Regional ocean model uncertainties using stochastic parameterizations and a global atmospheric ensemble

Vassilios Vervatis, Pierre de Mey-Frémaux, John Karagiorgos, Bénédicte Lemieux-Dudon, Nadia K. Ayoub, Sarantis Sofianos

To cite this version:

Vassilios Vervatis, Pierre de Mey-Frémaux, John Karagiorgos, Bénédicte Lemieux-Dudon, Nadia K. Ayoub, et al.. Regional ocean model uncertainties using stochastic parameterizations and a global atmospheric ensemble. Ocean Modelling, 2025, 194, pp.102501. 10.1016/j.ocemod.2025.102501 . hal-04921616

HAL Id: hal-04921616
https://hal.science/hal-04921616v1

Ocean Modelling 194 (2025) 102501

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Regional ocean model uncertainties using stochastic parameterizations and a global atmospheric ensemble

Vassilios D. Vervatis [a,*], Pierre De Mey-Frémaux [b], John Karagiorgos [a], Bénédicte Lemieux-Dudon [b], Nadia K. Ayoub [b], Sarantis Sofianos [a]

[a] *Department of Physics, National and Kapodistrian University of Athens, Athens, Greece*
[b] *LEGOS (CNRS/IRD/CNES/UPS), University of Toulouse, France*

ARTICLE INFO

Keywords:
Ocean model NEMO
Biogeochemical model PISCES
Empirical consistency analysis
Ensemble data assimilation
Array modes
Bay of Biscay

ABSTRACT

A Bay of Biscay model configuration is used as a test case to assess the data-based consistency of ensemble-based ocean model uncertainties of several types: [A] built-in stochastic parameterizations at regional ocean scales, [B] ocean model response to a global atmospheric model ensemble and [C] both A and B simultaneously. Ensembles of varying length were generated. In addition to a seasonal-range ensemble, three medium-range ensembles were carried out over successive overlapping segments permitting to compare consistency metrics for different lead times. The largest spread was obtained for the C case, although most of the model uncertainties were attributable to the stochastic ocean parameterizations in A. We addressed the question of which ensemble type and lead time was able to provide the most realistic model uncertainties given observations of SST, sea level, and Chlorophyll a, using a theoretical and diagnostic consistency analysis framework expanded from Vervatis et al. (2021a). In our results, consistency was satisfactory for the stochastic ensembles of types A and C, for the "aged" error cases (but only marginally with respect to the "young" error cases), and whenever physical and biogeochemical uncertainty processes were active in the region and could be detected by the observational networks, such as the onset of the spring shoaling of the thermocline and the phytoplankton abundance primary bloom. Sea level empirical consistency was improved when a wide range of low- to high-frequency errors were included in the signal of dynamic atmospheric process in the data and in the model inverse barometer. These findings provide additional insight that can help configure ensemble-based methods in academic studies and in operational ocean forecasting systems.

1. Introduction

During the first two decades of the 21st century, ensemble methods have been progressively adopted by the ocean modelling community, following marked advances in numerical weather prediction (Palmer, 2018). Despite the progress, the number of operational ensemble-based ocean forecasting systems is still limited; this is expected to change in the forthcoming years (ETOOFS, 2022).

The generation of ensemble-based model error estimates, and in particular the quantitative verification of their fitness for both data assimilation and ocean applications, are both an engineering activity at forecasting centers and an active field of research, the latter involving varied model-data synergy approaches (e.g. De Mey-Frémaux et al., 2019; Moore et al., 2021). The background formalism which can be used for verification is given by data assimilation, which combines a prior model state and observations to generate a posterior state with lower error – for instance, within the family of ensemble-based approaches, let us cite the Ensemble Kalman Filter (EnKF; Evensen, 2003). In the EnKF, estimates of model (prior) uncertainties are obtained by means of an ensemble.

Ocean model errors can be estimated by several categories of ensembles. One approach is to combine information from multi-model multi-physics ensembles, e.g. multi-model products via the Copernicus Marine Service (CMEMS) (http://marine.copernicus.eu/) or products for climate studies via the Coupled Model Intercomparison Project (CMIP6) (https://www.wcrp-climate.org/). Another method is to perform stochastic modelling, in which the state probability density function (*pdf*) can be approximated by the solution of a partial differential Fokker-Planck equation, including processes of stochastic diffusion, advection, and model tendencies (Bessières et al., 2017). In

---

* Corresponding author.
E-mail address: vvervatis@phys.uoa.gr (V.D. Vervatis).

https://doi.org/10.1016/j.ocemod.2025.102501
Received 2 January 2024; Received in revised form 10 January 2025; Accepted 21 January 2025

geophysical fluids, stochasticity can be introduced dynamically (Holm, 2015; Lang et al., 2023). A plethora of studies in the literature demonstrate the use of stochastic methods for ocean models (cf. recent studies and references within Storto and Andriopoulos, 2021; Leroux et al., 2022). A third approach is to generate ocean model ensembles by providing boundary conditions from another ensemble prediction system. These can be lateral ocean boundary conditions from a parent ocean model to a nested child domain (Vandenbulcke and Barth, 2019; Ghantous et al., 2020) or from surface forcing boundary conditions of an atmospheric model ensemble (e.g., from ECMWF) (Zuo et al., 2019). By "verifying" those ensembles, we mean checking [A] whether ensemble-modelled uncertainties are representative of the relevant error processes at hand, and [B] whether the ensemble variability (a proxy of model error) is observable by the available measurements and is consistent with *empirical* estimates of model error ("empirical" being meant as a synonym of "data-based") i.e., the model-data "*misfits*". The latter category of checks ([B]), which we will call *empirical consistency assessment* of the ensembles, critically depends on the availability of "reasonable" estimates of the uncertainties of observations.

This work builds upon, and expands, two previous studies by Vervatis et al. (2021a, 2021b). The authors investigate the ability to generate meaningful ocean model ensembles, by perturbing different sources of physical and biogeochemical model errors. Their results indicate that physical stochastic parameterizations have a larger impact on augmenting ecosystem model errors, than perturbing the intrinsic ecosystem model properties themselves. In those studies, the authors review and apply several approaches towards Ensemble-based Empirical Consistency Analysis (EECA): bias and variance analysis, rank histograms, and an early form of the so-called Array Modes approach which we will expand in this work. As illustrated in several publications (Le Hénaff et al., 2009; Charria et al., 2016; Lamouroux et al., 2016; Vervatis et al., 2021a), it is advantageous to carry out ensemble assessment in the space generated by the eigenvectors of the ensemble-based "representer matrix" (the prior error covariance matrix in data space), named "Array Modes". Indeed, as stated in Vervatis et al. (2021a), Array Modes ("ArM") provide patterns which are hierarchized (via the associated eigenvalues), representative of prior error covariances (as estimated from the ensemble) and observable by the measurement array.

In this study, we will carry out EECA in the space of the Array Modes (the "*array space*"). Using a novel set of empirical consistency metrics, we will compare different ensembles, one using a stochastic regional ocean model, one forced by a global atmospheric model ensemble, and one "hybrid" ensemble adopting both approaches simultaneously, within short- to medium-range "forecasts". We will assess the impact of the "age of errors" on empirical consistency by using overlapping time segments for our medium-range ensemble simulations. Model forecast performance will be assessed upon the verification of EECA metrics in "array space". We also investigate the sea level dynamic atmospheric correction (dac) and its relevance to the modelled uncertainties of the inverse barometer (IB) effect.

The study is organized as follows. In Section 2, we describe the EECA metrics, the physical-biogeochemical model, the stochastic parameterizations of our ensemble experiments, and the observations we will use for analysis. The results are presented in Section 3. Section 3.1 focuses on the generation of regional and coastal ocean model uncertainties, using [A] stochastic modelling together with Monte Carlo techniques to sample the *pdf* of model states, and [B] surface boundary conditions from a global atmospheric model ensemble. Approaches [A] and [B] are used independently and combined, and all approaches are assessed with variance analysis (3.1) and with our set of EECA metrics (3.2). The concluding remarks of this study are discussed in Section 4.

2. Methods and data

2.1. Empirical ensemble-based consistency analysis (EECA) framework

As in Vervatis et al. (2021a), we aim to evaluate the empirical consistency of model ensembles, i.e., we wish to characterize the consistency between model ensembles and observations. This will be done here in the space of Array Modes, which is one of the paths followed in Vervatis et al. (2021a), with some extensions. We now briefly present the mathematical concepts behind the consistency diagnostic metrics which we will use, and the metrics themselves. Some of our ideas were inspired by previous studies, notably those of Anderson (2003), Desroziers et al. (2005), in the context of evaluating the well-posedness of data assimilation schemes. A more complete presentation of the concepts is available in De Mey-Frémaux (2023). Wherever mathematical objects are not explicitly defined, we will use the standard notations from Ide et al. (1997). As above, the abbreviations "CM" and "ECM" respectively denote a Covariance Matrix and an Error Covariance Matrix.

The background idea here (and classic result) is that, assuming that prior/forecast errors and observational errors are uncorrelated to each other, in the space of observations, the ensemble estimate of the CM of innovations $\mathbf{d}$ (= observed values minus prior/forecast values) should be close to the ensemble estimate of the prior/forecast ECM, within observational ECM $\mathbf{R}$:

$$\langle \mathbf{d}\mathbf{d}^T \rangle \leftrightarrow \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} \tag{1}$$

where $\mathbf{H}$ is the observation operator and $\mathbf{P}^f$ is the ensemble covariance of prior/forecast state anomalies with respect to the ensemble mean in state space. The $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ matrix is the ensemble prior/forecast ECM in data space – this is how the measurement array "views" prior/forecast uncertainties as estimated from the ensemble.

As illustrated in several publications (Le Hénaff et al., 2009; Charria et al., 2016; Lamouroux et al., 2016; Vervatis et al., 2021a), it is advantageous to examine Eq. (1) in the space generated by the eigenvectors of $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$, named "Array Modes" in Le Hénaff et al. (2009) and hereafter. Indeed, as stated in Vervatis et al. (2021a) and in the introduction, Array Modes ("ArM") provide patterns which are hierarchized (via the associated eigenvalues), representative of prior error covariances (as estimated from the ensemble) and observable by the measurement array.

Array Modes have historically been first used to characterize the *performance of measurement arrays* at detecting prior errors (Le Hénaff et al., 2009; Charria et al., 2016; Lamouroux et al., 2016; Vervatis et al., 2021a). We will use the following stochastic *array performance* metric, as drawn from these references:

(ArM1) *Array performance* diagnostic metric:

Count singular values of $\mathbf{S}$ larger than 1, with $\mathbf{S} \overset{\text{def}}{=} \frac{1}{\sqrt{m-1}}\mathbf{R}^{-0.5}\mathbf{Y}^f$.

Above, m is the number of ensemble members, and $\mathbf{Y}^f$ is the matrix of ensemble anomaly samples (one $\delta\mathbf{y}_k$ per column k) in data space. The Array Modes, forming the $\boldsymbol{\mu}$ matrix, are defined as the singular vectors of $\mathbf{S}$; they form an orthonormal basis in data space in the sense of the scalar product $\langle \delta\mathbf{y}_1, \delta\mathbf{y}_2 \rangle = \delta\mathbf{y}_1^T\mathbf{R}^{-1}\delta\mathbf{y}_2$. The Array Mode spectrum $\{\sigma_k\}$ is composed of the squares of the singular values of $\mathbf{S}$; it is a by-product of the calculation of metric (ArM1). Examples of Array Mode spectra will be shown in Section 3.

In order to go further and examine the array-space equivalent of Eq. (1), let us note that because $\boldsymbol{\mu}$ is an orthogonal matrix it is possible to go back and forth between data space and array space. Any multivariate data-space sample $\delta\mathbf{y}_k$ can be written as a scaled linear combination of Array Modes – and *vice versa*:

$$\delta\mathbf{y}_k = \mathbf{R}^{0.5}\boldsymbol{\mu}\delta\mathbf{z}_k \tag{2}$$

$$\delta\mathbf{z}_k = \boldsymbol{\mu}^T\mathbf{R}^{-0.5}\delta\mathbf{y}_k \tag{3}$$

where the array-space sample $\delta \mathbf{z}_k$ is nondimensional and univariate, while being characteristic of the whole array. Samples can be examined in array-space given transformations (Eqs. (2) and (3)):

– Array-space observation samples:

$$\text{Matrix of samples}: \ \mathbf{Z}^o \stackrel{\text{def}}{=} \left[\delta \mathbf{z}_1^o \cdots \delta \mathbf{z}_m^o\right] = \boldsymbol{\mu}^T \mathbf{R}^{-0.5} \mathbf{Y}^o. \tag{4}$$

where $\mathbf{Y}^o$ is the matrix of perturbed observation anomaly samples. It is easy to show (e.g., De Mey-Frémaux, 2023) that the observational ECM in array space is the identity:

$$\mathbf{E}^o \stackrel{\text{def}}{=} Cov\left(\mathbf{Z}^o\right) \stackrel{\text{def}}{=} \frac{1}{m-1} \mathbf{Z}^o \mathbf{Z}^{o^T} = \mathbf{I}. \tag{5}$$

– Array-space prior/forecast samples:

$$\text{Matrix of samples}: \ \mathbf{Z}^f \stackrel{\text{def}}{=} \left[\delta \mathbf{z}_1^f \cdots \delta \mathbf{z}_m^f\right] = \boldsymbol{\mu}^T \mathbf{R}^{-0.5} \mathbf{Y}^f. \tag{6}$$

By construction, the prior/forecast ECM in array space is diagonal:

$$\mathbf{E}^f \stackrel{\text{def}}{=} Cov\left(\mathbf{Z}^f\right) \stackrel{\text{def}}{=} \frac{1}{m-1} \mathbf{Z}^f \mathbf{Z}^{f^T} = \boldsymbol{\sigma}. \tag{7}$$

– Array-space innovation samples:

$$\text{Matrix of samples}: \ \mathbf{Z}^i \stackrel{\text{def}}{=} \left[\delta \mathbf{z}_1^i \cdots \delta \mathbf{z}_m^i\right] = \boldsymbol{\mu}^T \mathbf{R}^{-0.5} \left(\mathbf{Y}^o - \mathbf{Y}^f\right). \tag{8}$$

The innovation CM in array space is generally full:

$$\mathbf{E}^i \stackrel{\text{def}}{=} Cov\left(\mathbf{Z}^i\right) \stackrel{\text{def}}{=} \frac{1}{m-1} \mathbf{Z}^i \mathbf{Z}^{i^T}. \tag{9}$$

We now have all elements to derive EECA diagnostic metrics in array space. As Eq. (1) states, if the Ensemble representation of uncertainties was consistent with innovation statistics, the innovation CM should be close to the sum of prior/forecast ECM and observational ECM. In array space, this comes back to comparing:

$$\mathbf{E}^i \leftrightarrow \mathbf{E}^f + \mathbf{E}^o \tag{10}$$

where $\mathbf{E}^o$, $\mathbf{E}^f$ and $\mathbf{E}^i$ are given by Eqs. (5), (7) and (9), respectively. Using those expressions:

$$e_{j,k}^i \leftrightarrow \delta_{j,k}(\sigma_k + 1) \tag{11}$$

where $\delta_{j,k}$ is a Kronecker delta. For the comparison in (Eqs. (10) and (11)) to be successful, we need [A] the diagonal elements $e_{k,k}^i$ to be close to $(\sigma_k + 1)$, and [B] the extra-diagonal elements $e_{j,k}^i (j \neq k)$ to be close to zero. Criterion [A] can be seen as a "*gain*" criterion, as we are examining the innovation variance in the space of array modes drawn from the ensemble: this variance must be of the same order as that drawn from the ensemble i.e., the *gain* must be close to 1. Criterion [B] can be seen as an "*alignment*" criterion of the modes of the innovation covariance matrix with the array modes drawn from the ensemble: if the extra-diagonal terms are non-zero, then the error modes on both sides of (Eq. (10)) are *misaligned* with each other.

Let us first turn to the diagonal elements. The following metric follows (as first introduced in Vervatis et al., 2021a):

(ArMCA1) *Gain consistency* diagnostic metric:

Under tolerance $\tau < 1$:

$$\text{Count array} - \text{space ranks } k \text{ for which } 1 - \tau \leq \frac{var_k\left(\mathbf{Z}^i\right)}{\sigma_k + 1} \leq 1 + \tau. \tag{12}$$

Two criteria are proposed in De Mey-Frémaux (2023) for the consistency check of non-diagonal terms. We will use the following synthetic metric, derived from their (ArMCA2.1) criterion:

(ArMCA2) *Alignment consistency* diagnostic metric:

Let us define $\mathbf{F}_k^i$ as the leading (upper left) submatrix of $\mathbf{F}^i$ at rank $k > 1$, with $\mathbf{F}^i = \mathbf{E}^i - diag(\mathbf{E}^i)$ only composed of the extra-diagonal elements of $\mathbf{E}^i$ (with zeroes on the diagonal).

Under tolerance $\tau' < 1$:

$$\text{Count array} - \text{space ranks } k > 1 \text{ for which } \tau' \geq \frac{1}{k-1}\rho\left(\mathbf{F}_k^i\right). \tag{13}$$

Above, $\rho(\mathbf{A})$ denotes the spectral radius of any positive definite matrix **A,** defined as the largest of its (positive) eigenvalues.

By-products of Eqs. (12) and (13) are the lists of ranks for which each criterion passes. Examples will be shown in Section 3.

2.2. Ocean model

We used the ocean model NEMOv3.6 (Nucleus for European Modelling of the Ocean; http://www.nemo-ocean.eu/; Madec, 2012) and its biogeochemical component PISCESv2 (Pelagic Interactions Scheme for Carbon and Ecosystem Studies volume 2; Aumont et al., 2015). The physical and biogeochemical models were coupled online, using the same grid at 1/36° horizontal resolution. The domain encompasses the Bay of Biscay and the western part of the English Channel (Fig. 1). The configuration, named BISCAY36, is a subgrid of the operational CMEMS Iberia-Biscay-Ireland (IBI) system (Sotillo et al., 2015) previously applied and validated (Maraldi et al., 2013; Quattrocchi et al., 2014; Vervatis et al., 2016; 2021a; 2021b).

The physical-biogeochemical model set-up is described in Vervatis et al. (2021b); only a few relevant details will be given here. The initial state and open boundary conditions were acquired from the CMEMS daily archives for physics and from the weekly archives for

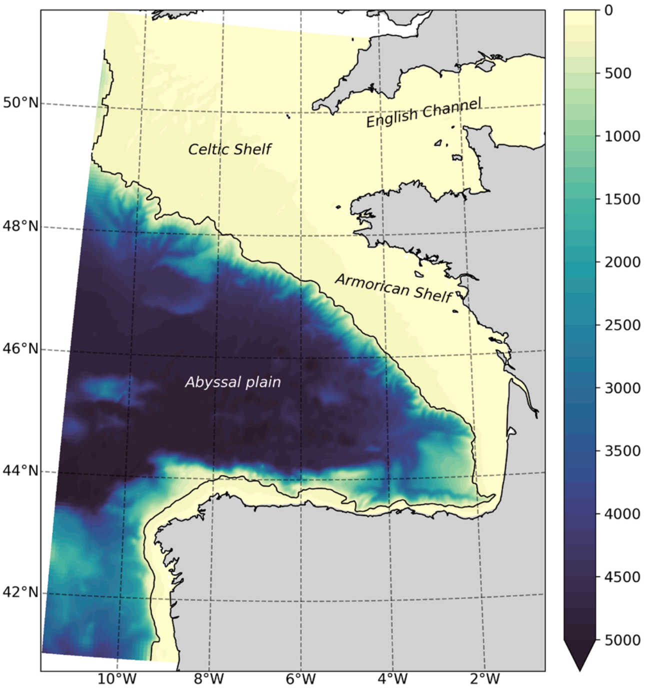


Fig. 1. Ocean model domain and bathymetry (m). The isobath of 200 m is shown. Abyssal plain: ~ 43–49 °N, depths > 200 m; Armorican Shelf: ~ 44–48 °N, depths ≤ 200 m (as referred to in subsequent figures).

biogeochemistry (http://marine.copernicus.eu/). The meteorological fields were provided by the ECMWF (European Center for Medium-Range Weather Forecasts; https://apps.ecmwf.int/archive-catalogue/). An inverse barometer signal was added to the sea level at the open boundary. The model included astronomical tidal potential and tidal harmonic forcing in the open boundaries, provided by the TPXO 7.1 global tide model (Egbert et al., 1994) as the sum of 11 constituents (Karagiorgos et al., 2020). Boundary fluxes of nutrient inputs from the atmosphere and in the river mouths were also considered (Aumont et al., 2015). The river nutrients discharges were derived from the Global News 2 dataset (Mayorga et al., 2010), interpolated and collocated on the three main rivers in the BISCAY36 domain (i.e., Loire, Gironde and Adour) (https://wwz.ifremer.fr/).

We run ensembles of deterministic physical-biogeochemical ocean model simulations, generating stochastic model simulations in which perturbations (as described below) are based on first-order autoregressive Markov processes AR(1), i.e. a statistical model operating under the hypothesis that the current state is based on the immediately preceding state within some error. We perturbed several quantities considered as sources of model error. These quantities, detailed in Section 2.2, consist of unresolved scales, parametrized tendencies and model parameters.

Three schemes are used to generate the stochastic model perturbations following recent advances on NEMO (Brankart et al., 2015; Bessières et al., 2017) namely: (a) stochastic perturbed parameterized tendencies (SPPT) (Buizza et al., 1999), (b) stochastic perturbed parameters (SPP) (Ollinaho et al., 2017), and (c) stochastic parameterization of unresolved fluctuations (SPUF) (Brankart, 2013). The SPPT scheme perturbs the net parametrized model tendencies, assumed to contain upscaled ocean physics errors due to sub-grid parameterizations. A random spatiotemporal stochastic pattern based on AR(1) processes is added collinearly to the net model tendencies. A kernel method is applied to generate random stochastic fields, with specific horizontal correlation structures. The stochastic pattern per member consisted of truncated Gaussian distributions (Vervatis et al., 2021b), defined by stochastic parameterizations controlling the uncertainty amplitude and the spatiotemporal correlations for each tendency. Ocean model parameters are perturbed with the SPP scheme, in a similar way to SPPT, but with one main difference: instead of perturbing the net result of the model parameterizations, we perturbed the parameters estimated during model integration within the parameterization schemes. The SPUF implementation is based on random walks sampling gradients from the state vector (which represent the sub-grid unresolved scales) and adding them to the models' solution. The random walks consisted of independent consecutive steps in all directions including horizontal and vertical steps, and their lengths are controlled by AR(1) processes and their stochastic parameterizations.

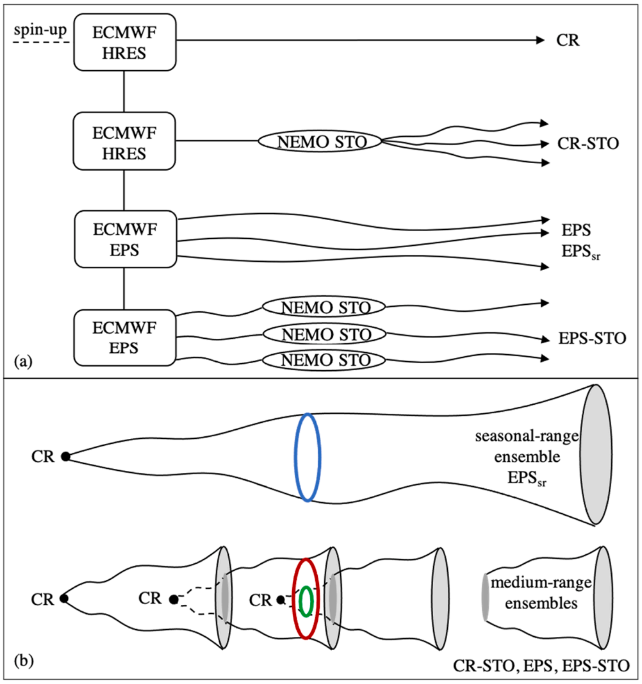


Fig. 2. Schematic of (a) ensemble simulations (cf. Table 1) and (b) initialization techniques; coloured circles correspond to different forecast lead time (blue LT 1–211, red LT 11–30, green LT 1–10 in days) for the same date.

Table 1
Ensemble simulations.

Experiment	Model errors	ECMWF atmospheric forcing and AR(1) ocean stochastic parameterizations (uncertainty, temporal corr., spatial scales)
CR	n/a	Control Run using the ECMWF-HRES atmospheric forcing
CR-STO	Physics & Biogeochemistry	SPPT-AR(1) ECMWF-HRES atmospheric forcing $U_{air}(0.4, 3 days, 1^\circ)$ $T_{air}(0.1, 15 days, 2^\circ)$ $SLP(0.01, 5 days, 3^\circ)$ SPP-AR(1) NEMO physical parameters $c_d, c_e, c_h(0.1, 3 days, 0.5^\circ)$ $c_b(0.2, 30 days, 0.2^\circ)$ SPUF-AR(1) NEMO equation of state T, S (2 *random walks*, 5 *days*, 1 *horiz*. & 0.5 *vert.grid*) SPPT-AR(1) PISCES sources-minus-sinks $SMS(C_{i=1,\dots,24})$ $(0.8, 5\ days, 0.5^\circ)$
EPS $\mathrm{EPS_{sr}}$	Physics	ECMWF-EPS atmospheric forcing U_{air}, T_{air}, SLP, *rh, prec, swr, lwr*
EPS-STO	Physics & Biogeochemistry	EPS atm. forcing and ocean stochastic parameterizations

abbreviations: corr. - correlation; SPPT - stochastic perturbed parameterized tendencies; SPP - stochastic perturbed parameters; AR(1) - first-order autoregressive processes; HRES - high resolution system; EPS - ensemble prediction system; U_{air} - 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; $SMS(C_{i=1,\dots,24})$ - sources-minus-sinks of 24 biogeochemical prognostic variables *C; rh* - relative humidity; *prec* - precipitation; *swr* - short wave radiation; *lwr* - long wave radiation. Lognormal perturbations were applied for the sources-minus-sinks and Gaussian for all other ocean stochastic parameterizations.

2.3. Ensemble simulations

A schematic of the simulations performed in this study is shown in Fig. 2. Table 1 summarizes our choice of perturbed variables and our settings for the stochastic parameterizations in the ensemble experiments. Two atmospheric forcings were used, one being the ECMWF High-Resolution System (HRES) and the other the ECMWF Ensemble Prediction System (EPS), at 9 km and 18 km resolution respectively. The stochastic representations of the ECMWF-EPS global atmospheric model are presented in Leutbecher et al. (2017). As in Vervatis et al. (2021b), a control run was carried out with no stochastic parameterizations using the ECMWF-HRES forcing (referred as CR in Fig. 2 and in Table 1). Starting from the CMEMS ocean global analysis at 1/12° resolution (Lellouche et al., 2013; 2018; http://www.mercator-ocean.fr) the control run was used to spin-up the BISCAY36 configuration from January 2015 to November 2016 and allow the model to develop coherent structures representative of the Bay of Biscay. The control run was then extended from December 2016 to June 2017 and served as reference for all other experiments.

Three approaches were envisaged to generate regional ocean model uncertainties, one was the use of the ECMWF-EPS global atmospheric forcing, another was the implementation of the NEMO ocean stochastic parameterizations, and the third was their combination (cf. Table 1 for the perturbed variables and parameters). In order to test which error sources are most significant for regional ocean model uncertainties, we designed the following experiments. The post spin-up solution of the control run was used to initialize a seven-month seasonal-range ocean ensemble using the ECMWF-EPS forcing, without ocean model stochastic parameterizations (referred as $\mathrm{EPS_{sr}}$ in Fig. 2 and Table 1). In addition to the seasonal-range ensemble, three medium-range ensembles were carried out by performing successive forecasting cycles of one-month, all initialized from the control run and spanning the same period as the seasonal-range ensemble from December 2016 to June 2017: (a) the first ensemble used again the ECMWF-EPS forcing, without ocean model stochastic parameterizations (referred as EPS in Fig. 2 and Table 1), (b) the second ensemble used the ECMWF-HRES forcing including ocean model stochastic parameterizations (referred as CR-STO in Fig. 2 and Table 1), and (c) the third ensemble combined the methods, using the ECMWF-EPS forcing and the ocean model stochastic parameterizations together (referred as EPS-STO in Fig. 2 and Table 1). Each ensemble experiment for the regional ocean model consisted of 50 members, as is the case for the ECMWF-EPS global atmospheric forcing. Hereafter, we refer to CR-STO and EPS-STO as “stochastic” ensembles, and to EPS and $\mathrm{EPS_{sr}}$ as “non-stochastic” ensembles (instead of “non-ocean-stochastic” omitting the word “ocean” for simplicity, since there is stochastic input for the atmosphere).

The medium-range successive segments of ensemble forecasts were chosen to overlap each other for a few days upon initialization from the control run (Fig. 2b). This approach permitted to take into account a short ensemble spin-up period meant to bring the model up to empirical consistency with observations (Vervatis et al., 2021a). The length of the ensemble spin-up period was adjusted to ten days i.e., lead time (LT) 1–10 in days and the rest of the medium-range forecasts gave access to “aging errors” up to a month i.e., lead time 11–30 in days. With this experimental protocol we were able to compare consistency diagnostics for different lead times within the same period and for the same date.

The notion of “aging errors” is meant in a relative way, with respect to any time t_0 within the range. For an Extended Kalman Filter (EKF; e.g., Eq. (4b) in Ide et al., 1997), using the linearized form of error propagation, the “prior” (forecast) error covariance matrix (ECM) at time t within our range writes:

$$\mathbf{P}^{f}(t) = \mathbf{M}(t_0, t)\mathbf{P}^{a}(t_0)\mathbf{M}^{T}(t_0, t) + \mathbf{Q}(t_0, t) \tag{14}$$

where $\mathbf{M}(t_0,t)$ is the transition matrix of the linearized model from time t_0 to $t > t_0$, $\mathbf{P}^{a}(t_0)$ is the “posterior” ECM after analysis at t_0, $\mathbf{Q}(t_0,t)$ is the modelling ECM, which is expected to grow with time t, and $(t - t_0)$ is the forecast range or lead time. In this study, we do not assimilate, but expression (Eq. (14)) is nevertheless useful to illustrate at a more general level how model errors “age” from any time t_0: part of their evolution is due to the model dynamics acting on “older”, pre-existing errors up to time t_0 (first term), and another “fresher” part is due to forcing errors and to model-related errors e.g., from parameter errors, acting from t_0 to t (second term). In this context, $\mathbf{P}^{f}$ is expressed in the square-root form $\mathbf{P}^{f} = \boldsymbol{S}^{f}\boldsymbol{S}^{fT}$ (cf. Eqs. (10) and (11) in Vervatis et al., 2021b).

In the following, we concentrate on the evolution of model uncertainties of age $(t - t_0)$ under the influence of the model error sources in Table 1, where t_0 will be the initial time of each ensemble simulation.

2.4. Observations

The datasets were accessed via the CMEMS infrastructure (https://marine.copernicus.eu/) and are summarized in Table 2. We considered global and regional satellite data from multi-missions for both near real-time and delayed modes. We focused on upper-ocean properties for sea surface temperature (SST), sea level anomaly (SLA) and surface total chlorophyll. The observational errors were available with a spatial distribution for the SST and the ocean colour products. A representative value from the literature (Vervatis et al., 2021a; 2021b) was used for the along-track SLA error. All datasets were provided at a daily frequency.

Table 2
Observational networks.

CMEMS Product Identifiers (http://marine.copernicus.eu/)[a]		Error[b]
OSTIA SST gridded ~ 5 *km* subsampling rate 3	SST_GLO_SST_L4_NRT_OBS_010_001	[0.1, 0.5 [^{o}C
SLA along-track ~ 14 *km* no subsampling	SEALEVEL_GLO_PHY_L3_MY_008_062 altimetry missions: Sentinel-3A, Jason-3, Cryosat-2	0.05*m*
Chlorophyll *a* gridded ~ 1 *km* subsampling rate 13	OCEANCOLOUR_ATL_BGC_L4_NRT_009_116	[1, 300 [%

[a] All datasets were provided at daily frequency.
[b] The gridded datasets were provided with spatial errors.

We used the OSTIA SST L4 gap-free gridded dataset provided on a high-resolution regular grid at 5 km (10.48670/moi-00165). The product is a foundation SST free of diurnal variability and the model proxy was set to be the temperature interpolated at 10 m depth. The SST data are computed via multi-scale analysis of an optimal interpolation scheme and provided with an uncertainty estimate at each grid point (referred also as the analysis error). The data error estimate is a combined field of background error variances and observational uncorrelated errors. For our period of investigation, the OSTIA SST data errors varied spatially from 0.1 °C in the open ocean to about 0.5 °C in the coastal (maximum values up to 0.65 °C for a few days during early-summer).

For the sea-level, we used the CMEMS L3 along-track product based on Sentinel-3A, Jason-3 and Cryosat-2 altimetry missions (10.48670/moi-00146). We used the filtered sub-sampled data with 14 km distance between successive points along the altimetry track. The observational error (i.e., standard deviation) was set at 0.05 m, assuming uncorrelated data errors. By default, the SLA data retrieved by CMEMS is already corrected using a dynamic atmospheric correction (dac) term, which removes the aliasing in the altimetry measurements based on two components. The first component is the static inverse barometer response of low-frequency elevations from the atmospheric pressure variations (Roblou et al., 2011) and the second component is the non-static response of the ocean to fast changes in the wind and pressure provided by a surge model (Carrère and Lyard, 2003). Additional corrections include the high-frequency elevations from the ocean tides and the long wavelength errors (lwe) remaining in the signal (Le Traon et al., 1998). In this study, we considered both corrected SLA and uncorrected $\mathrm{SLA_{unc}}$ data, using the equation:

$$SLA_{unc} = SLA + dac - lwe \tag{15}$$

The SLA model equivalent was estimated by subtracting the model mean sea surface height (MSSH) computed over the control run period 2015–2017. The model during integration included an inverse barometer and tides. The tidal signal was removed from the daily averaged model outputs and therefore was not considered also in the observations. For the model-data comparisons we considered two cases: (a) we removed the model inverted barometer from the ensemble outputs when the corrected SLA dataset was used, and (b) we included the model inverted barometer when we used the uncorrected sea level product $\mathrm{SLA_{unc}}$.

We selected the specific ocean colour L4 product regionally tuned over the Atlantic made available via CMEMS (10.48670/moi-00288). This is a “merged data record” product collected from multiple satellite missions, provided in a high-resolution gridded format at 1 km. The satellite chlorophyll *a* data was provided with a spatial scaled (%)

observational error (uncorrelated at each grid point) in comparison to their signal, varying from 1 % in the open ocean to more than 300 % in the coastal and during phytoplankton blooms (i.e., approximately more than 2 mg/m^3). The chlorophyll *a* model proxy was set to be the mean value of the first 5 m of the water column (Vervatis et al., 2021b).

For the consistency analysis framework presented in Section 2.1, the following protocol elements applied. All observations were perturbed (randomized) across the ensembles using appropriate pseudo-random number generators, fed by the uncertainty standard deviations in Table 2, so as to produce an observational spread consistent with the observation ECM **R**. For the gridded OSTIA SST (resp. ocean colour) datasets, we used the spatial distribution of their error standard deviations, as mentioned above, to draw Gaussian (resp. lognormal) quasi-random numbers. For along-track sea level datasets, we used a Gaussian pseudo-random generator, and the resolution of observations was kept at 14 km. Thinning was applied to the gridded OSTIA SST and ocean colour datasets to match an O(14km) resolution. This serves also to avoid including noise in the Array Mode space for the scales not well represented in the data. A subsampling rate of one-third was applied to the OSTIA SST and one-thirteenth to the satellite chlorophyll *a* (i.e., one data point every third point and thirteenth point was retained respectively in each direction), resulting in a number of observations in the model domain on the order of $\sim \left(3{\cdot}10^3\right)$ per day, as if 1-day assimilation cycle had been performed. Approximately the same number of sea level observations was obtained for a 10-day period, as if 10-day assimilation cycle was performed.

3. Results

3.1. Ensemble model spread

We focus on two Bay of Biscay areas, namely the Abyssal plain and the Armorican shelf (cf. Fig. 1), to illustrate model spread dependency based on geographical region. In Fig. 3a-b, we show the EPS-STO ensemble envelope by means of inter-quantile range. Fig. 3c-d shows SST model spread to be larger when using both an atmospheric model ensemble and an ocean stochastic model (i.e., experiment EPS-STO) simultaneously, rather than using only the former or the latter. Most of the model errors are attributed to the ocean stochastic model itself CR-STO, whilst the atmospheric ensemble experiment EPS and EPS$_{\mathrm{sr}}$ have only a moderate impact, being rather small in the shelf at all lead times. The SST model spread is larger in late-spring and early-summer compared to winter conditions, because of the shoaling of the seasonal thermocline. Model errors reach up to 0.8 °C and are evident in early-summer for the medium-range stochastic ensembles EPS-STO and CR-STO. During the same period, the non-stochastic ensembles EPS and EPS$_{\mathrm{sr}}$ show moderate SST model error estimates with values close to 0.4 °C and often (but not always) with little dependency on the forecast range. During winter, when the mixed layer is deeper than the Ekman layer, our experimental protocol potentially underestimates upper-ocean model errors, resulting in small SST spread, especially in the open ocean. A possible explanation is that, wind perturbations during winter are less effective than those during summer, in shaping the error regimes through dynamic processes (e.g., divergence) in the upper-layer. Overall, in most cases, a ten-day ensemble spin-up appears to be sufficient to obtain a reasonable SST model spread comparable to the magnitude of the observational errors.

Sea surface height model errors ought to be investigated with respect to the atmospheric IB signal and for this, we considered two approaches: the one was to include the sea level response to the IB signal in the daily model outputs and the other was to remove it in post-processing. This investigation is necessary because first, the sea level pressure (SLP) perturbations in the stochastic model ensembles are applied per model timestep, therefore the SSH response varies at high-frequency; second, the SLP patterns (and thus the IB response) are different between members when the ECMWF-EPS atmospheric forcing is used; and third, because such an investigation is relevant when we compute the SLA model equivalent and estimate the model-data misfits, using either the corrected or the uncorrected SLA altimetry data.

Fig. 4 shows the ocean model SSH ensemble envelope (Fig. 4a–d) and spread (Fig. 4e–h) as a response to SLP model errors. These errors are imposed by stochastic methods in the CR-STO experiment, by the ECMWF-EPS atmospheric forcing in the EPS and EPS$_{\mathrm{sr}}$ experiments, and by their combination in the EPS-STO experiment. The SLP perturbations trigger an isostatic IB signal dominant at large scales and a non-IB signal dominant at local scales depending on the geographic region. The impact of the latter can be seeing when the IB response is subtracted from the SSH model outputs in post-processing (Fig. 4c,d and g,h). When using only the ECMWF-EPS atmospheric forcing, the model SSH spread is found to be rather small i.e., less than 1 cm (not shown). An exception to this finding is the seasonal-range model SSH spread in the EPS$_{\mathrm{sr}}$ ensemble, due to the mesoscale decorrelation of eddies among members (not constrained by data assimilation) in the Abyssal plain (Fig. 4g), with a time scale of the order of ∼ 2 months (Vervatis et al., 2016), as

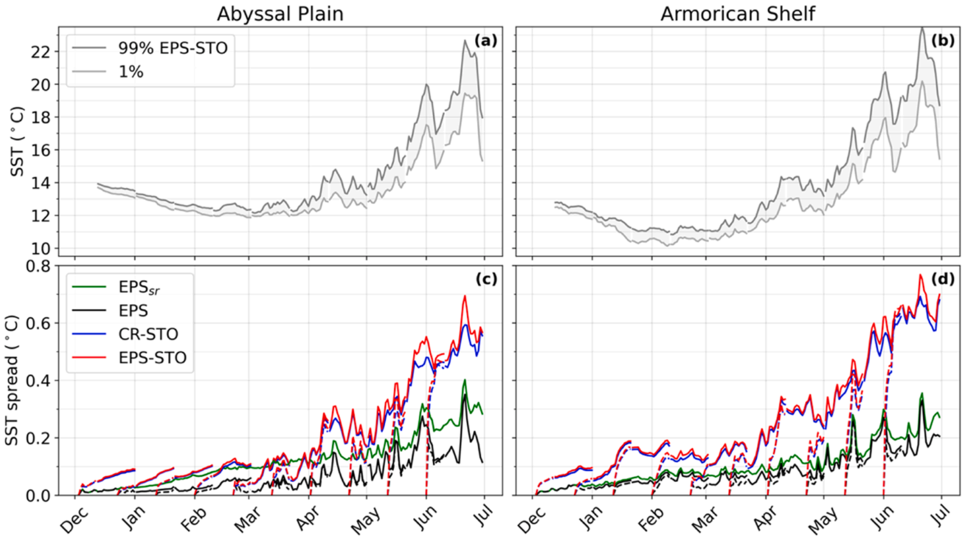


Fig. 3. (a-b) SST (°C) model ensemble EPS-STO inter-quantile range in the Abyssal plain and Armorican shelf; medium-range successive segments of forecasts with lead time 11–30 in days, (c-d) model spread for the ensembles in Fig. 2 and Table 1; the dashed lines in the medium-range ensembles correspond to a spin-up period with lead time 1–10 in days.

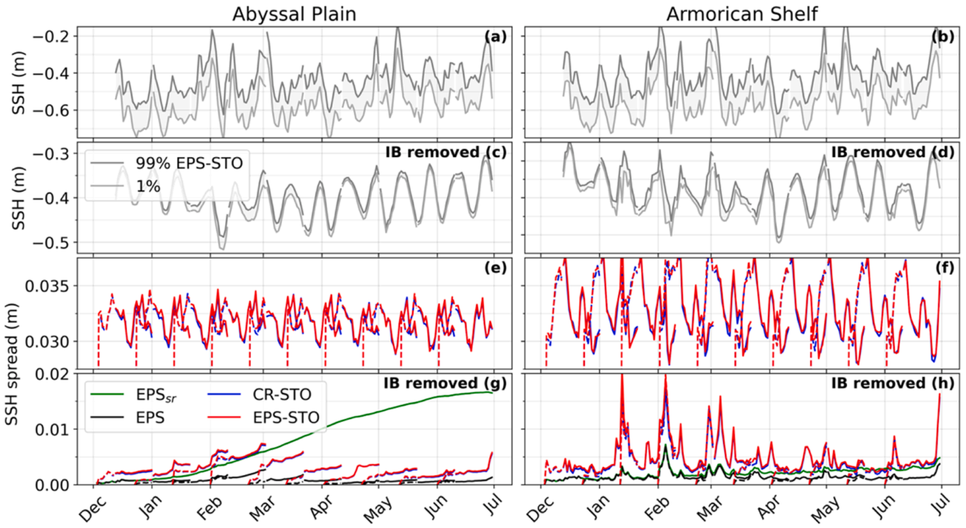


Fig. 4. Same as Fig. 3 for SSH (m). The inverse barometer is: (a-b, e-f) included, and (c-d, g-h) removed from the model solution in post-processing the model outputs. (e-f) The small spreads of the non-stochastic ensembles (i.e., green and black lines) are not shown.

opposed to the medium-range ensembles which do not have time to fully build mesoscale decorrelation. In the medium-range ensembles, the coastal areas show larger SSH model errors compared to the open ocean with limited dependence on the forecast range (Fig. 4h vs. 4 g), and associated mostly to processes in the high-frequency band in the shelves. For the stochastic ensembles, a reasonable SSH model error up to nearly 4 cm is reached after a short period of about two days and then varies, declining also in some occasions, as lead time increases (Fig. 4e-f).

Two periods of phytoplankton bloom are observed in the Bay of Biscay, associated with the two chlorophyll classes in PISCES and with their geographical distribution in the region (Fig. 5). The first bloom occurs in late-winter, early-spring in the Abyssal plain due to the abundance of nanophytoplankton, and the second in late-spring, early-summer in the coastal due to the abundance of diatoms (Vervatis et al., 2021b). The sum of the two classes yields the model equivalent of the surface total chlorophyll *a* concentration (hereafter chlorophyll). For a positive oriented logarithmic variable such as chlorophyll, the ensemble spread increases in periods of chlorophyll abundance with bloom-related model errors being large. The stochastic ensembles EPS-STO and CR-STO show larger model errors up to 0.05 mg/m^3 in comparison to the non-stochastic ensembles EPS and EPS$_{sr}$ (Fig. 5c-d). A spin-up of ten days in the medium-range ensembles appears to be sufficient to generate a noticeable chlorophyll model spread, comparable (and somewhat larger) with the seasonal-range ensemble EPS$_{sr}$.

In Fig. 6, we show the control run and the model error patterns for SST, SSH and total surface chlorophyll on April 11, 2017, for the different ensembles and lead times. Here, we mostly focus on the impact of the ECMWF-EPS atmospheric forcing to generate ocean model uncertainties. For this, we compare the non-stochastic seasonal-range ensemble EPS$_{sr}$ and the medium-range ensemble EPS-STO including ocean stochastic physics and biogeochemistry. In the seasonal-range ensemble model errors correspond to different periods and overlap each other, perhaps also cancelling each other, contaminating their statistical properties and showing a “blurring” effect in their spatial error patterns (Fig. 6b, f and g). Model errors from short- to medium-range forecasts in EPS-STO are flow-dependent, with a signal of the “errors of the day” imposed by our stochastic protocol, and their error patterns being relevant to the region’s dynamics. For instance, model errors are linked with processes being different in the open-ocean e.g., filament-like error patterns observed in mesoscale eddies, with respect to the shelves e.g., coastal river runoff processes (Fig. 6c-d, g-h and k-l). The model error imprint of the Ushant thermal front, associated with the

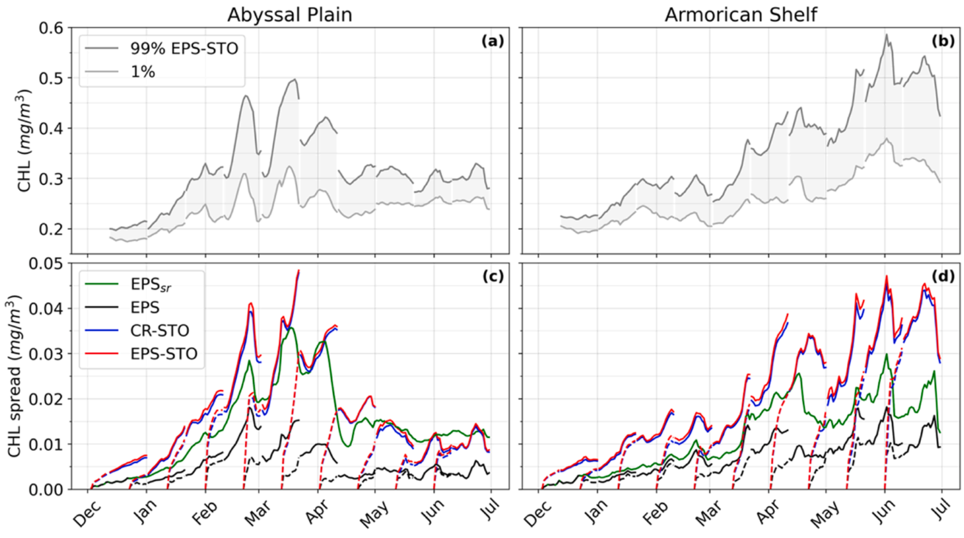


Fig. 5. Same as Fig. 3 for the model equivalent of the surface total chlorophyll *a* concentration (mg/m^3).

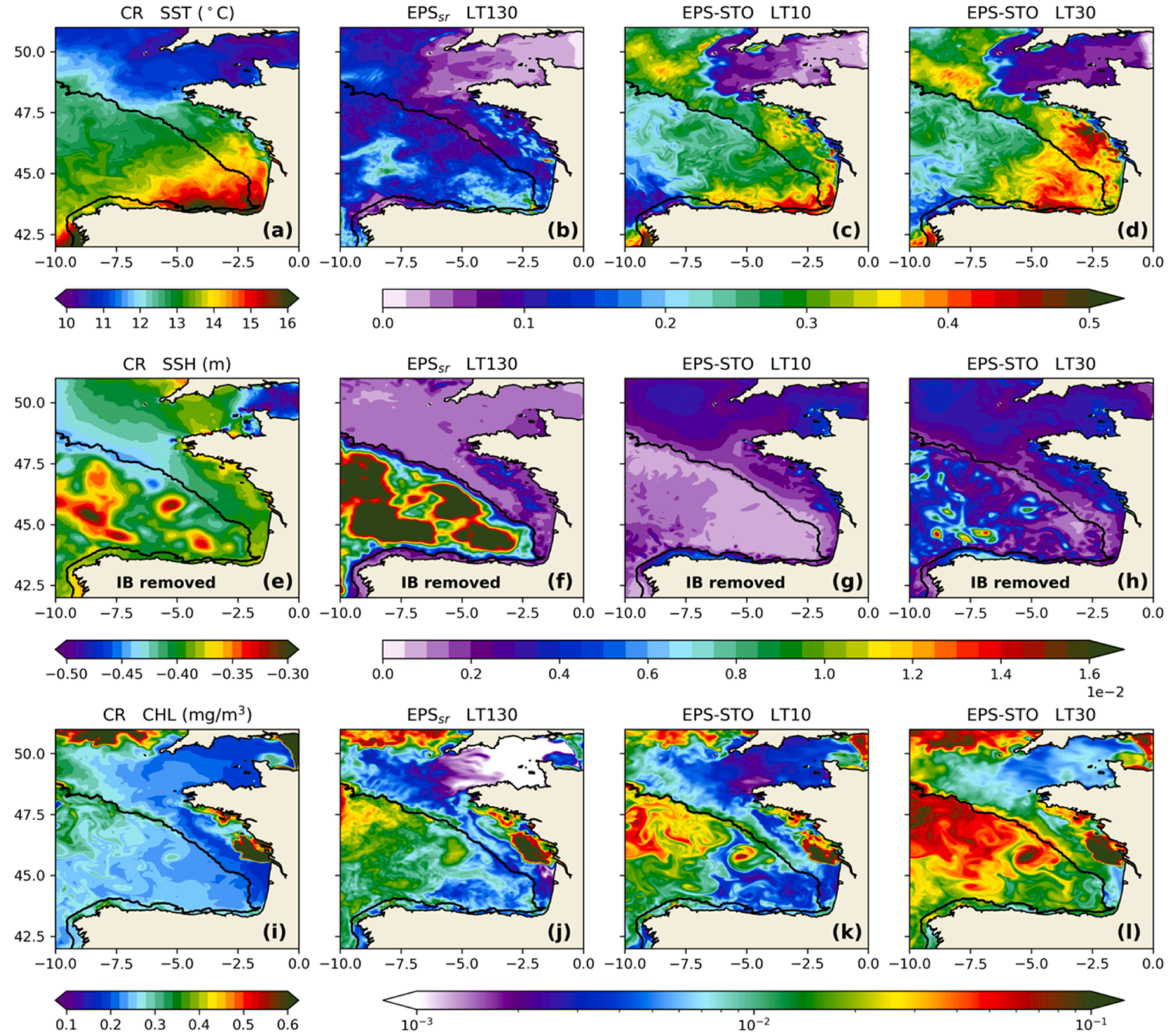


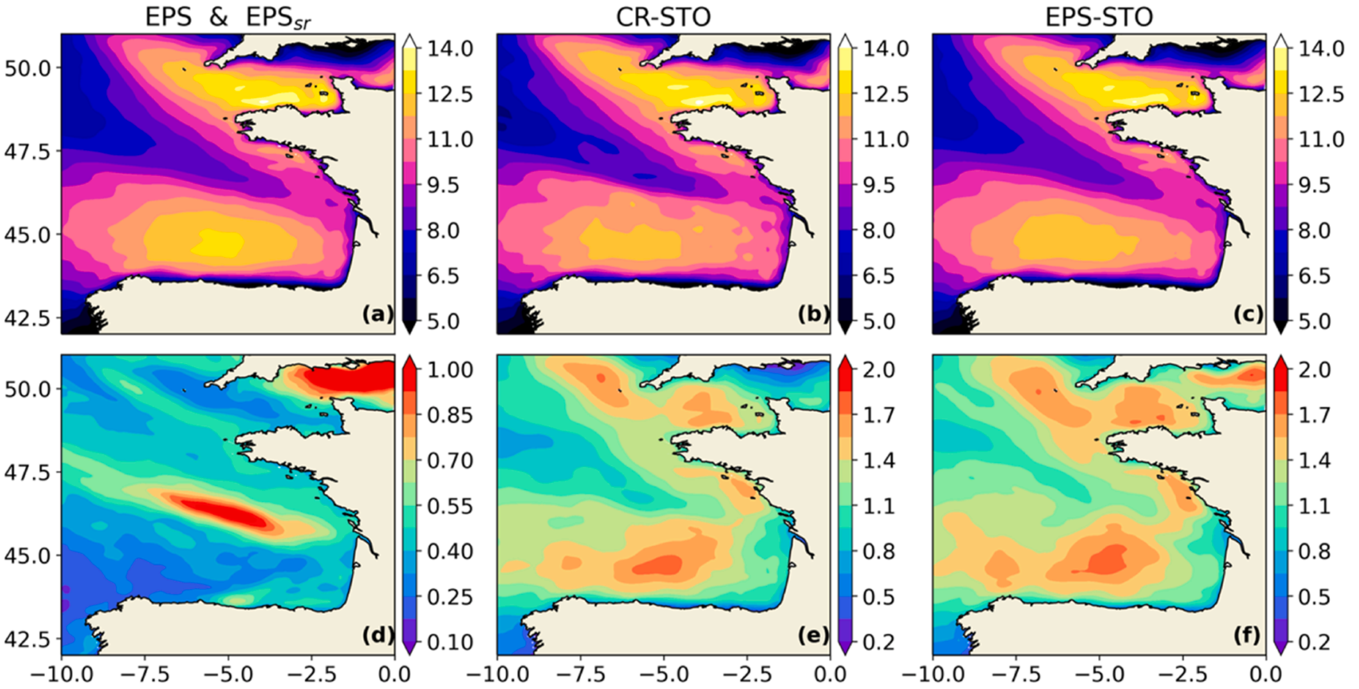
Fig. 6. (a–d) Control run and SST model spread (°C) on April 11, 2017, for the ensembles EPS$_{\text{sr}}$ (lead time: 130 day) and EPS-STO (lead time: 10 and 30 days respectively); (e–h) SSH (m) with the inverse barometer removed from the model solution; (i–l) surface total chlorophyll *a* (mg/m^3).

bottom roughness of the seabed and the tidal dynamics of the region, is revealed at the south-western edges of the English Channel. The latter is apparent especially during spring where the SST model spread is observed everywhere except in that region (Fig. 6b–d).

In Fig. 7, we investigate the efficacy of our stochastic perturbations to generate ocean model errors in comparison to the relatively small spread obtained when using only the ECMWF-EPS atmospheric forcing. The main source of ocean model errors in our forecasting system is the wind forcing, with the other perturbed variables having a moderate impact (e.g., air temperature, momentum/heat drag coefficients as in Vervatis et al., 2021b), and with an exception being the isostatic effect of the IB model errors on sea level. Fig. 7 shows the ensemble mean and spread of the wind modulus for all ensemble experiments on May 1, 2017. We note that for the atmospheric forcing the definition of the lead time is not applicable, as this is a static library of model outputs retrieved by the ECMWF global atmospheric model, used here as surface boundary conditions to our regional ocean model.

An important attribute observed in all experiments, is that the ensemble means of the wind pattern are similar to each other (Fig 7a–c) and to the deterministic control run (not shown). The latter suggests that the perturbed wind is not artificially modified when adding the SPPT-AR (1) stochastic processes to the net model tendencies. The ECMWF-HRES 9 km resolution forcing shows a few local scales in our Bay of Biscay domain for the CR-STO experiment (Fig. 7b), whereas the ECMWF-EPS 18 km coarser resolution forcing shows, as expected, broader wind patterns (Fig. 7a and c).

A marked difference is observed when comparing the model uncertainty of the wind patterns between the stochastic and non-stochastic ensembles (Fig. 7e-f vs. 7d). The model errors in the non-stochastic ensembles EPS and EPS$_{\text{sr}}$ are located in areas with marked spatial gradients in the wind forcing (Fig. 7a and d). This is due to model uncertainties of the order of ~ 1000 *km* (Leutbecher et al., 2017) in the location of large-scale weather patterns among ensemble members, imposed from the global atmospheric system to our regional ocean model domain. In contrast, in the CR-STO and EPS-STO experiments, the wind model errors are introduced by stochastic processes of the order of ~ 100 *km* (Table 1), taking under account the regional and local scales of our ocean model domain. In these experiments, the stochastic wind


EPS & EPS_{sr}

CR-STO

EPS-STO

Fig. 7. (a–c) Ensemble mean and (d–f) spread of the wind forcing modulus (m/s) for the EPS (and EPS_{sr}), CR-STO and EPS-STO ensembles on May 1, 2017. Note the different colour bars between the EPS and the two stochastic simulations CR-STO and EPS-STO.

model errors appear to collocate with the main Bay of Biscay weather patterns, which in turn are captured by the spatial truncated Gaussian distributions applied within the perturbed tendencies scheme. This opens up the question of which uncertainty estimate is the most realistic. In the next section, we will attempt to address that question by checking the consistency of both types of error estimates (stochastic vs. non-stochastic) with respect to observations, using the consistency analysis protocol of Section 2.1.

3.2. *Empirical consistency in array space*

We are now interested in assessing and comparing the consistency of those ensembles, meant as uncertainty estimates, with respect to observations within their own uncertainties (Section 2.4). The consistency analysis will be conducted in array space, which is derived from the space of observations by projecting on array modes, using a set of diagnostic metrics (Section 2.1). Also, by construction, our ensemble experiments with temporal overlap give access to consistency categorization based on the "age of errors". Here we use daily averaged fields from the ensembles.

Model ensemble averages were removed prior to consistency analysis, with a 50-member debiased ensemble corresponding to 49 degrees of freedom and thus 49 array modes. Model and data chlorophyll samples were log-transformed (Song et al., 2012) prior to computing array modes. The tolerance values τ and τ' in Eqs. (12) and (13) were set at 0.1 and 0.05, respectively. Thinning techniques were applied on high-density gridded observations (e.g., data sub-sampling rates as described in Section 2.4) for efficiency at the possible expense of array pattern consistency at very fine scales. Finally, the methodology allows for nondiagonal observational ECM, meaning that observational errors can be correlated, although we did not use that option here.

A first step is to examine how the observational arrays are able to capture the ensemble variability in the uncertainty estimates, and whether they can "detect" that variability on top of their own observational errors. To that end, we calculate the array mode eigenspectrum and the (ArM1) *array performance* metric defined in Section 2.1.

We start by examining the OSTIA SST *array performance* as per (ArM1). Fig. 8 shows Hovmöller plots of daily variations of array mode spectra vs. modal rank. For each ensemble, the spectra are calculated on the same dates, but at different lead times. Three periods of 10 days are selected in different months spanning different seasons and ocean state conditions: February 1–10, 2017, April 2–11, 2017, and June 1–10, 2017. Within those 10-day periods, we illustrate two categories of lead times: 1–10 and 21–30 days, using the successive forecast segments which overlap each other in the medium-range ensembles EPS, CR-STO and EPS-STO, and three categories of lead times: 61–70, 121–130 and 181–190 days in the seasonal-range ensemble EPS_{sr}.

The spectra in Fig. 8 are sorted down and always positive by construction. They are almost always quite red with larger dominant eigenvalues in the seasonal-range ensemble and in the stochastic medium-range ensembles. Eigenvalues smaller than 1 (grey pixels in Fig. 8) indicate reduced SST *array performance* at detecting the ensemble variability (as uncertainty estimates) above observational noise, which can be tracked by a "loss" in metric (ArM1) which becomes lesser than full rank. Please also note that the white pixel colour used for the consistency metrics described further below can hide a "grey" pixel. This relative degeneracy is mainly evident for the non-stochastic medium-range ensemble (Fig. 8i-j and p-q), with a few exceptions in early-summer (Fig. 8b-c) under the influence of the seasonal thermocline varying quite energetically across the oceanic ensemble in response to the atmospheric ensemble. As a general rule, *array performance* improves in late-spring, early-summer, reflecting the onset of the mixed layer shoaling. Also, in early-June an event appears to be captured by array mode increasing spectra, associated with the development of a low-pressure atmospheric system over the Bay of Biscay. During this period, there is a discontinuity in the way the ensemble variability (measured in observation space) is spread over the array modes. Array modes that were previously not or marginally activated show an increased variability, even at the tail of the spectra. In the same period, the ECMWF forcing shows evidence of the development of a low-pressure system over the Bay of Biscay, evolving from West to East (increased winds, drop of sea level pressure). The impact of using "aged" errors appears to be significant and positive in the stochastic medium-range cases (e.g., compare Fig. 8d, k and r to Fig. 8e, l and s), but also

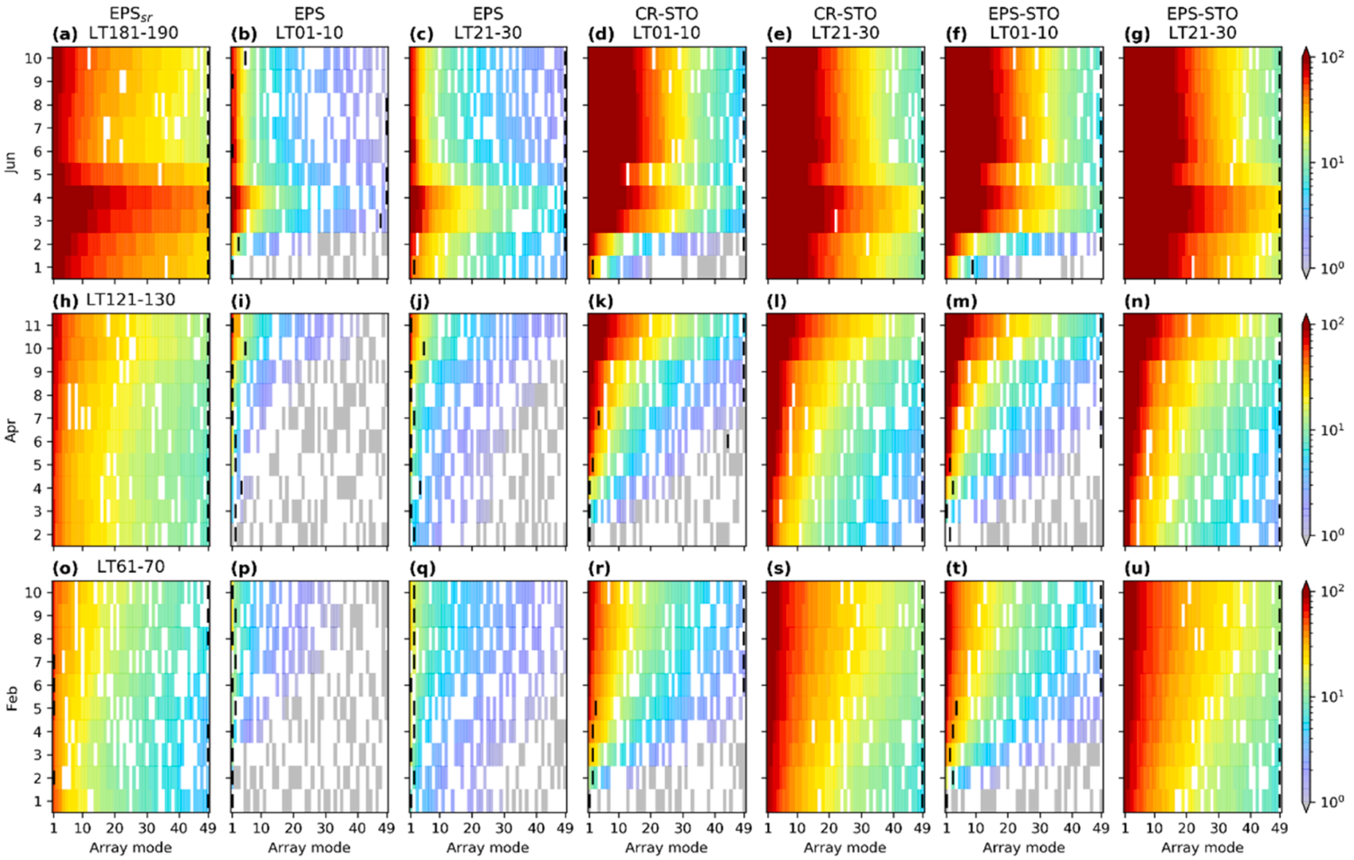


Fig. 8. Hovmöller plot of SST array mode spectra vs. modal rank against OSTIA. Three periods of 10 days are shown (from bottom to top panels): February 1–10, 2017, April 2–11, 2017, and June 1–10, 2017, corresponding to lead times 1–10 and 21–30 for the medium-range ensembles, and lead times 61–70, 121–130 and 181–190 for the seasonal-range ensemble. Colour bar: array mode spectra (array performance metric ArM1). Eigenvalues smaller than one (grey pixels) denote error modes marginally detectable by the array. White pixels depict inconsistent array modes (*gain consistency metric* ArMCA1). The vertical black bars indicate the number of consistent array modes (*alignment consistency metric* ArMCA2).

somewhat in the non-stochastic ones: the "aged" cases show higher spectral power and the absence of a visible ensemble spin-up period, compared to the "younger" cases. This would tend to confirm the need for sophisticated error-propagating data assimilation schemes if one wants to make the most efficient use of observations. Overall, the medium-range cases with the "aged" errors appear as the richest, especially in the stochastic ensembles (Fig. 8e, l and s).

We now address the empirical ensemble consistency analysis itself with respect to observed OSTIA SST values, randomized as explained in Section 2.4. We use both consistency metrics defined in Section 2.1, namely (ArMCA1) and (ArMCA2), verifying the degree of empirical consistency of our ensembles in array space, the first metric for the diagonal terms (*gain consistency*), and the second for the extradiagonal terms (*alignment consistency*). The loss of *gain consistency* as per (ArMCA1) along a particular OSTIA SST array mode appears as a white pixel in Fig. 8. In that same Figure, the number of *consistently aligned* array modes as per (ArMCA2) appears as a vertical bar, in the range [0–49] for 50 members.

Gain consistency for SST appears to be generally quite high along the dominant array modes in the spectra, with increasingly inconsistent cases observed as one moves to the tail of the spectra, i.e., where (ArM1) also fails. Overall, *gain consistency* roughly follows the rules observed for *array performance* above: it is better for the seasonal-range and for the stochastic ensembles, and better for the "aged" error cases. It is insufficient for the EPS ensembles, but results are slightly better for EPS-STO than for CR-STO.

The criterion Eq. (13) of *alignment consistency* as per (ArMCA2) was found to be quite selective, often passing for either *only dominant/no* modes or for *all* modes. Again, rules similar as for *array performance* and *gain consistency* seem to apply, but perhaps with more exceptions: *alignment* is generally good for the seasonal-range and for the stochastic ensembles, and better for the "aged" error cases. In particular, all ranks pass for the stochastic "aged" error cases, and for the seasonal-range ensemble beyond the initial spin-up. For the stochastic "young" error cases, all ranks pass beyond ensemble spin-up, especially during periods of upper-ocean re-stratification (Fig. 8d, k, f and m) – EPS-STO performing slightly better than CR-STO. For the non-stochastic cases, the *alignment* criterion passes only during re-stratification, mostly for "aged" errors (Fig. 8b-c).

We now present the same analyses on sea level, focusing on 10 consecutive days during the period April 2–11, 2017, assembling all tracks occurring in that 10-day period into a multi-track array spanning both space and time. Since there is only one occurrence of this space-time array, we will not be able to present those results in Hovmöller form, and have chosen a different visualization. As previously, this period is covered by ensembles with lead times 1–10 and 21–30 days for the medium-range ensembles, and with lead time 121–130 days for the seasonal-range ensemble. In Fig. 9, analysis results are shown for both the uncorrected and corrected sea level data against the model SLA equivalents, respectively including and excluding ("IB removed") the IB signal in post-processing the model outputs.

Applying our *array performance metric* to the stochastic medium-range ensembles including IB and using the SLA_{unc} data product from Eq. (15), most of the eigenvalues of the array mode spectra are found to be above observational noise except for a few tail modes (Fig. 9b-c). For those ensembles, and for the modes satisfying (ArM1), both *gain*

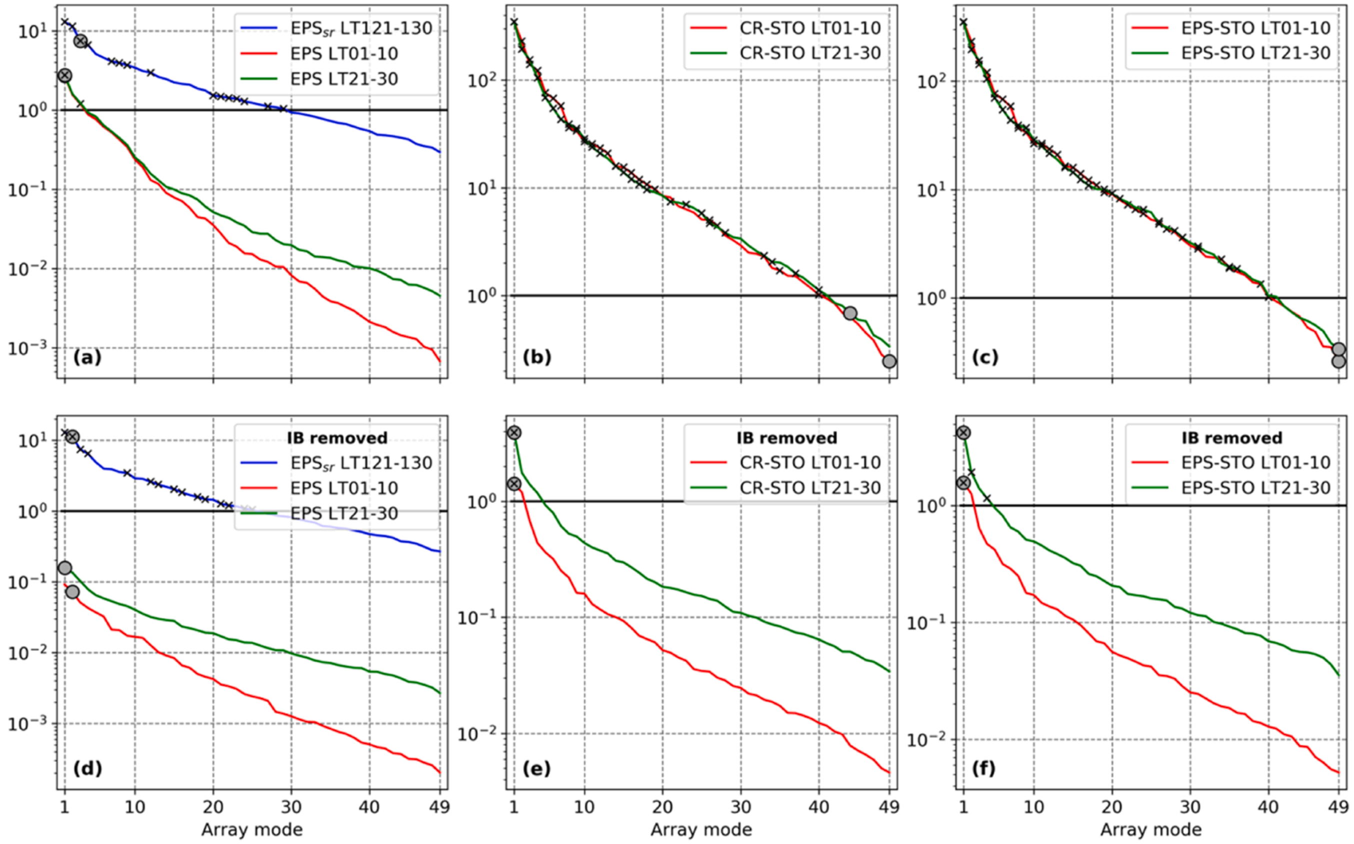


Fig. 9. Along-track SLA L3 array mode spectra for the period Apr 2–11, 2017, for the ensembles and lead times as in Fig. 8h–n, considering the array made of all tracks contained in that 10-day period. The crosses mark those eigenvalues above the observational noise floor (≥ 1) which are associated with consistent modes as per the *gain consistency metric* (ArMCA1). Greyed circles indicate in abscissa the number of consistent array modes as per the *alignment consistency metric* (ArMCA2). Panels (a–c) are for uncorrected sea level SLA_{unc} against the model equivalent including IB, and (d–f) for the corrected sea level SLA against the model equivalent with IB removed from model outputs.

consistency and *alignment consistency* are largely verified, even for short ensemble spin-up periods up to 10 days (“young” errors, red curves), almost identically for CR-STO and EPS-STO. To the contrary, the experiments in which the signal of the dynamic atmospheric process is removed (i.e., the dac term is included in the *SLA* observations retrieved from the CMEMS archives and removing the model IB in post-processing) show severely degraded *array performance* and *consistency* of both types (Fig. 9e-f). A possible reason is that, when removing the dac term from the observations, we also remove the signal from the model IB response, which in turn spans a wide range of low- to high-frequencies in a stratified ocean (Wunsch and Stammer, 1997). In the high-frequency band, the non-isostatic response to SLP is still present in the model sea level. “Flatter” spectra and lesser *array performance* can be observed for the seasonal-range ensemble compared to the medium-range stochastic ensembles (Fig. 9a–d), probably in part because of the much longer spin-up; this is observed for all periods irrespectively of the dac term. In the seasonal-range ensemble, the *consistency* for SLA is found to be average (*gain*) and mediocre (*alignment*), again irrespective of the dac term. Finally, the non-stochastic medium-range ensembles (EPS) fail all criteria for SLA.

Finally, we present analysis results for total surface chlorophyll *a* (Fig. 10). The results do bear some similarity with the SST results, but with some notable differences: lower array mode spectra (e.g. in winter, cf. Fig. 10r–u, possibly due to limited primary production), consequently lesser array performance (more “grey” in the plots), longer ensemble statistical spin-up time for chlorophyll than for SST (as already observed in Vervatis et al., 2021b), larger day-to-day variations, following the chlorophyll abundance variations in the region, as seen from the model and data. The medium-range stochastic ensembles with “aged” errors, show in most cases good *array performance* (Fig. 10e, l, s and Fig. 10g, n, u) and fairly good *consistency* of both types for those array modes passing (ArM1). Using only the ocean response to the atmospheric model ensemble, without ocean stochastic modelling, fail all criteria for chlorophyll *a* in the medium-range ensembles (Fig. 10b-c, i-j and p-q). The slight advantage of EPS-STO over CR-STO which was detected for SST and SLA is almost invisible here. In contrast, results are quite good for the seasonal-range ensemble, especially beyond the initial ensemble statistical spin-up and during the bloom onset coinciding with the spring shoaling of the thermocline (Fig. 10a and h).

4. Concluding remarks

We carried out three types of ensemble experiments, one ensemble using ocean stochastic parameterizations at regional scales (CR-STO), another one being the ocean response to a global atmospheric ensemble (EPS), and a third one using a “hybrid” approach incorporating both methods simultaneously (EPS-STO). The differences observed between the types of ensemble experiments stem from different formulations of the modelled errors of surface fluxes. By construction, the ECMWF-EPS ensemble mostly results in large-scale atmospheric forcing errors, corresponding to synoptic broad weather patterns, suitable to generate ocean model errors in global and basin-scale domains. In contrast, when performing stochastic modelling in high-resolution ocean configurations, the built-in stochastic parameterizations can be tuned to be of local relevance to the area under investigation, and can therefore produce meaningful regional and coastal ocean model errors. A prominent variable to efficiently perturb and increase ocean model spread in short-range forecasts is the wind forcing, controlling both the Ekman

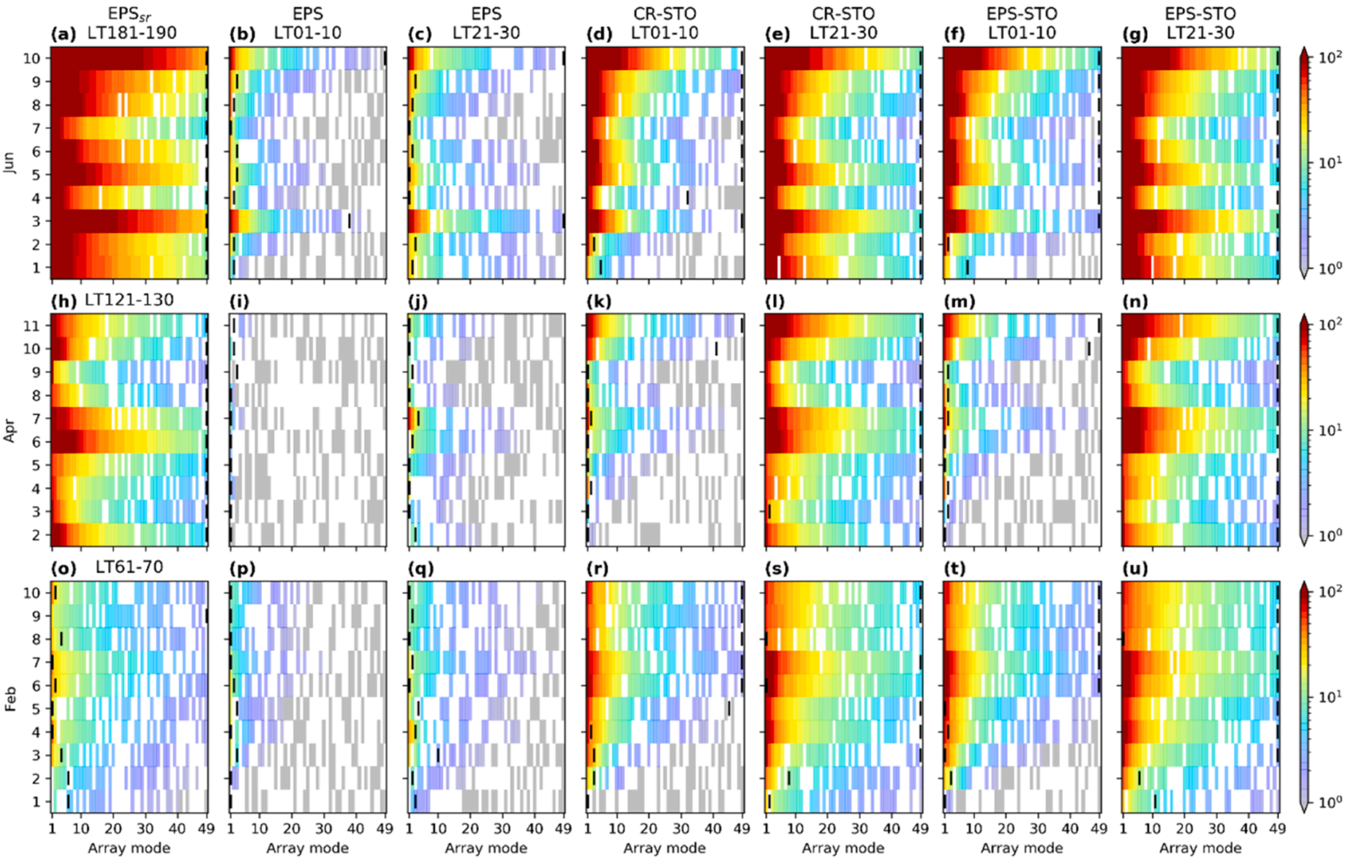


Fig. 10. Same as Fig. 8 for surface total chlorophyll *a* (distributions were log-transformed).

dynamics and the vertical mixing of the upper-ocean properties and the Sverdrup balance of the ocean interior. The efficacy of the wind perturbations to generate model uncertainties is also relevant to the SPPT method itself, in which autoregressive processes capture the main wind forcing patterns of the region. Stochastic perturbations for the SLP showed notable increase in SSH model spread via the IB effect, imposing isostatic model errors on sea level dominant at large scales, and a residual non-isostatic model error signal dominant at small scales. Stochastic perturbations in the heat fluxes can locally increase the ensemble spread for SST, sea level and chlorophyll, particularly over shelves and in coastal regions near river plumes. However, this spread is roughly an order of magnitude smaller than that caused by model uncertainties in the wind forcing (Vervatis et al., 2021b).

Ensembles of varying length were generated. In addition to a seasonal-range ensemble forced by ECMWF-EPS, three medium-range ensembles were carried out over successive segments of one month. The segments were chosen to overlap, permitting to compare our metrics for different lead times and "ages" of errors within the same period and at the same date. Each ensemble experiment, seasonal-range or medium-range, consisted of 50 members.

The largest spread was obtained when both perturbation approaches (stochastic parameterizations and ECMWF-EPS forcing) were used simultaneously, although most of the model uncertainties were attributable to the stochastic ocean parameterizations, the contribution of the ECMWF-EPS ensemble forcing being only incremental. The seasonal-range ensemble showed "blurred" error patterns and statistical contamination, because ocean model errors correspond to different "ages". In contrast, the short- to medium-range ensembles showed model errors relevant to the region's dynamics, because they are flow-dependent and include ocean "errors of the day" imposed by our stochastic protocol.

We attempted to address the question of which ensemble type and range was able to provide the most realistic model uncertainties. To that end, we chose to check the empirical consistency of both types of uncertainty estimates (stochastic vs. non-stochastic) at different lead times with respect to SST, sea-level, and chlorophyll *a* observations, using a consistency analysis protocol expanded from Vervatis et al. (2021a). By consistency check, we mean examining a covariance budget between the so-called *innovation* covariance matrix, on one side, and the sum of the observational error covariance matrix and the model error covariance matrix (as approximated from the ensemble) in the space of observations, on the other side. For the budget to be closed, and for consistency to be verified, the left-hand side must be of the order of the right-hand side.

As illustrated in several publications listed in Section 2.1 (Le Hénaff et al., 2009; Charria et al., 2016; Lamouroux et al., 2016; Vervatis et al., 2021a), it is advantageous to examine this covariance budget in the space generated by the eigenvectors of the so-called *representer matrix* (named *array modes*), that space being called *array space*. Our consistency analyses were conducted in that space. A first step, using an *array performance* metric (i.e., ArM1), was to examine how the observational arrays were able to capture the ensemble variability in the uncertainty estimates, and whether they could "detect" that variability on top of their own observational errors, which is a prerequisite if data assimilation was to be considered in a later step. Steps two and three (the empirical consistency analysis proper) aimed at evaluating the terms of the aforementioned covariance budget in array space, and involved the calculation of novel *gain consistency* (i.e., ArMCA1; Eq. (12)) and *alignment consistency* (i.e., ArMCA2; Eq. (13)) metrics.

It is important to recall that the results of the consistency analysis, and the conclusions we draw from it, largely depend on [A] the assumptions on the observational errors, and on the observational error models, which are in general uncertain and rarely subject to any form of validation in the literature, and [B] on the protocols used to generate our ensembles. Regarding [A], it should be noted that as Eqs. (1) and (10) show, our statistical consistency approach bears some form of symmetry: it can either be seen as testing ensemble-based uncertainties against observations, or testing observations and their error models against model priors. Also, our approach here cannot be used by itself to diagnose biases, to the contrary of some of the approaches of Vervatis et al. (2021a) for instance.

With some exceptions, we found the *performance* of our SST, SLA and chlorophyll *a* arrays quite satisfactory at detecting a significant number of degrees of freedom of our ensembles above observational error. However, the non-stochastic medium-range ensembles (forced by ECMWF-EPS with no other perturbations) offered subpar *array performance*. Also, a significant statistical spin-up period was observed for the "youngest errors" ensemble, affecting *array performance* as well. In contrast, *array performance* at detecting "aged" errors appears to be quite good, confirming the advantage of error-propagating data assimilation schemes if one wants to make the most efficient use of observations to correct a model.

In our results, *gain* and *alignment consistency* metrics were enhanced [A] for the stochastic ensembles (in particular with respect to the quite *inconsistent* medium-range ensembles forced by ECMWF-EPS), [B] for the "aged" error cases (but only marginally with respect to the "young" error cases), and [C] whenever physical and biogeochemical uncertainty processes were active in the region and could be detected by the observational networks. Fair consistency examples of SST and chlorophyll are observed during the spring shoaling of the thermocline and the phytoplankton abundance primary bloom, pertaining to the fact that these processes are susceptible to model error sources considered in our stochastic ensembles, and relevant to the spatial distribution of the data errors themselves. The leading modes of the eigenvectors, as observed by the operator $\boldsymbol{H}$ for SST and chlorophyll respectively, appear to be typical of large-scale and mesoscale (not shown; cf. Vervatis et al., 2021a). The sea-level *consistency* metrics are improved when model and data errors associated with high-frequency atmospheric processes are retained. In this case, both isostatic and non-isostatic IB model errors can be detected in array space by altimetric data, as long as the dynamic atmospheric correction term is omitted from the analysis. In the same line of thinking, retaining the high-frequency tidal signal (not the case here, since we used detided daily model outputs) could be expected to further improve the sea-level *consistency* metrics.

Empirical consistency was found to be subpar for the EPS ensembles, but results were slightly better for EPS-STO than for CR-STO, illustrating that ECMWF-EPS has the capacity to enrich the stochastic degrees of freedom of a regional model, even if it is more representative of global atmospheric uncertainties. However, our results have shown that for our study variables, ECMWF-EPS is not able by itself to generate consistent stochastic degrees of freedom in a regional model.

We have shown synthetic metrics in array space (a transformation of data space). It should be kept in mind that the *performance* and *consistency* analyses can vary spatially within the study domain. For instance, in areas close to the coast, the observational error is often higher, especially for satellite observations (this was the case here for OSTIA SST and ocean colour). This could have the effect of reducing the dimension of the detectable subspace, hence making *performance* and *consistency* goals less attainable and global metrics less relevant. As a consequence, the metrics described in this could be profitably applied in subdomains as a complement to the global metrics.

The approaches developed in this work do not allow us to directly address the question of the validity of the ensemble estimate of the prior state-space ECM, insofar as we are working in a subspace of the state space. The question of ensemble sampling errors, which is obviously of importance for data assimilation, would need to be studied specifically, for example via sensitivity studies. However, the methodology in this work allows us to compare the ensemble estimate of the prior ECM and an innovation-based estimate by means of gain and alignment metrics in the space of array modes. Anomalies detected in the hierarchical space of array modes, if relevant, will be symptomatic of anomalies in the state space. Consequently, it will probably make sense to undertake such an analysis when observational datasets are available at the scales where such spurious correlations are likely to occur.

Let us note that beyond mere consistency analysis, discrepancies in the empirical consistency analysis could possibly be transformed into useful information on the structure of the usually very poorly known observational ECM, e.g., in the form of an adaptive update while assimilating (in the spirit of Desroziers et al., 2005).

In a future work, we intend to expand the EECA formalism to include bias consistency analysis, and alternate approaches to compare ensembles and observations and their errors, such as those using probabilistic scores.

CRediT authorship contribution statement

Vassilios D. Vervatis: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pierre De Mey-Frémaux:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **John Karagiorgos:** Visualization, Validation, Software, Data curation. **Bénédicte Lemieux-Dudon:** Software, Methodology, Formal analysis. **Nadia K. Ayoub:** Methodology, Formal analysis, Conceptualization. **Sarantis Sofianos:** Resources, Conceptualization.

Declaration of competing interest

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.

Acknowledgments

This work was carried out as part of the Copernicus Marine Service (CMEMS) "Stochastic Coastal/Regional Uncertainty Modelling 2 (SCRUM2)" Service Evolution project. CMEMS is implemented by Mercator Ocean International in the framework of a delegation agreement with the European Union. 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 computing and archiving facilities in this research. The publication of the article in OA mode was financially supported by HEAL-Link. The model data will be made available by the corresponding author upon reasonable request. We thank two anonymous reviewers for their constructive comments.

Data availability

Data will be made available on request.

References

Andersson, E., 2003. Modelling of innovation statistics. In: Proceedings of the Workshop on Recent Developments in Data Assimilation for Atmosphere and Oceans. Reading, UK. ECMWF, pp. 153–164.

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

Bessières, L., Leroux, S., Brankart, J.M., Molines, J.M., Moine, M.P., Bouttier, P.A., Penduff, T., Terray, L., Barnier, B., 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 Model. 66, 64–76. https://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., 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., Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Q. J. R. Meteorol. Soc. 125, 2887–2908. https://doi.org/10.1002/qj.49712556006.

Carrère, L., Lyard, F., 2003. Modeling the barotropic response of the global ocean to atmospheric wind andpressure forcing - comparisons with observations. Geophys. Res. Lett. 30 (6), 1275. https://doi.org/10.1029/2002GL016473.

Charria, G., Lamouroux, J., De Mey, P., 2016. Optimizing observational networks combining gliders, moored buoys and FerryBox in the Bay of Biscay and English Channel. J. Mar. Syst. 162, 112–125. https://doi.org/10.1016/j.jmarsys.2016.04.003.

De Mey-Frémaux, P., 2023. Array-space analysis tools (ArM tools): a compendium (Version 3). Zenodo. https://doi.org/10.5281/zenodo.8208691.

De Mey-Frémaux, P., Ayoub, N., Barth, A., Brewin, R., Charria, G., Campuzano, F., Ciavatta, S., Cirano, M., Edwards, C.A., Federico, I., Gao, S., Garcia Hermosa, I., Garcia Sotillo, M., Hewitt, H., Hole, L.R., Holt, J., King, R., Kourafalou, V., Lu, Y., Mourre, B., Pascual, A., Staneva, J., Stanev, E.V., Wang, H., Zhu, X., 2019. Model-observations synergy in the Coastal Ocean. Front. Mar. Sci. 6, 436. https://doi.org/10.3389/fmars.2019.00436.

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.

Egbert, G.D., Bennett, A.F., Foreman, M.G.G., 1994. TOPEX/POSEIDON tides estimated using a global inverse model. J. Geophys. Res. Ocean 99, 24821–24852.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367.

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

Holm, DD., 2015. Variational principles for stochastic fluid dynamics. Proc. R. Soc. A 471, 20140963. https://doi.org/10.1098/rspa.2014.0963.

Ide, K., Courtier, Ph., Ghil, M., Lorenc, A., 1997. Unified notation for data assimilation: operational, sequential and variational. J. Met. Soc. Jpn. 75 (1B), 181–189. Special Issue on "Data Assimilation in Meteorology and Oceanography: Theory and Practice".

Implementing Operational Ocean Monitoring and Forecasting Systems, IOC-UNESCO, GOOS-275, 2022, https://doi.org/10.48670/ETOOFS.

Karagiorgos, J., Vervatis, V., Sofianos, S., 2020. The impact of tides on the bay of biscay dynamics. J. Mar. Sci. Eng. 8 (617).

Lamouroux, J., Charria, G., De Mey, P., Raynaud, S., Heyraud, C., Craneguy, P., Dumas, F., Le Hénaff, M., 2016. Objective assessment of the contribution of the RECOPESCA network to the monitoring of 3D coastal ocean variables in the Bay of Biscay and the English Channel. Ocean Dyn. 66 (4), 567–588. https://doi.org/10.1007/s10236-016-0938-y.

Lang, O., Crisan, D., Mémin, É., 2023. Analytical properties for a stochastic rotating shallow water model under location uncertainty. J. Math. Fluid Mech. 25, 29. https://doi.org/10.1007/s00021-023-00769-9.

Le Hénaff, M., De Mey, P., Marsaleix, P., 2009. Assessment of observational networks with the representer matrix spectra method – application to a 3-D coastal model of the Bay of Biscay. Ocean Dyn. 59, 3–20. https://doi.org/10.1007/s10236-008-0144-7.

Lellouche, J.M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., Benkiran, M., Testut, C.E., Bourdalle-Badie, R., Gasparin, F., Hernandez, O., Levier, B., Drillet, Y., Remy, E., Le Traon, P.Y., 2018. Recent updates to the Copernicus marine service global ocean monitoring and forecasting real-time 1/12° high-resolution system. Ocean Sci. 14, 1093–1126. https://doi.org/10.5194/os-14-1093-2018.

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., 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.

Leroux, S., Brankart, J.M., Albert, A., Brodeau, L., Molines, J.M., Jamet, Q., Le Sommer, J., Penduff, T., Brasseur, P., 2022. Ensemble quantification of short-term predictability of the ocean dynamics at a kilometric-scale resolution: a Western Mediterranean test case. Ocean Sci. 18, 1619–1644. https://doi.org/10.5194/os-18-1619-2022.

Le Traon, P.Y., Nadal, F., Ducet, N., 1998. An improved mapping method of multisatellite altimeter data. J. Atmos. Ocean. Technol. 15, 522–534.

Leutbecher, M., Lock, S.J., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H.M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R.M., Goddard, J., Haiden, T., Hogan, R.J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P.K., Subramanian, A., Vitart, F., Wedi, N., Weisheimer, A., 2017. Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. Q. J. R. Meteorol. Soc. 143, 2315–2339. https://doi.org/10.1002/qj.3094.

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

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

Mayorga, E., Seitzinger, S.P., Harrison, J.A., Dumont, E., Beusen, A.H.W., Bouwman, A. F., Fekete, B.M., Kroeze, C., Van Drecht, G., 2010. Global nutrient export from WaterSheds 2 (NEWS 2): model development and implementation. Environ. Model. Softw. 25, 837–853.

Moore, A.M., Levin, J., Arango, H.G., Wilkin, J., 2021. Assessing the performance of an ocean observing, analysis and forecast system for the Mid-Atlantic Bight using array modes. Ocean Model. 164, 101821. https://doi.org/10.1016/j.ocemod.2021.101821.

Ollinaho, P., Lock, S., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R.M., Haiden, T., Hogan, R.J., 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. https://doi.org/10.1002/qj.2931.

Palmer, T., 2018. The ECMWF ensemble prediction system: looking back (more than) 25 years and projecting forward 25 years. Q. J. R. Meteorol. Soc. 145, 12–24. https://doi.org/10.1002/qj.3383.

Quattrocchi, G., De Mey, P., Ayoub, N., Vervatis, V., Testut, C.E., Reffray, G., Chanut, J., 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. J. Oper. Oceanogr. 7 (2), 25–34, 10.

Chapter 9 Roblou, L., Lamouroux, J., Bouffard, J., Lyard, F., Le Hénaff, M., Lombard, A., Marsalaix, P., De Mey, P., Birol, F., 2011. Post-processing altimiter data towardcoastal applications and integration into coastal models. In: Vignudelli, S., Kostianov, A.G, Cipollini, P., Benveniste, J. (Eds.), Coastal Altimetry. Springer, Berlin Heidelberg.

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 Model. 54–55, 1–17. https://doi.org/10.1016/j.ocemod.2012.06.001.

Sotillo, M.G., Cailleau, S., Lorente, P., Levier, B., Aznar, R., Reffray, G., Amo-Baladrón, A., Chanut, J., Benkiran, M., Alvarez-Fanjul, E., 2015. The MyOcean IBI ocean forecast and reanalysis systems: operational products and roadmap to the future copernicus service. J. Oper. Oceanogr. 8 (1), 63–79. https://doi.org/10.1080/1755876X.2015.1014663.

Storto, A., Andriopoulos, P., 2021. A new stochastic ocean physics package and its application to hybrid-covariance data assimilation. Q. J. R. Meteorol. Soc. 147, 1691–1725. https://doi.org/10.1002/qj.3990.

Vandenbulcke, L., 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.

Vervatis, V.D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ciavatta, S., Brewin, R., Sofianos, S., 2021a. Assessment of a regional physical-biogeochemical stochastic ocean model. Part 2: empirical consistency. Ocean Model. 160, 101770. https://doi.org/10.1016/j.ocemod.2021.101770.

Vervatis, V.D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ghantous, M., Kailas, M., Testut, C.E., Sofianos, S., 2021b. Assessment of a regional physical-biogeochemical stochastic ocean model. Part 1: ensemble generation. Ocean Model. 160, 101781. https://doi.org/10.1016/j.ocemod.2021.101781.

Vervatis, V.D., Testut, C.E., De Mey, P., Ayoub, N., Chanut, J., 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 Model. 100, 1–19. https://doi.org/10.1016/j.ocemod.2016.01.003.

Wunsch, C., Stammer, D., 1997. Atmospheric loading and the oceanic "inverted barometer" effect. Rev. Geophys. 35, 79–107. https://doi.org/10.1029/96RG03037.

Zuo, H., Balmaseda, M.A., Tietsche, S., Mogensen, K., 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.