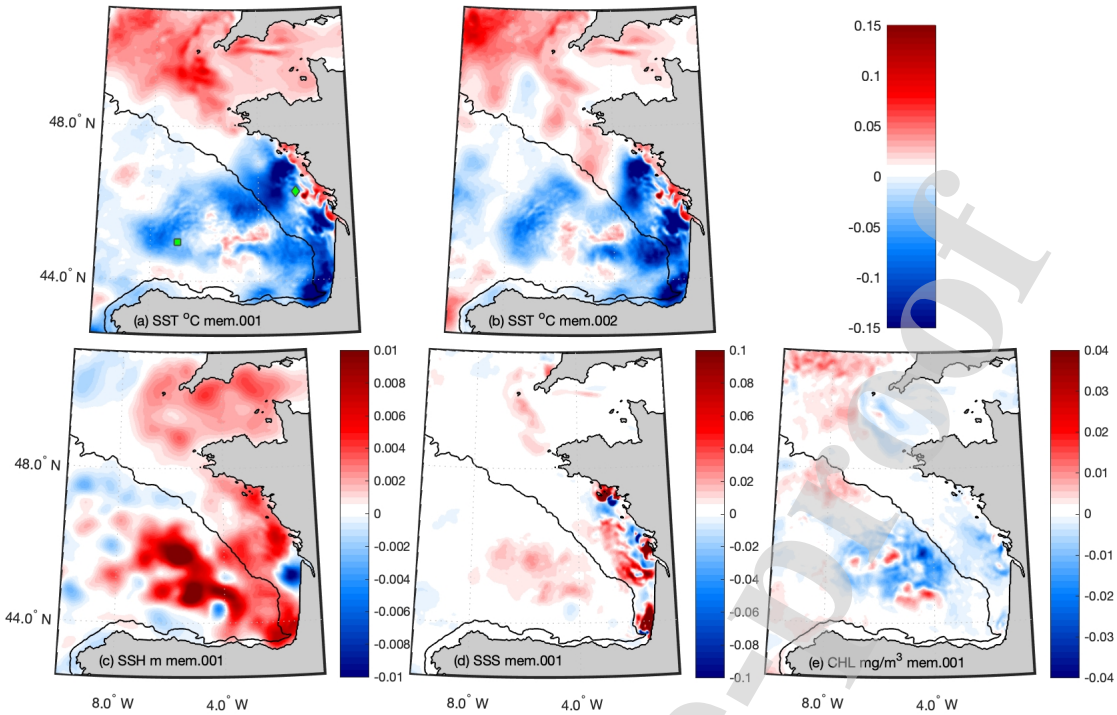


Figure 10 Incremental analysis using OSTIA SST L4 on May 07, 2012 and EnsP 40 members: (a, b) SST correction (°C) for the first two members 001 and 002, (c-e) correction for the first member for SSH (m), SSS and surface total chlorophyll (mg/m^3); (a) two locations illustrated in Fig. 11 for the abyssal plain (green square [7°W 45°N]) and the Armorican shelf (green rhombus [2.6°W 46.3°N]).

Journal Pre-proof

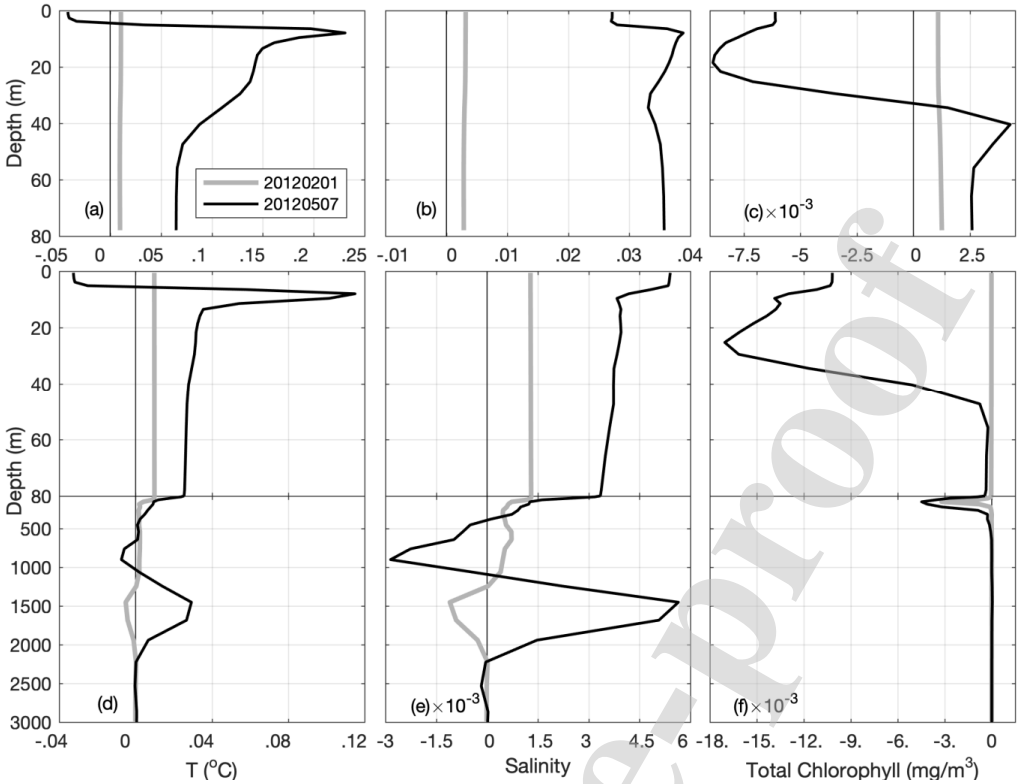


Figure 11 Incremental analysis using OSTIA SST L4 on February 01, 2012 (grey lines) and on May 07, 2012 (black lines), using EnsP 40 members; (a-c) vertical correction of member 001 on, respectively, T (°C), salinity and total chlorophyll (mg/m^3) over the Armorican shelf, and (d-f) the same for the abyssal plain (cf. Fig. 10a).

Journal Pre-proof

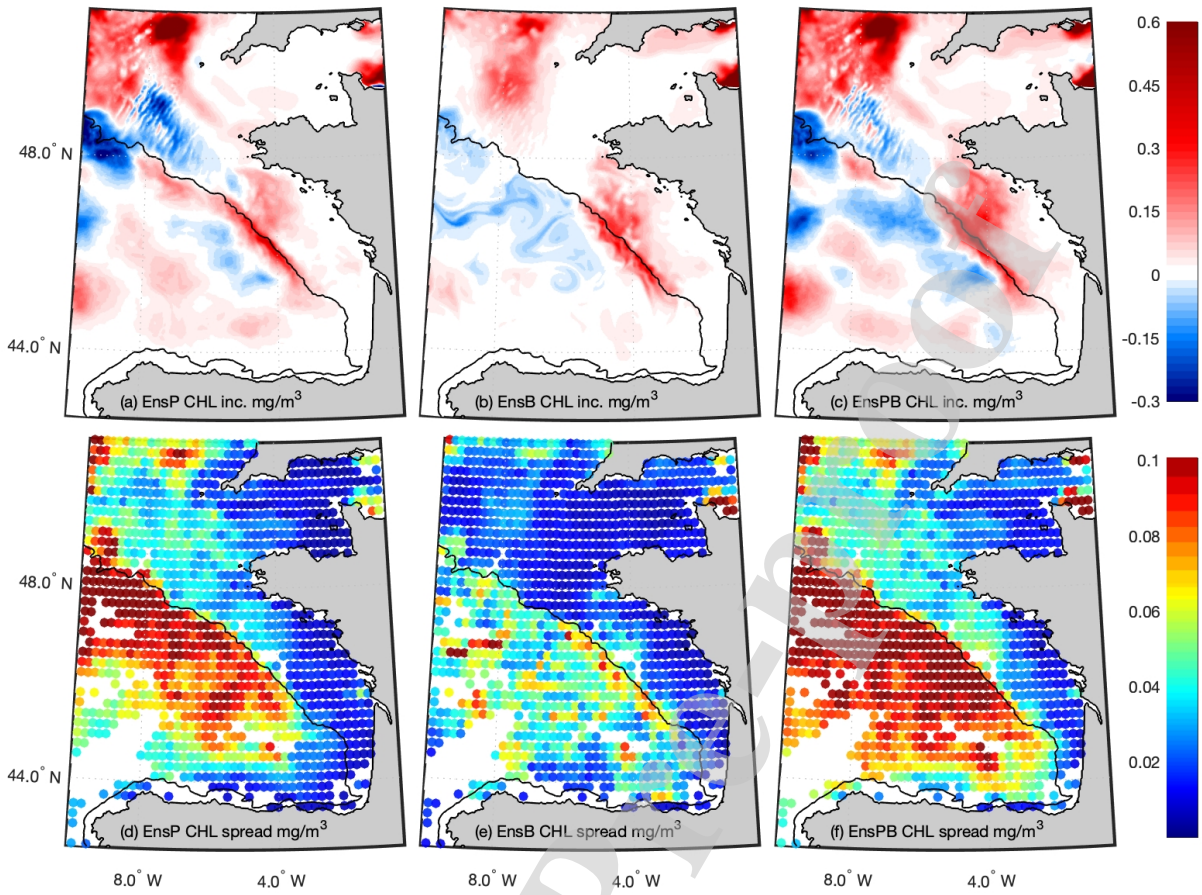


Figure 12 Incremental analysis using ocean colour L4 on May 07, 2012 and ensemble covariances from 40 members for three different ensembles, EnsP, EnsB, and EnsPB (left to right): (a-c) correction of the first member for total chlorophyll (mg/m^3) at a depth of 15 m, (d-f) prior model ensemble spread in data space of total chlorophyll (mg/m^3), as a mean value of the first 5 m of the water column.

Journal Pre-proof

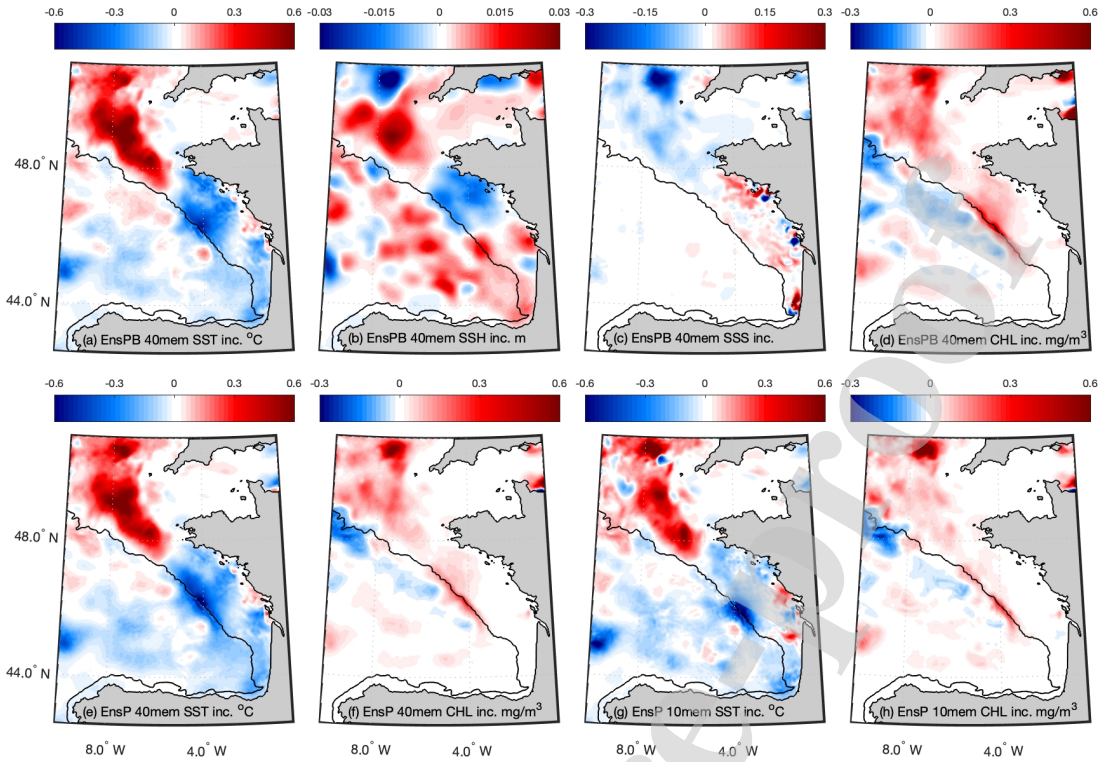


Figure 13 Incremental analysis using OSTIA SST L4 and ocean colour L4 on May 07, 2012: (a-d) ensemble covariances are calculated from EnsPB with 40 members; correction of member 001 on (from left to right) SST (°C), SSH (m), SSS, and surface total chlorophyll (mg/m^3), (e, f) same as (a) and (d) calculated from EnsP with 40 members, (g, h) same as (e, f) calculated from EnsP with 10 members.

Journal Pre-proof

Journal Pre-proof

Highlights

- Ocean model errors are mainly driven by wind forcing uncertainties
- Uncertainties on physical forcing have a large impact on ecosystem model errors
- Data assimilation on physics largely depends on multivariate analysis
- The biogeochemical model skill is mainly defined by the assimilation of chlorophyll
- Most of the chlorophyll correction arises from uncertainties in the physics

Journal Pre-proof

Journal Pre-proof

Authors' contribution statement using CRediT

1) Vassilios D. Vervatis: conceptualization, methodology, software, visualization, formal analysis, investigation, writing – original draft, writing – review and editing.
2) Pierre De Mey-Frémaux: conceptualization, methodology, software, formal analysis, writing – original draft, writing – review and editing.
3) Nadia Ayoub: conceptualization, formal analysis, writing – original draft, writing – review and editing.
4) John Karagiorgos: software, visualization, formal analysis.
5) Malek Ghantous: writing – original draft, writing – review and editing.
6) Marios Kailas: software, formal analysis.
7) Charles-Emmanuel Testut: conceptualization, methodology.
8) Sarantis Sofianos: conceptualization, resources.

The corresponding author,

Vassilios Vervatis

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Supplementary material

Fig. S1 presents the temporal variability of the PFT data in the Bay of Biscay, during the period December, 2011 to June, 2012. We also show the ocean colour L4 total chlorophyll product used in the Part 1 article (*Vervatis et al.*, 2021) and we verify that concentrations are of the same order to those of the PFT total chlorophyll. According to this, we categorised the four satellite PFT (pico, nano, diatoms and dino) into the two broad size groups in PISCES (nano or diatoms), in a manner most representative, ensuring that the total biomass (chlorophyll) from the model and satellite data can be compared like-for-like.

PFT diatoms and nanophytoplankton contribute together approximately more than 80% in total chlorophyll, whereas picoplankton contributes at about 10% and dinoflagellates less than 10% (Fig. S1). PFT diatom chlorophyll concentration is an order of magnitude larger compared with dinoflagellates and nanophytoplankton is about three times larger than picoplankton chlorophyll concentration (Fig. S1). In Fig. S2, we show the spatial distribution of the four satellite PFT and the total chlorophyll during the peak of the spring bloom on March 28, 2012. We confirm that the satellite micro class (i.e. diatoms and dino) is driven primarily by diatoms, far more abundant in the satellite data, with the two functional types being highly correlated in spatial. Fig. S3 presents scatter plots of combined vs. non-combined PFT chlorophyll, verifying the close relationship between functional types in a size class-based approach.

We also present results from one-on-one comparisons between model and PFT data, as opposed to the size class-based categorization merging different functional types. Figure S4 shows Hovmöller plot of rank histograms between EnsPB and PFT, in the same way as Figs. 5b-c, with one main difference: in Figs. S4a-b we do not combine the nano functional type with pico, nor we combine diatoms with dino. Rank histogram results for the nano class are degraded when pico and nano PFT data are not combined together in late-winter and early-spring when a primary bloom occurs (Fig. 5b vs. Fig. S4a). The latter may suggest that PISCES nano can be representative of a broader phytoplankton community, accounting also for smaller size classes. Rank histogram results are almost identical throughout the whole period for the micro class, regardless of whether dino and diatoms are combined together or not (Fig. 5c vs. Fig. S4b). Model-data one-on-one array mode consistency results (not shown) are in practice indistinguishable by visual inspection with the results presented in Figs. 10 and 11, confirming the validity of the size class-based approach.

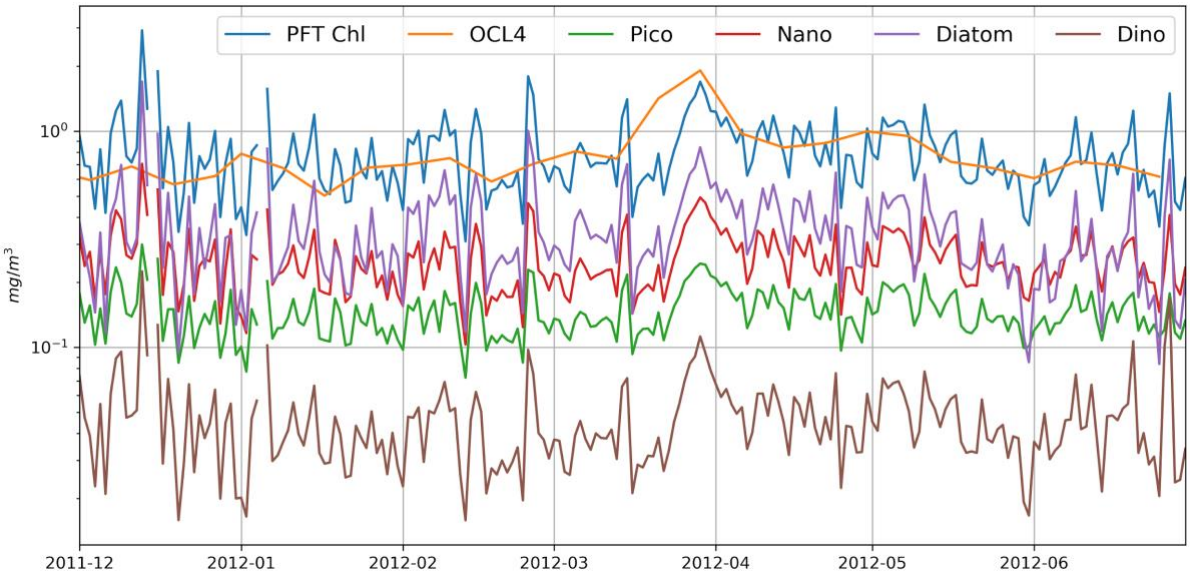


Figure S1 Ocean colour L4 (8-day frequency) and PFT (daily) chlorophyll concentration (mg/m^3) in the Bay of Biscay from December, 2011 to June, 2012.

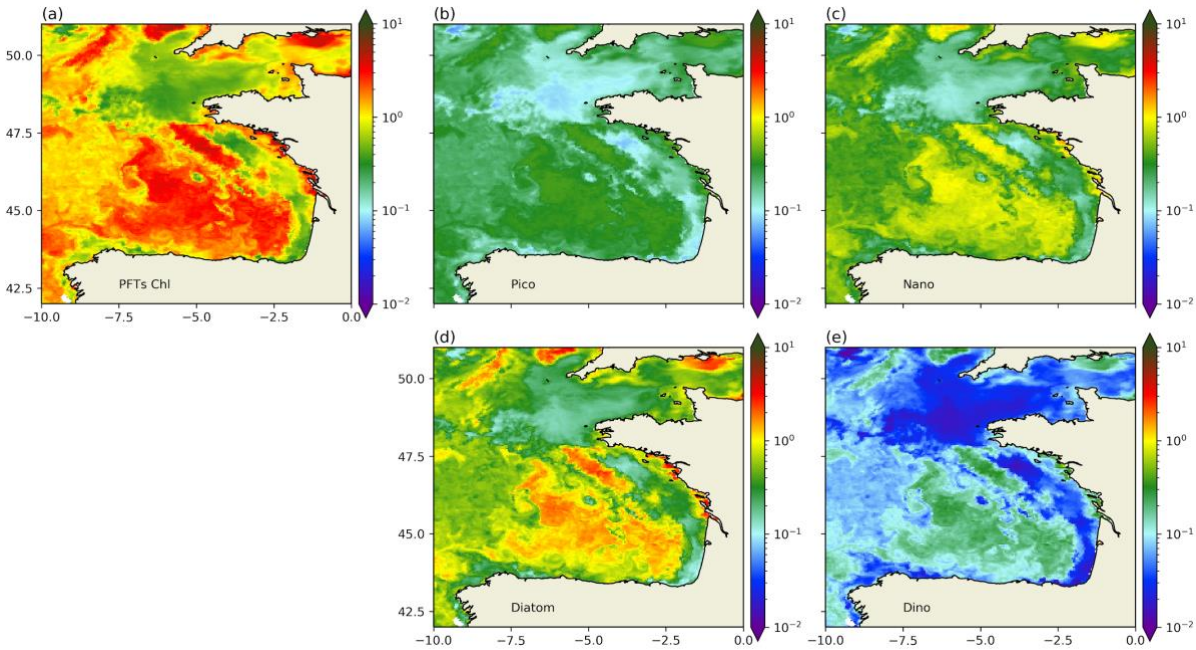


Figure S2 Spatial distribution of (a) PFT total chlorophyll and (b) pico, (c) nano, (d) diatoms, (e) dino concentrations (mg/m^3) on March 28, 2012.

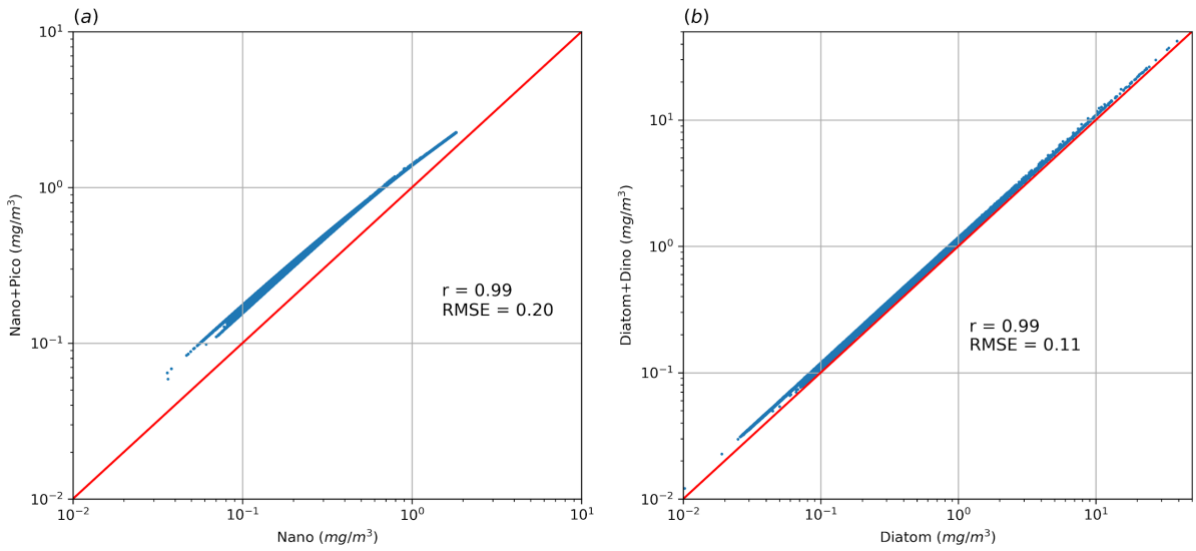


Figure S3 Scatter plots of chlorophyll concentrations in mg/m^3 on March 28, 2012: (a) PFT (nano and pico) vs. only nano, (b) PFT (diatoms and dino) vs. only diatoms. r is the correlation coefficient, RMSE the root mean square error (mg/m^3) and with red the 1:1 line.

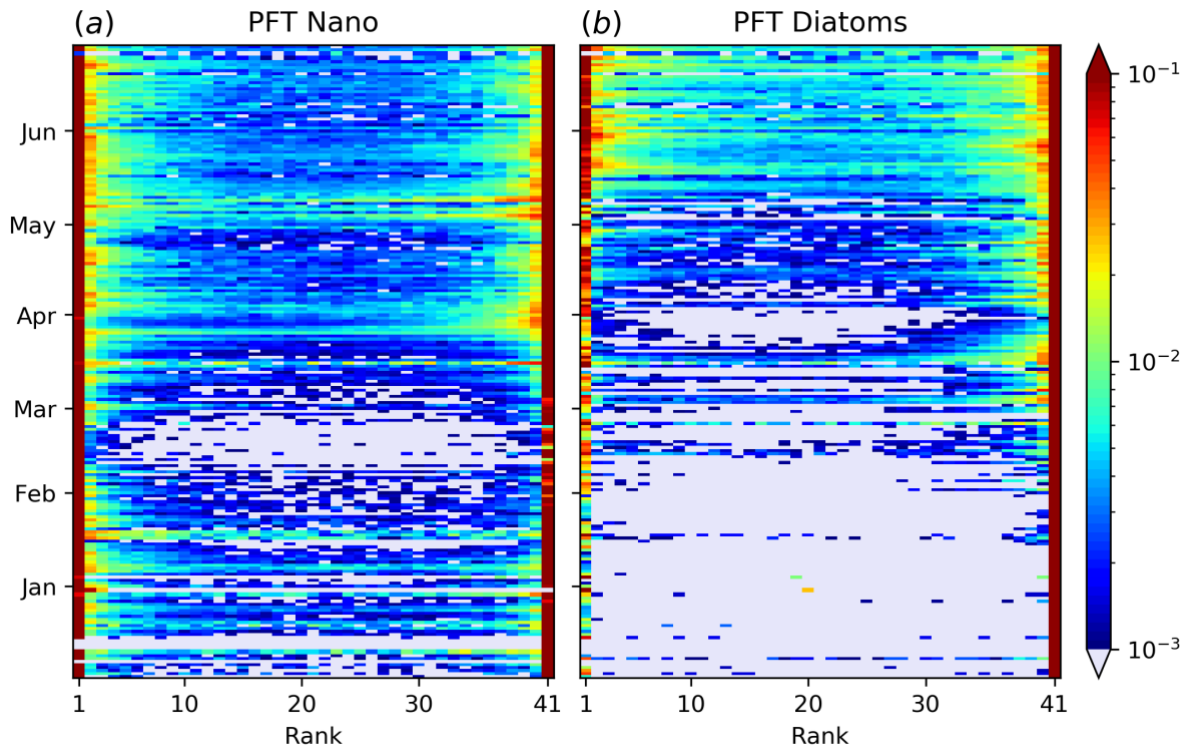


Figure S4 Hovmöller plot of rank histograms (same as in Figs. 5b-c) between EnsPB and PFT (a) nano not combined with pico, and (b) diatoms not combined with dino.