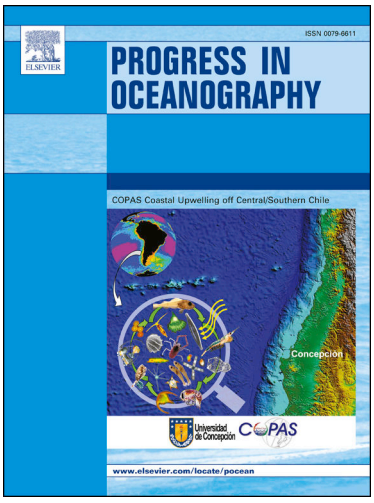
Journal Pre-proofs



Control of simulated ocean ecosystem indicators by biogeochemical observations

S. Ciavatta, P. Lazzari, E. Álvarez, L. Bertino, K. Bolding, J. Bruggeman, A. Capet, G. Cossarini, F. Daryabor, L. Nerger, M. Popov, J. Skákala, S. Spada, A. Teruzzi, T. Wakamatsu, V.Ç. Yumruktepe, P. Brasseur

PII: S0079-6611(24)00190-3
DOI: https://doi.org/10.1016/j.pocean.2024.103384
Reference: PROOCE 103384

To appear in: *Progress in Oceanography*

Received Date: 29 November 2023
Revised Date: 27 September 2024
Accepted Date: 15 November 2024

Please cite this article as: Ciavatta, S., Lazzari, P., Álvarez, E., Bertino, L., Bolding, K., Bruggeman, J., Capet, A., Cossarini, G., Daryabor, F., Nerger, L., Popov, M., Skákala, J., Spada, S., Teruzzi, A., Wakamatsu, T., Yumruktepe, V.Ccedil., Brasseur, P., Control of simulated ocean ecosystem indicators by biogeochemical observations, *Progress in Oceanography* (2024), doi: https://doi.org/10.1016/j.pocean.2024.103384

Control of simulated ocean ecosystem indicators by biogeochemical observations.

Ciavatta, S.[1], Lazzari, P.[2], Álvarez[2], E., Bertino[3], L., Bolding[4], K., Bruggeman[4], J., Capet[5], A., Cossarini, G. [2], Daryabor[6], F.[5], Nerger, L. [6], Popov, M.[7], Skákala, J.[8,9], Spada, S. [2], Teruzzi, A. [2], Wakamatsu, T.[3], Yumruktepe, V. Ç. [3], and P. Brasseur[7].

[1] Mercator Ocean international, France

[2] National Institute of Oceanography and Applied Geophysics, Italy

[3] Nansen Environmental and Remote Sensing Center, Norway

[4] Bolding and Bruggeman ApS, Denmark

[5] Royal Belgian Institute of Natural Sciences, Belgium

[6] Alfred-Wegener-Institut Helmholtz-Zentrum für Polar und Meeresforschung, Germany

[7] Université Grenoble Alpes, France

[8] Plymouth Marine Laboratory, United Kingdom

[9] National Centre for Earth Observation, United Kingdom

- Simulation of phytoplankton phenology was improved by merging observations with models
- Uncertainty on plankton community structure and production decreased
- Uncertainty of trophic efficiency and POC might be increased by chlorophyll assimilation
- Controllability depended on carbon to nutrient and to chlorophyll formulations
- Assessment of controllability should be a standard practice in operational modeling

Abstract

To protect marine ecosystems threatened by climate change and anthropic stressors, it is essential to operationally monitor ocean health indicators. These are metrics synthetizing multiple marine processes relevant to the users of operational services. Here we assess if selected ocean indicators simulated by operational models can be controlled (here meaning constrained effectively) by biogeochemical observations, by using a newly proposed methodological framework. The method consists in firstly screening the sensitivities of the indicators with respect to the initial conditions of the observable variables. These initial conditions are perturbed stochastically in Monte Carlo simulations of one-dimensional configurations of a multi-model ensemble. Then, the models are applied in three-dimensional ensemble assimilation experiments, where the reduction of the ensemble variance corroborates the controllability of the indicators by the observations. The method is applied for ten relevant ecosystem indicators (ranging from inorganic chemicals to plankton production), seven observation types (representing data from satellite and underwater platforms),

and an ensemble of five biogeochemical models of different complexity, employed operationally by the European Copernicus Marine Service. We demonstrate that all the indicators are controlled by one or more types of observations. In particular, the indicators of phytoplankton phenology are controlled and improved by the merged observations from the surface ocean colour and chlorophyll profiles. Similar observations also control and reduce the uncertainty of the plankton community structure and production. However, the uncertainty of the trophic efficiency and POC increases when assimilating chlorophyll-a data, though observations were not available to assess whether that was due to a worsen model skill. We recommend that the assessment of controllability proposed here becomes a standard practice in designing operational monitoring, reanalysis and forecast systems, to ultimately provide the users of operational services with more precise estimates of ocean ecosystem indicators.

Keywords: marine ecosystem models, ocean indicators, operational oceanography, data assimilation, biogeochemical observations, ocean colour, biogeochemical-Argo floats, Copernicus Marine Service, sensitivity analysis, controllability, plankton, carbon fluxes

1. Introduction

The ocean ecosystems provide us with vital services, by absorbing 30% of the anthropogenic carbon emissions and supporting the livelihoods of over three billion people (UNESCO-IOC, 2021). These services are endangered by human pressures and the climate crisis; therefore, the ocean health is routinely monitored by marine operational centers (Fennel et al., 2019). These centers use satellite sensors, *in situ* observing networks and operational marine ecosystem models to observe and predict the state of the ocean (Le Traon et al., 2019). The state can be characterized by means of ecosystem indicators, which are metrics synthetizing multiple marine processes and responding to the needs of stakeholders concerned by changes in ocean health and services. An example of biogeochemical indicator is the UN Sustainable Development Goal (SDG) Indicator 14.3.1 “Average marine acidity (pH) measured at agreed suite of representative sampling stations”. This indicator is used to monitor the ocean acidification caused by the human emission of carbon dioxide in the atmosphere (Barbiere et al., 2019).

Biogeochemical indicators are computed operationally by constraining model states with observations, by using data assimilation approaches (Le Traon et al., 2019). In such approaches, trusted-but-sparse observations correct continuous-but-uncertain model estimates of ocean biogeochemistry. This leads operational centers to provide more accurate and precise predictions of the “true value” of the biogeochemical states in the real system. For this to happen, however, the biogeochemical states need to be “controllable”, i.e. they can reach a subspace that includes the truth, by means of a suitable choice of values of the “control” variables. The latter can be the biogeochemical variables linked to the ocean observations that are assimilated in the operational model (see, e.g., Gelb et al., 1974, for the equivalence of the controllability problem in system engineering and data assimilation).

Most of the modern operational centers are assimilating ocean-colour chlorophyll-a, which is a proxy of the phytoplankton biomass driving the biogeochemical cycles in lower trophic level models (Sathyendranath et al., 2023). Such data are retrieved from satellite swathes, which typically cover large part of the ocean surface with high spatial and temporal resolution, if not covered by clouds (resolutions in the order 10^1-10^3 m and 1-3 days, Groom et al., 2019). The controllability of indicators by ocean colour, here defined as the effectiveness of the assimilated data to constrain simulated indicators, might be limited (i) to the ocean upper layers, and (ii) to the indicators with strong dynamical link to surface chlorophyll-a (Wang et al., 2020, 2022). These limits add up to the uncertainty of the ocean-colour data and the representation error of complex ecosystem dynamics, intrinsic to all types of observations and data assimilation systems (Janjić et al., 2018).

The limitations as well as benefits of observation alternative to ocean colour in controlling biogeochemical indicators have rarely been assessed in operational systems. Remarkable exceptions have led to breakthrough advances of these systems in estimating indicators, such as plankton community structure by assimilating remote sensing reflectance in the Great Barrier Reef (Jones at al., 2016), air-sea carbon flux by assimilating phytoplankton functional types in the North West European shelf-seas (Skákala et al., 2018) and vertical phenology of plankton by assimilating profiles of chlorophyll-a and nutrients from biogeochemical-Argo floats in the Mediterranean Sea (Teruzzi et al, 2021). However, these outcomes on controllability are hardly generalizable because they depend on the specific features of: (i) the investigated marine ecosystems (e.g. differences in open-ocean and shelf-sea trophic networks), (ii) the biogeochemical models (e.g. their level of details in representing the complex networks), (iii) their set-ups (e.g., hydrodynamic and climate forcings used to drive the models), and (iv) the assimilation methods (e.g., ensemble versus variational characterization of both model and data errors). Comparing how the above features i-iv impact the assimilative control of indicators is often prevented by the high computational cost of the three-dimensional assimilative models used in operational centers. One-dimensional configurations have therefore been used for objective multi-model comparisons, for example in parameter optimization studies (Friedrichs et al., 2006 and 2007). However, as the portability of optimized parameters from 1D to 3D configurations might be unstraight (e.g., Wang et al., 2020), so might be the still unexplored portability of control analysis from 1D to 3D systems. As an ultimate consequence, new opportunities to better simulate ecosystem indicators by integrating novel observing systems are currently hindered by a knowledge gap on the controllability of indicators by observations.

The objective of this work is to assess the controllability of ecosystem indicators simulated by assimilative operational models, by using a generalizable framework.

The framework includes ten indicators, seven observation types, and five biogeochemical models of different complexity, employed operationally in the European regional seas and global ocean by the Copernicus Marine Service (see Methods in Section 2). Firstly, one-dimensional model configurations underwent a controllability screening analysis, by using the same forcing and ensemble set-up for all the models (see results and discussion in Section 3.1). Then, the controllability of the most relevant indicators and observations was corroborated in three-dimensional configurations of the models used in real operational systems (see results and discussion in Section 3.2). The 1D screening and 3D corroboration results are compared and discussed in Section 3.3. Recommendations on exploiting the controllability assessment framework in operational centers are provided in the concluding Section 4.

2 Methods

2.1 The biogeochemical models

We used five well-established biogeochemical models which run operationally in the Monitoring and Forecasting Centres (MFCs) of the Copernicus Marine Service (Table 1). The models differ for complexity level, here approximated by the number of phytoplankton functional types (PFTs): two in the model PISCES, three in ECOSMO and ERGOM, four in ERSEM and BFM. Previous studies have defined model complexity as a function of the total number of state variables, but found that the number of PFTs was the critical feature influencing model skill and portability (Friedrichs et al., 2007) The models equations, implementations and validations have been described thoroughly in previous publications (see the review by Le Traon et al., 2019 and references therein) and we recall only the model features that are relevant here.

The PISCES model (Aumont et al. 2015) is an intermediate-complexity, carbon-based model that simulates marine biological productivity and carbon biomass based upon five main nutrients: nitrate, ammonium, phosphate, silicate and iron. Its architecture includes twenty-four biogeochemical

variables grouped into four main compartments: nutrients, phytoplankton, zooplankton, and detritus. The phytoplankton compartment is represented by two different PFTs, namely the nanophytoplankton and the diatoms, while the zooplankton compartment contains two size classes, namely micro- and mesozooplankton. The detritus compartment is divided in two pools: the dissolved organic matter (DOM) and the particulate organic matter (POM) including the small particles and the big particles which mostly differ by their sinking velocity. PISCES is used in both the Copernicus Marine Service MFCs of the global ocean (GLO) and of the Iberian-Biscay-Irish Seas (IBI).

The ECOSMO model (Daewel and Schrum, 2013) is an intermediate-complexity model with four nutrients (nitrate, ammonium, phosphate and silicate) and three PFTs (flagellates, diatoms and nitrogen fixing cyanobacteria, though the latter group was turned off in this study). Zooplankton are represented by two size classes (micro and meso). The model includes DOM, POM and oxygen. A simple representation of the sediment layer including sedimentation, resuspension, remineralization and denitrification under low oxygen conditions is included in the model. The model currency is carbon, and a fixed Redfield ratio is applied to the other elements. ECOSMO was originally developed for the North and Baltic Sea. It was then coupled to the ocean model HYCOM and adapted to the North Atlantic and Arctic with a new formulation for prognostic chlorophyll-a (ECOSMO II(CHL); Yumruktepe et al., 2022). ECOSMO II(CHL) is used by the Arctic (ARC) MFC and was investigated in this work.

The ERGOM model (Neumann, 2000; Lessin et al., 2014) is an intermediate complex nitrogen-based model that simulates the biogeochemical cycling in the coastal seas using three PFTs (cyanobacteria, flagellates, diatoms), two zooplankton size groups (micro and meso), four nutrient groups (nitrate, ammonium, phosphate, and silicate), two detritus groups (N-Detritus and Si-Detritus), oxygen and labile dissolved organic nitrogen in the water column. A carbonate system is included to compute partial pressure of CO_2 (pCO2), alkalinity and pH. In the original ERGOM formulation investigated here, chlorophyll-a concentration is computed diagnostically from the phytoplankton nitrogen biomass, with implications on the controllability discussed in Section 3.1. ERGOM was originally developed for the Baltic Sea and later extended for the North Sea (Maar et al., 2011). It is used by the Baltic (BAL) MFC in the configuration by Neumann et al., (2015).

The high-complexity models are ERSEM (Baretta-Bekker et al., 1995, Butenschön et al., 2016) and BFM (Vichi et al., 2015 and Salon et al., 2019). They have several features in common, both in their structures and outputs, making it convenient to list their difference rather than describing them separately, both here and in the section of the results. Both models distinguish between five chemical components: carbon, nitrogen, phosphorus, silicon and chlorophyll-a, using variable stoichiometry for the simulated plankton groups (e.g., Geider et al., 1997; Baretta-Bekker et al., 1995). The models split phytoplankton into four PFTs largely based on their size (e.g., Baretta-Bekker et al., 1995): picophytoplankton, nanophytoplankton, diatoms and dinoflagellates. Each PFT biomass is represented in terms of variable chlorophyll-a, carbon, nitrogen and phosphorous components, with diatoms also represented by silicon. Model predators are composed of three (ERSEM), or four (BFM) zooplankton types, with organic material being decomposed by one functional type of heterotrophic bacteria. The model inorganic component consists of carbon, nutrients (nitrate, phosphate, silicate, ammonium) and dissolved oxygen. One versus two particulate matter size-classes are included in BFM and ERSEM, respectively. The carbonate system is included in both ERSEM (Artioli et al., 2012) and BFM (Cossarini et al., 2015; Canu et al., 2015). Here we used the ERSEM configuration by Butenschön et al. (2016), with the addition of a bio-optical module and two-way physical-biogeochemical coupling in the three-dimensional configuration (Skákala et al., 2022). The BFM configuration is the one described in Salon et al. (2019), Lazzari et al. (2012, 2016). ERSEM and BFM are used in the North-West Shelf-Seas (NWS) and Mediterranean (MED) regions, respectively.

Table 1 Main differences in the features of the five biogeochemical models considered in this work

	PISCES	ECOSMO	ERGOM	BFM	ERSEM

Complexity	Intermediate	Intermediate	Intermediate	High	High
Number of phytoplankton types	Three	Three	Three	Four	Four
Number of zooplankton types	Two	Two	Two	Four	Thee
Number of size classes particulate matter	Two	One	One	One	Two
Number of bacteria functional types	None	None	None	One	One
Carbon:Nutrients	Fixed (C-based)	Fixed (C-based)	Fixed (N-based)	Variable	Variable
Chlorophyll-a:Carbon (or Nitrogen)	Variable	Variable	Fixed	Variable	Variable

2.2 The marine ecosystem indicators

In this work, we evaluated the controllability of ten ecosystem indicators that are relevant to operational systems because they are linked to the: (1) United Nations Sustainable Development Goals (UN SDGs; United Nations, 2023); (2) The Global Ocean Observing System Essential Ocean and Biodiversity Variables (EOVs and EBVs; Muller-Karger et al., 2018); (3) Copernicus Marine Service Ocean Monitoring Indicators (Le Traon et al., 2019) and Copernicus Marine Service user needs (European Commission, 2019). Additional criteria for the selection of the indicators were: (i) the capability to simulate them with operational models; (ii) the verifiable capability to constrain their simulation by using new operational monitoring infrastructures, besides ocean-colour chlorophyll-a normally used in operational models, such as biogeochemical-Argo floats and gliders. Here the indicators were defined and adapted to the features of the operational models (Section 2.1) as follows:

A) the Particulate Organic Carbon (POC) is defined here as the non-living carbon fraction of particulate organic matter, i.e., the detritus, and is computed as the average concentration of the 0-200m layer from the model output, or 0m-bottom in shallower areas. Three models (BFM, ERGOM and ECOSMO) have only one state variable for the particulate detritus, while ERSEM and PISCES have two detritus state variables, which are summed to compute POC. The unit is mmolC m^{-3}.

B) the trophic efficiency (TE) is the ratio of production at one trophic level to production at the next lower trophic level, here approximated by the ratio between the biomass of zooplankton and phytoplankton in carbon units for all the models, but using nitrogen biomass for ERGOM. In all the models, the sum of all the simulated types of phytoplankton (respectively zooplankton) provided the biomass of producers (respectively consumers). The only exceptions were the producers in BFM, which excluded heterotrophic nano-flagellates to avoid excessive fluctuations, and producers in ECOSMO, which included detritus grazed by the zooplankton. Biomasses were computed as the vertical integral in the 0-200m layer in deep ocean waters. Trophic efficiency is dimensionless.

C) the net primary production (NPP) is the synthesis of organic compounds from dissolved carbon dioxide through photosynthesis as source of energy. Models compute NPP as the difference between gross primary production (photosynthesis) and the internal phytoplankton respiration according to

their specific formulations. NPP is computed as the vertical integral of the 0-200m layer. The unit is mmolC m^{-2} d^{-1}.

D) the pH is a measure of the ocean acidity. All the models feature a carbonate system formulation that provides the pH as a diagnostic variable in total scale and at the *in-situ* condition. The pH is computed as the vertical average of the 0-200m layer from the model output. pH has no unit.

E) the dissolved oxygen (DO) is a state variable in all the models. The oxygen indicator refers to the concentration of oxygen at the depth of 150m (model output is averaged in the layer 145-155m) or at the bottom of the water column for areas shallower than 150 meters. The unit is mmolO_2 m^{-3}.

F) the Phytoplankton Functional Types (PFT ratio) indicator is defined here as the ratio between the large and total phytoplankton carbon biomasses over 0-200m. Given the different formulations of the models, the large phytoplankton consists of the diatoms group in ECOSMO, ERGOM and PISCES, and the sum of diatoms and dinoflagellates groups in ERSEM and BFM. The ratio is dimensionless.

G) the Phytoplankton phenology (PHE), based on chlorophyll-a (CHL), consists of four indicators that account for both surface blooms and deep chlorophyll-a maxima: (i) PHE_maxsc is the value of the maximum of chlorophyll-a concentration in the surface layer 0-5m computed from seven-day moving average of chlorophyll-a time series (mgChl m^{-3}) and (ii) PHE_maxst is the timing of the surface maximum (day), (iii) PHE_maxdc is the value of the maximum of chlorophyll-a concentration in the deeper layer 10-150m (mgChl m^{-3}) and (iv) PHE_maxdd is the depth of the deep maximum (m).

2.3 The observable variables

The controllability of the indicators was assessed with respect to biogeochemical observable variables, i.e., variables that are both simulated by the models, and can be observed by state-of-the-art operational monitoring platforms, including ocean colour, biogeochemical Argo floats and gliders. The list and computation of the observable variables depended upon (i) the model complexity (e.g., the number of PFTs contributing to total chlorophyll-a), (ii) the features of the assimilation systems of the Copernicus Marine Service MFCs (e.g. the plankton nutrient contents were observable if they are directly constrained in the chlorophyll-a analysis step, as in Skákala et al., 2018) and (iii) the observation capacity of the monitoring platform (e.g., ocean-colour chlorophyll-a refers just to the surface chlorophyll-a, while floats and gliders refer also to chlorophyll-a in the deeper layers) (Table 2).

Table 2. List and definition of the observable variables included in the controllability assessment. In the screening analysis, all the observable variables were perturbed from 0 to 1000 m depth (or to the bottom in shallower waters), except for chlorophyll-a from satellite ocean colour, which was perturbed from 0 the variable OC_{depth} specified in Table 3. The notation of the variables used in the figures is also reported.

Observable ocean variable	**Notation**	**Correspondent observable model variables**
1 Chlorophyll-a (from ocean colour)	OC	Surface total chlorophyll-a and carbon, nitrogen, phosphorus and (for diatoms only) silicate content of PFTs
2 Chlorophyll-a profiles	ARGO_CHL	Total chlorophyll-a and carbon, nitrogen, phosphorus and (for diatoms only) silicate content of all PFTs

3 Particulate Organic Carbon	ARGO_POC	Total carbon content of plankton and detritus < 20 micron
4 Nitrate	ARGO_NO3	Nitrate concentration
5 Phosphate	ARGO_PO4	Phosphate concentration
6 Oxygen	ARGO_O2	Dissolved oxygen concentration
7 Dissolved Inorganic Carbon	ARGO_DIC	Dissolved Inorganic Carbon (DIC) concentration

2.4 Controllability assessment

Controllability is here defined as the capacity of the observed variables to constrain the indicators, in a framework reproducing relevant features of operational assimilation systems. This definition is coherent with formal definitions in optimal control theory and data assimilation (e.g., Gelb, 1974). Data assimilation analyses typically compute new initial conditions of the observed variables, toward the assimilated observations of those variables. Re-initializations in assimilative analysis can impact also model variables that are not observed, either (i) directly, by computing initial conditions of unobserved variables by means of correlations with the observed variable, or (ii) indirectly, by integrating the model equations starting from re-initialized conditions of the observed variable. Ultimately, we expect that data assimilation reduces the uncertainty of the biogeochemical estimates. In the following, uncertainty and precision refers to the dispersion of the ensemble of biogeochemical estimates, while bias and accuracy refer to their closeness to the unknown "true" value of the biogeochemical state or, for practical purposes, the closeness to independent biogeochemical observations when available (see similar conventions in, e.g., Carrassi et al, 2018).

Based on the features above, our framework to assess the indicator controllability by the observable variables is two-pronged. Firstly, an overall controllability screening analysis is performed to rank the controllability of all the indicators, based on the Monte Carlo sensitivity analysis of the indicators with respect to the initial values of the observed variables, in computationally affordable, one-dimensional configurations of the models. Secondly, the controllability of a few top-ranked indicators is corroborated by assessing the reduction of their uncertainty with the computationally demanding, three-dimensional ensemble configurations of the operational models.

We note that formal definitions and approaches to assess controllability have been proposed in the scientific literature for ocean and biological models (e.g., Garcia et al., 2011, Gebbie and Hsieh 2017, Villaverde, 2019, Carrassi et al., 2022, Diaz-Seoane et al., 2023), but not for large operational biogeochemical models as the ones considered in this paper, to the authors' knowledge.

2.4.1 Controllability screening analysis

The controllability screening analysis consists in a sensitivity analysis of the indicators with respect to the initial conditions of the observable variables, in a Monte-Carlo simulation approach. This mimics reinitializing the observable variables in assimilative simulations. In the analysis presented here, (i) the initial conditions of the observed variable were perturbed stochastically, (ii) ensemble simulations were performed, and (iii) temporal averages of the ecosystem indicators were computed and

evaluated. Regarding point (i) above for chlorophyll-a, the original PFT chlorophyll-a ratios and internal nutrient ratios were preserved when perturbing the correspondent observable model variable (see Table 2).

The sensitivities are defined here as follows:

$$s_{i,j,k} = \frac{\sqrt{1/M \cdot \sum_{m=1}^{M} (\overline{\overline{y_{i,k,m}}} - <\overline{\overline{y_{i,k}}}>)^2}}{<\overline{\overline{y_{i,k}}}>} \Bigg/ \frac{\sqrt{1/M \cdot \sum_{m=1}^{M} (\overline{x_{j,m}^0} - <\overline{x_j^0}>)^2}}{<\overline{x_j^0}>} \qquad \text{eq. 1}$$

Where $s_{i,j,k}$ is the sensitivity of the ecosystem indicator *i*, output of the model *k* (i.e., $y_{i,k}$), with respect to the initial condition of the observable variable x_j^0. The single average bar (‾) indicates that the initial condition of the observable variable was averaged across its definition depth (see Section 2.3). The double average bars (=) indicate that the indicator was computed across its definition depth (see Section 2.2) and then averaged with respect to the time-length of the simulation period (see Table 3). In eq. 1, the symbol < > represents the mean value of the indicator or the observable variable across the Monte Carlo ensemble of m=1, …, M simulations.

The sensitivity $s_{i,j,k}$ normalizes the dispersion of the variables by their average values and is therefore dimensionless. The higher its value, the higher is the impact of the observed variable on the ecosystem indicator, i.e., the controllability of the indicator with respect to the observable variable simulated by one model k, in our approximation.

The generalized ensemble sensitivity of the indicator with respect to the observable variable is the mean of $s_{i,j,k}$ in eq. 1 computed across the ensemble of K=5 biogeochemical models:

$$< s_{i,j} > \; = \; 1/K \sum_{k=1}^{K} s_{i,j,k} \qquad \text{eq.2}$$

And the coefficient of variation of the sensitivity across the multi-model ensemble of is:

$$CV_{\,i,j} = \sqrt{1/K \cdot \sum_{k=1}^{K} (s_{i,j,k} - \; < s_{i,j} >)^2} \; / < s_{i,j} > \qquad \text{eq.3}$$

i.e. the standard deviation of the sensitivity across the multi-model ensemble, normalized by the mean value of the multi-model ensemble.

Our set-up of the sensitivity analysis followed well-established practices:

- The probability distributions of the observable variable were set as uniform, centred on the climatological value of the variable at the study sites, with a threshold of +/- 50% of the nominal values of all variables, to explore a relatively large variability space; the threshold was lowered to 10% for DIC, since its high nominal absolute value could induce unrealistic stochastic fluctuations (see comparable choices for Monte Carlo perturbations of ocean biogeochemical variables by Ciavatta et al., 2011)
- Monte-Carlo ensemble simulations of M = 100 members were performed to compute the distributions of the ten indicators for each of the seven observable variables perturbed separately, following the common practice of setting M one order of magnitude higher than the number of random variables (e.g., Saltelli et al., 2000)

The analysis was performed using the sensitivity analysis libraries of the *parsac* software (Bruggeman and Bolding; 2020; Andersen et al., 2021) extended for the purpose of SEAMLESS to do Monte Carlo-based sensitivity analysis.

2.4.2 Controllability corroboration analysis

The corroboration analysis of controllability assesses the reduction of the uncertainty of the indicators in ensemble assimilative simulations, by using three-dimensional model configurations, comparable to the operational ones. Here we corroborated the controllability of indicators selected in the screening analysis, by performing ensemble assimilation of ocean colour and/or biogeochemical profiles, with both high and intermediate complexity models (see Section 2.5 for details on the set-up of the ensemble assimilation simulations).

The reduction of the uncertainty of the indicator *i* was evaluated by computing the ratio of the coefficients of variation of the ensembles posterior (cv^{post}) and prior (cv^{prior}) the assimilation of the observations:

$$CV\ ratio_i = \frac{CV_i^{post}}{CV_i^{prior}} = \frac{\sqrt{1/N \cdot \sum_{n=1}^{N} (y_{i,n}^{post} - <y^{post}>)^2}}{<y^{post}>} \Bigg/ \frac{\sqrt{1/N \cdot \sum_{n=1}^{N} (y_{i,n}^{prior} - <y^{prior}>)^2}}{<y^{prior}>} \quad \text{eq. 4}$$

Where $y_{i,n}$ is the ecosystem indicator *i* output of the member *n* of the data assimilation ensemble of size N. The symbol $< y >$ represents the mean value of the indicator across the data assimilation ensemble, calculated for either the prior or the posterior. By scaling the standard deviation of the data assimilation ensemble by the mean value of the data assimilation ensemble, we focussed on the relative uncertainty of the model. Values of CV ratios below 1 indicate that the assimilated observation both controlled and benefited the simulated indicator by increasing its precision (i.e., the posterior distribution had a lower percentage spread that the prior distribution). Values of CV RATIO higher than 1 indicate that the assimilated observation controlled the indicator but deteriorated the precision. Values close to one suggest lack of control of the indicator by the observations.

We computed time series and vertical profiles of the members of the data assimilation ensemble, their mean and the CV ratio at single point locations, to assess the changes in uncertainty and controllability of the indicators with time and depth. Horizontal gradients of the controllability were evaluated by plotting regional maps of the prior mean, assimilation increments and the CV ratio of the indicators.

2.5 Set-up of the models and assimilation systems

In the controllability screening analysis (Section 2.4.1), all the five biogeochemical models were coupled via the Framework for Aquatic Biogeochemical Models (FABM, Bruggeman and Bolding, 2014) to the same hydrodynamic General Ocean Turbulence Model (GOTM, Umlauf and Burchard, 2005). GOTM is a one-dimensional (1D) water column model that uses traditional and state-of-the-art turbulence closure equations for the parameterisation of vertical turbulent fluxes of momentum, heat and dissolved and particulate matter.

The 1D ensemble of biogeochemical models was set-up in two contrasting marine ecosystems and contrasting starting times of the simulations, since these factors can affect the controllability of the indicators. The two sites are (1) the coastal mesotrophic site Station “L4”; and (2) the oligotrophic open ocean station “BATS”, which are characterized by different trophic regimes, phytoplankton cycles, and vertical hydrodynamics as synthetized in Table 3. The two starting times of the simulations at each Station were chosen to capture contrasting hydrodynamic conditions at the study sites, i.e. (1) fully mixed versus (2) stratified water column. The different hydrodynamic conditions are expected to influence the trophic state and disperse the magnitude and profiles of the biogeochemical variables perturbed in our simulations. These were therefore started at different seasons and run in three-month long periods characterized by contrasting hydrodynamic conditions, at both L4 and BATS (Table

3). The results are presented for the multi-model ensemble in the main text, and detailed for the single models in the Supplementary material (Figure S1). Selected forcings of the simulations in the two sites and seasons are presented in the Supplementary Material (Figure S2) along with simulated nutrient ratios computed at BATS (Figure S3), useful for the discussion of the results. To confirm further the dependence of the observability of the indicators on the start period of the simulations, we computed the sensitivity in eq. 1 for starting period lagged by two months, at one site (L4) and one model (ERSEM) and the results are presented in the supplementary material (Figure S4). To evaluate further the robustness of the screening analysis, we repeated the sensitivity analysis in eq. 1 at the station BOUSSOLE in the Mediterranean Sea for the model BFM, which is used in 3D both here for the corroboration analysis, and in the Copernicus Marine Service for operational predictions. The results are presented in the supplementary material (Figure S5).

All the biogeochemical models used the same climatological values for the initial conditions of the state variables, derived from public datasets of the two data-rich sites. Atmospheric reanalyses were used to force the coupled models (see Table 3). Multi-year long spin-ups were performed to stabilize the levels of the physical and biogeochemical variables prior the target year of the analyses (Table 3). During the spin-up simulations, both biogeochemical and physical variables were nudged to climatological monthly values derived from available data (Table 3), by using nudging routines available in GOTM (Umlauf and Burchard, 2005). During the target periods, only physical nudging was applied, to guarantee an unconstrained evaluation of the sensitivities of the biogeochemical indicators.

The multi-model ensemble and their one-dimensional configurations at BATS and L4 are available here: https://github.com/BoldingBruggeman/seamless-notebooks). The sensitivity analysis was run on the CINECA High-Performance Computer.

Table 3. Set-up of the multi-model simulations at the two contrasting sites BATS and L4. OC_{depth} is the depth of the profile initial conditions of ocean colour perturbed in the sensitivity simulations.

Site	**BATS**	**L4**
Location	Sargasso Sea (31°40'N 64°10'W)	English Channel (50°15' N 4°13' W)
Site main features	Oligotrophic open-ocean site; 4500 m deep; general strong stratification	Mesotrophic coastal site; 50 m deep; seasonally stratified
Initial conditions	Climatology from data of BATS station (www.bats.bios.edu/bats-data/)	Climatology from data of Western Channel Observatory (www.westernchannelobservatory.org.uk/)
Meteo forcings	ERA5 reanalysis of Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu)	ERA5 reanalysis of Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu)
Spin-up	5 years	7.5 years

Stratified period	15/06/2019-14/09/2019 (OC_{depth} = 5m)	01/06/2014-01/09/2014 (OC_{depth} = 5m)
Mixing period	01/01/2019-31/03/2019 (OC_{depth} =100m)	01/11/2014-31/01/2015 (OC_{depth} =50m)

In the corroboration analysis of the controllability (Section 2.4.2), we used three-dimensional (3D) configurations of two biogeochemical models of contrasting complexity (intermediate PISCES versus complex BFM) and different types of assimilated data (from satellite only versus multi-platform): see Table 4. Both the PISCES and BFM biogeochemical models were coupled to the Nucleus for European Modelling of the Ocean (NEMO; Madec et al., 2017) ocean model for the simulation of the three-dimensional ocean dynamics. The same coupled models are being run operationally in the Copernicus Marine Service for the simulation or assimilation of ocean colour in the GLO and IBI centers (PISCES, Lamouroux et al., 2022; Gutknecht et al., 2019) and for assimilation of both ocean colour and biogeochemical ARGO float data in the MED center (BFM, Salon et al., 2019; Teruzzi et al., 2021). Here, we employed the models with different flavours of ensemble approaches for generation and data assimilation (Table 4). Referring to the publications describing the two assimilation methods in detail, here we mention that the controllability corroboration in the North Atlantic sector of the global Copernicus operational system was performed by assimilating ocean colour data during 2019, by using a four-dimensional (4-D) stochastic ensemble methodology (Popov et al., 2024). A prior 40-member ensemble was first generated by using a stochastic version of the NEMO-PISCES global model at ¼° that accounts for uncertainties originating from biogeochemical model parameters, unresolved sub-grid scale processes and coupled physical/biogeochemical effects. The posterior distribution is then computed in the subdomain around PAP station using the Localized Ensemble Transform Kalman filter (LETKF) analysis scheme implemented in 4D and space-time ensemble correlations between state variables. In the Mediterranean Sea, we used a sequential ensemble assimilation approach (SEIK-OGSTM, Spada et al., 2023). This includes the three-dimensional OGSTM-BFM model of the Mediterranean Sea, at 1/24° resolution, which couples the BFM biogeochemical model forced offline by the NEMO model. The ensemble assimilative component is based on the Singular Evolutive Interpolated Kalman filter (SEIK, Pham et al., 2001). The SEIK-OGSTM system was applied in the corroboration analysis in two 2.5 month-long simulations, for the contrasting winter-mixed and summer-stratified seasons, with weekly sequential assimilation of satellite chlorophyll-a. An ensemble of 24 members was generated by perturbing seven biogeochemical model parameters, which were found the most relevant in a separate sensitivity analysis. The main difference in the two assimilation methods used here is them representing the two broad assimilation classes of (1) four-dimensional schemes (space plus time) which assimilate all the available data available in a time window and (2) three-dimensional schemes (space) that assimilate the data in sequential time steps, when they become available across the spatial domain. Crucially, the common feature of the two methods is them being ensemble approaches, which in our framework is essential to estimate the controllability of the simulated indicators through the analysis of changes in the uncertainty, i.e. in the distribution of the model ensembles (eq. 4). The sizes of the ensembles were chosen compatibly with the available computational resources and seminal papers (Spada et al., 2023 and Popov et al., 2024).

The controllability of the indicators in time and depth was corroborated at two additional sites with respect to the ones of the screening analysis, namely the PAP station in the North Atlantic (Hartman et al., 2021) and the BOUSSOLE station, in the Northwest part of the Mediterranean Sea (Antoine et al., 2002). The controllability at depth was evaluated in correspondence of the peak of ocean-colour

chlorophyll-a concentration at PAP (15th May), thus in conditions representative of spring phytoplankton blooms at the two sites. Horizontal patterns of the controllability were assessed at the scale of the whole Mediterranean Sea, and shown at two representative periods in a blooming winter (11th March) and summer (8th August) days, when the satellite coverage of the basin was relatively high and included the BOUSSOLE station.

Table 4**.** Set-up of the 3D corroboration analysis in two representative marine ecosystems.

	Atlantic Ocean	**Mediterranean Sea**
MFC domain	Global (GLO) Iberian-Biscay-Irish seas (IBI)	Mediterranean Sea (MED)
Model	NEMO-PISCES (medium complexity)	NEMO-BFM (high complexity)
Assimilated data	Chlorophyll-a from ocean colour (Copernicus Marine Service L3 product, i.e., gridded and quality-controlled data)	Chlorophyll-a from ocean colour (Copernicus Marine Service L3 product and a specific re-gridding and quality control); chlorophyll-a and nutrients from biogeochemical ARGO floats
Ensemble generation	Perturbation of biogeochemical parameters, sub-grid scales and location of features	Perturbation of initial conditions of the state variables and of the biogeochemical parameters
Assimilation method	Stochastic ensemble filter (Popov et al., (2024)	Singular Evolutive Interpolated Kalman filter (SEIK; Pham et al., 2001)
Multivariate analysis	Phytoplankton types and nitrate	Phytoplankton types, nitrate and phosphate
Assimilation time	Full year 2019 cycle (January-December)	Winter 2019 (February-March) Summer 2019 (July-August)
Ensemble members	40	24
Site of assessment	Porcupine Abyssal Plain (PAP), 16°30' W, 48°50'N, 4850 m deep, seasonal stratification, mesotrophic conditions	BOUSSOLE, 7°54’E ,43°22’N, 2400 m deep, seasonal stratification, oligotrophic conditions
Verification data	Chlorophyll-a from ocean colour (Copernicus Marine Service L4 product)	Same as the assimilated data

3 Results and discussion

3.1 Screening analysis

The controllability of the indicators was highly dependent on the observable variables, as shown by the large range of sensitivity values in Figure 1. High sensitivities and low CV indicate relatively high controllability and consistency among models, respectively. In general, the features of the marine ecosystems (L4 versus BATS) and the hydrodynamic conditions (mixed versus stratified water column) impacted the controllability. The controllability was influenced also by the specific models employed by the ensemble (see Fig S1 in the Supplementary Material), since the coefficient of variation of the multi-model ensemble was even higher than 2 (see yellow colour for BATS in Figure 1). However, the controllability was not clearly linked to the complexity of the models, here approximated by the number of PFTs, but rather to the specific formulation of the processes.

In more detail, all the indicators were controllable by at least one observable variable, with some notable examples. The maximum concentration of surface chlorophyll-a was controlled by all the observable variables (but DIC), although the magnitude of the controllability depended on the site and hydrodynamic conditions. Differently, the oxygen and pH indicators were controlled by only one observable variable, either oxygen or DIC, respectively. This latter result holds at both sites and seasons. Still, stratified conditions favoured the controllability of oxygen in the water column, arguably because the oxygen profile keeps memory of the initial conditions throughout the simulation period when the water column is stratified. Differently, when the water column is mixed, the air-sea flux forces the oxygen concentration throughout the water column, rather than initial conditions, likely explaining the lower controllability of oxygen.

The controllability of the indicators in Figure 1 was in general higher at BATS than at L4. This reflects the lower production levels and fluctuations in the oligotrophic ocean site than in the mesotrophic coastal site. This is shown by the one-order lower concentrations and fluctuations of chlorophyll-a at BATS than at L4, for the phenology indicators in Figure 2. These features imply that the models preserve the memory of the initial conditions for a longer time at BATS than at L4, explaining the higher controllability of the indicators at BATS.

The controllability of specific indicators depends also upon the trophic regimes of the two sites (Figure 1). At the oligotrophic BATS, the primary production is phosphate-limited in the simulations of most of the biogeochemical models exploited here (Figure S3 in the Supplementary Material), and this is coherent with previous literature findings (e.g., Salihoglu et al., 2008). That explains the control of phosphate observations (ARGO_PO4) upon all the ecosystem indicators simulated at BATS, in both mixed (more evident) and stratified conditions (less evident). Confirming indirectly the relevance of nutrient balance in determining the nutrient control on the simulated indicator, ARGO_PO4 and ARGO_NO3 have a comparable control for some of the indicators in the BATS simulations of PISCES and ERSEM (Figure S1), which indeed have the lowest N:P ratio among the models in Figure S3. On the other hand, in the mesotrophic site L4, nutrients limit production at the surface only in the summer stratified season. This explains the marked control of phosphate profiles on both the maximum and timing of the surface chlorophyll-a during the stratified season in summer (PHE_maxsc and PHE_maxdst, respectively). At site L4, nutrient profiles do not control indicators in the winter fully mixed conditions, when nutrient concentrations are higher and plankton production is mainly light-limited (Widdicombe et al., 2010).

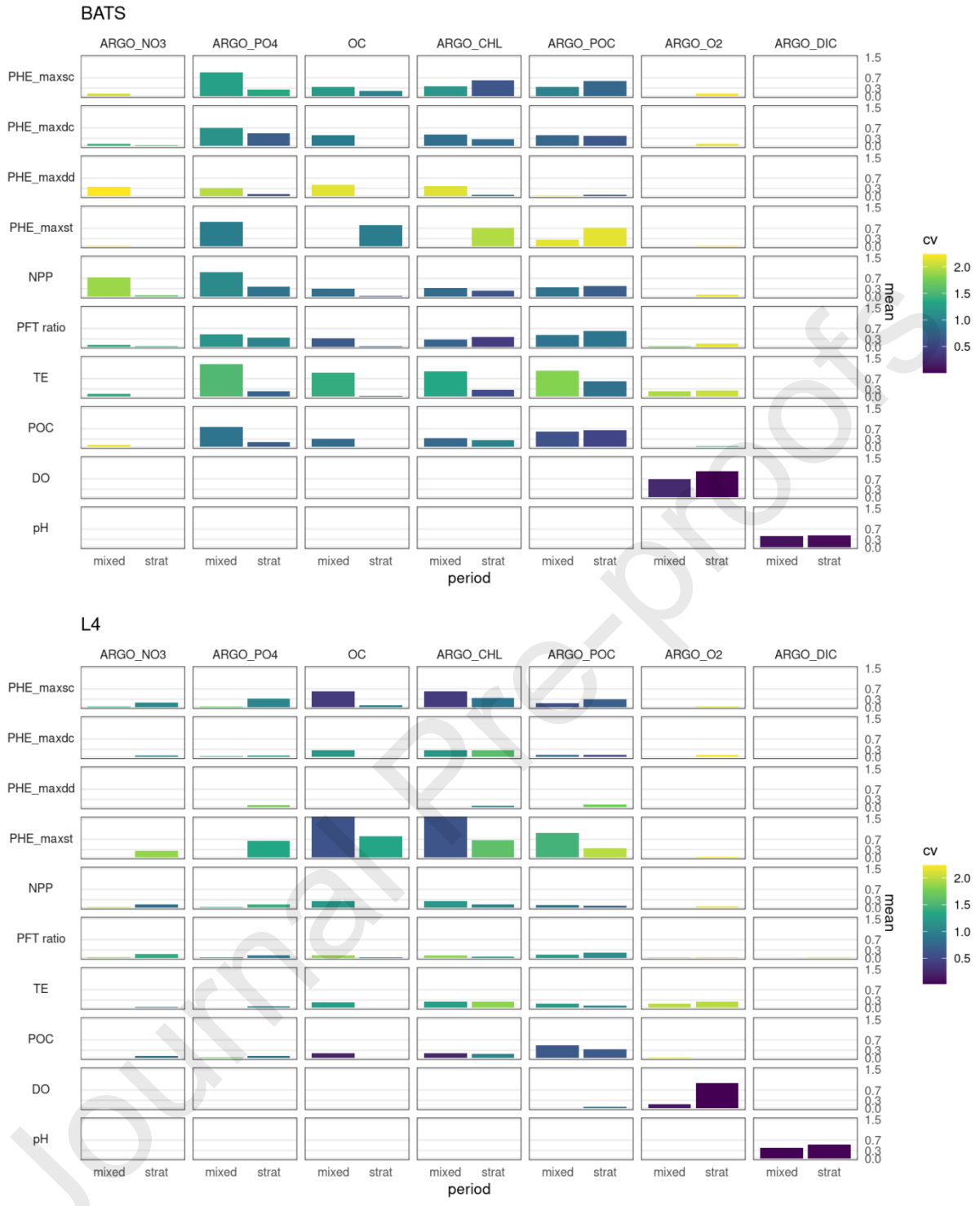


Figure 1. Results of the screening analysis of the controllability of the ecosystem indicators with respect to the observed variables at stations BATS (upper panel) and L4 (lower panel). The height of the bars shows the average magnitude of the controllability by the model ensemble, i.e., the sensitivities in eq. 2, for the ten ecosystem indicators (rows), with respect to the seven observed variables (columns), during two representative periods, i.e. hydrodynamic conditions (mixed and stratified water column). The colours represent the coefficient of variation (CV) of the sensitivities across the ensemble of 5 models, eq. 3.

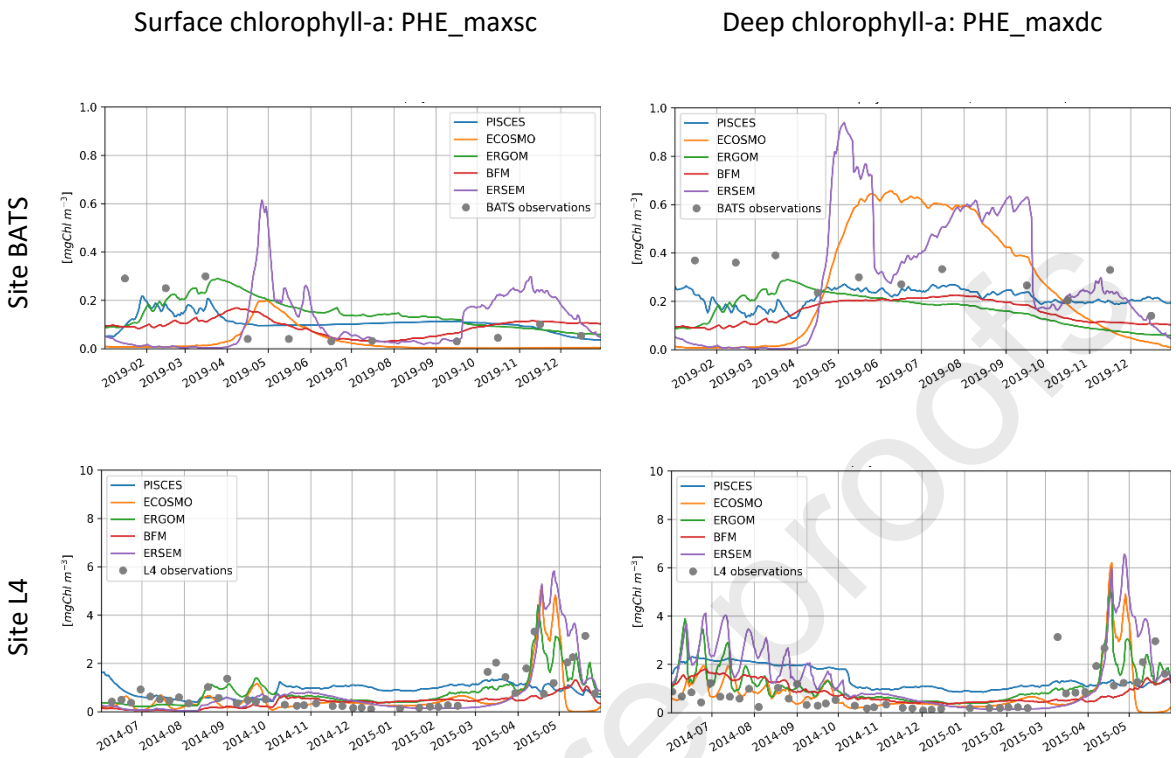


Figure 2. Multi-model ensemble reference simulations of the phenology indicators of maximum chlorophyll-a concentrations at the sites BATS (top plots,) and L4 (bottom), at both surface (PHE_maxsc 0-5 m; left) and depth (PHE_maxdc below 5 m; right). The colours of the lines distinguish the five different models of the ensemble (see section 2.1). The dots represent the chlorophyll-a observations available for comparison at the two sites. The one year-long simulation includes both the stratified and fully mixed three month-long periods targeted in the screening analysis.

The hydrodynamic conditions determine which specific observable variables can control the indicators. The most remarkable example is ocean colour, which is defined as chlorophyll-a in the surface layer (Section 2.2). When the column is stratified, ocean colour can control just the surface plankton phenology, i.e., the surface chlorophyll-a maximum concentration (PHE_maxsc) and timing (PHE_maxst), and not the deep chlorophyll-a maximum concentration (PHE_maxdc) and depth (PHE_maxdd). On the other hand, ocean colour can control most of the biogeochemical indicators when the water column is mixed, at both sites. At BATS, the simulated maximum depth of the mixed layer reached 200 m, and ocean colour can control the POC, NPP and the phytoplankton indicators that are integrated from the surface to that same depth (see the definitions in Section 2.2). At SITE L4, ocean colour can control indicators integrated from surface to the bottom at 50 m depth, because the mixed layer reaches the bottom of the water column when fully mixed in winter.

The dependency of the controllability of the indicators on the trophic regimes and hydrodynamic conditions implies that the sensitivities are clearly dependent upon the starting time of the simulations that are run to compute equation 1. Such dependency is evident in Figure 1, where different ensemble

sensitivities were computed for simulations started in the mixed and stratified season at the two sites. Such dependency is supported further by figure S4, in the Supplementary Material, which shows the fluctuations of the sensitivities computed at lags of two-month start time of the simulations. These results suggest that complementary biogeochemical observation platforms should be used at different periods of time to control the indicators simulated by biogeochemical models.

The controllability of the indicators depended upon the specific models employed in the screening analysis, as demonstrated by the relatively high coefficients of variation (CV) in Figure 1 (e.g. CV>2 at BATS for PHE_maxdd versus ARGO_NO3). The highest CVs were often linked to outlier models (Figure S1 in the Supplementary Materials). For example, linked to the outlier PISCES' controllability of the depth of chlorophyll-a maximum by nitrate at BATS, or to the outlier ERSEM's controllability of plankton-related indicators by oxygen profiles at both BATS and L4. Such latter controllability by oxygen, noticeable also for BFM in Figure S1, is possibly related to the oxygen control on bacteria and the microbial loop, which are represented only in ERSEM and BFM as path of nutrients alternative to the plankton herbivorous chain (Butenschön et al., 2016; Vichi et al., 2015). Besides for controllability, the models were in some instances rather different in their reference simulations of the indicators themselves, as exemplified in Figure 2. This shows the simulations of two phytoplankton-related indicators, along with observations of chlorophyll-a for comparison. At BATS the chlorophyll-a concentration (linked to the phenological indicators PHE_maxdc and PHE_maxdd) had comparable mean annual values in all the model simulations, but had larger seasonal oscillation in ECOSMO and ERSEM than in BFM, PISCES, ERGOM and the observations. For example, ERSEM and PISCES simulated a change of the average values of PHE_maxdc at depth at BATS, close to 0 mgChl m^{-3} between January-March, to 0.6 mgChl m^{-3} in May-September, matching the higher solar radiation in the summer months (see Figure S2 for the solar radiation at BATS in the stratified summer simulation), with superimposed fluctuations due to the simulated zooplankton grazing (see e.g. the drop of PHE_maxdc at BATS in the ERSEM simulation). Differences in the reference model simulations might lead to some differences in the single models to control the indicators, which might be in part captured by the CVs in Figure 1 (see for examples the relatively high values of the CVs of PHE_maxdd and PHE_maxdc across observable variables at BATS). The effect of such differences in the reference simulations might be not obvious in the stochastic Monte Carlo simulation framework of our controllability screening. These outcomes highlight the benefit of using a multi-model ensemble approach to screen the controllability. They also foster to quantify and possibly address the consistency of the models within the ensemble, e.g., by computing synthetic indexes of the dispersion of the simulated indicators.

The controllability was not linked clearly to the complexity of the models, here approximated by the number of plankton types (Section 1.1), but it was rather linked to model formulation of the plankton physiology. For example, ERSEM and BFM were neither systematically similar each other in controlling the indicators, nor systematically different from the simpler models, a part for their control by oxygen noted above. More than to complexity, the controllability was linked to chlorophyll-to-carbon and nutrient content formulations in phytoplankton, in either complex or simpler models. For example, the carbon-based indicators of plankton production (NPP) and structure (PFT ratio), and of trophic efficiency (TE) were often more sensitive to the POC profiles than to the chlorophyll-a ones, at both sites and both water column conditions (Figure 1). This is because chlorophyll-a is a prognostic variable in all the models (but ERGOM), and its ratio to carbon varies in ERSEM, BFM and PISCES as a function of the light conditions, based on the model of Geider et al. (1997). That suggests that the carbon-related observations of POC have a relative stronger control than chlorophyll-a observations on carbon-based production indicators. This is consistent with the conclusions of Wang et al. (2020), who demonstrated that a variable chlorophyll-to-carbon ratio model could better estimate primary production and deep carbon export fluxes when using BGC-Argo float backscattering observations, in addition to chlorophyll-a. In contrast, when the chlorophyll-carbon ratio was fixed, the controllability of NPP, PFT ratio, TE and POC itself by the chlorophyll-a profiles was comparable to the one by POC profiles (see ERGOM in Figure S1 of the Appendix). Including variable chlorophyll-to-carbon ratio in

phytoplankton not only improves the reliability of models (Geider, 1987), but also influence the controllability of the indicators simulated by the models. All the operational models of the Copernicus Marine Service represent this feature, including the version of the ERGOM model now used in the Baltic Monitoring and Forecasting center (Ringgaard et al., 2024).

We stress that the controllability of an indicator by a model does not automatically imply that the model is skilled or preferable to less controllable models. As an example, Ciavatta et al (2011) showed that chlorophyll-a assimilated into a simplified version of ERSEM model using a constant chlorophyll-to-carbon and Redfield formulation, was capable to control nutrients more strongly than the standard ERSEM formulation with variable chlorophyll-to-carbon and carbon-to-nutrients formulation. However, the model validation showed a deteriorated skill for ERSEM with the fixed chlorophyll-to-carbon ratio, even to the degree that the model dynamics became unstable eventually, because the model had lost plasticity. These findings are in line with the suggestions of Friedrichs at al. (2006 and 2007) that variable chlorophyll-to-carbon ratio, besides the number of phytoplankton functional types, can contribute to the higher portability of complex models across multiple locations, even though simple model can have better skill when optimized and applied at single locations.

Two supplementary sensitivity analyses (see Methods Section 2.5) confirmed further the dependency of the controllability on the trophic regime and hydrodynamic conditions, already shown in Figure 1. The first test showed that the controllability of the selected indicators with respect to the selected observed variables depended upon the two-month lagged starting times of the ERSEM simulations at L4 (Supplementary Material, Figure S4). In particular, the sensitivity of both NPP and PFT ratio were higher during the mixed, autumn/winter months (October, December, February) than during the stratified spring/summer months (April, June August), for perturbations of chlorophyll-a from both ocean colour and profiles. Winter controllability of POC increased more in winter than in summer for ocean colour only. Sensitivities to PO4 profiles were higher in spring/summer than in winter, though they laid one order magnitude lower than the sensitivities to chlorophyll-a. These supplementary results are coherent with the different sensitivities of ERSEM in the winter-mixed and summer-stratified periods at L4 detailed in Figure S1 and synthetized in Figure 1. Ultimately, the different seasonal hydrodynamic and trophic conditions simulated at L4, as well as at BATS, can be related to the differences of the meteorological forcings used in the simulations in the two periods, as exemplified in Figure S2. Relatively higher winds trigger water column mixing in winter, when plankton growth is limited by the low light but not by nutrients (hence the higher sensitivities of the indicators to the background values of chlorophyll-a but not to nutrients). Differently, in summer, the solar radiation drives stratification and fuel primary production, which however might be limited by nutrients (hence the higher controllability by nutrients).

The supplementary test presented in Figure S5 showed that the results on controllability in Figure 1 at BATS are portable to a different site, with analogous trophic and hydrodynamic regimes. In fact, the sensitivities of BFM at the oligotrophic, open-ocean Mediterranean site BOUSSOLE were consistent with those of BFM at BATS in Figure S1. The most striking examples are the similarly high sensitivities of the indicators to the profiles of phosphate at BATS and BOUSSOLE (Figures 1 and S5 respectively), for being both the sites phosphate-limited in the 1D BFM simulations. Interestingly, also the 3D Mediterranean model used in the corroboration analysis simulate phosphate-limited conditions in the BOUSSOLE area (Lazzari et al., 2012). This supplementary evidence suggests that the results in Figure 1 can represent the controllability in contrasting trophic and hydrodynamic conditions (open-ocean oligotrophic condition versus coastal mesotrophic conditions) that we can encounter in the North-East Atlantic and Mediterranean Sea considered in the corroboration analysis (see also the discussion in Section 3.3 on the portability of the control results from 1D to 3D model configurations).

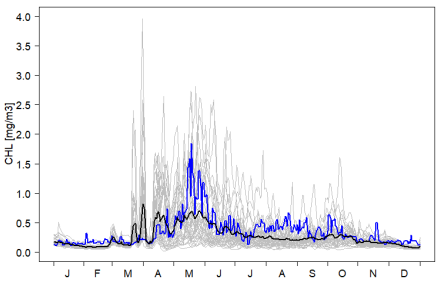
3.2 Corroboration analysis

3.2.1 Atlantic Ocean: PAP station

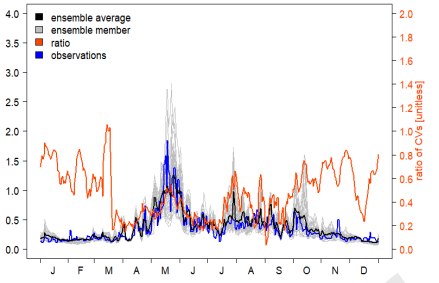
When corroborating the controllability of the indicators with the (pre)operational model of the Atlantic Ocean, the assimilation of surface ocean colour improved the simulation of chlorophyll-a observation at the PAP station (Figure 3). In addition, assimilation effectively controlled the temporal and vertical simulation of the unobserved indicators at the same site (Figure 3 and 4). This is shown by (i) CV ratios that are typically below 1 in both Figures 3 and 4, indicating assimilation decreased the ensemble spread (eq. 4), as well as by (ii) relevant shifts of the ensemble means of all the indicators, indicating significant assimilative changes of the indicator trajectories. The points (i)-(ii) are more evident from March to October in Figure 3 and confirmed by the vertical profiles shown in Figure 4.

The comparison of the prior and posterior time series in Figure 3a and 3b indicates that the assimilation of chlorophyll-a (L3 product) not only controlled, but also significantly improved the simulation of the phytoplankton phenology, i.e., the initiation and amplitude of the spring bloom from mid-March to May. This period had the highest uncertainty of the prior distribution (i.e. the largest spread of the ensemble of states enveloping the observations, before the assimilation of the observation), and the assimilation improved the model precision around the observations by halving the ensemble spread in the posterior distribution (see CV ratios = ~0.5 in March-April in Figure 3b), explaining the lowest values of the variance ratio. The assimilation corrected a negative bias of chlorophyll-a in summer from June to September, when the a priori uncertainty was halved. The CV ratio reached 0.6 in October (Fig 3b), when the posterior captured an observed chlorophyll-a peak neglected by the prior. This increase of phytoplankton in autumn is often observed at PAP in correspondence with seasonal water column mixing and surface nutrient repletion (e.g., Binetti et al., 2020). From November to February, the CV ratio approached 1 (Fig 3b), indicating a lower control of the assimilated ocean colour on the simulated chlorophyll-a in late autumn and winter than in spring and summer. This can be explained by the fact that the mean and variance of the prior ensemble were already in quite good agreement with the observations, requiring smaller correction for the bias and uncertainty components in winter than in summer (Figure 3a and 3b). Furthermore, at the latitude of the PAP station, the ocean colour observations are typically fewer in winter than in the summer, due to cloud cover and low solar zenith angle (Sathyendranath et al., 2019), implying a lower constrain in winter than in summer. The corroboration also indicated that ocean colour can control the indicators at the onset of stratification and phytoplankton bloom (April-May in Figure 3a and 3b), which is a transition phase that was not explored in the controllability screening analysis in Figure 1.

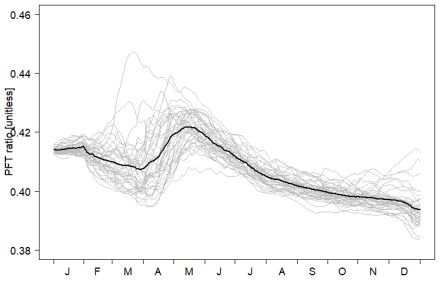
a



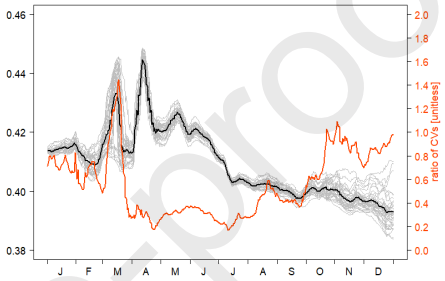
b



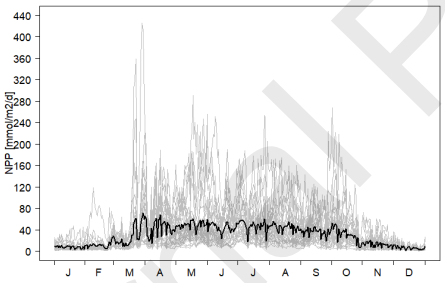
c



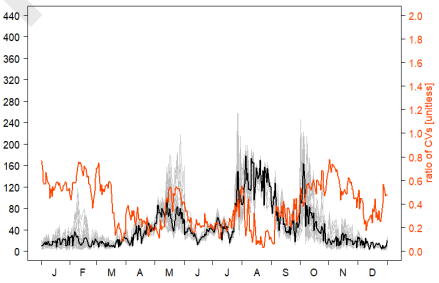
d



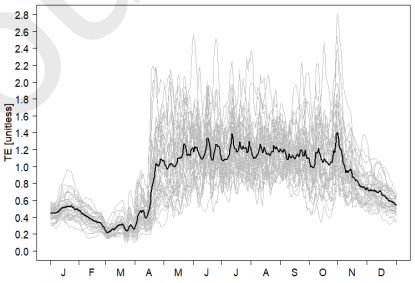
e



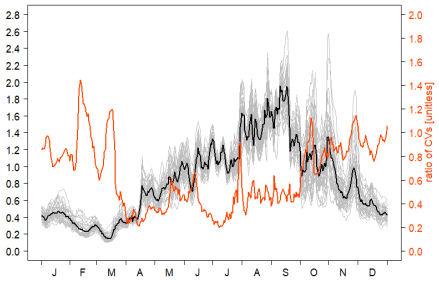
f



g



h



i

j

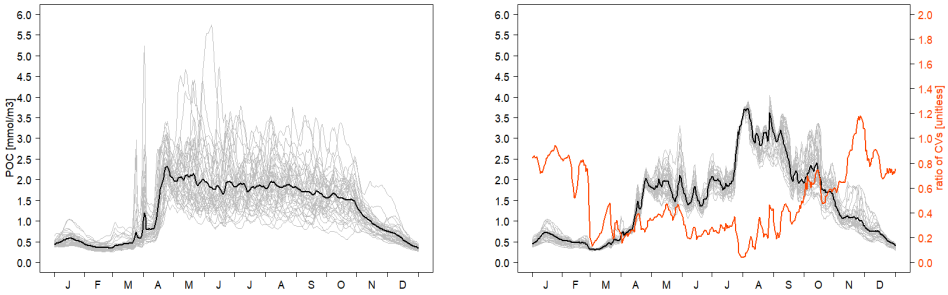


Figure 3. Time series of the indicators at PAP station, for the prior (left panel) and posterior (right panel) ensembles, in the months 2019. The dispersions of the 40 member ensembles (grey lines) are shown along with their mean value (black line) and the ratios of the coefficients of variation in eq. 4 (red lines in the plots of the posteriors on the right). The plots show chlorophyll-a (CHL, a and b), phytoplankton type ratio (PFT ratio, c and d), net primary production (NPP, e and f), trophic efficiency (TE, g and h) and particulate organic carbon (POC, i and j).

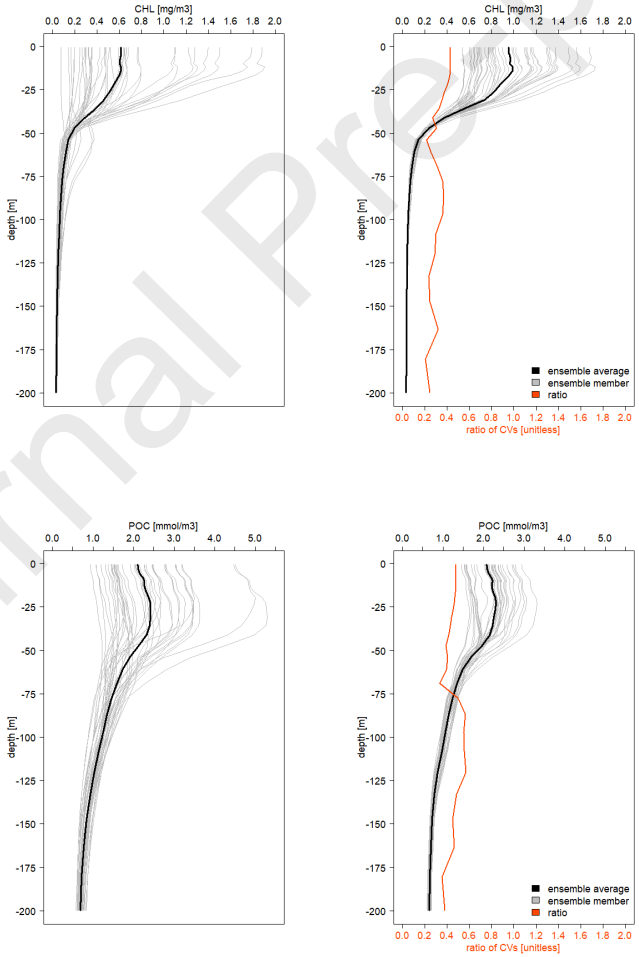


Figure 4. Vertical distribution of the chlorophyll-a concentration (CHL, upper panels) and POC (lower panels) at PAP station. For each indicator, the distributions of the prior (left graph) and posterior (right graph) and the ratio of their variance (eq. 4, right graph, red line) are presented for 15th May.

Ocean colour assimilation also markedly controlled the unobserved indicators in Figure 3, which all have CV ratios below one for most of the year (Fig. 3d, 3f. 3h, 3j). For these indicators, however, the skill of the analysis cannot be validated due to the lack of observations. The temporal patterns of the changes in both the ensemble mean and variance reflect the patterns of chlorophyll-a. In spring and summer, for all unobserved indicators, ocean colour assimilation reduced markedly the high uncertainty of the prior (see CV ratios below ~0.4, in general, from mid-March to September for all the indicators in Figure 3) and markedly shifted the ensemble mean, often in not obvious ways. For example, assimilation of total chlorophyll-a changed the plankton community structure differently depending on the season (PFT ratio in Figure 3c and 3d, i.e., the ratio of diatoms to diatoms plus nanophytoplankton in PISCES; see Section 2). The reduction of the unobserved peaks of chlorophyll-a in the prior in March and April (Fig 3b) led to peaks of PFT ratio in the posterior (Fig. 3d). These are due to a strong reduction of the nanophytoplankton biomass. On the other hand, large PFTs biomass slightly increased with respect to small PFTs (PFT ratio from ~0.422 in Fig 3c to ~0.427 in Fig 3d) to produce the chlorophyll-a bloom observed in May (Fig 3b). These changes are consistent with previous findings that total chlorophyll-a assimilation changes primarily the simulated dominant PFT (Ciavatta et al., 2011, 2018). Differently, the PFT ratio remained substantially unchanged, and in favour of the nanophytoplankton fraction, from June to September (Fig 3d), when the chlorophyll-a in the posterior (Fig 3b) corrected the summer negative bias of chlorophyll-a in the prior (Fig 3a). The correction of this bias was rather linked to the increase of primary and secondary production and POC in summer (see the maxima in August and September of the posterior of NPP: Fig 3f, TE: Fig 3h and POC: Fig 3j). All the above changes are consistent with biogeochemical and trophic dynamics that had been previously observed at PAP, i.e., shifts from large to small phytoplankton from spring to summer, accompanied by increased primary production, grazing and sinking of particulate organic carbon (Binetti et al, 2020).

Ocean colour assimilation produced smaller changes of the unobserved indicators in late autumn and winter, when CV ratio reached values around one in some instances (see, e.g., PFT ratio: Fig 3d, TE: Fig 3h and POC: Fig 3j in November and December). This corresponds to the small changes of chlorophyll-a due the close fit of the prior to the observations, as well as by the sparsity of chlorophyll-a data available for assimilation. Both factors led to smaller changes of the prior of the unobserved indicators in late autumn and winter.

Figure 4 shows that the assimilation of surface ocean colour had a significant impact on the estimated vertical distribution of chlorophyll-a during the spring bloom: the CV ratio (top-right panels of Fig. 4) is ~0.2 along the whole water column, indicating that the ocean colour's control extended below the simulated mixed layer depth (~20 m at the time of the year in Fig. 4). The posterior ensemble mean increased significantly in the top 40 meters, reproducing the bloom peak missed by the prior, as noted for the surface layer peak in mid-May shown in Figure 3a. While the prior was exhibiting a sub-surface maximum at ~ 60 meters for some ensemble members, such maximum is absent in the posterior ensemble. Instead, the posterior slightly enhanced the chlorophyll-a maximum at ~20-meter depth, more consistent with previous observational evidence at the study site (Binetti et al., 2020). Although the ensemble means of the posterior and prior are not significantly different below 70 meters, the CV ratio remains ~0.2 also in the deeper waters.

Figure 4 also shows that the assimilation of surface ocean colour controlled the ensemble spread of POC: the CV ratio is ~0.3 up to 70 m depth and ~0.4 in deeper waters. The prior and posterior ensemble mean, in this case, were not significantly different, similarly to the results for POC in May shown in Fig 3j.

3.2.2 Mediterranean Sea and BOUSSOLE station

The results of the corroboration of the indicator controllability with the 3D model of the Mediterranean Sea (Figures 5 and 6) shows that ocean colour assimilation reduced the ensemble dispersion in the simulation of the chlorophyll-a and phenological indicators at BOUSSOLE, as shown by the CV ratio below 1 for most of the simulated periods (Figure 5b). Assimilation also significantly impacted the simulation of all the other unobserved indicators, in both mean values and dispersion (fig 5d, f, h, and j). The ensemble spread, however, was increased in some instances leading the CV ratio to values higher than one (e.g., for POC in fig 5j). In general, we noted quite different effects on the indicators in the winter and summer (see, e.g., the CV ratio values above and below 1, respectively, for TE in Fig 5j).

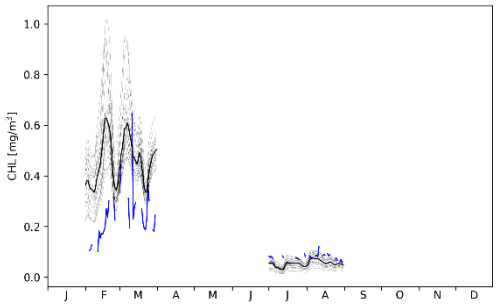
In winter, the assimilation controlled chlorophyll-a by shifting its peak form February to March, thus delaying the initiation of the simulated bloom (Fig. 5b). In this case the uncertainty was decreased, since the CV ratio was below 1, in February. In summer, the assimilation halved the CV.

Assimilation controlled also the estimation of the unobserved indicators in Figure 5. In some instances, assimilation decreased the uncertainty of the estimates, in particular for the PFT ratio and trophic efficiency in summer, with CV ratio of ~0.5 (Fig 5d) and 0.7 (Fig. 5h), respectively. In other instances, the strong control increased the uncertainty, in particular for POC and trophic efficiency in winter (CV>1 in Fig 5j and 5h, respectively). The temporal patterns of the changes in the mean values and the spread of the ensemble were linked to the ones described for chlorophyll-a (Fig. 5b).

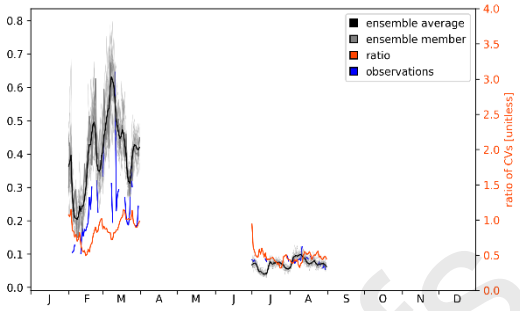
The capability of the assimilation to control indicators at BOUSSOLE (Figure 5) and at PAP (Figure 3) shows some similarities. For example, at both BOUSSOLE (Fig 5d) and PAP (Fig 3d) total chlorophyll-a assimilation impacted the PFT ratio differently in winter than in the summer. At BOUSSOLE, the phytoplankton bloom was corrected in February by lowering the dominant large phytoplankton type relatively to the small phytoplankton (the ratio decreased from 0.55 to 0.43 in February). On the other hand, in summer the mean PFT ratio was kept substantially unchanged in favour of the small fraction, while the ensemble spread halved. As in PAP, the correction of the summer chlorophyll-a bias substantially increased the primary and secondary production and POC (see the posterior maxima of NPP: Fig. 5f, TE: Fig. 5h and POC: Fig 5j in August and September). These changes in the structure of the plankton community in winter and in plankton production in summer are coherent with previous findings on plankton and biogeochemical dynamics at BOUSSOLE, i.e., shifts from large to small phytoplankton and increased primary production from winter to summer (Álvarez et al., 2023; Mayot et al., 2017).

Figure 6 shows that ocean colour assimilation controlled the vertical patterns of chlorophyll-a and POC at BOUSSOLE in both winter and summer. During the winter bloom (top-left panels) the assimilation reduced the uncertainty of the chlorophyll-a profile in the first 70 meters, while the mean value of the profile did not change significantly with respect to the prior non-assimilative ensemble simulation. Such persistence of the chlorophyll-a mean was contrasted by POC, which decreased in concentration (by one order magnitude) and increased in relative uncertainty (CV ratio up to 2) throughout the water column (bottom-left panels), possibly because of the continued decrease of POC during the previous months (Figure 5j). In summer (Fig. 6, right-panels) the ensemble mean and spread of the deep chlorophyll-a maximum shallowed and increased in magnitude, and this was reflected by analogous changes in the distribution and uncertainty of POC.

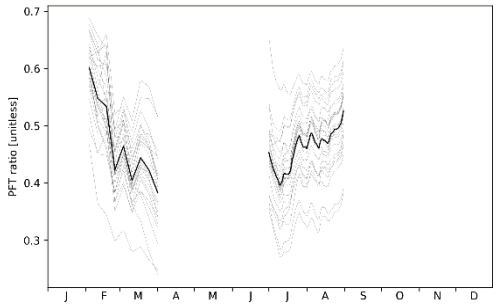
a



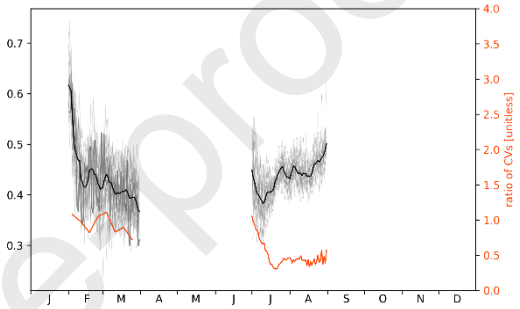
b



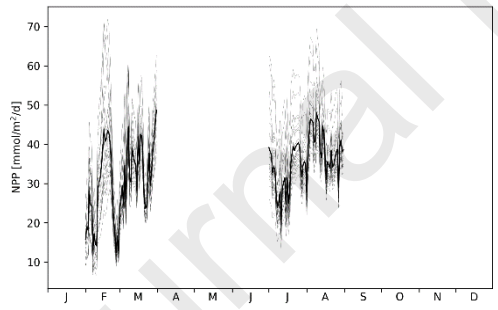
c



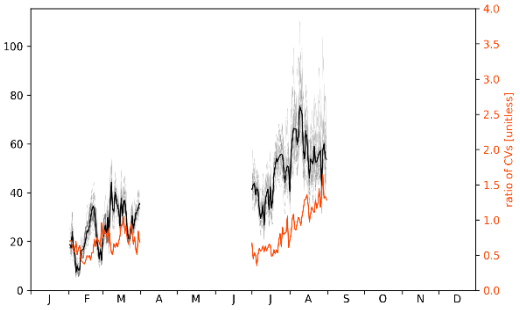
d



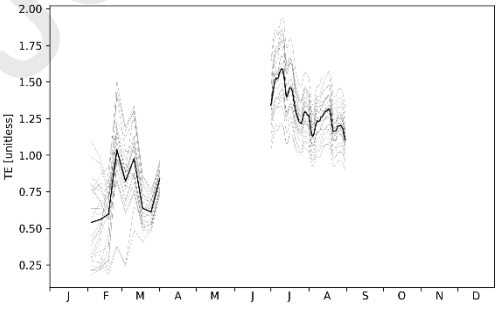
e



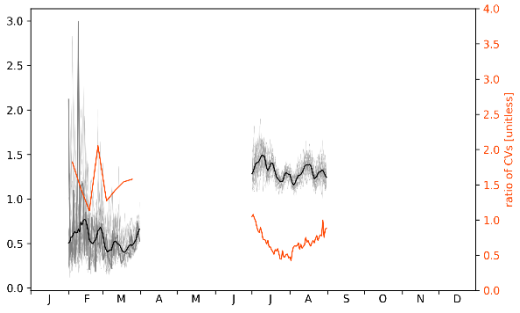
f



g



h



i j

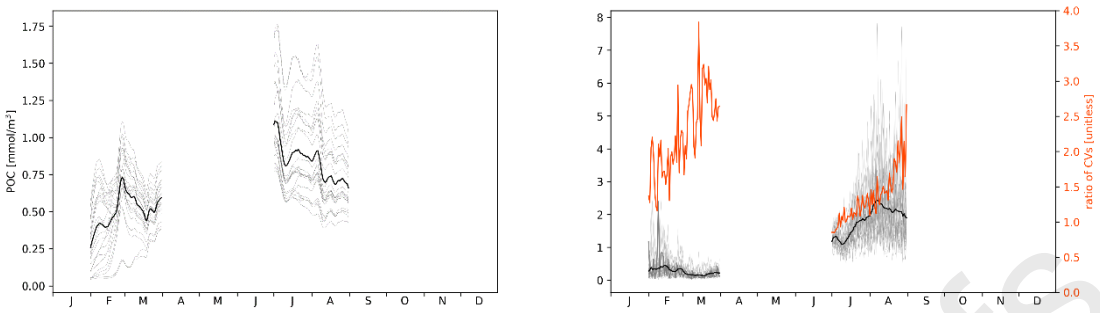


Figure 5. Time series of the indicators at BOUSSOLE station, for the prior (left panel) and posterior (right panel) ensembles. The simulations were performed for winter and summer 2019.

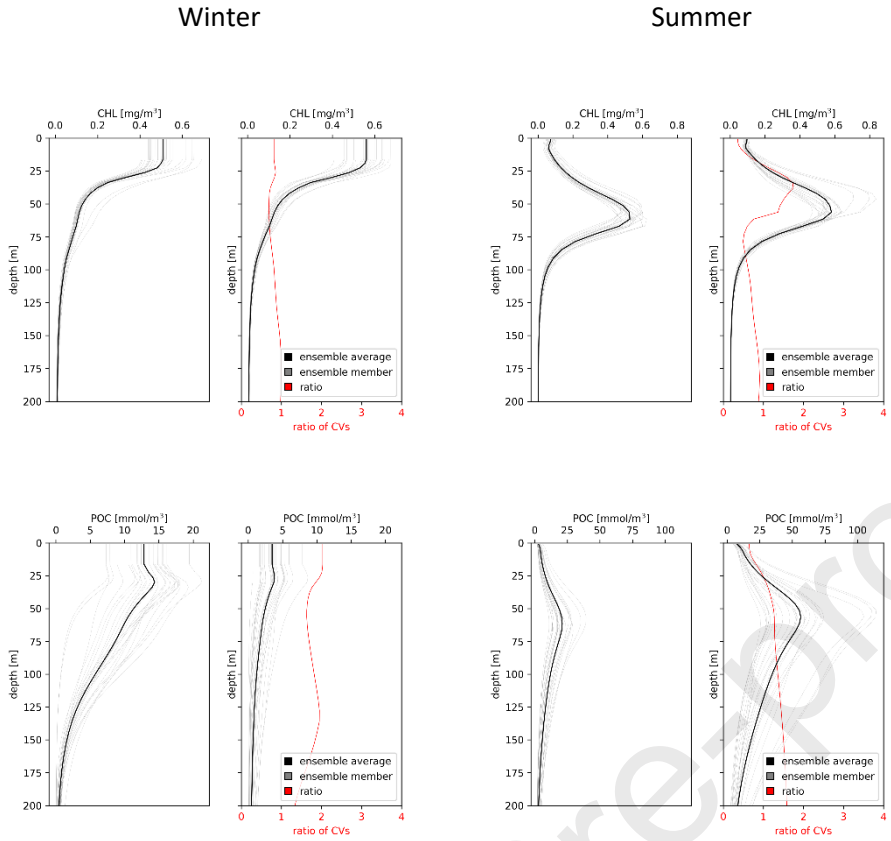


Figure 6. Vertical distribution of the chlorophyll-a concentration (upper panels) and POC (lower panels) at BOUSSOLE station. For each indicator, the distributions of the prior (left graph) and posterior (right graph) and the ratio of their variance (right graph, red line) are presented for winter (11th March 2019) and summer (8th August 2019).

The controllability of the indicators by multiplatform assimilation at the scale of the whole Mediterranean Sea is evaluated for winter in Figures 7 and summer in Figure 8.

In winter (11th of March, Figure 7), assimilation produced three different types of effect in the different Mediterranean sub-domains. Firstly, the satellite data strongly controlled the spatial distribution of chlorophyll-a in the most productive areas of the Mediterranean Sea (i.e, the northwestern basin) producing both positive and negative chlorophyll-a increments during the bloom (see Fig. 7b). Secondly, systematic positive increments occurred in the two marginal seas (i.e., Aegean and Adriatic Seas). Finally, the assimilation mostly decreased the chlorophyll-a concentration in the rest of the Mediterranean Sea. The map of the CV ratio of chlorophyll-a shows that corrections mostly lower the chlorophyll-a uncertainty (see the Aegean, Alboran, Levantine and northern Ionian seas in Fig. 7b). In the north-western Mediterranean, the control of the complex spatial pattern of the bloom impacts the spatial variability of the uncertainty of chlorophyll-a. The CV ratio of NPP decreased quite homogeneously in space (Fig. 7d). On the other hand, the uncertainty of the PFT ratio decreased spatially similarly to the uncertainty of chlorophyll-a, confirming the link between this indicator and chlorophyll-a (Figure 7c). In several areas, the decrease of chlorophyll-a favours the small phytoplankton groups which are characterized by faster growth rate (e.g. 3.5 d^{-1} for smaller picophytoplankton versus 1.5 d^{-1} for larger dinoflagellate in BFM: Lazzari et al, 2012), leading NPP to increase in Fig. 7d. Reducing biomass of the large phytoplankton group decreased the rate of transfer

of organic matter toward zooplankton and detritus (e.g., negative anomaly > 0.6 in the central Mediterranean, Fig. 7e). Such decrease over time decreased TE and POC and increased their uncertainty in large areas of the Mediterranean (Figures 7e and 7f, respectively). The concentration of large phytoplankton was increased by chlorophyll-a increments in March in some productive areas, e.g., in the northwestern Mediterranean as already noted in Figure 5. In this case, NPP increased in relation to the overall higher biomass of the primary community. Here POC still decreased, because it integrated the POC decrease in the previous simulated months (Fig. 5).

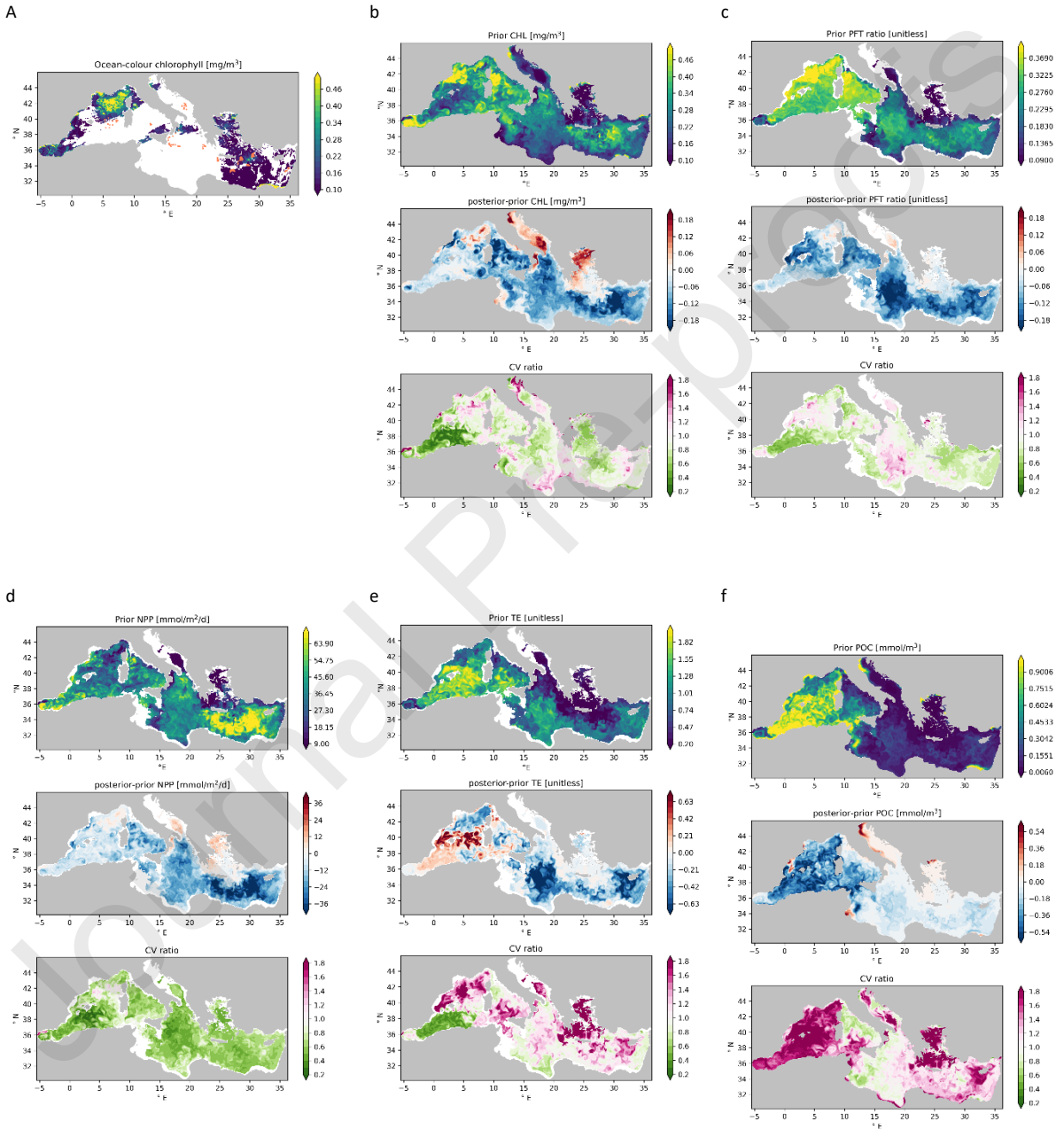
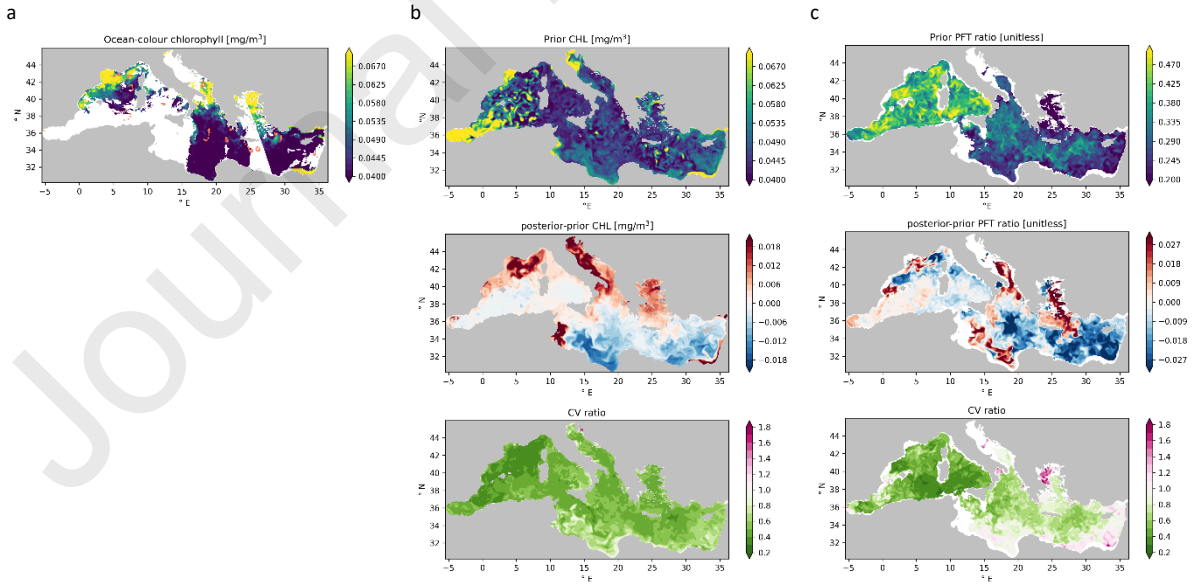


Figure 7. Ocean-colour chlorophyll-a [mg m^{-3}] (a) observed and (b) a-priori simulated in the Mediterranean Sea on 11th March. Second and third rows in (b) report the assimilation increments on chlorophyll-a and the coefficient of variation (CV) ratio. A priori, increments and CV ratio for large over all phytoplankton: PFT ratio, (c) primary production: NPP (d), trophic efficiency: TE (e), and POC (f).

In summer (8th of August, Figure 8), the assimilation controlled chlorophyll-a and reduced the CV ratio in the whole Mediterranean Sea (Fig. 8b). In summer, however, the impact of assimilation was lower than in the winter. This is due to the lower values of chlorophyll-a observed in summer than in winter in general (up to one order magnitude differences in chlorophyll concentration in Fig 7a and 8a) along with one order magnitude smaller bias in winter than in summer (Figure 5a). This led the increments of chlorophyll (Fig. 8b) to be ten-fold lower than those observed in the winter (Fig. 7b). Consequently, the impact of chlorophyll-a assimilation on the indicators was relatively low and the spatial patterns of the indicator increments are not straightforwardly linked to those of chlorophyll-a, see, e.g., the PFT ratio (Fig. 8c). Local trophic and physical conditions likely significantly changed the spatial structure of the ensemble covariances, hence the spatial changes of the plankton analysis driven by chlorophyll-a assimilation. For example, NPP had positive increments in the Northern part of the Mediterranean (in particular in the northwestern basin) and negative increments elsewhere with decreasing uncertainty (lower CV) across the whole Mediterranean Sea (8d).

The change in phytoplankton influenced the TE increments (Fig. 7e). For example, the increases of TE in the eastern Mediterranean Sea are linked to chlorophyll-a decreases. This counterintuitive feedback can be related to a delay in the response of zooplankton to phytoplankton changes. Higher uncertainties correspond to the positive TE increments, highlighting that small changes of phytoplankton dynamics in stratified and oligotrophic conditions can produce amplified responses in the higher trophic levels of the BFM model. In the rest of the Mediterranean Sea, the corrections of TE were very small, and TE uncertainty decreased, as the CV ratio of the other indicators.

POC corrections were not always linked to the increments of chlorophyll-a, and POC uncertainty mostly increased with assimilation (Fig. 8f). Interestingly, POC dispersion increased less in the areas where BGC-Argo profiles were available for assimilation, i.e., in the northwestern Mediterranean, in the areas east and south of Crete, south of Cyprus, and in the southern Adriatic Sea. This suggests that the assimilation of chlorophyll-a profiles might have mitigated the increase in POC dispersion.



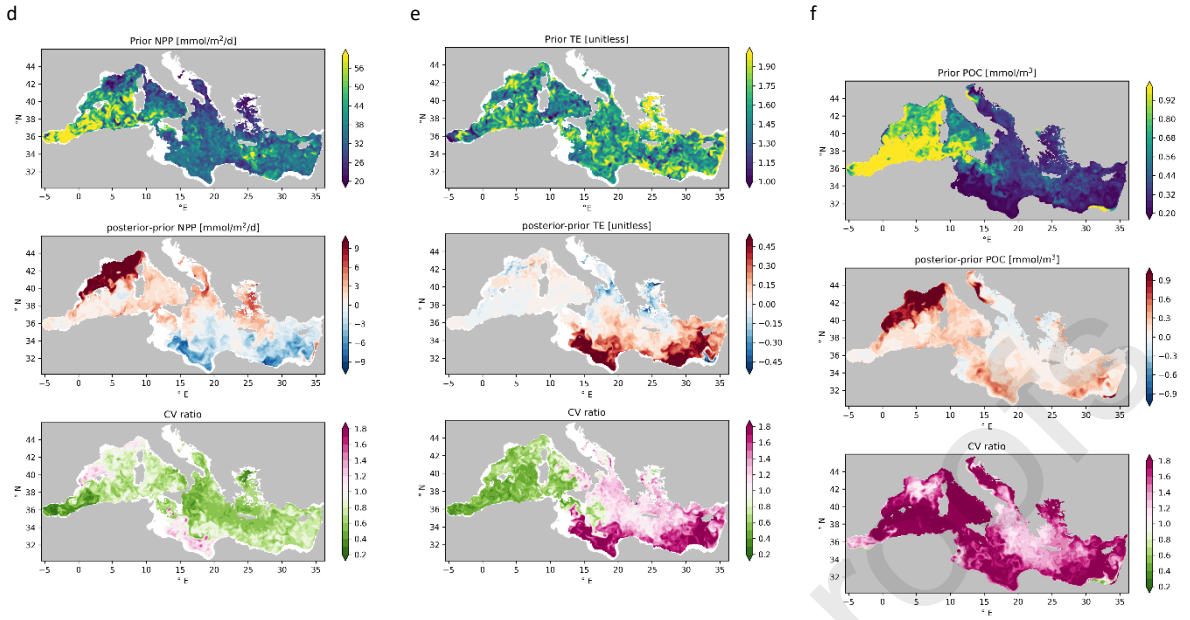


Figure 8. Ocean-colour chlorophyll-a [mg m^{-3}] (a) observed and (b) a-priori simulated in the Mediterranean Sea on 8th August 2019. Second and third rows in (b) report the assimilation increments on chlorophyll-a and the coefficient of variation (CV) ratio. A priori, increments and CV ratio for large over all phytoplankton (c) primary production: NPP (d), trophic efficiency: TE (e) and POC (f).

3.3. Portability of the control results from 1D to 3D systems

The 3D corroboration analysis overall confirmed the results of the 1D screening analysis, but it also showed that the actual assimilation of observations can expand the number of indicators that can be controlled operationally.

For example, in the North Atlantic modelling system, the 3D assimilative results in Figure 3 and Figure 4 corroborated that the ecosystem indicators are controlled by the assimilation of ocean colour during winter mixed conditions, similarly to the 1D sensitivity analysis (Figure 1). However, in Figure 3b we noted that the assimilated ocean colour chlorophyll had a lower control of chlorophyll-a in late autumn and winter than in spring and summer, in relation to the season-dependant misfit of the model to the ocean colour data, and the season-dependant availability of ocean-colour data for assimilation. This outcome differs from, but does not contradict, the screening in Figure 1, which showed OC controlling the phenology indicators more in the mixed winter season than in the stratified summer period (higher sensitivities of most indicators with respect to OC in “mixed” than “strat”, at both BATS and L4 in Figure 1). The difference in the control results for the 1D and 3D analyses are therefore explained by the “real operational situations” of a 3D simulation which might be biased, and ocean colour data that might be unavailable for assimilation. These situations encountered in the 3D North Atlantic simulation were not represented in the 1D screening analysis.

In the Mediterranean modelling system, the results in Figures 5-8 corroborate the outcome of the sensitivity shown in Figure 1 on the capacity of ocean colour chlorophyll-a to control the indicators. They confirm the ocean colour control of the surface concentration of chlorophyll-a (PHE_maxsc) in winter. However, ocean colour assimilation controlled the indicators also in summer in the 3D Mediterranean simulations (e.g. on POC in Figure 6 and 8), while the OC control of the indicators in

the 1D analyses was negligible in the summer ("strat" season in Figure 1). In the Mediterranean Sea, the above links between the controllability evaluated in 1D and 3D are further confirmed by the results of the supplementary 1D screening analysis of BFM at BOUSSOLE (Figure S5 in the Supplementary Material). Again, the ocean colour control on indicators was evident in winter mixed conditions but not in the summer stratified conditions in 1D (Fig. S5 in the supplementary material).

The portability of results from 1D model configurations to 3D set-ups is a relevant, open matter in marine ecosystem modelling, in particular in the framework of parameter optimization (e.g. Wang et al., 2020; Hoshiba et al., 2018; Kane et al., 2011; Friedrichs et al. 2007, Oschlies and Schartau, 2005). Computationally cheap 1D models allow much faster analysis than expensive 3D configurations. However, horizontal transport and ecosystem variability make challenging using 1D optimized parameters in 3D. For example, Hoshiba et al (2018) found that parameters optimized by data assimilation in 1D model and used in the 3D model led the 3D model to outperform the 1D, by benefiting from horizontal transport. On the other hand, Wang et al. (2020) found that the direct use of 1D optimized parameters led to some degradation of the 3D simulations, because of both advection and the ecosystem heterogeneity. This latest issue has been addressed successfully in 3D applications, by simultaneously optimizing parameters at different stations to obtain one single set for the 3D model (Kane et al., 2011; Oschlies and Schartau, 2005) or by optimizing the parameters at contrasting sites and use regionally variable parameters in the 3D model (Hoshiba et al., 2018).

Our analysis of portability of indicators control from 1D to 3D configurations is unprecedented, to the authors' knowledge. The results presented here suggest that some features of realistic operational systems such as time-variable model bias and data availability can lead assimilation to control ecosystem indicators differently than the simple perturbation of the initial conditions in the 1D screening analysis. Arguably, additional sources of difference between the screening and corroboration results are the 3D horizontal transport, the ecosystem variability, and the use of multivariate ensemble data assimilation methods. At one specific location in a 3D framework, horizontal transport of nearby water masses that underwent ocean colour analysis can preserve memory of the assimilative control through time, in particular when the gradients are relatively low in the assimilated two-dimensional ocean colour fields. This memory is not available in 1D configurations. Our screening at 1D sites with contrasting trophic and hydrodynamic conditions (i.e. L4 and BATS in the Atlantic, supplemented by BOUSSOLE in the Mediterranean, in both fully mixed and stratified seasons) took account of different conditions encounterable within the three-dimensional domains of the controllability corroboration, in analogy with Hoshiba et al., 2018, Kane et al., 2011; Oschlies and Schartau, 2005 for parameter optimization. Still, the 1D screening cannot capture the whole spatial-temporal gradients of the 3D simulated ecosystems. Finally, the ensemble multivariate assimilation schemes applied here in the 3D corroboration corrected a larger sub-set of model variables (phytoplankton and nutrients: see Table 4) than the initialization in the 1D screening (phytoplankton only: see Table 2). This leads the 3D ocean colour assimilation to correct different levels of the trophic web simultaneously, with an overall stronger control on the 3D simulation of the indicators than in the 1D screening.

4. Conclusions and recommendations

All the marine ecosystem indicators were constrained by the observable variables investigated in this work. The indicators of phytoplankton phenology were controlled, and also improved, when merging the biogeochemical observations with the models. The indicators of the plankton community structure and production were also controlled, and their uncertainty decreased in general. Trophic efficiency and POC were also controlled, but their uncertainty increased in some instances. The degree of controllability of the indicators was in general higher when internal trophic processes dominated

over external forcings in driving the variability of the indicators, i.e., dissolved nutrient limitation versus air-sea oxygen flux. The model formulations of the processes, rather than their complexity, influenced the controllability of the indicators.

The methodological framework proposed here is unique in the fact that it assembled five operational biogeochemical models to assess the controllability of the indicators, by using the same physical and forcing set-ups. Such multi-model ensemble approach makes the assessment of controllability more robust with respect to the potential differences among the specific biogeochemical models. The use of the one-dimensional configuration allowed us to perform an extended Monte Carlo-based sensitivity analysis, which would not be feasible with the three-dimensional operational systems. We note that the screening results were more conservative than those of the corroboration. In fact, indicators controllable in 1D were in general confirmed controllable in the 3D systems, while indicators non controllable in 1D were found, in several cases, controllable in the 3D corroboration analysis. This was related to differences between realistic operational situations (i.e., time-variable model bias and availability of data for assimilation), though horizontal processes, ecosystem variability, as well as multivariate assimilative analysis are likely additional sources of difference. Therefore, the portability of the results from one- to three-dimensional configurations is not straightforward and requires focussed corroboration analyses with the operational system, informed by preliminary screening analyses with computationally efficient model configurations.

The ensemble framework proposed here provides a tool to assess the uncertainty of the indicators. The definition of controllability based on the reduction of the ensemble spread is useful in providing a quantitative index synthetising results across models, domains, and seasons. Here we considered also shifts of the ensemble mean value as an indication of control in qualitative discussion of the results. Future applications should include shifts of the mean in the control-index quantitatively (Skákala et al., 2024).

Controllability can in some cases increase the uncertainty of the estimated indicators because non-linear dynamics can increase the spread of the ensemble fuelled by perturbations of the observable variables. Also, controllability does not necessarily imply model skill in estimating the indicators: this needs to be validated with independent observations. In our case the only available observations were for the plankton phenology indicators. We recommend increasing and sustaining the observation of biogeochemical and biological variables to be able to validate more extensively the indicators simulated by models. Of particular interest is the observation of biogeochemistry in the ocean interior and the acquisition of more data of plankton biomass, composition, and production.

Ultimately, we recommend that the assessment of controllability becomes a standard practice in the design of operational monitoring, reanalysis and forecast systems, along with other established practices such as product validation. The controllability analysis will define what observations should be monitored and assimilated to effectively constrain the indicators delivered to the users of operational marine services.

Acknowledgements

This work was funded by the project SEAMLESS (“Services based on Ecosystem data AssiMiLation: Essential Science and Solutions”), which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101004032. Part of the work was supported by the project NECCTON (“New Copernicus Capability for Trophic Ocean Networks”), which has received funding from Horizon Europe RIA under Grant Number 101081273. We also acknowledge support from the UK Natural Environment Research Council (NERC) single centre national capability

programme – Climate Linked Atlantic Sector Science (NE/R015953/1) and NERC supported National Centre for Earth Observation (NCEO), contract number PR140015.The multi-model ensemble and their one-dimensional configurations at stations BATS and L4 used in this work are available here: https://github.com/BoldingBruggeman/seamless-notebooks. The code of parsac is available here: https://doi.org/10.5281/zenodo.4280520.

References

Álvarez, E., Cossarini, G., Teruzzi, A., Bruggeman, J., Bolding, K., Ciavatta, S., Vellucci, V., D'Ortenzio, F., Antoine, D., Lazzari, P., 2023. Chromophoric dissolved organic matter dynamics revealed through the optimization of an optical-biogeochemical model in the northwestern Mediterranean Sea. *Biogeosciences, 20*(22), pp. 4591-4624. https://doi.org/10.5194/bg-20-4591-2023

Andersen, T.K., Bolding, K., Nielsen, A., Bruggeman, J., Jeppesen, E. and Trolle, D., 2021. How morphology shapes the parameter sensitivity of lake ecosystem models. Environmental Modelling & Software, 136, p.104945.

Antoine, D., Morel, A., Babin, M. and Claustre, H., 2002. Measurements and modeling of apparent optical properties of ocean waters in support to ocean color data calibration, validation and merging. *NASA Technical memorandum*, (2002-210005).

Artioli, Y., Blackford, J.C., Butenschön, M., Holt, J.T., Wakelin, S.L., Thomas, H., Borges, A.V. and Allen, J.I., 2012. The carbonate system in the North Sea: Sensitivity and model validation. *Journal of Marine Systems*, *102*, pp.1-13.

Aumont, O., Éthé, C., Tagliabue, A., Bopp, L. and Gehlen, M., 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development* , *8*(2), pp.1375-1509.

Baretta-Bekker, J.G., Baretta, J.W. and Rasmussen, E.K., 1995. The microbial food web in the European regional seas ecosystem model. *Netherlands Journal of Sea Research*, *33*(3-4), pp.363-379.

Barbière, J., Isensee, K., School K., 2019, Indicator methodology for 14.3.1, Information Document, UNESCO-IOC. https://oceanexpert.org/document/26155

Binetti, U., Kaiser, J., Damerell, G.M., Rumyantseva, A., Martin, A.P., Henson, S. and Heywood, K.J., 2020. Net community oxygen production derived from Seaglider deployments at the Porcupine Abyssal Plain site (PAP; northeast Atlantic) in 2012–13. *Progress in Oceanography*, *183*, p.102293.

Brasseur, P., Brankart, J. M., Popov, M., Skákala, J., Teruzzi, A., Spada, S., Cossarini, G., Wakamatsu, T., Bertino, L., Sun, Y., & Nerger, L. (2023). Observability of the target indicators in the 3D CMEMS MFC systems (D3.4) (Version 4). Zenodo. https://doi.org/10.5281/zenodo.7584865

Bruggeman, J. and Bolding, K., 2014. A general framework for aquatic biogeochemical models. *Environmental modelling & software*, *61*, pp.249-265.

Bruggeman, and Bolding, 2020. parsac: parallel sensitivity analysis and calibration (0.5.7). Zenodo. https://doi.org/10.5281/zenodo.4280520

Butenschön, M., Clark, J., Aldridge, J.N., Allen, J.I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S. and Lessin, G., 2016. ERSEM 15.06: a generic model for marine biogeochemistry

and the ecosystem dynamics of the lower trophic levels. *Geoscientific Model Development*, *9*(4), pp.1293-1339.

Canu, D.M., Ghermandi, A., Nunes, P.A., Lazzari, P., Cossarini, G. and Solidoro, C., 2015. Estimating the value of carbon sequestration ecosystem services in the Mediterranean Sea: An ecological economics approach. *Global Environmental Change*, *32*, pp.87-95.

Carrassi, A., Bocquet, M., Bertino, L. and Evensen, G., 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(5), p.e535.

Carrassi, A., Bocquet, M., Demaeyer, J., Grudzien, C., Raanes, P. and Vannitsem, S., 2022. Data assimilation for chaotic dynamics. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)*, pp.1-42.

Ciavatta, S., Torres, R., Saux-Picart, S. and Allen, J.I., 2011. Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?. *Journal of Geophysical Research: Oceans*, *116*(C12).

Ciavatta, S., Brewin, R.J.W., Skákala, J., Polimene, L., de Mora, L., Artioli, Y. and Allen, J.I., 2018. Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations. *Journal of Geophysical Research: Oceans*, *123*(2), pp.834-854.

Cossarini, G., Querin, S. and Solidoro, C., 2015. The continental shelf carbon pump in the northern Adriatic Sea (Mediterranean Sea): Influence of wintertime variability. *Ecological Modelling*, *314*, pp.118-134.

Daewel, U. and Schrum, C., 2013. Simulating long-term dynamics of the coupled North Sea and Baltic Sea ecosystem with ECOSMO II: Model description and validation. *Journal of Marine Systems*, *119*, pp.30-49.

Díaz-Seoane, S., Blas, A.B. and Villaverde, A.F., 2023. Controllability and accessibility analysis of nonlinear biosystems. *Computer Methods and Programs in Biomedicine*, *242*, p.107837.

d'Ortenzio, F., & Ribera d'Alcalà, M. (2009). On the trophic regimes of the Mediterranean Sea: A satellite analysis. Biogeosciences, 6(2), 139–148. https://doi.org/10.5194/bg-6-139-2009

European Commission, 2019. Expression of User Needs for the Copernicus Programme, https://www.copernicus.eu/sites/default/files/2019-10/STAFF_WORKING_PAPER_2019-394-Expression_of_User_Needs_for_the_Copernicus_Programme.pdf

Fennel, K., Gehlen, M., Brasseur, P., Brown, C.W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C.A., Ford, D., Friedrichs, M.A. and Gregoire, M., 2019. Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. *Frontiers in Marine Science*, *6*, p.89.

Friedrichs, M. A. M., Hood, R. R., & Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. Deep-Sea Research Part II: Topical Studies in Oceanography, 53(5–7), 576–600. https://doi.org/10.1016/j.dsr2.2006.01.026

Friedrichs, M. A. M., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., McGillicuddy, D. J., Moore, J. K., Schartau, M., Spitz, Y. H., & Wiggert, J. D. (2007). Assessment of skill and portability in regional marine biogeochemical models: Role of

multiple planktonic groups. Journal of Geophysical Research: Oceans, 112(8). https://doi.org/10.1029/2006JC003852

García, G.C., Osses, A. and Puel, J.P., 2011. A null controllability data assimilation methodology applied to a large scale ocean circulation model. *ESAIM: Mathematical Modelling and Numerical Analysis*, *45*(2), pp.361-386.

Gebbie, G. and Hsieh, T.L., 2017. Controllability, not chaos, key criterion for ocean state estimation. *Nonlinear Processes in Geophysics*, *24*(3), pp.351-366

Geider, R.J., 1987. Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton. *New Phytologist*, pp.1-34.

Geider, R.J., MacIntyre, H.L. and Kana, T.M., 1997. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature. *Marine Ecology Progress Series*, *148*, pp.187-200.

Gelb, A. 1974. Applied Optimal Estimation, MIT Press, Cambridge, MA.

Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C., Chauhan, P., Choi, J.K., Chuprin, A. and Ciavatta, S., 2019. Satellite ocean colour: Current status and future perspective. *Frontiers in Marine Science*, *6*, p.485.

Gutknecht, E., Reffray, G., Mignot, A., Dabrowski, T. and Sotillo, M.G., 2019. Modelling the marine ecosystem of Iberia–Biscay–Ireland (IBI) European waters for CMEMS operational applications. *Ocean Science*, *15*(6), pp.1489-1516.

Hartman, S.E., Bett, B.J., Durden, J.M., Henson, S.A., Iversen, M., Jeffreys, R.M., Horton, T., Lampitt, R. and Gates, A.R., 2021. Enduring science: three decades of observing the Northeast Atlantic from the Porcupine Abyssal Plain Sustained Observatory (PAP-SO). *Progress in Oceanography*, *191*, p.102508.

Hoshiba, Y., Hirata, T., Shigemitsu, M., Nakano, H., Hashioka, T., Masuda, Y., and Yamanaka, Y.: Biological data assimilation for parameter estimation of a phytoplankton functional type model for the western North Pacific, Ocean Sci., 14, 371–386, https://doi.org/10.5194/os-14-371-2018, 2018.

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A. and Weston, P., 2018. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), pp.1257-1278.

Jones, E.M., Baird, M.E., Mongin, M., Parslow, J., Skerratt, J., Lovell, J., Margvelashvili, N., Matear, R.J., Wild-Allen, K., Robson, B. and Rizwi, F., 2016. Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef. *Biogeosciences*, *13*(23), pp.6441-6469.

Kane, A., C. Moulin, S. Thiria, L. Bopp, M. Berrada, A. Tagliabue, M. Crépon, O. Aumont, and F. Badran (2011), Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations, *J. Geophys. Res.*, 116, C06011, doi:10.1029/2009JC006005.

Lamouroux, J., Perruche, C., Mignot, A., Paul, J., and Szczypta, C., 2022. Quality information document for the Global Biogeochemical Analysis and Forecast Product GLOBAL_ANALYSIS_FORECAST_BIO_001_028, doi:10.48670/moi-00015

Lazzari, P., Solidoro, C., Ibello, V., Salon, S., Teruzzi, A., Béranger, K., Colella, S. and Crise, A., 2012. Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: a modelling approach. *Biogeosciences*, *9*(1), pp.217-233.

Lazzari, P., Solidoro, C., Salon, S. and Bolzon, G., 2016. Spatial variability of phosphate and nitrate in the Mediterranean Sea: A modeling approach. *Deep Sea Research Part I: Oceanographic Research Papers*, *108*, pp.39-52.

Lessin G, Raudsepp U, Stips A (2014) Modelling the Influence of Major Baltic Inflows on Near-Bottom Conditions at the Entrance of the Gulf of Finland. PLoS ONE 9(11): e112881. https://doi.org/10.1371/journal.pone.0112881

Le Traon, P.Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V.E., Kreiner, M.B. and Benkiran, M., 2019. From observation to information and users: The Copernicus Marine Service perspective. *Frontiers in Marine Science*, *6*, p.234.

Maar, M., Møller, E.F., Larsen, J., Madsen, K.S., Wan, Z., She, J., Jonasson, L. and Neumann, T., 2011. Ecosystem modelling across a salinity gradient from the North Sea to the Baltic Sea. *Ecological Modelling*, *222*(10), pp.1696-1711.

Madec, G., Bourdallé-Badie, R., Bouttier, P.A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D. and Ethé, C., 2017. NEMO ocean engine. In Scientific Notes of IPSL Climate Modelling Center (v4.2, Number 27). Zenodo. https://doi.org/10.5281/zenodo.6334656

Mayot, N., D'Ortenzio, F., Uitz, J., Gentili, B., Ras, J., Vellucci, V., Golbol, M., Antoine, D., and Claustre, H. (2017), Influence of the Phytoplankton Community Structure on the Spring and Annual Primary Production in the Northwestern Mediterranean Sea, J. Geophys. Res. Oceans, 122, 9918–9936, doi:10.1002/2016JC012668.

Muller-Karger, F.E., Miloslavich, P., Bax, N.J., Simmons, S., Costello, M.J., Sousa Pinto, I., Canonico, G., Turner, W., Gill, M., Montes, E. and Best, B.D., 2018. Advancing marine biological observations and data requirements of the complementary essential ocean variables (EOVs) and essential biodiversity variables (EBVs) frameworks. Frontiers in Marine Science, p.211.

Neumann, T., 2000. Towards a 3D-ecosystem model of the Baltic Sea. *Journal of Marine Systems*, *25*(3-4), pp.405-419.

Neumann, T., Siegel, H. and Gerth, M., 2015. A new radiation model for Baltic Sea ecosystem modelling. *Journal of Marine Systems*, *152*, pp.83-91.

Oschlies, A. and Schartau, M.: Basin-scale performance of a locally optimized marine ecosystem model, J. Mar. Res., 63, 335–358, 2005.

Pham, D.T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly weather review*, *129*(5), pp.1194-1207.

Popov, M., Brankart, J.M., Capet, A., Cosme, E. and Brasseur, P., 2024. Ensemble analysis and forecast of ecosystem indicators in the North Atlantic using ocean colour observations and prior statistics from a stochastic NEMO–PISCES simulator. *Ocean Science*, *20*(1), pp.155-180.

Ringgaard, I., Korabel V., Spruch L, Lindenthal A, and Huess, V., 2024. Product User Manual For the two Baltic Sea Products: BALTICSEA_MULTIYEAR_BGC_003_012 &

BALTICSEA_MULTIYEAR_PHY_003_011,
(https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-BAL-PUM-003-011-012.pdf)

Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., Solidoro, C. and Crise, A., 2019. Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts. *Ocean Science*, *15*(4), pp.997-1022.

Saltelli, A., Chan, K., Scott, M., 2000. Handbook of Sensitivity Analysis. Probability and Statistics Series. John Wiley & Sons Publishers, New York, USA

Sathyendranath, S., Brewin, R.J., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A.B., Dingle, J., Doerffer, R. and Donlon, C., 2019. An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI). *Sensors*, *19*(19), p.4285.

Sathyendranath, S., Brewin, R.J., Ciavatta, S., Jackson, T., Kulk, G., Jönsson, B., Vicente, V.M. and Platt, T., 2023. Ocean biology studied from space. *Surveys in Geophysics*, pp.1-22.

Skákala, J., Ford, D., Brewin, R.J., McEwan, R., Kay, S., Taylor, B., de Mora, L. and Ciavatta, S., 2018. The assimilation of phytoplankton functional types for operational forecasting in the northwest European shelf. *Journal of Geophysical Research: Oceans*, *123*(8), pp.5230-5247.

Spada, S., Teruzzi, A., Maset, S., Salon, S., Solidoro, C., and Cossarini, G.: GHOSH v1.0.0: a novel Gauss-Hermite High-Order Sampling Hybrid filter for computationally efficient data assimilation in geosciences, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2023-170, in review, 2023.

Teruzzi, A., Bolzon, G., Feudale, L. and Cossarini, G., 2021. Deep chlorophyll maximum and nutricline in the Mediterranean Sea: emerging properties from a multi-platform assimilated biogeochemical model experiment. *Biogeosciences*, *18*(23), pp.6147-6166.

UNESCO-IOC (2021). The United Nations Decade of Ocean Science for Sustainable Development (2021-2030) Implementation plan – Summary. Paris, UNESCO. (IOC Ocean Decade Series, 19)

Salihoglu, B., Garçon, V., Oschlies, A. and Lomas, M.W., 2008. Influence of nutrient utilization and remineralization stoichiometry on phytoplankton species and carbon export: A modeling study at BATS. *Deep Sea Research Part I: Oceanographic Research Papers*, 55(1), pp.73-107.

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 Science*, *16*(5), pp.1297-1315.

Skákala, J., Bruggeman, J., Brewin, R.J., Ford, D.A. and Ciavatta, S., 2020. Improved representation of underwater light field and its impact on ecosystem dynamics: A study in the North Sea. *Journal of Geophysical Research: Oceans*, *125*(7), p. e2020JC016122.

Skákala, J., Bruggeman, J., Ford, D., Wakelin, S., Akpınar, A., Hull, T., Kaiser, J., Loveday, B.R., O'Dea, E., Williams, C.A. and Ciavatta, S., 2022. The impact of ocean biogeochemistry on physics and its consequences for modelling shelf seas. *Ocean Modelling*, *172*, p.101976.

Skákala, J., Ford, D., Fowler, A., Lea, D., Martin, M.J. and Ciavatta, S., 2024. How uncertain and observable are marine ecosystem indicators in shelf seas?. *Progress in Oceanography*, *224*, p.103249.

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

United Nations, 2023. The UN Sustainable Development Goals. United Nations, New York, 2015. Available at (accessed 26 October January 2023): www.un.org/sustainabledevelopment/summit/

Vichi, M., Lovato, T., Mlot, E.G. and McKiver, W., 2015. Coupling BFM with ocean models. *Nucleus for the European Modelling of the Ocean, Release*, *1*.

Villaverde, A.F., 2019. Observability and structural identifiability of nonlinear biological systems. *Complexity*, *2019*(1), p.8497093.

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*(15), pp.4059-4074.

Wang, B., Fennel, K., and Yu L. 2021. Can assimilation of satellite observations improve subsurface biological properties in a numerical model? A case study for the Gulf of Mexico. *Ocean Science* 17.4, pp 1141-1156.

Widdicombe, C. E., D. Eloire, D. Harbour, R. P. Harris, P. J. Somerfield, Long-term phytoplankton community dynamics in the Western English Channel, *Journal of Plankton Research*, Volume 32, Issue 5, May 2010, Pages 643–655, https://doi.org/10.1093/plankt/fbp127

Yumruktepe, V.Ç., Samuelsen, A. and Daewel, U., 2022. ECOSMO II (CHL): a marine biogeochemical model for the North Atlantic and the Arctic. *Geoscientific Model Development*, *15*(9), pp.3901-3921.

Supplementary material

Figure S1. Results of the sensitivity analysis of the ecosystem indicators with respect to the observed variables at station BATS (upper panel) and L4 (lower panel). The height of the bars shows the magnitude of the sensitivities in eq.1, for the ten ecosystem indicators (rows), with respect to the seven observed variables (columns), during two representative seasons (mixed and stratified water column), computed with each of the five biogeochemical models (colours).

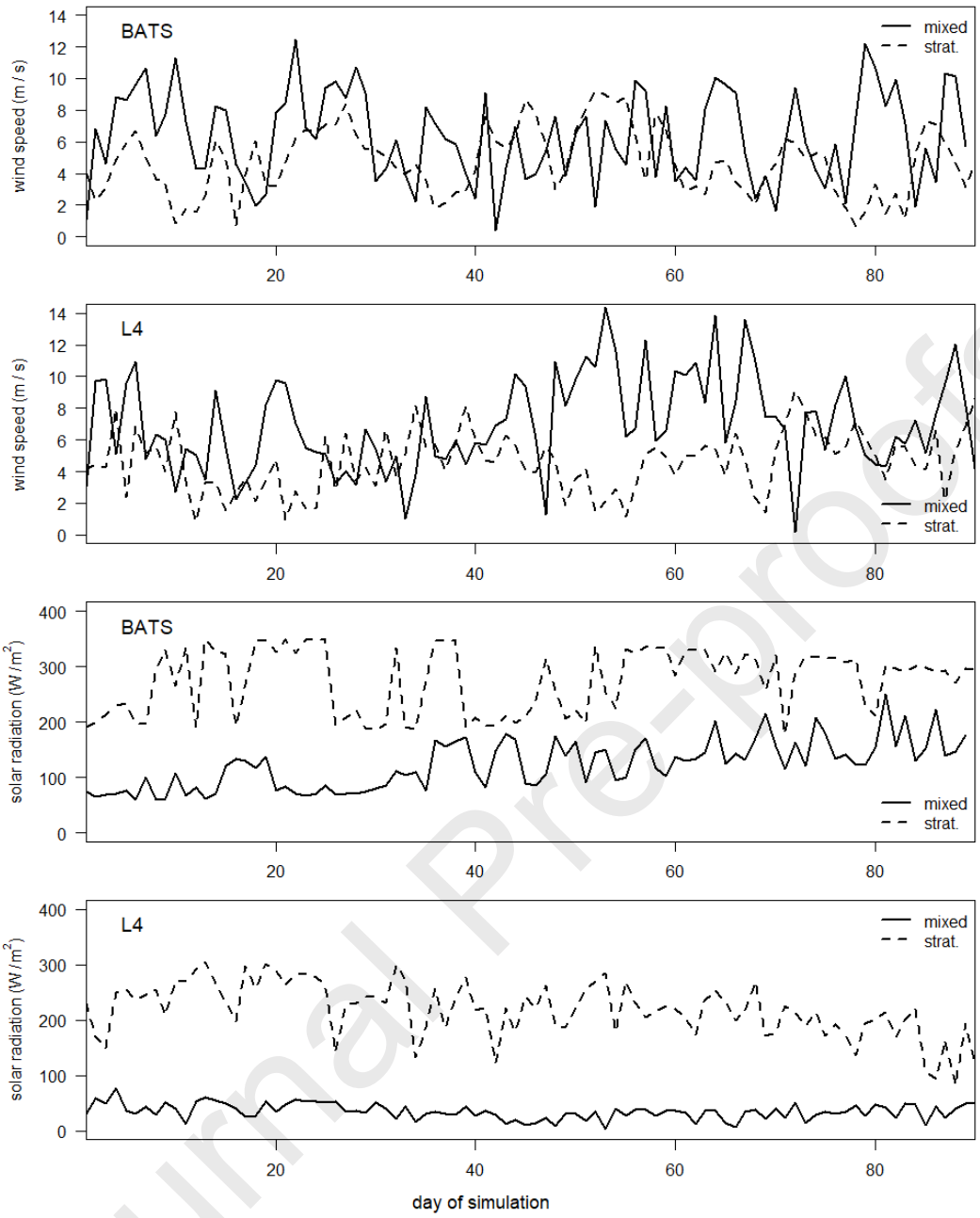


Figure S2. Wind speed and solar radiations data used as meteorological inputs in the screening analysis in the ninety-days long simulations of the mixed and stratified seasons at stations BATS and L4

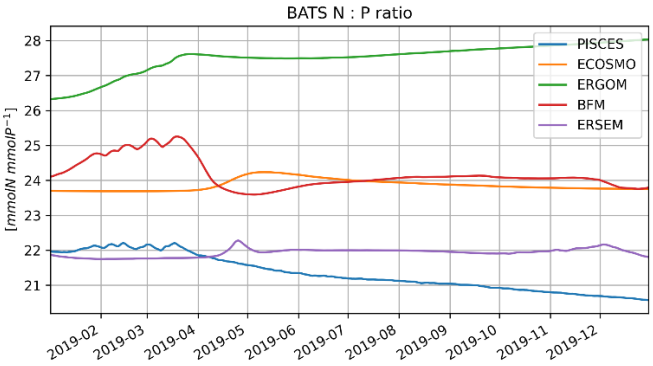


Figure S3 Multi-model reference simulations of the nitrogen to phosphate ratio at BATS. The colours of the lines distinguish the five different models of the ensemble (see section 2.1). The one year-long simulation includes both the stratified and fully mixed three month-long periods targeted in the screening analysis.

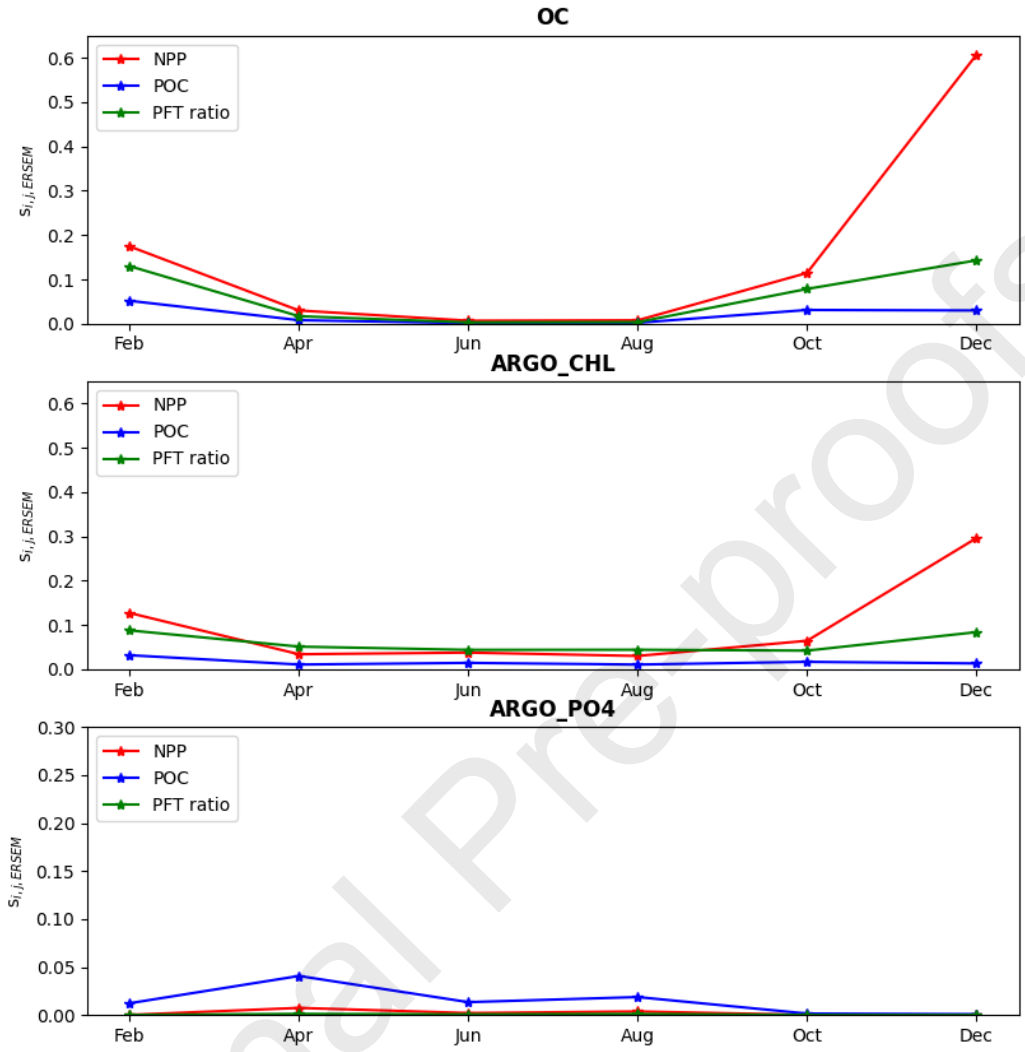


Figure S4. Dependency of the controllability of biogeochemical indicators on the time of the year when the model simulation is started. The plots show the sensitivities in eq. 1 for three selected indicators (net primary production: NPP, particulate organic carbon: POC, and plankton community structure: PFT ratio) with respect to three observable variables (surface ocean colour and profiles of chlorophyll-a and profiles of phosphate) computed in six simulations of the model ERSEM, starting with two-month lags, lasting for three-months, at Site L4.

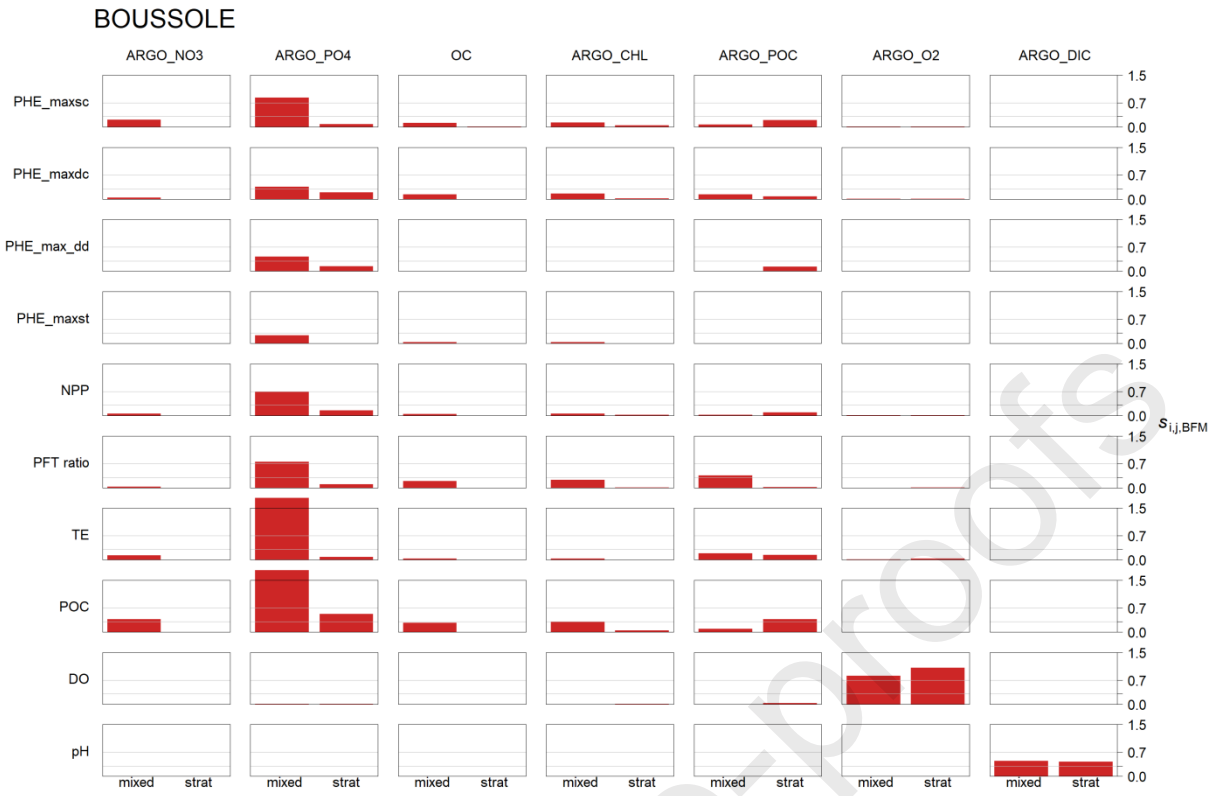


Figure S5. Results of the sensitivity analysis of the ecosystem indicators with respect to the observed variables at station BOUSSOLE, in the Northwestern Mediterranean Sea. The height of the bars shows the magnitude of the sensitivities in eq.1, for the ten ecosystem indicators (rows), with respect to the seven observed variables (columns), during two representative seasons (mixed and stratified water column), computed with the same model BFM of the 3D Mediterranean Sea corroboration analysis.

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: