Optimizing a large number of parameters in a Biogeochemical Model: A Multi-Variable BGC-Argo Data Assimilation Approach

Quentin Hyvernat[1,2], Alexandre Mignot[1], Elodie Gutknecht[1], Giovanni Ruggiero[1], Coralie Perruche[1], Guillaume Samson[1], Raphaëlle Sauzède[3], Olivier Aumont[4], Hervé Claustre[2], Fabrizio D'Ortenzio[2]

[1]Mercator Ocean International, 31400 Toulouse, France
[2]Laboratoire d'Océanographie de Villefranche, CNRS, Sorbonne Université, 06230 Villefranche-sur-Mer, France
[3]Institut de la Mer de Villefranche, CNRS, Sorbonne Université, 06230 Villefranche-sur-Mer, France
[4]Sorbonne Université (CNRS/IRD/MNHN), LOCEAN-IPSL, Paris, France

Correspondence to: Quentin Hyvernat (qhyvernat@mercator-ocean.fr)

Abstract. The predictive accuracy of marine biogeochemical models is fundamentally limited by uncertainty in their parameter values. We present a parameter ~~optimization~~ optimisation framework using iterative Importance Sampling (iIS) to constrain the PISCES ~~model~~ biogeochemical model within a one-dimensional representation of the ocean by leveraging the ~~rich~~ comprehensive, multi-variable dataset provided by Biogeochemical-Argo (BGC-Argo) floats. Using ~~data~~ seasonal observations from a single BGC-Argo float in the North Atlantic during the year 2015, we assimilate ~~a comprehensive suite of 20~~ twenty observational quantities derived from eight biogeochemical ~~metrics~~ tracers to ~~constrain~~ directly optimise all 95 poorly known model parameters ~~of the PISCES model within a 1D vertical configuration. Our.~~ A prerequisite global sensitivity analysis (GSA) identifies parameters controlling zooplankton dynamics as the dominant source of model sensitivity ~~for~~ at this ~~specific~~ site. We compare three strategies: (1) ~~optimizing~~ optimising a subset of parameters selected for their strong direct influence (Main effects); (2) ~~optimizing~~ optimising a larger subset that also includes parameters influential through non-linear interactions (Total effects); and (3) simultaneously ~~optimizing~~ optimising all 95 parameters. All three approaches achieve a statistically indistinguishable ~~and significant improvement in model skill~~ goodness of fit to the assimilated observations, reducing ~~Normalized Root Mean Square Error (NRMSE)~~ the median normalised RMSE across metrics by 54–56 % relative to the reference simulation with default PISCES parameters. The ~~rich~~ comprehensive, multi-variable ~~dataset provides sufficient orthogonal constraints to yield~~ observational constraint yields posterior parameter distributions with negligible inter-parameter correlation, shifting the long-standing challenge of correlated equifinality to uncorrelated equifinality~~, where a range of~~ : multiple optimal parameter sets ~~can be found~~ exist, but the individual parameters are independently constrained. Parameter uncertainty is reduced by 16–41~~%, and the optimized ensembles demonstrate strong portability.~~ % relative to the broad uniform prior distributions. While all strategies produce a similar~~, tightly constrained predictive spread~~ goodness of fit for the assimilated variables, they differ ~~significantly~~ in computational cost and in their estimation of uncertainty for unobserved parts of the model. The All-parameters strategy provides a more comprehensive quantification of predictive uncertainty for unassimilated variables, because it accounts for uncertainty in all parameters rather than artificially fixing poorly known parameters at their default values. The method is computationally tractable thanks to the use of a one-dimensional model

configuration, requiring approximately 24 CPU-hours for the optimisation step, whereas the prerequisite GSA was ~40 times more computationally expensive ~~than the optimization, while the All-parameters strategy, by exploring the full parameter space, provides a more comprehensive and robust quantification of the model's uncertainty in unassimilated variables. We therefore conclude that directly optimizing all model parameters is the recommended strategy. This work delivers~~ although nitrate and silicate show degraded skill in the Mediterranean. Surface chlorophyll-a is also better reproduced across the domain, when evaluated against a multi-observation reprocessed product.~~, parameter set for the North Atlantic and demonstrates a scalable framework to advance biogeochemical modeling from using static, globally-uniform parametrization to developing a map of regionally-tuned parameters.~~ The generality of the approach should be tested by applying it to additional BGC-Argo floats in other oceanic regions.

1. Introduction

~~Since the beginning of industrialization, the world's oceans have absorbed approximately 26% of anthropogenic carbon dioxide (CO_2) emissions (Friedlingstein et al., 2023), leading to profound changes in marine ecosystems. This increased CO_2 uptake is driving ocean acidification, which alters ocean chemistry (Doney et al., 2009), harms calcifying organisms (Orr et al., 2005), and disrupts broader biogeochemical processes (Hoegh-Guldberg et al., 2017). Concurrently, global deoxygenation, exacerbated by rising temperatures and ocean stratification (Keeling et al., 2010), is expanding oxygen minimum zones (Stramma et al., 2008), posing significant threats to marine life (Breitburg et al., 2018; Limburg et al., 2020). These challenges are further compounded by more direct anthropogenic impacts, such as plastic pollution (Wilcox et al., 2015) and overfishing (Pauly et al., 2002) which disrupt food webs (Dulvy et al., 2021) and degrade marine habitats (MacLeod et al., 2021).~~

~~Numerical biogeochemical (BGC) models are essential tools for understanding, monitoring, predicting, and mitigating these human-induced changes (Fennel et al., 2022)~~ Since the beginning of industrialization, the world's oceans have absorbed approximately 26% of anthropogenic carbon dioxide (CO_2) emissions (Friedlingstein et al., 2023), leading to profound changes in marine ecosystems. This increased CO_2 uptake is driving ocean acidification, which alters ocean chemistry (Doney et al., 2009), harms calcifying organisms (Orr et al., 2005), and disrupts broader biogeochemical processes (Hoegh-Guldberg et al., 2017). Concurrently, global deoxygenation, exacerbated by rising temperatures and ocean stratification (Keeling et al., 2010), is expanding oxygen minimum zones (Stramma et al., 2008), posing significant threats to marine life (Breitburg et al., 2018; Limburg et al., 2020). These challenges are further compounded by more direct anthropogenic impacts, such as plastic pollution (Wilcox et al., 2015) and overfishing (Pauly et al., 2002) which disrupt food webs (Dulvy et al., 2021) and degrade marine habitats (MacLeod et al., 2021).

Numerical biogeochemical (BGC) models are essential tools for understanding, monitoring, predicting, and mitigating these human-induced changes (Fennel et al., 2022). These models simulate key three-dimensional ocean processes, such as nutrient and plankton dynamics and carbon cycling, for past, present, and future ocean conditions. By capturing the complex interactions among physical, chemical, and biological components of the marine environment, BGC models produce critical outputs that supports scientific research, inform environmental management strategies, and guides policy development to conserve marine ecosystems ~~(Fennel et al., 2019)~~ (Fennel et al., 2019).

The accuracy of Ocean BGC models is fundamentally limited by the inherent uncertainty in their parameter values. This significantly affects their predictive skill, particularly when it comes to monitoring the ocean carbon sink ~~(Mayot et al., 2024) and for climate projections of the biological carbon pump (Löptien et al., 2021; Tagliabue et al., 2021; Henson et al., 2022; Rohr et al., 2023; Wang and Fennel, 2024; Doléac et al., 2025). This parameter uncertainty stems from multiple sources. Many parameters are extrapolated from laboratory experiments using a limited selection of representative species or even specific laboratory strains; while this approach provides valuable insights, it falls short in simulations of diverse oceanic bioregions that host a vast diversity of organisms (Ward et al., 2010; Schartau et al., 2017). Furthermore, some parameters cannot be experimentally determined and are thus assigned wide plausible ranges. The resulting parametric uncertainty can be substantial (Schartau et al., 2017)~~ (Mayot et al., 2024) and for climate projections of the biological carbon pump framework. This parameter uncertainty stems from multiple sources. Many parameters are extrapolated from laboratory experiments using a limited selection of representative species or even specific laboratory strains; while this approach provides valuable insights, it falls short in simulations of diverse oceanic bioregions that host a vast diversity of organisms (Ward et al., 2010; Schartau et al., 2017). Furthermore, some parameters cannot be experimentally determined and are thus assigned wide plausible ranges. The resulting parametric uncertainty can be substantial (Schartau et al., 2017), leading to significant uncertainty in model predictions. Therefore, refining these model parameters is essential to improve the accuracy and reliability of BGC models.

~~To address parameter uncertainty, the ocean biogeochemical modeling community has increasingly relied on data assimilation techniques that leverage observational data to constrain model parameters (Dowd et al., 2014; Schartau et al., 2017; Fennel et al., 2022). This process typically begins with a sensitivity analysis to identify the most influential parameters (Chu et al., 2007). Parameters with minimal influence on model outcomes are fixed at reference values, while only the influential control parameters are subsequently adjusted, usually around their nominal values. This common strategy, often based on local sensitivity analysis, is computationally efficient but fails to capture behavior across the entire parameter space or account for parameter interactions and other non-linear effects.~~

~~Global sensitivity analysis (GSA) is a method that captures complex nonlinear interactions among parameters (Homma and Saltelli, 1996; Sobol′, 2001)~~ It should be noted that structural uncertainty — arising from simplifications in the model's process formulations — is an additional source of error that parameter optimisation alone cannot address. The present study focuses on parametric uncertainty, which can be systematically reduced through data assimilation.

To address parameter uncertainty, the ocean biogeochemical modeling community has increasingly relied on data assimilation techniques that leverage observational data to constrain model parameters (Dowd et al., 2014; Schartau et al., 2017; Fennel et al., 2022). This process typically begins with a sensitivity analysis to identify the most influential parameters (Chu et al., 2007). Parameters with minimal influence on model outcomes are fixed at reference values, while only the influential control parameters are subsequently adjusted, usually around their nominal values. This common strategy, often based on local sensitivity analysis, is computationally efficient but fails to capture behavior across the entire parameter space or account for parameter interactions and other non-linear effects.

Global sensitivity analysis (GSA) is a method that captures complex nonlinear interactions among parameters (Homma and Saltelli, 1996; Sobol′, 2001). Historically, the extensive computational demands of GSA restricted its practical use, but recent increases in computing power have greatly broadened its accessibility. Sobol' indices, a key GSA technique, quantify the influence of individual parameters and their interactions by apportioning the model output's variance among them. A key study by ~~Prieur et al., (2019)~~ Prieur et al., (2019) applied this method to ~~the PISCES~~ the MODECOGeL biogeochemical model which has 74 parameters (Lacroix and Grégoire, 2002). They found that nearly every parameter significantly affected model behavior, largely through intricate non-linear interactions rather than solely through direct effects on model outputs. By contrast, a gradient-based analysis of the same model highlighted only two influential parameters, overlooking the critical dependencies revealed by the GSA. This finding implies that parameter-optimization efforts should include a large number of parameters to effectively reduce model-observation misfit. Furthermore, it underscores the limitations of traditional gradient-based optimization methods, which may fail to account for the broad impact of nonlinear parameter interactions.

Fulfilling the need to optimize many parameters, however, requires ~~information-rich observations. Traditional data sources, such as time-series stations and ocean-color satellite products, often lack the information content needed to constrain the large parameter sets of modern BGC models (Matear, 1995; Fennel et al., 2001; Friedrichs et al., 2007; Ward et al., 2010; Mamnun et al., 2022). Their vertical and temporal resolutions are too coarse, and the suite of measured variables too narrow, to fully inform the models (Fennel et al., 2022). Consequently, parameter-estimation studies based on such data frequently yield strong parameter correlations (Matear, 1995; Fennel et al., 2001; Mamnun et al., 2022), large posterior uncertainties (Ward et al., 2010), and optimized parameter sets that fail to reproduce independent observations (Friedrichs et al., 2007)~~comprehensive, multi-variable observations. Traditional data sources, such as time-series stations and ocean-color satellite products, often lack the information content needed to constrain the large parameter sets of modern BGC models (Fennel et al., 2001; Friedrichs et al., 2007; Hurtt and Armstrong, 1996; Mamnun et al., 2022; Matear, 1995; Ward et al., 2010). Their vertical and temporal resolutions are too coarse, and the suite of measured variables too narrow, to fully inform the models (Fennel et al., 2022). Consequently, parameter-estimation studies based on such data frequently yield strong parameter correlations (Matear, 1995; Fennel et al., 2001; Mamnun et al., 2022), large posterior uncertainties (Ward et al., 2010), and optimized parameter sets that fail to reproduce independent observations (Friedrichs et al., 2007). Fundamentally, any one of these outcomes indicates that the assimilated data lack sufficient independent information to robustly constrain each parameter.

~~BGC-Argo floats have revolutionized the acquisition of biogeochemical data by providing unprecedented data density at high vertical and temporal resolutions, while complementing the broad spatial coverage of traditional data sources (Claustre et al., 2020). These floats directly measure variables that correspond to key model state variables, including nitrate, chlorophyll-a, oxygen (Mignot et al., 2023), and particulate organic carbon (POC) – which represents the total biomass of phytoplankton, micro-zooplankton, and small detritus (Galí et al., 2022). Furthermore, the CANYON-B neural network (Bittig et al., 2018) can infer silicate, phosphate, dissolved inorganic carbon, and alkalinity from direct BGC-Argo measurements (oxygen, temperature, salinity), significantly expanding the suite of available float-derived biogeochemical information. These high-resolution profiles also illuminate seasonally variable phenomena such as the precise onset of the North Atlantic bloom (Mignot et al., 2018). Because the profiles extend well below the euphotic zone, BGC-Argo data also resolve key vertical structures like the nitracline depth, deep-chlorophyll maxima (DCM), and oxygen-minimum zones (Mignot et al., 2014, 2023; Cornec et al., 2021; Bock et al., 2022; Liu et al., 2024). Each of these emergent properties provides a distinct constraint on different sets of model parameters. Crucially, dedicated scientific studies have quantified uncertainty estimates for each variable (Johnson et al., 2017; Mignot et al., 2019), supplying the error statistics required for robust data assimilation. Overall, this superior temporal-and-vertical coverage supplies a richer information set than other types of in situ datasets and should allow many more model parameters to be constrained. However, to date, parameter optimization studies have used BGC-Argo data primarily for chlorophyll-a and/or POC (Wang et al., 2020; Galí et al., 2022; Shu et al., 2022) no study has yet exploited the full suite of variables~~

~~In this study, we develop and apply a framework to optimize the 95 parameters of the Pelagic Interaction Scheme for Carbon and Ecosystem Studies (PISCES) biogeochemical model (Aumont et al., 2015) by assimilating the full suite of observations from a BGC-Argo float in the North Atlantic. Our focus on this region is motivated by its critical role in the global carbon cycle (Takahashi et al., 2009; DeVries et al., 2014) , and by the fact that PISCES, despite its extensive use in operational (e.g., Copernicus Marine Service) and climate (e.g., CMIP6) applications, exhibits systematic biases there. These biases are particularly evident in the seasonal cycles of key biogeochemical processes such as net primary production and pCO_2 (Rodgers et al., 2023; Hieronymus et al., 2024).~~

~~A central objective of this work is to identify the most effective and efficient parameter-selection strategy for such a high-dimensional problem. We therefore evaluate and compare three distinct approaches:~~

- ~~Main Effects: Optimizing a parameter subset based on first-order Sobol' indices from a GSA. This strategy targets parameters that have a significant direct effect on model output, independent of other parameters.~~

- ~~Total Effects: Optimizing a larger subset based on total-order Sobol' indices. This strategy also includes parameters that are influential primarily through non-linear interactions with other parameters.~~
- ~~All Parameters: Optimizing all 95 model parameters simultaneously, providing a computationally simple alternative that bypasses the prerequisite GSA.~~

BGC-Argo floats have revolutionized the acquisition of biogeochemical data by providing unprecedented data density at high vertical and temporal resolutions, while complementing the broad spatial coverage of traditional data sources (Claustre et al., 2020). These floats directly measure variables that correspond to key model state variables, including nitrate, chlorophyll-a, oxygen (Mignot et al., 2023), and particulate organic carbon (POC)—which represents the total biomass of phytoplankton, micro-zooplankton, and small detritus (Galí et al., 2022). Furthermore, the CANYON-B neural network (Bittig et al., 2018) can infer silicate, phosphate, dissolved inorganic carbon, and alkalinity from direct BGC-Argo measurements (oxygen, temperature, salinity), significantly expanding the suite of available float-derived biogeochemical information. These high-resolution profiles also illuminate seasonally variable phenomena such as the precise onset of the North Atlantic bloom (Mignot et al., 2018). Because the profiles extend well below the euphotic zone, BGC-Argo data also resolve key vertical structures like the nitracline depth, deep-chlorophyll maxima (DCM), and oxygen-minimum zones (Mignot et al., 2014, 2023; Cornec et al., 2021; Bock et al., 2022; Liu et al., 2024). Each of these emergent properties provides a distinct constraint on different sets of model parameters. Crucially, dedicated scientific studies have quantified uncertainty estimates for each variable (Johnson et al., 2017; Mignot et al., 2019), supplying the error statistics required for robust data assimilation. Overall, BGC-Argo floats supply a more comprehensive information set than other types of in situ datasets, combining multi-variable breadth, high temporal resolution (~5-day profiling versus monthly bottle sampling), and high vertical resolution (1 m grid versus discrete depths). This greater information content should allow many more model parameters to be constrained. However, to date, parameter-optimization studies have used BGC-Argo data primarily for chlorophyll-a and/or POC (Wang et al., 2020; Galí et al., 2022; Shu et al., 2022) no study has yet exploited the full suite of variables

In this study, we develop and apply a framework to optimise the 95 parameters of the Pelagic Interaction Scheme for Carbon and Ecosystem Studies (PISCES) biogeochemical model (Aumont et al., 2015) by assimilating the full suite of observations from a single BGC-Argo float over one seasonal cycle (January 2015–January 2016) in the North Atlantic. Our focus on this region is motivated by its critical role in the global carbon cycle (Takahashi et al., 2009; DeVries et al., 2014), and by the fact that PISCES, despite its extensive use in operational (e.g., Copernicus Marine Service) and climate (e.g., CMIP6) applications, exhibits systematic biases there. These biases are particularly evident in the seasonal cycles of key biogeochemical processes such as net primary production and pCO_2 (Rodgers et al., 2023; Hieronymus et al., 2024).

The optimisation is performed using iterative Importance Sampling (iIS), a Bayesian method that iteratively refines an ensemble of model simulations by reweighting parameter sets according to their agreement with the observations. At each iteration, the ensemble is concentrated towards the regions of parameter space that best reproduce the data, yielding full posterior distributions of parameters, including uncertainties and correlations.

A central objective of this work is to identify the most effective and efficient parameter-selection strategy for such a high-dimensional problem. We therefore evaluate and compare three distinct approaches:

- Main Effects: Optimizing a parameter subset based on first-order Sobol' indices from a GSA. This strategy targets parameters that have a significant direct effect on model output, independent of other parameters.
- Total Effects: Optimizing a larger subset based on total-order Sobol' indices. This strategy also includes parameters that are influential primarily through non-linear interactions with other parameters.
- All Parameters: Optimizing all 95 model parameters simultaneously, providing a computationally simple alternative that bypasses the prerequisite GSA.

Finally, a critical challenge in biogeochemical optimization is determining whether parameters estimated in simplified 1D configurations are robust enough to improve fully three-dimensional models, or if they merely compensate for missing physical processes (Löptien and Dietze, 2019; Pasquier et al., 2023). To address this, we assess the transferability of the optimization results by implementing the fully optimized parameter set into a high-resolution (1/36°) 3D regional model of the Iberian-Biscay-Irish (IBI) sector; the model configuration of the IBI Monitoring and Forecasting Center of the Copernicus Marine Service. We validate this 3D simulation against an independent multi-observation reprocessed product of chlorophyll-a and independent BGC-Argo nutrients and carbonates observations to verify that skill improvements persist in a realistic dynamic environment.

Section 2 describes the data sets and model configurations used for the sensitivity analysis and subsequent optimization. Section 3 details the assimilation framework and the sensitivity-analysis methods. Section 4 reports the sensitivity results, defines the parameter subsets retained for optimization, ~~and~~ presents the resulting model ~~skill~~ performance, and assesses the transferability of the optimized parameters in the three-dimensional IBI regional configuration. Section 5 discusses the implications, and Sect. 6 summarizes the main conclusions.

2. Data and Model Configuration

2.1. BGC-Argo Data

The primary dataset for this study comes from BGC-Argo float with World Meteorological (WMO) number #5904479. Operating in the North Atlantic from April 2014 to December 2017 (Fig. 1), the float collected vertical profiles of temperature, salinity, and four biogeochemical variables: chlorophyll-a concentration (Chl-a), particulate backscattering coefficient (bbp), nitrate concentration (NO_3^-), and dissolved oxygen (O_2). The Chl-a data were taken from a curated, quality-controlled archive (https://www.seanoe.org/data/00911/102324/), in which raw fluorescence was corrected for dark counts, non-photochemical quenching, and physiological calibration.

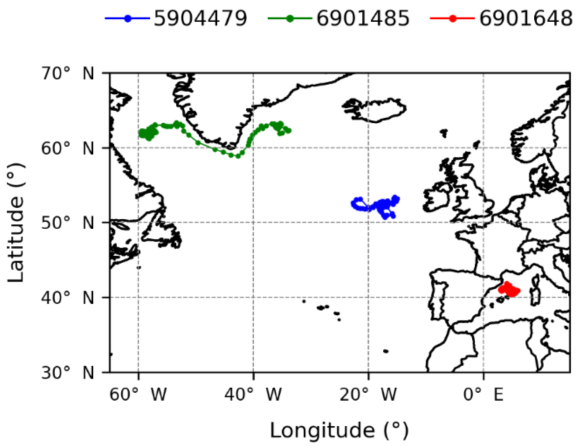


Figure 1. Trajectories of the BGC-Argo floats used in this study, identified by their World Meteorological Organization (WMO) numbers: 5904479 (blue), 6901485 (green), and 6901648 (red). Observations from float WMO 5904479, covering January 2015 to January 2016, served as the primary dataset for parameter optimization. Data from the additional floats (WMO 6901485 and WMO 6901648), sampling diverse regions of the North Atlantic basin and the Mediterranean Sea, were used to assess the ~~portability~~ generalizability of the optimized parameter set.

Two independent BGC-Argo floats #6901485 and #6901648, are used to ~~validate~~ evaluate the ~~portability~~ accuracy and calibration of the optimized ~~parameter sets~~ ensembles against independent observations. Both measure the same variables as the primary float (#5904479) but operate in fundamentally different environments, providing robust tests for generalization.

The first validation float #6901485, also profiled in the North Atlantic, specifically in the western part of the Subpolar Gyre. While this region also features a prominent spring bloom, it is characterized by deeper winter convection and the influence of colder, fresher Arctic waters, providing a test of parameter robustness to different physical forcing within the same ocean basin.

The second float #6901648, operated in the oligotrophic Mediterranean Sea. This basin is defined by much lower nutrient availability, being particularly limited in phosphate relative to nitrate, yet its deep winter convection still drives a distinct, albeit less intense, spring bloom. This provides a challenging test of whether biological rate parameters tuned on a high-nutrient system can generalize to a low-nutrient one.

CTD profiles, additional BGC data and trajectory information were retrieved from the Argo CORIOLIS Global Data Assembly Centre (ftp://ftp.ifremer.fr/argo, accessed February 2025). ~~These data were quality-controlled according to the standard Argo~~

~~procedures (Wong et al., 2020). Biogeochemical variables (O_2, NO_3^-, bbp) were processed with the established BGC-Argo procedures (Schmechtig et al., 2015, 2023; Thierry et al., 2018; Johnson et al., 2025). Particulate organic carbon (POC) concentration was estimated from bbp following (Mignot et al., 2023)~~These data were quality-controlled according to the standard Argo procedures (Wong et al., 2020). Biogeochemical variables (O_2, NO_3^-, bbp) were processed with the established BGC-Argo procedures (Schmechtig et al., 2015, 2023; Thierry et al., 2018; Johnson et al., 2025). Particulate organic carbon (POC) concentration was estimated from bbp following Mignot et al., (2023).

~~To supplement the directly measured variables, we incorporated pseudo-observations of phosphate (PO_4^{3-}), silicate (Si), total alkalinity (TA), and dissolved inorganic carbon (DIC). These were derived from the Copernicus Marine Service "Nutrient and Carbon Profiles Vertical Distribution" product (2025; https://doi.org/10.48670/moi-00048). This product supplies vertical profiles of nutrient concentrations (NO_3^-, PO_4^{3-}, and Si) and carbonate-system variables (TA, DIC) for every Argo float equipped with an O_2 sensor. These concentrations are estimated using the CANYON-B neural-network for nutrients and the CONTENT algorithm for DIC and TA (Bittig et al., 2018); both algorithms were trained on ~30 years of quality-controlled profiles from the GLODAPv2 database (Olsen et al., 2016).~~
To supplement the directly measured variables, we incorporated pseudo-observations of phosphate (PO_4^{3-}), silicate (Si), total alkalinity (TA), and dissolved inorganic carbon (DIC). These were derived from the Copernicus Marine Service "Nutrient and Carbon Profiles Vertical Distribution" product (2025; https://doi.org/10.48670/moi-00048). This product supplies vertical profiles of nutrient concentrations (NO_3^-, PO_4^{3-}, and Si) and carbonate-system variables (TA, DIC) for every Argo float equipped with an O_2 sensor. These concentrations are estimated using the CANYON-B neural-network for nutrients and the CONTENT algorithm for DIC and TA (Bittig et al., 2018); both algorithms were trained on ~30 years of quality-controlled profiles from the GLODAPv2 database (Olsen et al., 2016).

First, to ensure the validity of the 1D model assumption, the time window was selected such that the float remained within a single water mass exhibiting weak horizontal gradients (Mignot et al., 2018). The prevalence of these quasi-one-dimensional dynamics is confirmed by the near-constancy of temperature and salinity below the mixed layer during the selected periods (e.g., Figs. 2a, 2c for float #5904479). This selection is crucial, as pronounced T-S changes at depth would otherwise imply movement across water-mass boundaries and lateral biomass variability that our 1-D PISCES configuration cannot represent.

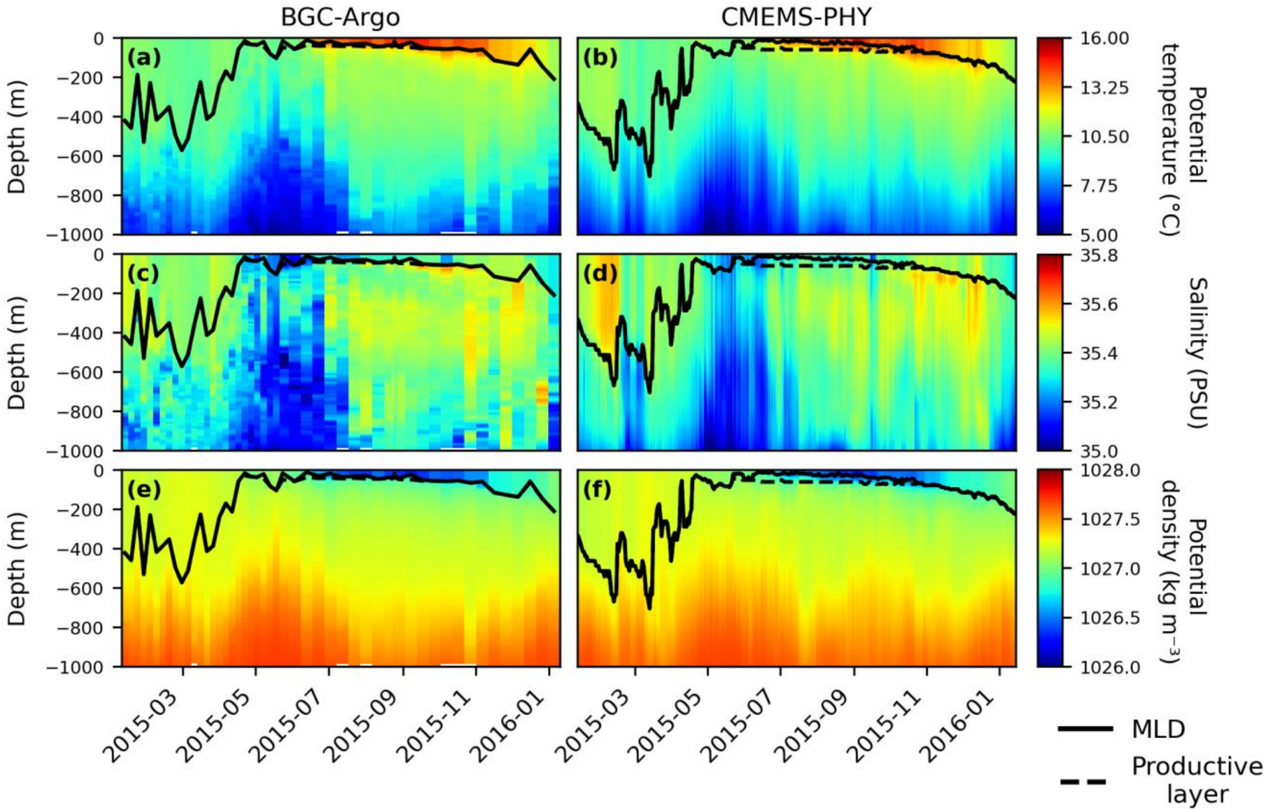


Figure 2. Comparison of observed and modelled vertical profiles of physical properties along the trajectory of BGC Argo float #5904479 (January 2015-January 2016). Panels show in situ Argo observations (left column: a, c, e) and CMEMS-PHY output (right column: b, d, f) for: (a, b) potential temperature (°C), (c, d) salinity (PSU), and (e, f) potential density kg m-³). The x-axis represents time and the y-axis depth (0-1000 m). Solid black lines in panels indicate the mixed layer depth, calculated using a potential density threshold criterion of $\Delta\sigma\ \theta = 0.03$ kg m-³. The dashed black lines indicate the depth of the productive layer.

Second, to minimize bias from initial conditions, the precise start date within this window was chosen to correspond to the time of minimum Root Mean Square Error (RMSE) calculated across all available direct and pseudo-observations, between the float data and the Copernicus Marine Service Global Ocean Biogeochemistry Analysis and Forecast (CMEMS-BGC) (https://doi.org/10.48670/moi-00015). The final, optimized start dates and resulting one-year analysis periods are: 11 Jan 2015–14 Jan 2016 for float #5904479; 26 Nov 2013–6 Dec 2014 for float #6901485; and 21 Jan 2015–15 Feb 2016 for float #6901648.

In addition to the three floats used for optimization and 1D independent ensemble evaluation, a broader set of BGC-Argo profiles from the IBI domain was used for the independent 3D validation. Profiles of O_2 with CANYON-B and CONTENT derived variables (NO_3^-, PO_4^{3-}, Si, DIC, TA) were extracted over the 0–500 m depth range. A total of 5,436 unique profiles from 2017 were retained after quality control.

2.2. Ocean color data

We estimated the observational euphotic depth ($Z_{Euphotic}$) from the Copernicus Marine Service GlobColour product (https://doi.org/10.48670/moi-00281, accessed 31 Jan 2024). This product delivers monthly 4-km maps of the diffuse-attenuation coefficient at 490 nm~~,~~ ($K_d(490)$)~~,~~ obtained by merging multi-mission ocean-colour sensors. ~~Following (Morel et al., 2007)~~ Following (Morel et al., 2007), we converted $K_d(490)$ to the photosynthetically available radiation attenuation coefficient, $K_d(PAR)$. The euphotic depth, $Z_{Euphotic}$, is defined as the depth where 1% of surface PAR persists and was calculated as:

$$Z_{Euphotic} = \frac{-\log(0.01)}{K_d(PAR)} \tag{1}$$

2.3. Multi-observation reprocessed product of chlorophyll-a

Surface chlorophyll-a concentrations for the 3D validation were obtained from the CMEMS Multi-Observation Global Ocean 3D Biochemistry Reprocessed Product (MULTIOBS_GLO_BIO_BGC_3D_REP_015_010; hereafter MOB-TAC; https://doi.org/10.48670/moi-00046). This product combines satellite ocean-color observations with in-situ hydrological properties from BGC-Argo floats through a neural network (Sauzède et al., 2017) to provide weekly three-dimensional fields of chlorophyll-a at 0.25° resolution from the surface to 1000 m. Only surface values (depth = 0 m), over the period January 2017–December 2019 were retained to match the 3D IBI simulation period.

2.4. Biogeochemical Model ~~description, framework and configurations~~ (PISCES)

The Ocean circulation and thermodynamics are simulated ~~with~~ using the Nucleus for European Modelling of the Ocean (NEMO) platform, version 4.2 ~~(Madec et al., 2023)~~ (Madec et al., 2023). The lower-trophic-level biogeochemistry is ~~handled~~ represented by PISCES version 2 ~~(Aumont et al., 2015)~~ (Aumont et al., 2015), which is distributed with NEMO 4.2.

PISCES is a mechanistic ecosystem model that resolves 24 prognostic state variables. Phytoplankton growth can be limited by five nutrients: nitrate, ammonium, phosphate, silicate, and iron. The plankton community is represented by four functional types defined by size: two phytoplankton groups (nanophytoplankton and diatom-dominated microphytoplankton) and two zooplankton groups (micro- and mesozooplankton). The model simulates phytoplankton biomass through its carbon, iron, and Chl-a content (and silicate for diatoms), while zooplankton biomass is simulated in carbon only. Three non-living organic carbon pools—semi-labile dissolved organic matter, small particles, and large particles—are distinguished by size and reactivity. Particles carry both carbon and iron, while large particles also include calcium carbonate and biogenic silica.

PISCES additionally resolves the seawater carbonate system (DIC and TA) and dissolved oxygen. In this study, we adopt the default parameter set supplied with PISCES ~~v2~~ 4.2 as our reference configuration.

To conduct our analysis, we employ ~~two~~ three complementary configurations of the NEMO-PISCES ~~v2~~ 4.2 model. A computationally lightweight 1-D column setup is used for the large optimization ensembles, while a global 3-D configuration supplies the state variable estimates needed to calculate the representativeness error (as described in Sect. 3.7). Finally, a regional 3-D configuration is used to assess the transferability of the optimized parameter set from 1D to 3D framework.

2.4.1. ~~2.3.1~~ 1D configuration

~~The parameter optimization method used in this study, Iterative Importance Sampling (iIS), is computationally intensive because it requires thousands of simulations to explore large parameter spaces, rendering its direct application to fully three-dimensional (3-D) configurations prohibitive in terms of CPU time and storage. To reduce cost, we follow the common practice of carrying out the optimization in a lightweight one-dimensional (1-D) set-up (Schartau and Oschlies, 2003; Hoshiba et al., 2018).~~

~~This strategy is valid only if the assimilated BGC-Argo profiles meet a quasi-1-D assumption; we ensure this by selecting floats located in water masses with weak lateral gradients (Mignot et al., 2018)~~The parameter optimization method used in this study, iIS is computationally intensive because it requires thousands of simulations to explore large parameter spaces, rendering its direct application to fully three-dimensional (3-D) configurations prohibitive in terms of CPU time and storage. To reduce cost, we follow the common practice of carrying out the optimization in a lightweight one-dimensional (1-D) set-up (Schartau and Oschlies, 2003; Hoshiba et al., 2018).

This strategy is valid only if the assimilated BGC-Argo profiles meet a quasi-1-D assumption; we ensure this by selecting floats located in water masses with weak lateral gradients (Mignot et al., 2018). In the cost function we combine observational errors with a representativeness term that captures unresolved 3-D variability present in the observations but absent from the 1-D model. Including this error prevents the optimization from compensating for missing physics and yields a more reliable calibration.

The 1-D setup neglects both horizontal and vertical advection, retaining only vertical turbulent diffusion and surface flux boundary conditions ~~(Reffray et al., 2015)~~ (Reffray et al., 2015). The biogeochemical model runs offline, forced by two sets of data: (1) daily profiles of temperature, salinity, and vertical diffusivity extracted from the 1/12°Copernicus Marine Service global physical (CMEMS-PHY) analysis (https://doi.org/10.48670/moi-00016), and (2) the same surface atmospheric fluxes used to drive the CMEMS-PHY product.

The 1-D configuration simulates a Lagrangian water column that follows each of the three floats described in Sect. 2.1 (one for optimization and two for validation). For each of these simulations, the vertical grid comprises 75 levels, with 24 in the upper 100 meters and a surface resolution of 1 meter that progressively decreases with depth. A flat bottom at 3000 m represents a consistent bathymetry, matching the deep-water environments sampled.

The model is initialized at the optimal start date for each float using a hybrid approach that combines data from three primary sources (Table 1). A key challenge is initializing the model's Plankton Functional Types (PFTs) using bulk observations from the BGC-Argo float. The PISCES model simulates distinct pools for Chl-a and Particulate Organic Carbon (POC) associated with different PFTs, whereas the float provides only a single, integrated measurement for each. To bridge this gap, we disaggregated the bulk observational data using component ratios derived from the CMEMS-BGC analysis.

Specifically, the total observed Chl-a was distributed between the model's nanophytoplankton and diatom pools based on their relative proportions in CMEMS-BGC. A similar procedure was applied to the total observed POC, which was partitioned among its four constituent pools in the model: nanophytoplankton, diatoms, microzooplankton, and small detritus. This method ensures that the sum of the initialized components for both Chl-a and POC precisely matches the total value observed by the float.

Table 1. Initialization methods and data sources for state variables in the one-dimensional (1D) biogeochemical model simulation.

Initialization Method/Source	State variable
From BGC-Argo observations	NO_3^-, PO_4^{3-}, Si, O_2, TA, DIC
From BGC-Argo profile structure : inter-variable ratios constrained by CMEMS-BGC model	POC = Small detrital particles +PHY+PHY2+ZOO CHL = NCHL+DCHL
From CMEMS-BGC model output	BFe, DFe, NFe, SFe, Fe, $CaCO_3$, DOC, GOC, DSi, Gsi, NH_4^+, PHYC, ZOO2

For variables not directly measured but inferred from float data, initial values were taken from the CANYON-B/CONTENT neural network products (pseudo-observations). Finally, initial conditions for any remaining unobserved variables were prescribed from the CMEMS-BGC analysis.

~~**2.3.2 3D configuration**~~

~~A global NEMO-PISCES configuration is run to estimate the representativeness error associated with unresolved three-dimensional processes in the 1-D setup. To ensure methodological consistency, the global system uses the same NEMO 4.2-PISCES v2 code base as the 1-D configuration, but with the biogeochemistry running online. The grid has a 1/4° horizontal resolution and 75 vertical levels. The simulation spans 2010–2022 and is forced with ERA5 atmospheric reanalysis (Brodeau et al., 2010; Hersbach et al., 2020) plus a terrestrial heat-flux correction from (Lucazeau, 2019). We refer to this experiment as the '3D-Free' simulation, which is a fully free-running integration without any assimilation of physical or biogeochemical data.~~

2.4.2. 3D Global Configuration

A global NEMO-PISCES configuration is run to estimate the representativeness error associated with unresolved three-dimensional processes in the 1-D framework. To ensure methodological consistency, the global system uses the same NEMO-PISCES 4.2 code base as the 1-D configuration, but with the biogeochemistry running online with the physics. The model grid has a horizontal resolution of 1/4° and 75 vertical levels. The simulation covers the period 2010–2022 and is forced with ERA5 atmospheric reanalysis (Brodeau et al., 2010; Hersbach et al., 2020), complemented by a correction of terrestrial heat-flux (Lucazeau, 2019). We refer to this experiment as the '3D-Free' simulation, which is a fully free-running integration without assimilation of physical or biogeochemical data.

2.4.3. 3D IBI Regional Configuration

To assess the transferability of the optimized parameters in a fully three-dimensional framework, we used the IBI (Iberian–Biscay–Irish) regional, which is based on the NEMO ocean engine coupled with PISCES, implemented at a horizontal resolution of 1/36° (approximately 2–3 km), with 50 vertical levels and enhanced resolution near the surface. Two simulations were performed:

(i) REF: using the default PISCES parameter values.

(ii) OPTI: using the best-member parameter set from the All-Parameters optimization.

Both simulations span January 2017 to December 2019. Initial conditions for the biogeochemical variables were obtained from the CMEMS-BGC analysis, while physical variables (temperature, salinity, currents) were initialized from the CMEMS-PHYS analysis. Atmospheric forcing was provided by ECMWF at hourly temporal resolution. Model outputs were saved as daily averages.

3. Method

3.1. Metrics for sensitivity analysis and parameter optimization

~~To calibrate PISCES and conduct our global sensitivity analysis, we employ twenty metrics (Mignot et al., 2023) that quantify both layer-averaged concentrations and emergent aspects of the vertical structure of biogeochemical variables. A concise description of all twenty metrics is provided in Table 2.~~ We define twenty observational metrics — layer-mean concentrations and vertical-structure diagnostics — derived from eight biogeochemical tracers measured or inferred from the BGC-Argo float (Mignot et al., 2023). These metrics serve as individual targets for the sensitivity analysis and optimisation. Within the iIS framework, they enter the observation likelihood function, which determines the weight of each ensemble member based on how well it reproduces the observations across all 20 metrics simultaneously, accounting for the observation errors associated with each metric (Sect. 3.5). The prior parameter distributions are uniform, with bounds and inter-parameter constraints defined in consultation with the PISCES model developer (Sect. 3.4;-Table S1). A concise description of all twenty metrics is provided in Table 2. Each layer-averaged metric is computed within two depth domains—the productive layer and the mesopelagic layer—using both BGC-Argo observations and the 1-D model output. The following subsections detail the four steps required to calculate and compare these metrics: (1) the definition of the vertical layers used for analysis; (2) the calculation of layer-mean concentrations; (3) the diagnosis of emergent vertical features; and (4) the spatiotemporal interpolation procedure used to ensure a consistent comparison.

Table 2. Metrics used for sensitivity analysis (SA) and data assimilation (DA), showing the correspondence between observations and PISCES model variables. Only metrics designated as state variables (S) were used for the SA. Key: S = State variable; E = Emergent property.

Process	Observed variables	PISCES variables	Metrics	Units	Obs. error (%)	Assessment level	Application
Carbonate Chemistry	DIC	DIC	DIC_{Prod}	$\mu mol\ kg^{-1}$	0.46	S	SA/DA
			DIC_{Meso}	$\mu mol\ kg^{-1}$	0.46	S	SA/DA
	TA	TA	TA_{Prod}	$\mu mol\ kg^{-1}$	0.41	S	SA/DA
			TA_{Meso}	$\mu mol\ kg^{-1}$	0.41	S	SA/DA
Biological carbon pump	$Chl-a$	NCHL + DCHL	$Chl-a_{Prod}$	$mg\ m^{-3}$	13.4	S	SA/DA
			Chl_{DCM}	$mg\ m^{-3}$	13.4	E	DA
			$HChl_{DCM}$	m	/	E	DA
	PO_4^{3-}	PO_4^{3-}	$PO_4^{3-}{}_{Prod}$	$\mu mol\ kg^{-1}$	8.45	S	SA/DA
			$PO_4^{3-}{}_{Meso}$	$\mu mol\ kg^{-1}$	8.45	S	SA/DA

	Si	Si	Si_{Prod}	µmol kg^{1}	46.8	S	SA/DA
			Si_{Meso}	µmol kg^{1}	46.8	S	SA/DA
	POC	Small detrital particles +PHY+PHY2+ZOO	POC_{Prod}	mg m^{-3}	40	S	SA/DA
			POC_{Meso}	mg m^{-3}	40	S	SA/DA
	NO_3^-	NO_3^-	${NO_3^-}_{Prod}$	µmol kg^{-1}	7	S	SA/DA
			${NO_3^-}_{Meso}$	µmol kg^{-1}	7	S	SA/DA
			$H_{Nitracline}$	m	/	E	DA
Oxygens levels	O_2	O_2	${O_2}_{Prod}$	µmol kg^{-1}	3	S	SA/DA
			${O_2}_{Meso}$	µmol kg^{-1}	3	S	SA/DA
			${O_2}_{min}$	µmol kg^{-1}	3	E	DA
			$H{O_2}_{min}$	m	/	E	DA

3.1.1. Layer definitions and depth diagnostics

For our analysis, we subdivide the water column into two biogeochemically distinct layers. The productive layer extends from the surface down to the deeper of either the mixed-layer depth (MLD) or $Z_{Euphotic}$. The mesopelagic layer spans from the base of the productive layer down to 1,000 m. These layers are defined by the MLD and $Z_{Euphotic}$, which are determined as follows for both the model and observations.

The MLD is diagnosed as the shallowest depth where potential density exceeds its surface value by 0.03 kg·m^{-3} ~~(de Boyer Montégut et al., 2004)~~ (de Boyer Montégut et al., 2004). For the model, daily MLD is computed from the potential density profiles from the CMEMS-PHY analysis. For observations, MLD is calculated from BGC-Argo temperature and salinity profiles using the same density threshold.

The euphotic depth, the depth where 1% of surface photosynthetically available radiation (PAR) persists, is taken directly from the model's internal radiation scheme. For observations, the 1% PAR depth is calculated from satellite-derived diffuse-attenuation coefficients, as detailed in Sect. 2.2.

3.1.2. Layer-mean metrics

Following ~~(Mignot et al., 2023)~~ (Mignot et al., 2023), we calculate layer-mean concentrations for a suite of eight state variables. For seven of these—DIC, TA, O_2, NO_3^-, PO_4^{3-}, Si, and POC—we compute the mean concentration in both the productive and mesopelagic layers. For Chl-a, the metric is computed for the productive layer only, as this layer contains the vast majority of its biomass and variability. This procedure yields a total of fifteen layer-mean metrics (eight for the productive layer and seven for the mesopelagic). To account for their lognormal distribution, the Chl-a and POC metrics were subsequently $\log_{10}$-transformed.

It is important to note that while model state variables are initialized directly from observational profiles, the initial values of the layer-mean metrics can still differ between the model and the BGC-Argo data. This discrepancy arises because the layer boundaries are diagnosed independently for the model and the observations, which can result in slightly different layer depths. Consequently, integrating a tracer profile with a strong vertical gradient over these different depth ranges will produce different initial layer-averaged values

3.1.3. Emergent vertical-structure metrics

To constrain processes that depend on vertical structure rather than bulk averages, we include five additional metrics that diagnose the depth and magnitude of key biogeochemical features:

- DCM: We record both the depth of the DCM (H_{DCM}) and the Chl-a concentration at that depth (Chl_{DCM}). The DCM is crucial for phytoplankton growth and nutrient cycling in low latitude environment.
- Nitracline Depth ($H_{nitracline}$): This is defined as the first depth where NO_3^- exceeds 1 µmol·kg^{-1}, a threshold corresponding to the upper limit of BGC-Argo nitrate accuracy ~~(Johnson et al., 2017; Mignot et al., 2019, 2023)~~ (Johnson et al., 2017; Mignot et al., 2019, 2023). This metric captures surface nitrate limitation, a key factor controlling primary production (Cermeño et al., 2008; Bendtsen et al., 2023)
- Oxygen Minimum: We identify both the depth of the minimum oxygen concentration (H_{O_2min}) and the corresponding oxygen value (O_{2min}). These quantities serve as proxies for mesopelagic remineralization intensity and ventilation (Stramma et al., 2008; Schmidtko et al., 2017).

3.1.4. Interpolation procedure

Before computing any metrics, both the simulated data and the float observations are processed and interpolated onto a common spatio-temporal grid. This procedure involves two steps. ~~First, all vertical profiles are linearly interpolated onto a uniform 1-meter vertical grid from the surface to 1000 m from which each metric is then computed.~~ First, all vertical profiles are linearly interpolated onto a uniform 1-meter vertical grid from the surface to 1000 m, from which each averaged metric within the productive and mesopelagic layers is then computed. Second, we address the irregular sampling of the float by linearly interpolating the observations to create a regular 5-day time series. ~~Although the model provides daily outputs, we interpolate its time series onto this same 5-day grid to ensure a direct, point-for-point comparison with the regularized observations.~~ Although the model provides daily outputs, we interpolate the time series of each metrics onto this same 5-day grid to ensure a direct, point-for-point comparison with the regularized observations. ~~Time series for each of the twenty metrics~~

~~are then calculated from these temporally aligned datasets.~~ Finally, to reduce short-term variability and better highlight seasonal dynamics, a 6-point moving average is applied to smooth each metric's time series.

3.2. Parameter optimization method: Iterative Importance Sampling

This section describes the parameter optimisation method. The central idea is to draw a large ensemble of model simulations from broad prior parameter distributions, then iteratively reweight the ensemble so that parameter sets producing good agreement with the observations receive higher weight. After several iterations, the resulting weighted ensemble provides an approximation of the posterior distribution, from which parameter estimates, uncertainties, and correlations can be extracted. The mathematical formulation is detailed below; the Bayesian likelihood function that drives the reweighting is presented in the Supplementary Information (Eqs. S14).

To quantify posterior parameter correlations, uncertainties, and the predictive spread for both assimilated and unconstrained variables, we adopt iIS, an optimization scheme built on a particle-filter assimilation framework. Particle filters, widely used in biogeochemical modelling for state and parameter estimation ~~(Mattern et al., 2013)~~ (Mattern et al., 2013), are gradient-free ensemble methods that assimilate observations into a population of particles—each a unique, evolving model state—thereby generating full probability distributions for states and parameters. This approach is well suited to the strong non-linearities and non-Gaussian errors characteristic of marine biogeochemistry ~~(Ristic et al., 2004)~~ (Ristic et al., 2004).

Iterative Importance Sampling is a technique for estimating a probability density function (PDF) using a weighted combination of samples drawn from a different, known PDF ~~(Raices Cruz et al., 2022)~~ (Raices Cruz et al., 2022). In this work, this concept is applied within the framework of Bayesian inference to estimate model parameters. More formally, the goal is to approximate the posterior PDF of the model state variables and parameters, conditioned on observational data.

The Bayesian framework provides an approach for calculating the posterior PDF using prior knowledge and observational evidence. According to Bayes' theorem, the posterior distribution can be expressed as a function of the prior distribution and the observational likelihood ~~(Wikle et al., 2013)~~ (Wikle et al., 2013):

$$p(\mathbf{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\mathbf{x})p(\mathbf{x})}{\int p(\boldsymbol{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \tag{2}$$

where $\mathbf{x} = (\boldsymbol{s}, \boldsymbol{\theta})$ is a random vector containing model state variables ($\boldsymbol{s}$) and model parameters ($\boldsymbol{\theta}$), $\boldsymbol{y}$ is the observation vector, $p(\boldsymbol{y}|\mathbf{x})$ is the likelihood of the observations given the model outcome and $p(\mathbf{x})$ is the probability of the model state. Here, following van Leeuwen et al. (2019) the prior probability is represented by an ensemble of N particles as:

$$p(\mathbf{x}) = \sum_{j=1}^{N} \frac{1}{N}\,\delta(\mathbf{x} - \mathbf{x}_j) \tag{3}$$

where $\mathbf{x}_j$ represents the jth member of the ensemble, and $\delta(\mathbf{x} - \mathbf{x}_j)$, is the Dirac delta function that evaluates to 1 if $\mathbf{x} = \mathbf{x}_j$ and 0 otherwise. Equation (3) means that the probability density function is a weighted sum of the delta functions centered at each member of the ensemble, where each member has an equal weight of N^{-1}.

The algorithm described in this article makes use of a self-normalized importance sampling techniques to estimate the posterior $p(\mathbf{x}|\boldsymbol{y})$ using weighted samples of $p(\mathbf{x})$. Importance sampling generally involves the use of an auxiliary PDF, called proposal density function, ~~which has as the~~ whose main objective ~~restraint the~~ is to restrict sampling to regions of the state space ~~of~~ with high probability ~~(Owen and Zhou, 2000)~~ (Owen and Zhou, 2000). In our setting, we used the prior $p(\mathbf{x})$ as the proposal density since there is no reasonable estimate of the posterior distribution of the parameters. In this case the weights of each particle j is simply calculated using the observation likelihood ~~(van Leeuwen et al., 2019)~~ (van Leeuwen et al., 2019) :

$$\boldsymbol{\omega}_j = \frac{p(\boldsymbol{y}|\mathbf{x}_j)}{\sum_{j=1}^{N} p(\boldsymbol{y}|\mathbf{x}_j)} \tag{4}$$

The posterior is then written as a weighted combination of the prior states:

$$p(\mathbf{x}|\boldsymbol{y}) = \sum_{j=1}^{N} \boldsymbol{\omega}_j \, \delta(\mathbf{x} - \mathbf{x}_j) \tag{5}$$

Estimating model parameters from observations of the state is challenging because we often lack reliable prior information about the parameter distributions, and the relationship between the state and the parameters is typically highly non-linear. Moreover, models generally exhibit biases, partly due to uncertainty in the parameters and partly because of structural deficiencies, which include necessary simplifications, incomplete knowledge of key processes, and uncertainties in their mathematical representation. In the context of importance sampling, the small magnitude of observational noise significantly reduces the probability of obtaining model trajectories with high likelihood, thereby amplifying the impact of model imperfections.

Consequently, plain Monte Carlo sampling would produce many samples with negligible weights, making the estimation inefficient. To address this, the proposed algorithm evaluates equations (4) and (5) iteratively, modifying the likelihood $p(\boldsymbol{y}|\mathbf{x})$ to $p(\boldsymbol{y}|\mathbf{x})^{\alpha}$, where α is an inflation parameter. This parameter is dynamically adjusted based on the effective sample size (ESS) defined as (Martino et al., 2017) :

$$\cancel{ESS = \frac{1}{\sum_{j=1}^{N} \boldsymbol{\omega}_j^2} \quad (6)}$$

~~At each iteration, both the prior and the likelihood function are updated. The goal of each step is to construct a more informative prior, which allows for a gradual reduction of the inflation parameter α. Importantly, α plays a critical role in preserving parameter diversity throughout the iterations. It can also be interpreted as an inflation factor accounting for unresolved~~

~~dynamics, analogous to error inflation techniques commonly used in the data assimilation community (see, e.g., Minamide and Zhang, 2017; Ohishi et al., 2022).~~

~~The sampling procedure is designed to efficiently explore the high-dimensional parameter space while keeping computational costs manageable. Model parameter samples are first drawn using Sobol's sequence, a low-discrepancy Quasi-Monte Carlo method (Sobol' et al., 2011).~~

$$ESS = \frac{1}{\sum_{j=1}^{N} \omega_j^2} \quad (6)$$

At each iteration, both the prior and the likelihood function are updated. The goal of each step is to construct a more informative prior, which allows for a gradual reduction of the inflation parameter α. Importantly, α plays a critical role in preserving parameter diversity throughout the iterations. It can also be interpreted as an inflation factor accounting for unresolved dynamics, analogous to error inflation techniques commonly used in the data assimilation community (see, e.g., Minamide and Zhang, 2017; Ohishi et al., 2022).

The sampling procedure is designed to efficiently explore the high-dimensional parameter space while keeping computational costs manageable. Model parameter samples are first drawn using Sobol's sequence, a low-discrepancy Quasi-Monte Carlo method (Sobol' et al., 2011), implemented through the scipy.stats.qmc.Sobol function in the SciPy library ~~(Virtanen et al., 2020)~~ (Virtanen et al., 2020). Low-discrepancy sampling ensures more uniform coverage of the parameter space and reduces the risk of missing key regions of variability ~~(Renardy et al., 2021)~~ (Renardy et al., 2021). This method requires the sample size (N) to be a power of two.

To balance computational expense with sampling density, a dynamic ensemble size is used. For the first iteration, which must sample the entire broad prior parameter space, a larger ensemble of N = 8,192 is used. In subsequent iterations, where the sampling is focused on a narrower region of interest, the ensemble size is reduced to N = 2,048 to lower the computational cost.

The model is then integrated with these parameter sets to generate an ensemble of trajectories. Self-normalized importance weights are computed using the adaptive inflation factor (α), which is adjusted to maintain an effective sample size (ESS) of at least 25% of the current ensemble size. In the final iteration of the iIS algorithm, the resampling process retains the top-ranked particles corresponding to the ESS, which resulted in a final optimized ensemble of 512 members.

For emergent metrics related to depth (H_{O_2min}, $H_{nitracline}$, and H_{DCM}), we apply a uniform error distribution around the observational values. In practice, if a model realization produces a value outside the specified error range for any of these metrics, it is assigned a weight of zero. Additionally, some observed variations in these depth metrics are not captured by any simulation in the ensemble. This is particularly true for variations in the depth of the oxygen minimum, which appear to be driven more by physical than biogeochemical processes. To account for this, an additional condition is introduced: If more than 25% of the simulations fall outside the error range for a given observation, that observation is excluded from the analysis. This approach assumes that such observations are influenced by physical processes not well represented in the model. Finally, DCM are not present in all vertical profiles, resulting in discontinuities in the corresponding time series. By applying this method, simulations that produce an unobserved DCM, as well as those that fail to reproduce an observed one, are filtered out.

To ensure biologically realistic growth rates of mesozooplankton in the model ensemble, an additional constraint based on temperature-dependent physiological dynamics has been implemented. ~~Specifically, building on the work of (Gillooly et al., 2002)~~ Specifically, building on the work of (Huntley, 1992) and using the temperature measured by float #5904479, the maximum generation time of mesozooplankton should not exceed 30 days. Accordingly, the minimum plausible accumulation rate was set to 0.01 day^{-1}. To prevent the regeneration of ensemble members yielding unrealistically low mesozooplankton accumulation rates, any member with a maximum accumulation rate below 0.01 day^{-1} was assigned a weight of zero and therefore excluded from subsequent resampling.

To approximate the posterior distribution $p(\mathbf{x}|\boldsymbol{y})$, a kernel density function (KDF) is fitted using the KernelDensity class from the sklearn.neighbors module of the scikit-learn library ~~(Pedregosa et al., 2011)~~ (Pedregosa et al., 2011). The top-ranked particles corresponding to the ESS are retained, while the remaining particles are resampled from the fitted KDF. All weights are reset to N^{-1} for the next iteration, and the PISCES model is re-run for the newly resampled particles (Fig. 3).

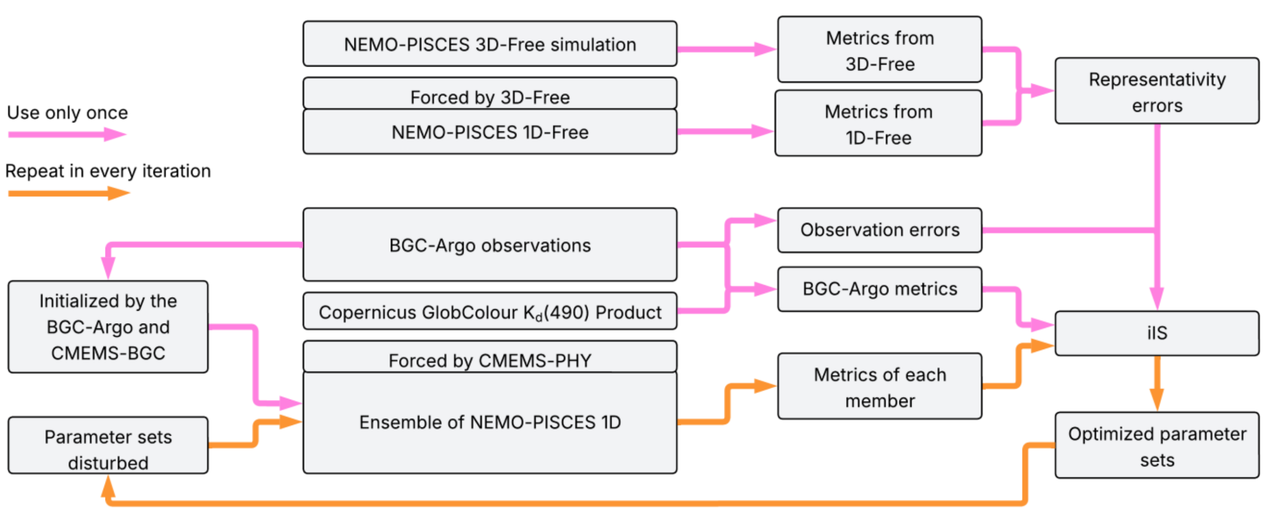


Figure 3. Schematic overview of the different datasets used during the parameter optimization process, referred to as iterative Importance Sampling (iIS).

This procedure is repeated iteratively until the inflation factor α converges or a predefined maximum number of iterations is reached. The proposed algorithm can also be interpreted as an adaptative IS method, following the classification of (Elvira and Martino, 2021), due to the use of a new proposal distribution, here, the prior $p(\mathbf{x})$, at each iteration.

3.3. ~~3.3~~ Sobol Indices

The purpose of the sensitivity analysis is to identify which of the 95 PISCES parameters have a measurable influence on the assimilated metrics, and to quantify whether this influence operates directly (first-order effect) or through interactions with other parameters (higher-order effects). This information is used to define two reduced parameter subsets for comparison with the full 95-parameter optimisation. The following subsections describe the theoretical framework (Sect. 3.3.1), the quantities of interest used to measure parameter influence (Sect. 3.3.2), and the selection criterion applied to define the parameter subsets (Sect. 3.3.3)

3.3.1. Theoretical framework

Sensitivity analysis plays a key role in the calibration of marine biogeochemical models such as PISCES, as it identifies the parameters that most strongly influence model outputs. For the purpose of this study, 'influential parameters' are defined as those whose variation accounts for a significant portion of the model output variance. By highlighting these parameters, SA enables a targeted reduction of the parameter space, thereby decreasing the computational cost of optimization. Given the inherent complexity and nonlinearity of modelled biogeochemical processes, the choice of an appropriate SA method is particularly important. In this study, we use a global method based on first-order and total-order Sobol' indices ~~(Sobol′, 2001)~~ (Sobol′, 2001). This section outlines the theoretical background, implementation, and criteria used for parameter selection.

Global Sensitivity Analysis using Sobol' indices offers a robust framework for quantifying the contribution of individual input parameters and their interactions to the variance in model outputs. Unlike local sensitivity methods, Sobol' indices account for nonlinear effects and interactions across the entire parameter space. This makes them particularly well-suited for complex models like PISCES, which involves 95 parameters with potentially intricate dependencies ~~(Prieur el al., 2019; Issan et al., 2023)~~ (Prieur el al., 2019; Issan et al., 2023).

Sobol' indices are derived from a functional Analysis Of Variance (ANOVA) decomposition of a scalar Quantity of Interest (QoI) of the model output, denoted by Y, as follows:

$$Y = f_o + \sum_i f_i(\theta_i) + \sum_{i<j} f_{ij}(\theta_i, \theta_j) + \cdots + f_{1,2,\dots,d}(\theta_1, \theta_2, \dots, \theta_d) \ , \tag{7}$$

where f_o is the mean output, and each term f_i or f_{ij} represents contributions of individual parameters or their interactions. $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ is the parameter vector, and d is the number of parameters.

Sobol' indices distinguish between direct parameter effects and higher-order interactions. The first-order Sobol' index quantifies the proportion of output variance attributable to a single parameter while accounting for variations in all other inputs ~~(Sobol′, 2001)~~ (Sobol′, 2001). In contrast, the total-order index captures both the direct contribution of a parameter and all its interactions with other parameters, including second-order and higher-order dependencies, offering a comprehensive measure of its influence. For the sake of simplicity, first-order Sobol indices will be referred to as ~~“Main effects”,~~ 'Main effects', and total-order Sobol indices as ~~“Total effects”~~ 'Total effects' throughout this paper. The mathematical definitions of $\boldsymbol{S}_i$ and $\boldsymbol{T}_i$ are given in the Supplementary Information (Eqs. S1–S2).

~~The first-order index is calculated as~~

$$\boldsymbol{S}_i = \frac{\mathbb{V}ar_{\mathbf{X}_i}\left[\mathbb{E}_{\mathbf{X}_{\sim i}}\left(f(\boldsymbol{\theta})|(\boldsymbol{\theta}_i)\right)\right]}{\mathbb{V}ar[f(\boldsymbol{\theta})]}, i = 1, \dots, d \tag{8}$$

~~and the total-order index, capturing both direct and interaction effects, is calculated as:~~

$$\boldsymbol{T}_i = 1 - \frac{\mathbb{V}ar_{\mathbf{X}_i}\left[\mathbb{E}_{\mathbf{X}_i}\left(f(\boldsymbol{\theta})|(\boldsymbol{\theta}_{\sim i})\right)\right]}{\mathbb{V}ar[f(\boldsymbol{\theta})]} = \frac{\mathbb{E}_{\mathbf{X}_{\sim i}}\left[\mathbb{V}ar_{\mathbf{X}_i}\left(f(\boldsymbol{\theta})|(\boldsymbol{\theta}_i)\right)\right]}{\mathbb{V}ar[f(\boldsymbol{\theta})]} \tag{9}$$

A key step in implementing Sobol' analysis is the construction of input sampling matrices, which are used to generate different model simulations. Monte Carlo methods or quasi-random sampling techniques, such as Sobol' sequences, are commonly used to estimate these indices, requiring multiple model evaluations ~~(Campolongo et al., 2000).~~ (Campolongo et al., 2000). Following the approach described by ~~(Issan et al., 2023)~~ (Issan et al., 2023), the process involves:

1. Generating a baseline matrix ($\mathbf{A}$): A matrix where each row represents a parameter set sampled from its respective probability distribution. This matrix has dimension $N \times d$, where N represents the number of samples per parameter and d the number of parameters studied.
2. Generating a perturbed matrix ($\mathbf{B}$): A second independent sample matrix of the same size as $\mathbf{A}$.
3. Constructing hybrid matrices ($\mathbf{C}$): Matrix $\mathbf{C}$ is generated by replicating matrix ($\mathbf{A}$) d times. In the ith copy, the ith line of $\mathbf{B}$ is replaced with the corresponding line from matrix $\mathbf{A}$, denote as $\mathbf{A}(:, i)$. This procedure ensures that each parameter is individually perturbed while keeping all others unchanged.

$$\mathbf{C}^{(i)} = \begin{bmatrix} & \mathbf{B}(:,1) & \\ & \vdots & \\ - & \mathbf{A}(:,i) & - \\ & \vdots & \\ & \mathbf{B}(:,d) & \end{bmatrix} \in \mathbb{R}^{N\times d}, i = 1, \dots, d$$

In this study, both the first-order and total-order Sobol sensitivity indices are estimated for 95 PISCES model's parameters. Using Sobol's method, the total-order sensitivity can be estimated with $N \times (d+2)$ model evaluations. To ensure a robust estimation of the sensitivity indices, the parameter space must be sampled with a sufficiently large number of points (N). The sample size was chosen to be consistent with established practice for such high-dimensional models ~~(Prieur et al., 2019; Issan et al., 2023).~~ (Issan et al., 2023; Prieur et al., 2019). Therefore, N was set to 2^{14}, as the Sobol' sequence sampling method is most efficient when the number of samples is a power of two. Additionally, the same parameter ranges as those defined during the initialization of the iIS (see Sect. 3.f) are imposed.

~~To perform sensitivity analysis efficiently, the Python Sensitivity Analysis Library (SALib) enables both parameter space sampling and the computation of first-order and total-order Sobol indices (Iwanaga et al., 2022). The sampling is conducted using the Sobol sequence, a quasi-random method with low discrepancy (Sobol′, 2001).~~

To perform sensitivity analysis efficiently, the Python Sensitivity Analysis Library (SALib) enables both parameter space sampling and the computation of first-order and total-order Sobol indices (Iwanaga et al., 2022). The sampling is conducted using the Sobol sequence, a quasi-random method with low discrepancy (Sobol′, 2001).

3.3.2. Definition of the Quantity of Interest (QoI)

For each metric defined in Table 2, except for those describing emergent properties, two QoIs are computed. The first is the RMSE, which quantifies the discrepancy in amplitude between the modeled and observed metrics. The second is the temporal correlation between the modeled and observed metrics, intended to assess the sensitivity of the system's temporal dynamics to variations in the model parameters.

More precisely, ~~the RMSE~~ for ~~a~~ each metric M ~~and for a sample j is computed as:~~

$$\cancel{Y_j^{RMSE,M} = \sqrt{\frac{1}{T}\sum_{t\in\mathcal{T}}\left(s_{j,t}^{M} - y_t^{M}\right)^2} \qquad (10)}$$

~~Where $\mathcal{T}$ is~~ ensemble member j, we compute the root-mean-square error between the ~~set of observation~~ simulated and observed time ~~indices and $T = |\mathcal{T}|$ is the number of available observations. Here, $s_{j,t}^{M}$~~ series, and ~~y_t^{M} denote~~ the Pearson correlation coefficient between them (Eqs. S3–S4 in the ~~restrictions of the model state vector and the observation vector, respectively, to metric M at time index t.~~ Supplementary Information).

~~The Correlation Coefficient (ρ) for each metric and sample is computed as follow:~~

$$Y_j^{\rho,M} = \frac{\sum_{t\in\mathcal{T}}\left(s_{j,t}^{M} - \overline{s_j^{M}}\right)\left(y_{j,t}^{M} - \overline{y_t^{M}}\right)}{\sqrt{\sum_{t\in\mathcal{T}}\left(s_{j,t}^{M} - \overline{s_j^{M}}\right)^2 \sum_{t\in\mathcal{T}}\left(y_{j,t}^{M} - \overline{y_t^{M}}\right)^2}} \tag{11}$$

~~Where the $\overline{\mathbf{s}_j^{M}}$ and $\overline{\mathbf{y}^{M}}$ denote the time average of the model and observation metrics.~~

3.3.3. Parameter selection for optimization

The aim is to include in the iIS procedure only those parameters that have an impact on the assimilated metrics, and therefore only those parameters that can be significantly constrained by the metrics defined in Table 2. Parameters that do not significantly influence these metrics, either in terms of phenology or amplitude, are not perturbed. For such non-influential parameters, we assume that their reference values should be retained, as there is insufficient information to constrain their uncertainties.

To identify influential parameters, we select all parameters with a Sobol’ sensitivity index (first-order or total-order; see Sect. 3.3 for details) greater than 0.02 for at least one quantity of interest (QoI), following the approach of (Prieur et al., 2019). This filtering step yields a reduced parameter set, thereby lowering the dimensionality of the parameter uncertainty space.

3.4. Parameter perturbation

We have shown in Sect. 3.2 and 3.3 that both iIS and Sobol' sensitivity analysis require a prior estimate of the parameters' probability density functions. In practice, however, limited information about these parameters is available in the literature; most often, only plausible value ranges are reported ~~(Denman, 2003)~~ (Denman, 2003).

Given this lack of detailed prior knowledge, we assume uniform distributions. This choice reflects a non-informative prior assumption, treating all values within the specified interval as equally likely. It provides a conservative starting point for inference, minimizing subjective bias.

To account for the possibility that fitting the ensemble of assimilated metrics may require the model to reach a substantially different equilibrium state, we define broad perturbation intervals, ranging from one-hundredth to twice the reference values. To ensure physical consistency, certain parameters are subject to additional constraints to prevent unrealistic phenomena such as the artificial generation or loss of matter for example (Table S1 in the Supplement).

3.5. Observation errors

In the Sect. 3.2 each particle is weighted using the observation likelihood. As noted by ~~(van Leeuwen et al., 2019)~~ (van Leeuwen et al., 2019), the likelihood function $p(\boldsymbol{y}|\mathbf{x})$ quantifies the probability of obtaining the observation $\boldsymbol{y}$ assuming that $\mathbf{x}$ represents the true state. Given that observations are modeled as $\boldsymbol{y} = H(\mathbf{x}) + \boldsymbol{\epsilon}$, with H a possibly non-linear observation operator mapping from the model state to the observation space and $\boldsymbol{\epsilon}$ a random noise following a known distribution, the likelihood becomes a function of the discrepancy between the observed and model-predicted values, shifted by the noise distribution p_ϵ :

$$\cancel{p(\boldsymbol{y}|\mathbf{x}) = p_{\boldsymbol{\varepsilon}}(\boldsymbol{y} - H(\mathbf{x})) \quad (12)}$$

$$p(\boldsymbol{y}|\mathbf{x}) = p_{\boldsymbol{\varepsilon}}(\boldsymbol{y} - H(\mathbf{x})) \tag{8}$$

Therefore, accurately estimating observational errors is crucial, as these errors directly influence the weighting of particles in the importance sampling procedure. We consider two main sources of observational error: measurement error and representativity error. Measurement error accounts for uncertainties in mapping from the model state to the observation space. This includes uncertainties associated with the observation operator (H) itself, the instrumental precision of direct sensor measurements, and the inferred uncertainty of variables derived from neural networks (e.g., CANYON-B). Representativity error, in contrast, arises from discrepancies between what the 1D model can simulate and the finer-scale or unresolved 3D features present in the observations.

The observation error $\boldsymbol{\epsilon}$ is assumed to follow a Gaussian distribution for most metrics. However, for metrics related to the depth of the nitracline, the DCM, and the depth of the oxygen minimum, a uniform distribution is used. Within the iIS procedure, this choice effectively excludes particles that produce unrealistic values for these depth-related variables. The explicit form of the likelihood function is given in the Supplementary Information (Eq. S14).

3.6. Measurement errors

The observation likelihood in iIS requires an estimate of the measurement uncertainty for each metric. We used four distinct methods to compute ~~the measurement~~ these errors, ~~with~~ each ~~method~~ tailored to a specific group of metrics. The resulting observation errors, expressed in percentages, are shown for each metric in Table 2.

~~3.6.1 Direct observations (Chl-a, O_2, NO_3^-)~~

~~For variables directly measured by BGC-Argo (Chl-a, O_2, NO_3^-), we defined the measurement error based on the mean RMSE values reported in (Mignot et al., 2019). That study calculated the RMSE between float observations and co-located, ship-based measurements. To convert these absolute RMSE values into the relative percentage errors required for our study, we~~

~~normalized them by the average of the same ship-based reference dataset used in (Mignot et al., 2019). For this normalization, we used a robust mean, calculated after excluding the top and bottom 5% of the reference data to reduce the influence of outliers.~~

3.6.1. ~~3.6.2~~ Direct observations (Chl-a, O_2, NO_3^-)

For variables directly measured by BGC-Argo (Chl-a, O_2, NO_3^-), we defined the measurement error based on the mean RMSE values reported in (Mignot et al., 2019). That study calculated the RMSE between float observations and co-located, ship-based measurements. To convert these absolute RMSE values into the relative percentage errors required for our study, we normalized them by the average of the same ship-based reference dataset used in (Mignot et al., 2019). For this normalization, we used a robust mean, calculated after excluding the top and bottom 5% of the reference data to reduce the influence of outliers.

3.6.2. Particulate Organic Carbon

For the POC concentrations derived from particulate backscatter, we adopt a fixed relative measurement error of 40%. This choice is based on the error model of ~~(Johnson et al., 2017)~~ (Johnson et al., 2017), which recommends using the greater of an absolute error of 35 mg C·m^{-3} or a relative error of 20%. For our study region, the mean observed POC concentration in the productive layer is approximately 79 mg C·m^{-3}. At this concentration, the absolute error threshold of 35 mg C·m^{-3} is equivalent to a relative error of ~44% (i.e., 35/79). Since this is greater than the 20% relative error threshold, it becomes the dominant error source. We therefore adopt a rounded, conservative value of 40% for the POC measurement error.

3.6.3. ~~3.6.3~~ Error representation for log-transformed metrics

The metrics for Chl-a and POC concentration are $\log_{10}$-transformed to account for their lognormal distribution. Consequently, their respective relative percentage errors are converted into a fixed, additive error in log-space. This error is calculated using a first-order error propagation formula; (error% / 100) / ln(10).

3.6.4. ~~3.6.4~~ Uncertainty of Neural Network-Derived Variables

The measurement error for variables derived from the CANYON-B and CONTENT neural networks (i.e., PO_4^{3-}, Si, TA, and DIC) was estimated by quantifying and combining two main sources of uncertainty: the propagated uncertainty from input oxygen measurements, and the intrinsic uncertainty of the network algorithms themselves.

This analysis was performed on a computationally feasible subset of the data (1,313 points, selected by taking one of every 20 points from the full profiles). For each of these data points, we first calculated a total measurement error. The propagated uncertainty was quantified via a Monte Carlo experiment in which 300 perturbed oxygen values (assuming a 3% input error) were passed through the networks to find the standard deviation of the outputs. This was combined in quadrature with the intrinsic uncertainty, which is a direct output of the neural networks.

This process yielded a set of 1,313 total error estimates. To derive a single, robust value for each variable, we then calculated the mean of these 1,313 estimates after excluding the top and bottom 5% as outliers. Finally, this mean total error was normalized by the mean of the corresponding 1,313 variable observations to yield the final percentage error reported in Table 2.

3.6.5. Uncertainty in depth-based metrics

For depth-based metrics (H_{DCM}, $H_{nitracline}$, H_{O_2min}), we used a gradient-based approach. For each individual profile, we first estimated a depth uncertainty by dividing the known concentration error of the relevant variable (e.g., Chl-a error for H_{DCM}) by the local vertical concentration gradient at that feature's depth. This procedure yielded a time series of individual depth error estimates for each metric.

To derive a single, robust error value for the entire time series, we used the interquartile range (IQR). After calculating all the individual depth errors, the final representative uncertainty (in meters) was defined as the width of the interquartile range (i.e., the 75th percentile minus the 25th percentile) of these estimates. This method was chosen over a simple mean or median because the distribution of depth errors was highly skewed, and the IQR provided a more stable and representative measure of the typical uncertainty, avoiding inflation from extreme outliers.

BGC-Argo floats typically do not sample the top few meters of the water column, creating an "unseen" surface layer. If a feature, was detected at the shallowest point of a profile, it is impossible to know if the true maximum was at that depth or shallower, within the unsampled layer. To account for this ambiguity, we introduced an additional uncertainty term in these specific cases. This term was set equal to the depth of the shallowest observation itself, effectively representing the possibility that the true feature depth could be anywhere between the surface and the first measurement. This additional term was combined in quadrature with the gradient-based depth error.

3.7. Representativity errors

Using a 1D model to represent a 3D ocean introduces a 'representativity error' due to neglected 3D physical processes. To account for this, the cost function used in our assimilation combines observational errors with this representativeness term. Including this error is crucial, as it prevents the optimization from compensating for missing physics and yields a more reliable calibration. The magnitude of this error was quantified at each time step as the absolute difference between the value from the full ‘3D-Free’ simulation (sampled along the float's trajectory) and the value from the corresponding 1D simulation (‘1D-Free’, initialized and forced with the same 3D fields).

This time-dependent representativity error is then added in quadrature to the measurement error at each corresponding time step, creating the total observational error used in the likelihood calculation. This approach allows for greater uncertainty during periods when the 1D assumption is weakest, prevents overconfidence in observational constraints, and avoids forcing the parameter optimization to compensate for missing physics with unrealistic parameter values.

3.8. Assessment of Neglected Error Sources

In addition to measurement and representativity errors, we also evaluated two other potential sources of uncertainty: (i) error covariances among the observed and derived variables, and (ii) grid discretization errors.

First, to assess the impact of error covariances, we estimated the full error covariance matrix for the observation vector. We propagated a 3% perturbation in the input oxygen observations through the CANYON-B and CONTENT algorithms to quantify the covariance terms both among the network-derived variables (e.g., DIC, TA) and between those variables and the uncertain input (O_2). A comparative analysis demonstrated that including these off-diagonal covariance terms in the likelihood calculation had a negligible effect on the final results; the RMSE between the optimized solution and the observations changed by less than 2%.

Second, we assessed the impact of grid discretization error. This error arises because the 1D model is forced by a single grid point from the 1/4° physical model, while the true float position varies within that grid cell. To test the sensitivity to this choice, we ran an ensemble of 1D simulations where each member was forced by a different, randomly selected 1/12° sub-grid point from within the same 1/4° grid cell. By comparing simulations with identical parameters but different physical forcing, we could quantify the influence of this sub-grid variability. The analysis indicated that the resulting grid-induced differences were minor compared to measurement and representativity errors; including this physical uncertainty in the optimization process changed the final RMSE between the optimized solution and the observations by less than 2%.

Given the significant computational cost required to quantify these two error sources and their minor impact on the results, both were excluded from our final error budget.

3.9. Framework for Performance Evaluation

We evaluate the effectiveness of the three optimization strategies using statistical criteria that assess three key aspects of performance:

- ~~Model Skill~~ Goodness of fit: The improvement in skill against the assimilated data is quantified using the Normalized Root Mean Square Error (NRMSE).

- ~~Portability and Robustness~~ Ensemble evaluation: The solution's performance against independent data is tested using the reduced centred random variable metric (RCRV).
- Parameter Constraints: The impact on the model parameters is assessed by measuring the reduction in their posterior uncertainty via the Highest Density Interval (HDI) and by calculating the correlations among them to test for independence.

The ~~Normalized~~ Normalised Root Mean Square Error (NRMSE) is our primary metric for quantifying the misfit between simulated and observed metrics. ~~It is the standard~~ We calculate two versions for each metric M: one for the single best ensemble member and one for the weighted-mean ensemble. The NRMSE normalises the RMSE ~~normalized~~ by the total observation error ~~associated with the metrics, which is the quadratic sum of the ε~~ ε_t^M (measurement and representativity errors~~. An NRMSE~~ combined in quadrature) so that a value of ~1 indicates ~~that the~~ model-data misfit ~~is, on average, as large as~~ comparable to the expected observation error, while values less than 1 indicate a good fit.

~~We calculate two versions of the NRMSE for each metric M: one for the single best ensemble member~~ ~~s_{Best}^M~~ ~~and one for the full weighted-mean ensemble~~ ~~$\overline{s_t^M}^{\omega}$~~~~. The NRMSE for the best member is:~~

$$\cancel{NRMSE_{Best}^M = \sqrt{\frac{1}{T}\sum_{t\in\mathcal{T}}\left(\frac{y_t^M - s_{Best,t}^M}{\varepsilon_t^M}\right)^2}} \qquad \cancel{(13)}$$

~~For computing the NRMSE for the weighted-mean of the new subset of 512 members we first define~~ ~~p_j~~ ~~:~~

$$\cancel{p_j = \frac{\omega_j}{\sum_{j\in\mathcal{N}}\omega_j}} \qquad \cancel{(14)}$$

~~Where~~ ~~p_j~~ ~~is the normalized weight of each simulation for the jth ensemble member. With~~ ~~$\mathcal{N}$~~~~, the set of simulation member indices and~~ ~~$N = |\mathcal{N}|$~~ ~~is the number of available simulations.~~

~~We also define the weighted average~~ ~~$\overline{s_t^M}^{\omega}$~~ ~~:~~

$$\cancel{\overline{s_t^M}^{\omega} = \sum_{j\in\mathcal{N}} s_{j,t}^M * p_j} \qquad \cancel{(15)}$$

~~Where~~ ~~$s_{j,t}^M$~~ ~~is the value of metric M for the jth ensemble member at time t.~~

~~The NRMSE for the weighted-mean ensemble is given by:~~

$$NRMSE^{M}_{Ensemble} = \sqrt{\frac{1}{T}\sum_{t\in\mathcal{T}}\left(\frac{y^{M}_{t} - \overline{s^{M}_{t}}^{\omega}}{\varepsilon^{M}_{t}}\right)^{2}} \tag{16}$$

~~Where $s^{M}_{j,t}$ is the value of metric M for the jth ensemble member at time t, y^{M}_{t} is the corresponding observation, ε^{M}_{t} is the total observational and representativity error, p_j is the normalized weight of the jth member, and T is the number of time points. For the depth-based metrics (H_{DCM}, $H_{nitracline}$, and H_{O2min}), the error ε^{M}_{t} used is the average of the asymmetric errors above and below the observed value.~~ (Eqs. S5–S8 in the Supplementary Information). The performance of the ~~optimized~~ optimized simulations is then evaluated by comparing their NRMSE values to that of the reference simulation ~~(run with default parameters).~~with default PISCES parameters (hereafter REF). The optimised simulations are referred to as OPTI. For both the best-member and ensemble NRMSE, the relative change in performance is quantified as $\Delta\%NRMSE = (NRMSE_{OPTI} - NRMSE_{REF}) / NRMSE_{REF} \times 100\%$, so that negative values indicate improvement.

To assess the performance of the ~~optimized~~ optimised ensembles against independent ~~validation data~~ observations, we use ~~a diagnostic based on~~ the Reduced ~~Centered~~ Centred Random Variable (RCRV). At each time, ~~This metric standardizes the model-data misfit at point by accounting for both observational and model uncertainty, allowing us to test for both systematic bias and the robustness of the ensemble spread.~~

For computing step, the RCRV ~~with the weighted standard deviation we first define $\sigma^{M,\omega}_{t}$:~~

$$\sigma^{M,\omega}_{t} = \sqrt{\sum_{j\in\mathcal{N}}\left(s^{M}_{j,t} - \overline{s^{M}_{t}}^{\omega}\right)^{2} * p_j} \tag{17}$$

~~Where $\overline{s^{M}_{t}}^{\omega}$~~ is normalises the residual between the weighted ensemble mean ~~simulation, p_j is the normalized weight of the jth member. With $\mathcal{N}$, the set of simulation member indices and $N = |\mathcal{N}|$ is the number of available simulations.~~

~~For each metric M at each time point t, the **RCRV** is calculated as:~~

$$RCRV^{M}_{t} = \frac{y^{M}_{t} - \overline{s^{M}_{t}}^{\omega}}{\sqrt{{\varepsilon^{M}_{t}}^{2} + {\sigma^{M,\omega}_{t}}^{2}}} \tag{18}$$

~~Where y^{M}_{t} is the~~ and the observation, ~~$\overline{s^{M}_{t}}^{\omega}$ is the weighted ensemble mean simulation, ε^{M}_{t} is~~ by the total observational uncertainty, combining observation error ~~(measurement + representativity), and $\sigma^{M,\omega}_{t}$ is the weighted ensemble standard deviation at that time. For the depth-based metrics (H_{DCM}, $H_{nitracline}$, and H_{O2min}), the~~ and ensemble spread in quadrature (Eqs. S9–S12 in the Supplementary Information). The time-mean RCRV (bias) measures systematic error ~~ε^{M}_{t} used is the average of the asymmetric errors above and below the observed value.~~

We then summarize the full **RCRV** time series using two key statistics:

- Bias: Calculated as the mean of the **RCRV** time series. A value : values near zero ~~indicates~~ indicate an unbiased ensemble.

$$RCRV^{M}_{Bias} = \frac{1}{T}\sum_{t \in T} RCRV^{M}_{t} \qquad (19)$$

~~Dispersion: Calculated as the~~. The standard deviation of the RCRV ~~time series. A~~(dispersion) measures calibration: a value ~~near one~~of 1.0 indicates a ~~well-~~perfectly calibrated ensemble~~, where the model's predictive spread is consistent with the actual model-data errors. Values significantly below one suggest overconfidence (spread is too small), while values above one suggest underconfidence (spread is too large).~~values below 1.0 indicate a conservative (over-dispersive) ensemble, and values above 1.0 indicate an overconfident one.

$$RCRV^{M}_{std} = \sqrt{\frac{1}{T}\sum_{t \in T}\left(RCRV^{M}_{t} - RCRV^{M}_{Bias}\right)^{2}} \qquad (20)$$

We quantify the reduction in parameter uncertainty by ~~computing~~ the percentage change in the ~~width of the~~ 67% Highest Density Interval (HDI~~.~~) between the prior and posterior distributions (Eq. S13 in the Supplementary Information). The HDI represents the narrowest credible interval containing 67% of the probability mass, ~~analogous to a one-standard-deviation range. The percentage reduction ($\Delta HDI_{67\%}$) is calculated as:~~

$$\Delta HDI_{67\%} = \left(1 - \frac{HDI^{posterior}_{67\%}}{HDI^{prior}_{67\%}}\right) * -1 * 100 \qquad (21)$$

~~Here, $HDI^{prior}_{67\%}$ and $HDI^{posterior}_{67\%}$ are the widths of the 67% HDI for the initial and optimised distributions, respectively. A negative value indicates a reduction (a tightening) of~~. Negative values indicate that the posterior is narrower than the prior, i.e., the observations have reduced parameter uncertainty.

Finally, we calculate the pairwise correlation coefficients between all optimized parameters using the 512 best-weighted ensemble members from each strategy. This analysis tests whether the assimilation framework successfully found an independent solution for each parameter, a key indicator of a well-constrained system.

~~**4 Results**~~

~~4.1~~

3.10. Three-dimensional validation framework

For the-surface chlorophyll-a validation against MOBTAC, we adopt the five regional boxes defined by Gutknecht et al. (2019) in their formal skill assessment of the CMEMS IBI biogeochemical system (Fig. 10, panel b): Box 1 (Celtic Sea / Rockall Trough), Box 2 (Canary / Madeira region), Box 3 (Bay of Biscay), Box 4 (Gulf of Lion), and Box 5 (Balearic Sea). Using the same geographical framework ensures direct comparability with the established IBI validation protocol.

These boxes are, however, too small to yield robust statistics for the subsurface BGC-Argo validation, where float coverage is much sparser — particularly in the North Atlantic, which concentrates only 10–15 % of the matched profiles. The IBI domain was therefore divided into three broader sub-regions for the BGC-Argo assessment (Fig. 10): (i) Northern basin (20°W–17°E, 48–65°N), (ii) Iberian Atlantic (20–5°W, 26–48°N), and (iii) Mediterranean & Bay of Biscay (5°W–17°E, 26–48°N). For the aggregated analysis presented in Sect. 5, the Northern and Iberian Atlantic basins are further combined into a single "North Atlantic" region.

The technical details of the model–observation colocation procedure and the adaptation of the NRMSE metric for the 3D assessment are provided in the Supplementary Information (Sect. S2). In brief, subsurface BGC-Argo profiles are matched to the IBI model in space, depth, and time over the 0–500 m range, with model values interpolated horizontally, vertically, and temporally to each float observation. Surface chlorophyll-a is compared on the MOBTAC 0.25° grid using monthly means in $\log_{10}$-transformed space, displayed on Taylor diagrams. Model performance is assessed using the same NRMSE framework as in Sect. 3.9, with two differences: the representativity error term is omitted (since the 3D model resolves the processes it was designed to account for), and the observation uncertainty uses the point-by-point error field associated with each BGC-Argo measurement rather than the propagated Monte Carlo estimates. The Δ%NRMSE is computed as defined in Sect. 3.9. When both NRMSE values fall below unity, a positive Δ%NRMSE does not constitute a meaningful degradation, since both simulations already reproduce observations within their associated uncertainties.

4. Results

4.1. Global Sensitivity Analysis

The GSA revealed that the parameter controlling organic matter recycling (e.g., the half-saturation constant for DOC remineralization, K_{DOC}) exerted the strongest first-order influence, primarily through its impact on mesopelagic concentrations of DIC, nitrate, phosphate, and dissolved oxygen (Fig. 4a). This parameter, however affected only a limited subset of metrics. In contrast, the phytoplankton light response (e.g., the P-I slope for nanophytoplankton, α^{Nano}) showed significant effects on almost all state variables, including those in the mesopelagic layer. Zooplankton parameters also emerged as important drivers, particularly those related to grazing and prey preference (e.g., the microzooplankton preference for nanophytoplankton, $P_{Nano}^{MicroZoo}$, and the half-saturation constant for microzooplankton grazing, $K_G^{MicroZoo}$).

The total-order analysis demonstrated that many of the most influential parameters, especially for zooplankton, acted primarily through interactions rather than direct effects (Fig. 4b). For instance, the non-assimilated fractions of nanophytoplankton consumed by microzooplankton ($\sigma^{MicroZoo}$) and mesozooplankton ($\sigma^{MesoZoo}$) were ranked as the first and second most sensitive parameters overall, despite having only weak first-order effects (Fig. 4a). In contrast, K_{DOC} displayed similar sensitivity in both the first-order and total-order analyses, suggesting its influence is largely independent of parameter interactions (Fig. 4a). When accounting for interactions, seven parameters had significant impacts on all model metrics (α^{Nano}, $P_{Nano}^{MicroZoo}$, $K_G^{MicroZoo}$, $K_G^{MesoZoo}$, $\theta_{Max}^{Fe\ Nano}$, $G_m^{MicroZoo}$, $G_m^{MesoZoo}$). These correspond to parameters of microzooplankton, mesozooplankton and nanophytoplankton, highlighting their central role in biogeochemical dynamics (Fig. 4b).

By contrast, nine parameters related to processes irrelevant to the float's open-ocean location—such as the half-saturation constant for anoxia, coastal iron release, and iron concentration in sea ice—showed no measurable influence.

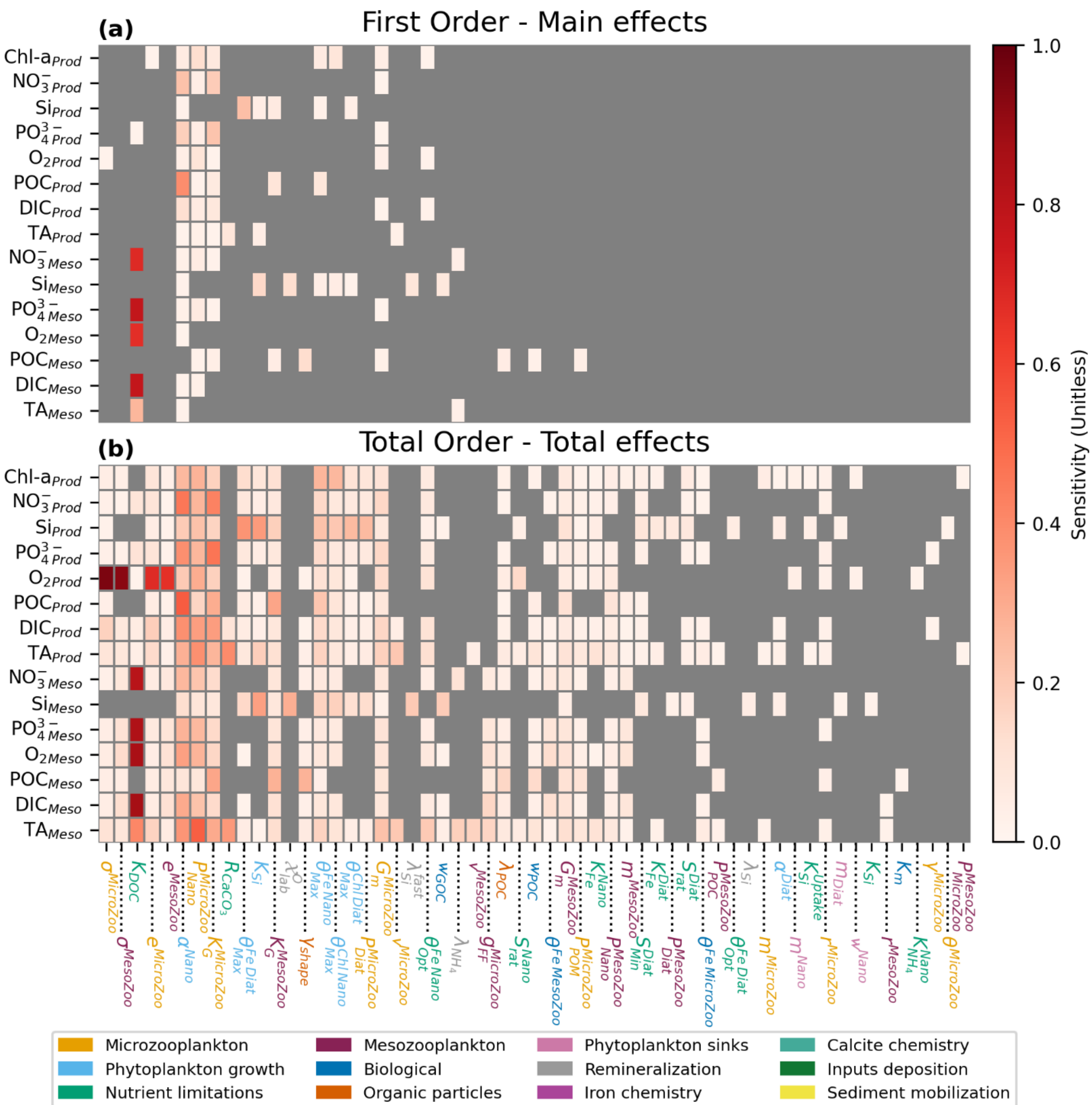


Figure 4. First-order and Total-order Sobol sensitivity indices of the parameters across the assimilated metrics. Sensitivity is defined as the maximum between the index estimated via RMSE and that estimated via correlation. Only parameters with a sensitivity index greater than 0.02 for at least one metric are shown. Panel (a) displays the first-order Sobol indices, while panel (b) shows the total-order Sobol indices. Grey cells indicate a sensitivity index below 0.02, which is considered non-significant in this study. Parameters are ranked in descending order based on their maximum Sobol indices. They are grouped according to the same categories defined in the PISCES model. Metrics defined as emerging properties are not used in this study.

4.2. ~~4.2~~ Productive Layer Skill

All three parameter optimization strategies significantly improved the simulation of the productive-layer seasonal cycle, correcting major biases present in ~~the reference model~~ REF. A visual comparison reveals that the optimized ensembles for the All-parameters (Fig. 5), Total effects (Fig. S2, in the Supplement), and Main effects (Fig. S5, in the Supplement) strategies produce nearly identical corrections. In all cases, the ensembles successfully capture the magnitude and timing of the spring phytoplankton bloom, as indicated by Chl-a concentrations (Fig. 5a). This improvement is consistently reflected across related variables: the model now reproduces the observed seasonal drawdown of (NO_3^-, PO_4^{3-}) and DIC, while the concentration of Chl-a and POC aligns closely with float data (Fig. 5a-d, f).

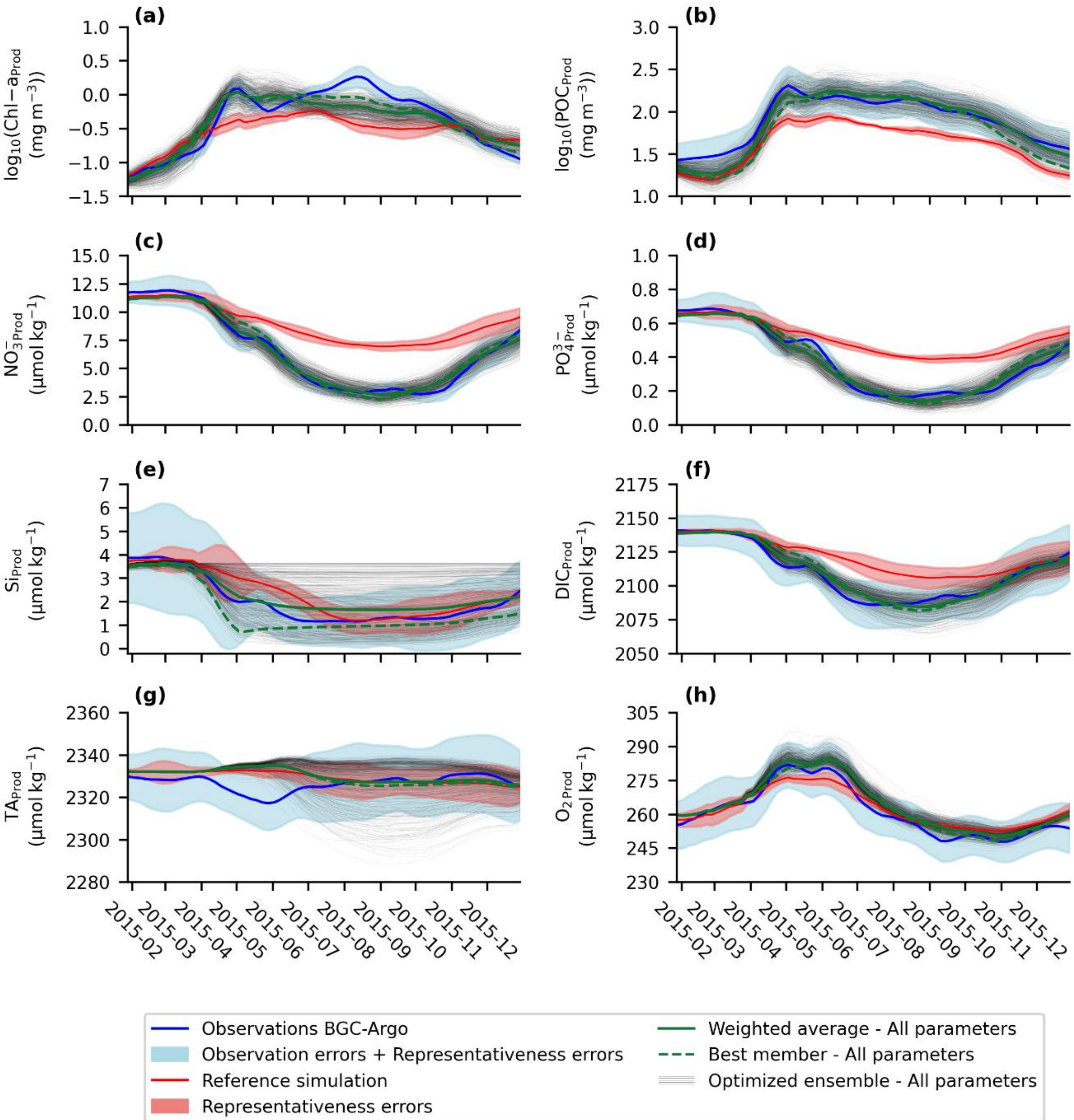


Figure 5. Seasonal cycle of assimilated metrics in the productive layer. Panels show: (a) $\log_{10}(\mathrm{Chl-a_{Prod}})$, (b) $\log_{10}(\mathrm{POC_{Prod}})$, (c) $\mathrm{NO_3^-}_{\mathrm{Prod}}$, (d) $PO_4^{3-}{}_{\mathrm{Prod}}$, (e) $\mathrm{Si_{Prod}}$, (f) $\mathrm{DIC_{Prod}}$, (g) $\mathrm{TA_{Prod}}$, and (h) $\mathrm{O_{2\,Prod}}$. The blue curve represents observations from BGC-Argo float #5904479, with the blue shading indicating the combined observations and representativity errors. The red curve corresponds to ~~the reference simulation~~ REF from PISCES-1D, with the red shading representing representativity errors. Green line indicate the weighted means of the ensemble optimized using All parameters of the PISCES model. The black curves represent the ensemble of selected members obtained by optimizing all parameters of the PISCES model. A six-point moving average was applied to all time series to smooth short-term fluctuations.

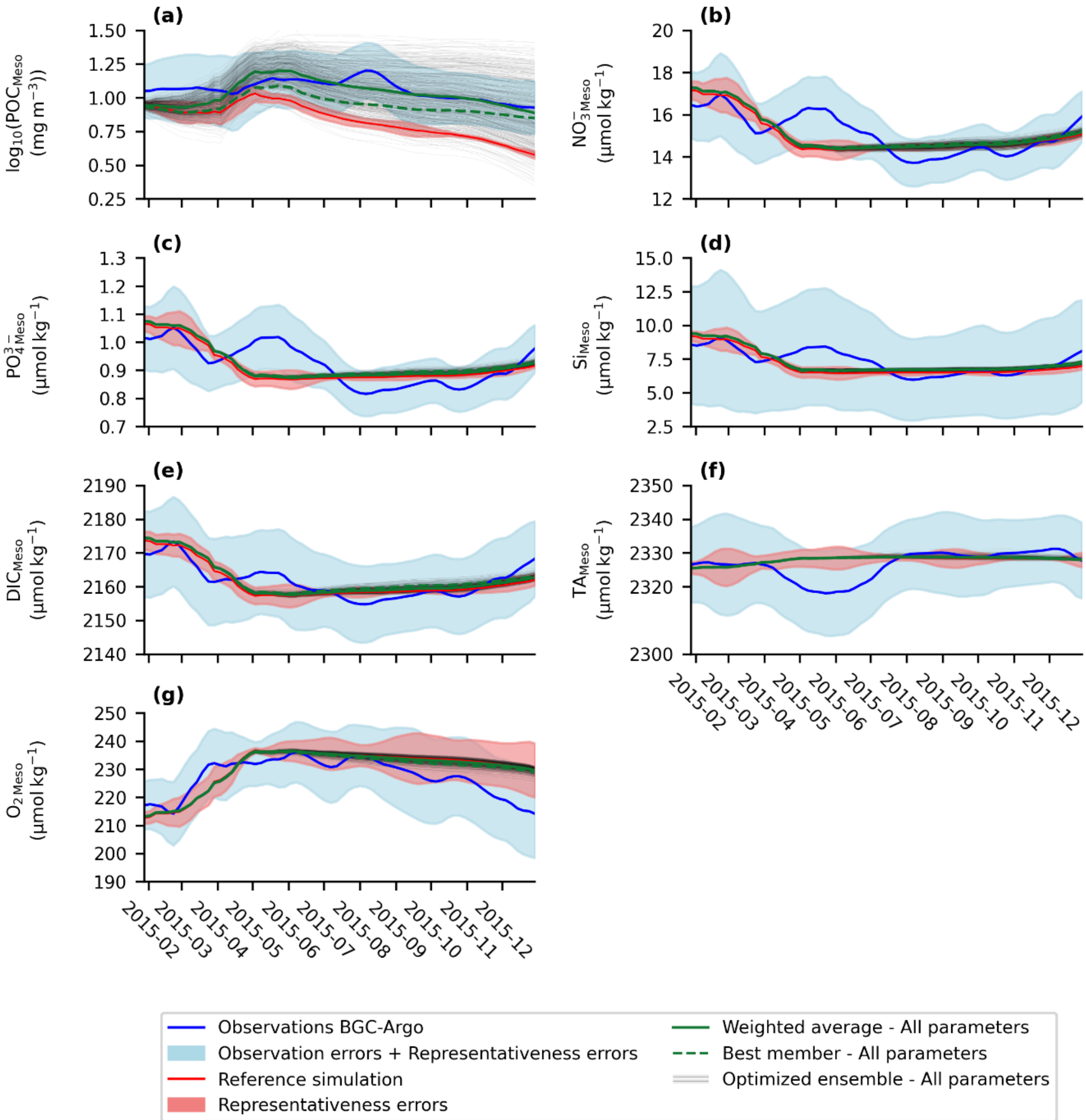


Figure 6. Seasonal cycle of assimilated metrics in the mesopelagic layer. Panels show : (a) $\log_{10}(\mathrm{POC_{Meso}})$ (b) $\mathrm{NO^-_{3\,Meso}}$, (c) $PO_4^{3-}{}_{\mathrm{Meso}}$, (d) $\mathrm{Si_{Meso}}$, (e) $\mathrm{DIC_{Meso}}$, (f) $\mathrm{TA_{Meso}}$, and (g) $\mathrm{O_{2\,Meso}}$. The blue curve represents observations from BGC-Argo float #5904479, with the blue shading indicating the combined observations and representativity errors. The red curve corresponds to ~~the reference simulation~~ REF from PISCES-1D, with the red shading representing representativity errors. Green line indicate the weighted means of the ensemble optimized using All parameters of the PISCES model. The black curves represent the ensemble of selected members obtained by optimizing all parameters of the PISCES model. A six-point moving average was applied to all time series to smooth short-term fluctuations.

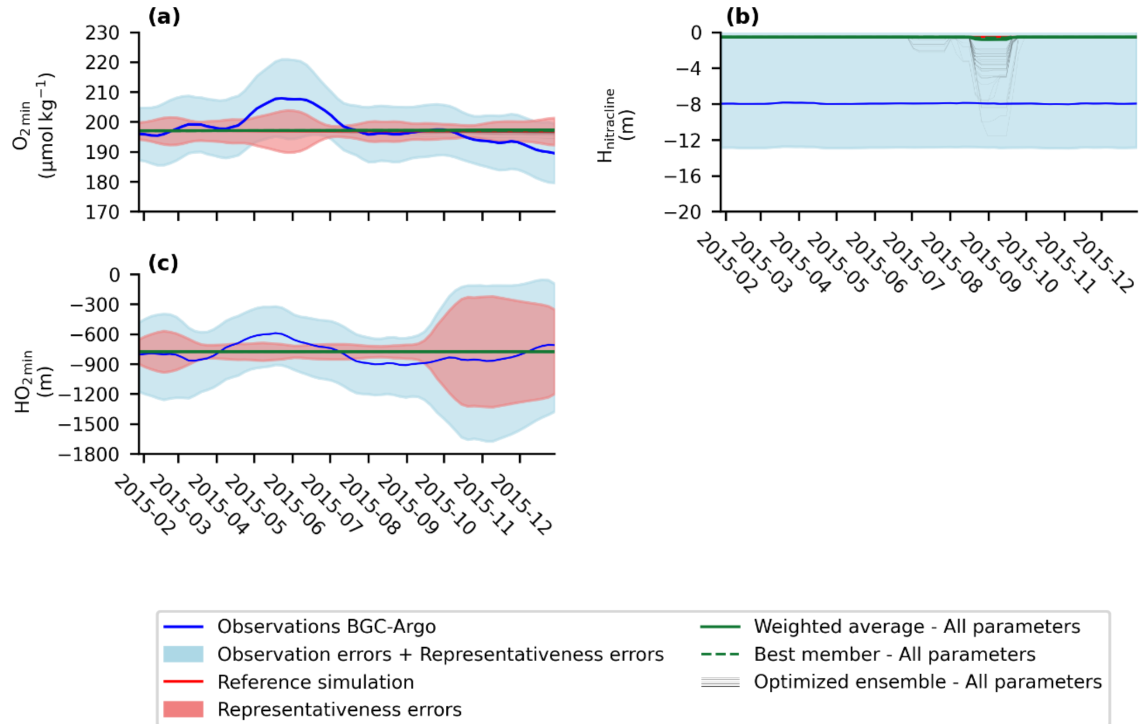


Figure 7. Seasonal cycle of assimilated emerging metrics. Panels show : (a) $O_{2\,min}$, (b) $H_{Nitracline}$, (c) $HO_{2\,min}$. Chl_{DCM} and H_{DCM} are not shown, as there were not enough DCM observations to reconstruct these metrics. The blue curve represents observations from BGC-Argo float #5904479, with the blue shading indicating the combined observations and representativity errors. The red curve corresponds to ~~the reference simulation~~ REF from PISCES-1D, with the red shading representing representativity errors. Green line indicate the weighted means of the ensemble optimized using All parameters of the PISCES model. The black curves represent the ensemble of selected members obtained by optimizing all parameters of the PISCES model. A six-point moving average was applied to all time series to smooth short-term fluctuations.

These visual improvements are confirmed by a large and statistically robust reduction in the model-data misfit (Table 3a). For the assimilated float, the three strategies achieved a comparable median NRMSE reduction across the eight productive-layer metrics: -55.6% (±36.1%) for Main effects, -54.2% (±24.2%) for Total effects, and -53.6% (±29.9%) for All-parameters. In all cases, the median improvement is substantially larger than the associated interquartile uncertainty, indicating a significant enhancement of ~~model skill.~~the goodness of fit. Given the small sample size (n=8 metrics), we used a non-parametric Kruskal-Wallis H-test to formally compare the distributions of NRMSE reductions. For the weighted-mean ensembles, the test confirmed that there is no statistically significant difference among the three strategies (p = 0.99). This remarkable similarity in performance demonstrates that while a small subset of highly sensitive parameters drives most of the improvement, perturbing all 95 parameters achieves the same high level of skill.

Table 3. Normalized Root Mean Square Error (NRMSE) across the metrics, with IQR/2 shown in parentheses. RMSE values are normalized by the combined observational and representativity errors. Percentage improvements are calculated as the relative

difference between the NRMSE of ~~the optimized simulations and the reference simulation.~~OPTI and REF. Negative values indicate improved performance. Results are presented for the assimilated float (#5904479). Columns compare ~~the reference simulation~~ REF against three parameter selection strategies. 'Main effects' and 'Total effects' refer to parameter selections based on first-order and total-order Sobol indices, respectively, while 'All Parameters' corresponds to optimization using the full parameter set. For each method, 'Best' denotes the simulation with the highest weight, and 'Ensemble' represents the weighted mean across all ensemble members. Values represent the median percentage improvement in metrics related to the productive layer (a), and for all remaining metrics (b). Note that metrics related to the DCM were excluded due to the lack of long-term observational data for this metrics. Uncertainty on the median improvement is estimated using half the interquartile range of the percentage improvements across the metrics.

(a)

	Main effects Best	Total effects Best	All Parameters Best	Main effects Ensemble	Total effects Ensemble	All Parameters Ensemble
5904479 Improvement (%)	-53.7 (±38.6)	-54.5 (±29.3)	-50.4 (±27.1)	-55.6 (±36.1)	-54.2 (±24.2)	-53.6 (±29.9)

Table 3. Continued

(b)

	Main effects Best	Total effects Best	All Parameters Best	Main effects Ensemble	Total effects Ensemble	All Parameters Ensemble
5904479 Improvement (%)	-0.25 (±1.6)	0.0 (±2.2)	-0.35 (±0.7)	-0.31 (±1.3)	-0.38 (±1.2)	-0.36 (±1.4)

The median NRMSE of the weighted-mean ensemble for each strategy was comparable to that of its single best-performing member (Table 3a). However, the ensemble weighted mean (solid lines, Fig. 5) consistently provided a smoother and more physically plausible representation of the seasonal cycle than any individual "best" member simulation (dashed lines), which is a known advantage of the ensemble approach ~~(Germineaud et al., 2019). Regarding the parameters themselves, while the overall posterior distributions were largely similar across the three strategies, the specific values for the single best-performing member of each strategy show notable differences. These parameter sets are provided in Table 4~~(Germineaud et al., 2019). Regarding the parameters themselves, while the overall posterior distributions were largely similar across the three strategies, the specific values for the single best-performing member of each strategy show notable differences. These parameter sets are provided in Table S1 as a concrete and reusable outcome for future studies.

~~Table 4. Parameter optimization results. (a) Model parameters for microzooplankton. (b) Model parameters for mesozooplankton. (c) Model parameters for phytoplankton growth. (d) Model parameters for nutrient limitations. (e) Model parameters biological. (f) Model parameters for organic particles. (g) Model parameters for phytoplankton sinks. (h) Model parameters for remineralization. (i) Model parameters for iron chemistry. (j) Model parameters for calcite chemistry. (k) Model parameters for input deposition. (l) Model parameters for sediment mobilization. The 'Ref' column indicates the reference values used in the standard configuration of the PISCES model. The 'Best' column reports the parameter value implemented in the member with the lowest cost function during the optimization procedure. A dash in any column indicates that the corresponding parameter was not modified as part of the~~

~~optimization of the parameters selected for this sensitivity test. An estimate of the uncertainty associated with the "Best" values obtained when optimizing all parameters simultaneously is provided by the lower and upper bounds of the 67% HDI.~~

~~(a)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters (HDI)~~
$p_{Diat}^{MicroZoo}$	~~MicroZoo. preference for Diatoms~~	~~0.8~~	~~0.13 (8e-3 – 8.58e-2)~~	~~8.44e-2 (8e-3 – 9.57e-2)~~	~~0.29 (8e-3 – 0.12)~~
$p_{Nano}^{MicroZoo}$	~~MicroZoo. Preference for NanoPython~~	~~1~~	~~0.32 (0.17 – 0.41)~~	~~0.97 (0.16 – 0.64)~~	~~0.6 (0.17 – 0.58)~~
$K_{G}^{MicroZoo}$	~~H-Sat. constant for MicroZoo. grazing (µmol C L⁻¹)~~	~~2e-5~~	~~3.49e-5 (2.78e-5 – 3.98e-5)~~	~~2.26e-5 (1.45e-5 – 3.16e-5)~~	~~2.65e-5 (2.21e-5 – 3.74e-5)~~
$e^{MicroZoo}$	~~Efficiency of MicroZoo. growth~~	~~0.4~~	~~0.5 (0.15 – 0.46)~~	~~0.23 (8.76e-2 – 0.46)~~	~~0.37 (0.23 – 0.59)~~
$e_{Min}^{MicroZoo}$	~~Min efficiency of MicroZoo. growth~~	~~0.4~~	~~0.5 (0.15 – 0.46)~~	~~0.23 (8.76e-2 – 0.46)~~	~~0.37 (0.23 – 0.59)~~
$G_{m}^{MicroZoo}$	~~Max. MicroZoo. grazing rate (d⁻¹)~~	~~2~~	~~2.19 (1.69 – 2.81)~~	~~1.96 (1.7 – 3.25)~~	~~1.81 (1.69 – 3.06)~~
$\gamma^{MicroZoo}$	~~Fraction of MicroZoo excretion as DOM~~	~~0.6~~	-	~~0.27 (6e-3 – 0.56)~~	~~0.4 (6e-3 – 0.47)~~
$\upsilon^{MicroZoo}$	~~Part of calcite not dissolved in MicroZoo~~	~~0.75~~	~~0.12 (7.5e-3 – 0.48)~~	~~0.4 (7.5e-3 – 0.51)~~	~~0.5 (7.5e-3 – 0.55)~~
$m^{MicroZoo}$	~~MicroZoo. mortality rate ((mol L⁻¹)⁻¹ d⁻¹)~~	~~0.005~~	-	~~8.61e-3 (4.4e-3 – 8.93e-3)~~	~~9.4e-3 (4.42e-3 – 9.46e-3)~~
$p_{POM}^{MicroZoo}$	~~MicroZoo. preference for POM~~	~~0.15~~	~~6.44e-2 (1.5e-3 – 8.51e-2)~~	~~0.19 (1.5e-3 – 0.11)~~	~~0.12 (1.5e-3 – 0.11)~~
$\phi_{Scalling}^{MicroZoo}$	~~Predation window size scalling MicroZoo~~	~~1~~	-	-	~~0.54 (1e-2 – 0.94)~~
$F_{Tresh}^{MicroZoo}$	~~Food threshold for feeding (µmol C L⁻¹)~~	~~3e-7~~	-	-	~~2.01e-7 (8.12e-8 – 3.8e-7)~~
$\phi^{MicroZoo}$	~~Predation window size MicroZoo~~	~~0.5~~	-	~~0.23 (5e-3 – 0.63)~~	~~0.76 (0.21 – 0.86)~~
$\sigma^{MicroZoo}$	~~Non assimilated fraction of NanoPhyto. by MicroZoo~~	~~0.3~~	~~6.31e-3 (3e-3 – 0.29)~~	~~7.6e-2 (0.14 – 0.41)~~	~~0.25 (0.13 – 0.47)~~
$r^{MicroZoo}$	~~Linear mortality rate of MicroZoo (d⁻¹)~~	~~0.02~~	-	~~3.29e-2 (1.89e-2 – 3.41e-2)~~	~~2.95e-2 (1.69e-2 – 3.36e-2)~~

~~**Table 4. Continued**~~

~~(b)~~

Symbol	Description (Units)	Ref	Best Main effects	Best Total effects	Best All parameters
$p_{POC}^{MesoZoo}$	MesoZoo. preference for POC	0.3	-	5.5e-2 (3e-3 – 8.83e-2)	0.32 (3e-3 – 0.1)
$p_{Nano}^{MesoZoo}$	MesZoo. preference for NanoPhyto	0.3	-	4.24e-2 (3e-3 – 9.5e-2)	3.96e-2 (3e-3 – 6.03e-2)
$e^{MesoZoo}$	Efficiency of MesoZoo growth	0.4	-	0.55 (0.38 – 0.63)	0.48 (0.36 – 0.67)
$e_{Min}^{MesoZoo}$	Min Efficiency of MesoZoo growth	0.4	-	0.55 (0.38 – 0.63)	0.48 (0.36 – 0.67)
$G_{m}^{MesoZoo}$	Max. MesoZoo. grazing rate (d^{-1})	0.5	0.4 (0.57 – 0.95)	0.83 (0.49 – 0.89)	0.53 (0.54 – 0.9)
$\sigma^{MesoZoo}$	Non assimilated fraction of NanoPhyto. by MesoZoo.	0.3	-	0.24 (0.13 – 0.35)	0.32 (7.83e-2 – 0.32)
$K_{G}^{MesoZoo}$	H-Sat. constant for MesoZoo. grazing (µmol C L^{-1})	2e-5	8.91e-6 (1.46e-5 – 2.73e-5)	3.13e-5 (5.18e-6 – 2.26e-5)	1.38e-5 (5.89e-6 – 2.32e-5)
$p_{MicroZoo}^{MesoZoo}$	MesoZoo. preference for MicroZoo	1	-	0.13 (0.43 – 0.88)	0.52 (0.64 – 1.)
$p_{Diat}^{MesoZoo}$	MesoZoo. Preference for diatoms	1	-	0.41 (0.22 – 0.77)	0.88 (0.33 – 0.95)
$g_{FF}^{MicroZoo.}$	Flux-feeding rate ((mmol L^{-1})$^{-1}$)	3e3	-	5.04e3 (1.57e3 – 5.37e3)	3.46e3 (1.58e3 – 4.63e3)
$\phi_{Scalling}^{MesoZoo}$	Predation window size scalling MesoZoo	1	-	-	1.83 (0.62 – 1.85)
$F_{Tresh}^{MesooZoo}$	Food threshold for grazing (µmol C L^{-1})	3e-7	-	-	4.89e-8 (3.e-9 – 2.93e-7)
$\upsilon^{MesoZoo}$	Part of calcite not dissolved in MesoZoo	0.75	-	0.41 (7.5e-3 – 0.55)	0.71 (0.31 – 0.95)
$m^{MesoZoo}$	MesoZoo. mortality rate ((mol L^{-1})$^{-1}$ d^{-1})	0.01	-	1.61e-2 (6.13e-3 – 1.89e-2)	6.84e-3 (2.17e-3 – 1.23e-2)
$\phi^{MesooZoo}$	Predation window size MesoZoo	0.5	-	-	0.9 (5e-3 – 0.56)
$\gamma^{MesoZoo}$	Fraction of MesoZoo excretion as DOM	0.6	-	-	0.93 (6.e-3 – 0.49)
$r^{MesoZoo}$	Exsudation rate of MesoZoo (d^{-1})	0.005	-	1.85e-3 (5.67e-4 – 6.78e-2)	8.02e-3 (3.2e-3 – 9.4e-3)

Table 4. Continued

~~(c)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~$\theta^{Chl-Nano}_{Max}$~~	~~Max. Chl-a/C in NanoPhyto (mg Chl-a (mg C)$^{-1}$)~~	~~0.033~~	~~1.1e-2 (7.23e-3 – 1.26e-2)~~	~~6.15e-3 (5.52e-3 – 1.42e-2)~~	~~1.11e-2 (6.77e-3 – 1.42e-2)~~
~~α^{Nano}~~	~~P-I slope NanoPhyto ((W m^{-2})$^{-1}$ d^{-1})~~	~~2~~	~~1.2 (0.86 – 1.39)~~	~~1.87 (0.92 – 1.68)~~	~~1.34 (1.07 – 1.88)~~
~~$\theta^{Chl-Diat}_{Max}$~~	~~Max. Chl-a /C in Diatoms (mg Chl-a (mg C)$^{-1}$)~~	~~0.05~~	~~3.65e-2 (2.66e-2 – 6.4e-2)~~	~~2.88e-2 (1.55e-2 – 5.14e-2)~~	~~9.13e-2 (1.61e-2 – 5.8e-2)~~
~~$\theta^{Chl-Phyto}_{Min}$~~	~~Min. Chl-a /C in Phyto (mg Chl-a (mg C)$^{-1}$)~~	~~0.003~~	-	-	~~3.53e-3 (3.15e-3 – 5.8e-3)~~
~~$\theta^{Fe-Diat}_{Max}$~~	~~Max. Fe/C in Diatoms (µmol Fe mol C^{-1})~~	~~6e-5~~	~~5.16e-5 (2.88e-5 – 8.28e-5)~~	~~1.09e-4 (1.29e-5 – 7.99e-5)~~	~~4.16e-5 (2.59e-5 – 9.03e-5)~~
~~$\theta^{Fe-Nano}_{Max}$~~	~~Max. Fe/C in NanoPhyto (µmol Fe mol C^{-1})~~	~~6e-5~~	~~3.14e-5 (1.31e-5 – 6.9e-5)~~	~~6.02e-5 (4.92e-5 – 1.07e-4)~~	~~4.27e-5 (4.31e-5 – 1.07e-4)~~
~~$\gamma^{Diatoms}$~~	~~Excretion ration of Diatoms~~	~~0.05~~	-	-	~~3.38e-2 (1.63e-2 – 7.22e-2)~~
~~α^{Diat}~~	~~P-I slope for Diatoms ((W m^{-2})$^{-1}$ d^{-1})~~	~~2~~	-	~~1.69 (1.69 – 3.78)~~	~~3.3 (1.27 – 3.71)~~
~~K_{Si}~~	~~Mean Si/C ratio (µmol Si mol C^{-1})~~	~~0.13~~	~~0.25 (0.14 – 0.25)~~	~~3.25e-2 (5.51e-2 – 0.22)~~	~~0.21 (7.67e-2 – 0.25)~~
~~γ^{Nano}~~	~~Excretion ration of NanoPhyto~~	~~0.05~~	-	-	~~9.96e-2 (3.76e-2 – 9.47e-2)~~
~~b_{resp}~~	~~Basal respiration rate (d^{-1})~~	~~0.033~~	-	-	~~5.41e-2 (2.2e-2 – 5.9e-2)~~

~~Table 4. Continued~~

~~(d)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~$\theta^{Fe-Nano}_{Opt}$~~	~~Optimal quota of NanoPhyto. (µmol Fe (mol C)$^{-1}$)~~	~~10e-6~~	~~2.63e-6 (1e-7 – 1.65e-6)~~	~~2.12e-6 (1.e-7 – 4.22e-6)~~	~~6.39e-7 (1e-7 – 6.23e-6)~~
~~K_{DOC}~~	~~H-Sat. constant of DOC remineralization (µmol C L^{-1})~~	~~417e-6~~	~~7.77e-4 (4.54e-4 – 7.87e-4)~~	~~6.32e-4 (4.89e-4 – 7.62e-4)~~	~~5.39e-4 (5.46e-4 – 7.89e-4)~~
~~R_{CaCO_3}~~	~~Mean rain ratio~~	~~0.2~~	~~5.34e-2 (2e-3 – 9.84e-2)~~	~~0.12 (2e-3 – 0.21)~~	~~5.54e-2 (2e-3 – 0.19)~~
~~K^{Nano}_{Fe}~~	~~Iron H-Sat. for NanoPhyto (nmol Fe L^{-1})~~	~~1.7e-9~~	-	~~3.27e-9 (1.68e-11 – 1.94e-9)~~	~~2.01e-9 (7.14e-10 – 2.33e-9)~~

~~K_{Si}^{Uptake}~~	~~H-Sat. constant for Si uptake (μmol Si L^{-1})~~	~~8e-6~~	-	~~5.96e-6 (2.77e-6 – 1.2e-5)~~	~~6.43e-6 (8e-8 – 9.82e-6)~~
~~$\theta_{Opt}^{Fe\text{-}Diat}$~~	~~Optimal quota of Diatoms (μmol Fe (mol C)$^{-1}$)~~	~~10e-6~~	-	~~1.19e-5 (1.e-7 – 9.36e-6)~~	~~1.01e-5 (6.07e-6 – 1.88e-5)~~
~~S_{Min}^{Diat}~~	~~Min. size criteria for Diatoms (m)~~	~~1e-6~~	-	~~1.73e-6 (1.11e-7 – 1.35e-6)~~	~~9.e-7 (7.47e-7 – 1.89e-6)~~
~~K_{Fe}^{Bact}~~	~~Iron H-Sat. for for DOC remin (nmol Fe L^{-1})~~	~~3e-11~~	-	-	~~5.91e-11 (1.6e-11 – 5.22e-11)~~
~~K_{Si}~~	~~H-Sat. constant for Si/C (μmol Si L^{-1})~~	~~20e-6~~	-	~~9.92e-6 (6.37e-6 – 3.11e-5)~~	~~4.39e-6 (2e-7 – 2.6e-5)~~
~~S_{rat}^{Nano}~~	~~Size ratio for NanoPhyto~~	~~3~~	-	~~5.58 (1.6 – 5.04)~~	~~3.91 (2.52 – 5.68)~~
~~S_{Min}^{Nano}~~	~~Min. size criteria for NanoPhyto (m)~~	~~1e-6~~	-	-	~~6.45.e-7 (1.65e-7 – 1.37e-6)~~
~~O_2^{min}~~	~~H-Sat. constant for anoxia (μmol O_2 L^{-1})~~	~~1e-6~~	-	-	~~7.72.e-7 (1e-8 – 1.24e-6)~~
~~$K_{Fe}^{Diatoms}$~~	~~Iron H-Sat. for Diatoms (nmol Fe L^{-1})~~	~~5e-9~~	-	~~8.91e-9 (2.09e-9 – 7.39e-9)~~	~~4.17e-10 (2.38e-9 – 7.9e-9)~~
~~$S_{rat.}^{Diat.}$~~	~~Size ratio for Diatoms~~	~~4~~	-	~~7.84 (3.58 – 8.)~~	~~1.37 (3.39 – 7.48)~~
~~$K_{NO_3}^{Nano}$~~	~~NO_3^- HS of NanoPhyto (mol N L^{-1})~~	~~1e-6~~	-	~~1.75e-6 (8.42e-7 – 1.83e-6)~~	~~4.79e-7 (8.86e-7 – 2e-6)~~
~~$K_{NH_4}^{Nano}$~~	~~NH_4 HS of NanoPhyto (mol N L^{-1})~~	~~1e-6~~	-	~~1.75e-6 (8.42e-7 – 1.83e-6)~~	~~4.79e-7 (8.86e-7 – 2e-6)~~
~~$K_{NO_3}^{Bact}$~~	~~NO_3^- HS for DOC remin (mol N L^{-1})~~	~~3e-7~~	-	~~5.73e-7 (3.39e-7 – 5.68e-7)~~	~~4.52e-7 (3.8e-7 – 5.6e-7)~~
~~$K_{NH_4}^{Bact}$~~	~~NH_4 HS for DOC remin (mol N L^{-1})~~	~~3e-7~~	-	~~5.73e-7 (3.39e-7 – 5.68e-7)~~	~~4.52e-7 (3.8e-7 – 5.6e-7)~~
~~$K_{NH_4}^{Diatoms}$~~	~~NH_4 HS of Diatoms (mol N L^{-1})~~	~~3e-6~~	-	~~3.54e-6 (2.97e-6 – 5.68e-6)~~	~~1.88e-6 (2.92e-6 – 5.99e-6)~~
~~$K_{NO_3}^{Diatoms}$~~	~~NO_3^- HS of Diatoms (mol N L^{-1})~~	~~3e-6~~	-	~~3.54e-6 (2.97e-6 – 5.68e-6)~~	~~1.88e-6 (2.92e-6 – 5.99e-6)~~

~~Table 4. Continued~~

~~(e)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~$\theta^{Fe\text{-}MesoZoo}$~~	~~Fe/C in MesoZoo (μmol Fe (mol C)$^{-1}$)~~	~~15e-6~~	-	~~2.22e-5 (1.5e-8 – 1.25e-5)~~	~~6.16e-6 (1.5e-8 – 1.27e-5)~~
~~K_m~~	~~H-Sat. constant for mortality~~	~~1e-7~~	-	~~1.04e-7~~	~~1.09e-7~~

	~~(µmol C L^{-1})~~			~~(3.32e-8 – 1.36e-7)~~	~~(9.22e-8 – 2e-7)~~
~~W_{POC}~~	~~POC sinking speed (m d^{-1})~~	~~2~~	~~0.28 (0.02 – 1.19)~~	~~0.61 (0.02 – 1.87)~~	~~0.3 (0.02 – 1.33)~~
~~$\theta^{Fe,MicroZoo}$~~	~~Fe/C in MicroZoo (µmol Fe mol C^{-1})~~	~~10e-6~~	-	~~1.39e-5 (9.54e-6 – 1.9e-5)~~	~~1.48e-5 (8.17e-6 – 1.99e-5)~~
~~W_{GOC}~~	~~Big particles sinking speed (m d^{-1})~~	~~50~~	~~87.88 (5.57 – 67.72)~~	~~84.03 (11.25 – 68.22)~~	~~18.86 (0.5 – 61.45)~~
~~$W_{GOC\,Max}$~~	~~Max big particles sinking speed (m d^{-1})~~	~~50~~	~~87.88 (5.57 – 67.72)~~	~~84.03 (11.25 – 68.22)~~	~~18.86 (0.5 – 61.45)~~
~~W_{GOC}^{length}~~	~~Big particles length scale of sinking (m)~~	~~5e3~~	-	-	~~6.77e3 (2.11e3 – 8.31e3)~~

~~Table 4. Continued~~

~~(f)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~γ_{shape}~~	~~Shape of the gamma function~~	~~1~~	~~0.55 (0.22 – 0.85)~~	~~0.84 (0.29 – 1.37)~~	~~0.82 (0.53 – 1.7)~~
~~λ_{POC}~~	~~Remineralisation rate of POC (d^{-1})~~	~~0.035~~	~~5.52e-2 (1.13e-2 – 4.89e-2)~~	~~2.16e-2 (1.14e-2 – 4.79e-2)~~	~~3.78e-2 (1.51e-2 – 5.1e-2)~~

~~Table 4. Continued~~

~~(g)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~m^{Nano}~~	~~NanoPhyto mortality rate (d^{-1})~~	~~0.01~~	-	~~1.4e-2 (7.19e-3 – 1.89e-2)~~	~~1.44e-2 (1.01e-2 – 1.91e-2)~~
~~W^{Nano}~~	~~Quadratic mortality of NanoPhyto ((µmol C L^{-1})$^{-1}$L^{-1})~~	~~0.01~~	-	~~1.9e-2 (1.e-4 – 1.21e-2)~~	~~1.05e-2 (1.e-4 – 9.3e-3)~~
~~$W^{Diatoms}_{Max}$~~	~~Maximum quadratic mortality of Diatoms ((µmol C L^{-1})$^{-1}$L^{-1})~~	~~0.03~~	-	~~2.11e-2 (1.92e-2 – 4.74e-2)~~	~~2.01e-2 (2.22e-2 – 5.13e-2)~~
~~m^{Diat}~~	~~Diatoms mortality rate (d^{-1})~~	~~0.01~~	-	~~7.32e-3 (7.63e-3 – 1.89e-2)~~	~~1.39e-2 (5.18e-3 – 1.78e-2)~~

~~Table 4. Continued~~

~~(h)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~

~~λ_{NH_4}~~	~~NH_4 nitrification rate (d^{-1})~~	~~0.05~~	~~4.78e-2 (5e-4 – 5.06e-2)~~	~~6.5e-3 (5e-4 – 5.89e-2)~~	~~2.11e-2 (5e-4 – 5.11e-2)~~
~~λ_{Si}~~	~~Remineralization rate of Si (d^{-1})~~	~~0.003~~	~~3.17e-3 (3e-5 – 3.13e-3)~~	~~5.05e-3 (3e-5 – 2.81e-3)~~	~~1.59e-3 (3e-5 – 3.14e-3)~~
~~$\theta^{Fe,Bact}$~~	~~Fe/C quota in Bacteria (µmol Fe mol C^{-1})~~	~~60e-6~~	-	-	~~9.25e-6 (6e-7 – 6.7e-5)~~
~~λ_{Si}^{fast}~~	~~Fast remineralization rate of Si (d^{-1})~~	~~0.03~~	~~9.77e-3 (3e-4 – 3.16e-2)~~	~~1.86e-2 (9.62e-3 – 4.67e-2)~~	~~1.04e-2 (6.63e-3 – 3.94e-2)~~
~~K_{Fe}^{Bact}~~	~~H-Sat. constant for bacteria Fe/C (µmol Fe mol C^{-1})~~	~~4e-10~~	-	-	~~3.42e-10 (7.81e-11 – 5.07e-10)~~
~~χ_{lab}^{0}~~	~~Fraction of labile biogenic silica~~	~~0.5~~	~~0.98 (5e-3 – 0.67)~~	~~0.25 (0.35 – 0.89)~~	~~0.85 (0.37 – 0.9)~~

~~Table 4. Continued~~

~~(i)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~λ_{Fe}^{dust}~~	~~Savenging rate of iron by dust (d^{-1} mg^{-1} L)~~	~~150~~	-	-	~~2.14e2 (64.2 – 2.21e2)~~
~~λ_{Fe}^{POFe}~~	~~Fraction of scavenged Fe that goes to POFe~~	~~1~~	-	-	~~0.34 (0.16 – 0.79)~~
~~λ_{Fe}~~	~~Scavenging rate of iron by biogenic particles (d^{-1})~~	~~0.02~~	-	-	~~2.83e-3 (2e-4 – 2.27e-2)~~
~~$\lambda_{Nanopart}$~~	~~Nanoparticle formation rate constant (s^{-1})~~	~~0.01~~	-	-	~~4.31e-3 (1e-4 -1.21e-2)~~

~~Table 4. Continued~~

~~(j)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~λ_{CaCO_3}~~	~~Calcite dissolution rate constant (d^{-1})~~	~~100~~	-	-	~~1.46e2 (1. – 1.32e2)~~

~~Table 4. Continued~~

~~(k)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~W_{dust}~~	~~Dust sinking speed (m s^{-1})~~	~~2~~	-	-	~~2.27 (2e-2 – 2.39)~~
~~υ^{Fe}_{dust}~~	~~Fe mineral fraction of dust~~	~~0.035~~	-	-	~~6.32e-2 (1.16e-2 – 5.17e-2)~~
~~$\upsilon^{Fe}_{^3He}$~~	~~Fe to ^{3}He ratio assumed for vent iron supply (mol Fe mol ^{3}He^{-1})~~	~~1e7~~	-	-	~~9.92e6 (7.99e6 – 1.89e7)~~
~~$Fe^{sed}_{Fe,min}$~~	~~Coastal release of Iron (mol Fe m^{-2} s^{-1})~~	~~2e-9~~	-	-	~~7.73e-10 (2e-11 – 2.32e-9)~~
~~Fe_{ice}~~	~~Iron concentration in sea ice (mol Fe L^{-1})~~	~~15e-9~~	-	-	~~4.78e-9 (9.53e-9 – 2.89e-8)~~

~~Table 4. Continued~~

~~(l)~~

~~Symbol~~	~~Description (Units)~~	~~Ref~~	~~Best Main effects~~	~~Best Total effects~~	~~Best All parameters~~
~~N_{fix}~~	~~Nitrogen fixation rate (µmol N L^{-1} d^{-1})~~	~~2e-7~~	-	-	~~1.75e-7 (2e-9 – 2.74e-7)~~
~~$K^{Diazotrophs}_{Fe}$~~	~~Diazotrophs H-Sat constant for iron (nmol Fe L^{-1})~~	~~1e-10~~	-	-	~~4.6e-11 (1.53e-11 – 1.26e-10)~~
~~$b^{Diazotrophs}_{light}$~~	~~Diazotrophs sensitivity to light (W m^{-2})~~	~~30~~	-	-	~~36.1 (0.3 – 37.6)~~

4.3. Deeper properties skill

In contrast to the productive layer, the parameter optimization had a limited impact on the biogeochemistry of the mesopelagic layer and on emergent vertical properties (Figs. 6, 7 and S3, S4, S6, S7, in the Supplement). For most deeper metrics, such as nutrients, oxygen and carbonate chemistry, the optimized ensembles remain nearly indistinguishable from ~~the reference simulation.~~ REF. This outcome can be attributed to several factors: ~~the reference model~~ REF already simulated these variables with reasonable skill, leaving little mean bias for the optimization to correct (e.g., Fig. 7a, c); mesopelagic properties have long adjustment timescales, making a one-year assimilation period insufficient to induce substantial changes; and the 1D model configuration neglects the deep advective processes that drive much of the variability in the ocean interior.

A key exception is mesopelagic POC. Because its dynamics are directly and rapidly forced by particle export from the surface, its response timescale is much shorter. For this variable, the optimization successfully mitigated a bias present in the reference run, significantly improving the simulation skill (Fig. 6a).

The primary role of the optimization for these deeper properties was therefore not to correct large mean-state biases, but rather to act as a constraint that prevented the model from drifting away from the observations. For instance, the assimilation effectively penalized any parameter set that produced an unrealistic nitracline depth or oxygen minimum, ensuring the optimized ensembles remained consistent with the float data throughout the simulation period.

This limited overall impact is confirmed by the error statistics. For the seven non-productive-layer metrics, the median NRMSE improvements for the weighted-mean ensembles were negligible: -0.31% (±1.3%) for Main effects, -0.38% (±1.2%) for Total effects, and -0.36% (±1.4%) for All-parameters (Table 3b). The associated uncertainty (IQR) for each strategy is several times larger than the median improvement, and a Kruskal-Wallis H-test confirmed that the differences among the strategies are not statistically significant ($p = 0.99$). These results suggest that while the framework can successfully constrain specific, well-observed features at any depth, the one-year assimilation period is insufficient to correct for potential systemic biases in slower, deeper biogeochemical processes.

As with the productive layer, the median NRMSE for the weighted-mean ensemble of each strategy was statistically indistinguishable from that of its corresponding single "best" member (Table 3b). For these deeper properties, where model-data misfits are often smaller and less dynamic, the visual difference between the smoother ensemble mean and the more variable best-member simulation is less pronounced than in the surface layer, but the principle remains that the ensemble mean provides a more robust estimate.

4.4. Uncertainty in Unobserved Variables

While all three optimization strategies showed comparable skill against assimilated data, their performance diverged significantly when estimating the predictive uncertainty for unobserved variables (Fig. 8). The Main effects ensemble, which perturbed only 29 parameters, produced the smallest uncertainty spread, with a median standard deviation of 0.17 (±0.16) relative to the seasonal cycle of the unobserved metrics. In contrast, the Total effects (66 parameters) and All-parameters (95 parameters) ensembles produced substantially larger spreads, with median values of 0.29 (±0.21) and 0.27 (±0.26), respectively.

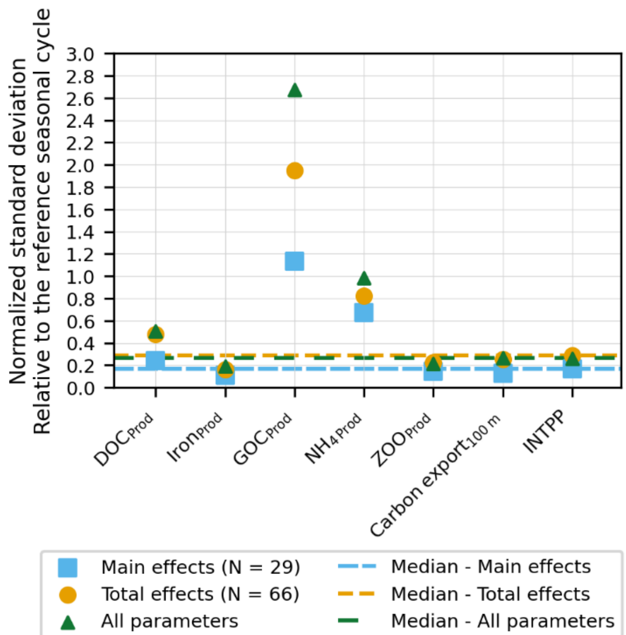


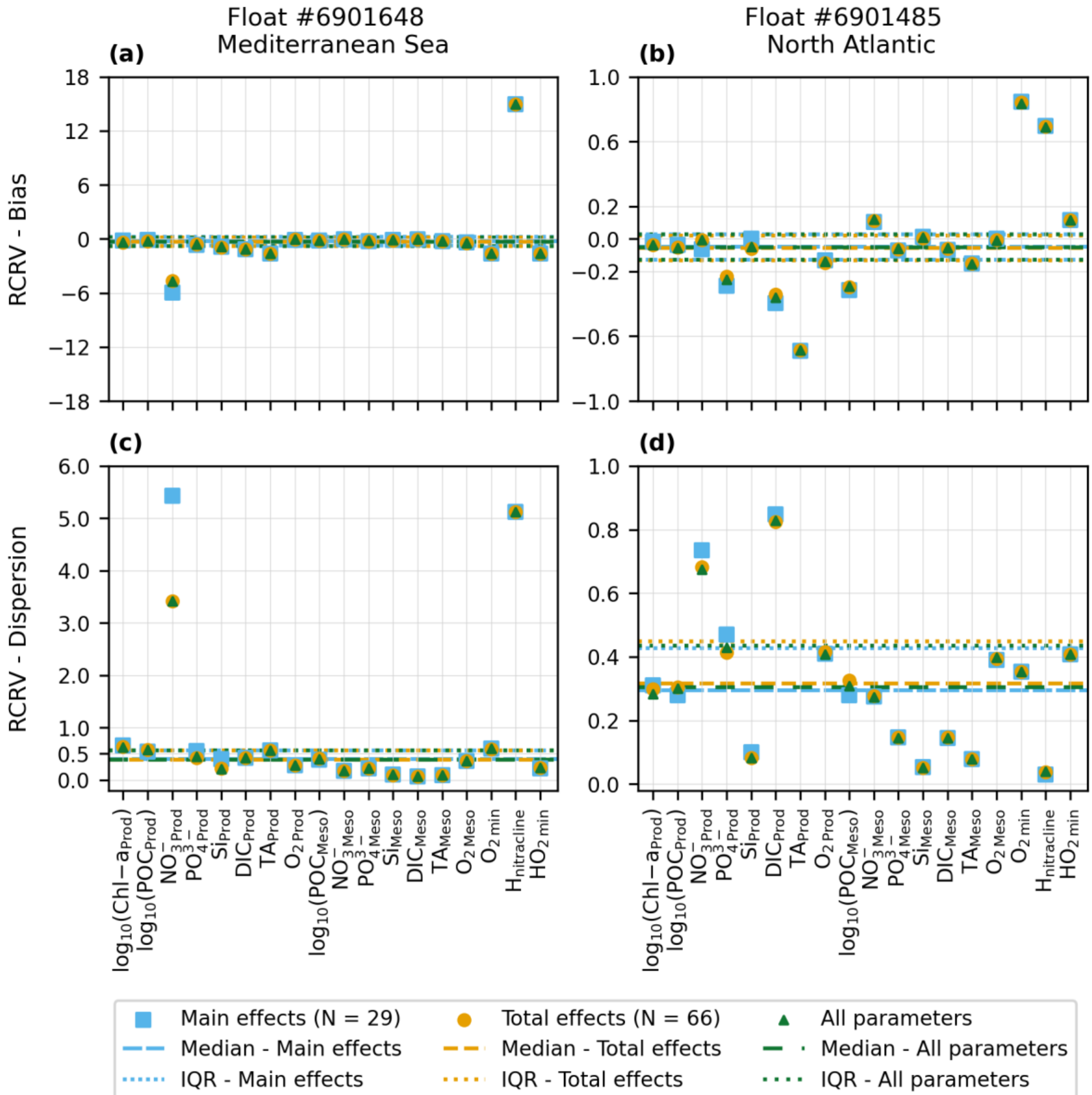
Figure 8. Standard deviation (unitless) of the ensemble optimized using different parameter selection strategies, normalized by the variability of the seasonal cycle in ~~the reference simulation~~ REF. The X-axis indicates non-assimilated state variables as well as two derived outputs : carbon export at 100 m depth and integrated net primary production (INTPP). Results are shown for the three parameter selection strategies : 'Main effects (N=29)', 'Total effects (N=66)' and 'All parameters'. Each point corresponds to a specific metric, with dashed lines indicating the median value.

Formally, a Kruskal-Wallis H-test on these distributions of uncertainty spreads does not indicate a statistically significant difference between the three strategies at the α=0.05 level (p = 0.26). However, the clear difference in the median values is not spurious; it is a direct and necessary consequence of the experimental design. A fundamental principle of uncertainty quantification is that a credible estimate of predictive uncertainty must account for all known, significant sources of error. The GSA itself demonstrated that many parameters exert influence primarily through interactions, yet the Main effects strategy deliberately holds these and dozens of other parameters fixed at their default values. This approach, therefore, knowingly omits many sources of parametric uncertainty from the analysis. The resulting narrow spread is an artifact of this omission and must be considered an underestimate of the true model uncertainty.

Conversely, the Total effects and All-parameters strategies honor the principle of comprehensive error accounting by allowing a much larger set of influential parameters to vary. The wider predictive spreads they produce are not a sign of a worse model, but rather a more honest and scientifically robust reflection of the model's true confidence limits. Therefore, we conclude that while any of the strategies can produce a skillful "best-guess" simulation, approaches that perturb more parameters provide a more trustworthy quantification of predictive uncertainty, which is essential for the reliable use of the model in forecasting or climate projection.

~~4.5 Portability of the Optimized Ensembles to Other Bioregions~~

~~To test the portability of the optimized parameter sets, the three ensembles were used to simulate conditions along the trajectories of two independent BGC-Argo floats from different bioregions: one in the North Atlantic Subpolar Gyre (#6901485) and one in the oligotrophic Mediterranean Sea (#6901648) (Fig. 1). A visual comparison of the model performance for these validation floats is provided in the Supplement (Figs. S8-S25). We then assessed the performance using the RCRV bias and dispersion metrics, which evaluate the accuracy and robustness of the ensembles against these out-of-sample observations (Fig. 9).~~



~~**Figure 9. RCRV values across all assimilated metrics. Each column corresponds at one of the two test floats (#6901648 and #6901485). The bias is estimated as the mean RCRV, while the dispersion is quantified using the standard deviation of the RCRV values. The scatter plot shows results obtained using three different parameter selection strategies : 'Main effects' and 'Total effects' correspond to selections based on first-order and total-order Sobol indices, respectively, while 'All Parameters' refers to optimization using the full set of available parameter. The median value across metrics is indicated by the dashed lines. DCM observations did not allow for the reconstruction of the corresponding observational metrics. As a result, RCRV values could not be computed for the DCM depth and DCM intensity metrics.**~~

~~The RCRV bias analysis confirmed that all three optimization strategies produced robust ensembles that were not over-fitted to the original assimilated data. For both validation floats, the median ensemble bias was low across most metrics, indicating~~

~~that the optimized simulations accurately represent the independent observations (Fig. 9a, b). A Kruskal-Wallis H-test confirmed that the bias distributions were statistically indistinguishable among the three strategies (p = 0.99). All ensembles did, however, produce the same two large, predictable biases when applied to the Mediterranean float. The first was in the nitracline depth. In the North Atlantic, the assimilated float observed a consistently shallow nitracline. The optimization algorithm successfully tuned the model's biological rate parameters to maintain this structure. However, the Mediterranean is an oligotrophic environment with a much deeper nitracline. These North Atlantic-tuned parameters caused the model to incorrectly simulate a shallow nitracline when applied there. The second large bias appeared in the productive layer nitrate concentration. This is a statistical artifact of the RCRV metric, which normalizes the model-data difference by the observational error. In the oligotrophic Mediterranean, observed nitrate concentrations and their associated errors are extremely small, causing the RCRV calculation to mathematically amplify a physically negligible model-data mismatch into a large bias score. These results do not indicate a failure of the optimization, but rather highlight the regional-specificity of certain ecosystem parameters.~~

~~Finally, the analysis of the RCRV dispersion metric assesses the realism of the quantified uncertainty. In this framework, a dispersion of 1.0 indicates a perfectly calibrated ensemble, where the predicted uncertainty matches the actual model-data error. Values significantly less than 1.0 indicate that the ensemble is over-dispersive (i.e., the predicted spread is larger than the observed error), while values greater than 1.0 indicate an under-dispersive (overconfident) ensemble.~~

~~For both validation floats, the median dispersion scores were consistent across all three strategies, approximately 0.3 for the North Atlantic and 0.4 for the Mediterranean (Fig. 9c, d). A Kruskal-Wallis test confirmed no significant difference among the strategies (p = 0.99). These values, being well below 1.0, indicate that the optimized ensembles are somewhat over-dispersive, meaning they represent a conservative (under-confident) estimate of the model's true uncertainty~~

~~Although not perfectly calibrated, the assimilation has constrained the model (dispersion is significantly greater than zero, thus avoiding ensemble collapse). Furthermore, a conservative estimate of uncertainty is often preferable to an overconfident one in forecasting applications. This balance confirms that the framework yields a portable parameter set with low bias and a reliable, albeit cautious, uncertainty envelope.~~

4.5. ~~4.6~~ Parameter Uncertainty and Correlation

The optimization framework successfully leveraged the BGC-Argo data to significantly reduce parameter uncertainty, a process we quantified by the percentage reduction in the 67% HDI of the parameter distributions (Table ~~5~~ 4). The results demonstrate the framework's ability to intelligently target and robustly constrain parameter uncertainty. This targeted nature is most evident in the All-parameters experiment. In this run, the median HDI reduction showed a clear cascade based on parameter sensitivity: the reduction was greatest for the Main effects subset (-27.4%), smaller for the broader Total effects subset (-20.4%), and smallest when averaged over all 95 parameters (-16.3%). Using a significance level of $\alpha = 0.10$, a Kruskal-Wallis test confirms that this observed cascade is statistically significant ($p = 0.08$). This demonstrates that the algorithm intelligently allocates its constraining power to the parameters that most influence the assimilated observations. The strongest overall constraint was achieved, as expected, in the most focused experiment: when only the 29 Main effects parameters were optimized, their median HDI shrank by a remarkable 41.3%.

~~Table 5~~.Furthermore, the optimization produced posterior parameter ensembles that were effectively decorrelated (Table 4). This is a notable finding, as strong posterior correlations are a common challenge in data assimilation, often indicating that the available data are insufficient to constrain parameters independently. In contrast, our analysis reveals an absence of significant linear dependencies. Across all three optimization strategies, the maximum correlation coefficient observed between any two parameters was low (peaking at 0.34), while the median correlation was statistically indistinguishable from zero (0.032–0.04). These consistently low correlations indicate that the comprehensive, multi-variable BGC-Argo dataset provided sufficient orthogonal constraints to allow the framework to find a solution where parameters were constrained independently of each other.

4.6. Ensemble Evaluation Against Two Independent BGC-Argo Floats

To evaluate the accuracy and robustness of the optimized ensembles against independent observations, the three ensembles were used to simulate conditions along the trajectories of two BGC-Argo floats that were not involved in the optimization: one in the North Atlantic Subpolar Gyre (#6901485) and one in the oligotrophic Mediterranean Sea (#6901648) (Fig. 1). A visual comparison of the model performance for these floats is provided in the Supplement (Figs. S8–S25). Ensemble accuracy and reliability were assessed using the RCRV bias and dispersion metrics (Fig. 9).

The RCRV bias analysis shows that all three optimization strategies produce ensembles with low systematic error against these out-of-sample observations. For both validation floats, the median ensemble bias is small across most metrics, confirming that the optimized simulations accurately represent independent data (Fig. 9a, b). A Kruskal-Wallis H-test confirms that the bias distributions are statistically indistinguishable among the three strategies ($p = 0.99$). All ensembles do, however, produce two large biases when applied to the Mediterranean float. The first concerns the nitracline depth: the optimization, constrained by a North Atlantic float with a consistently shallow nitracline, yields biological rate parameters that incorrectly maintain this shallow structure in the oligotrophic Mediterranean, where the nitracline is much deeper. The second concerns the productive-layer nitrate concentration, and is a statistical artifact of the RCRV metric: in the oligotrophic Mediterranean, observed nitrate concentrations and their associated errors are extremely small, causing the normalisation to amplify a physically negligible model–data mismatch into a large bias score.

The RCRV dispersion metric assesses whether the ensemble uncertainty is realistically calibrated. A dispersion of 1.0 indicates a perfectly calibrated ensemble, where the predicted spread matches the actual model–data error; values below 1.0 indicate an over-dispersive (conservative) ensemble, while values above 1.0 indicate an under-dispersive (overconfident) one. For both validation floats, the median dispersion scores are consistent across all three strategies, approximately 0.3 for the North Atlantic float and 0.4 for the Mediterranean float (Fig. 9c, d), with no significant difference among strategies (Kruskal-Wallis $p = 0.99$). These values, being well below unity, indicate that the optimized ensembles provide a conservative estimate of the model's true uncertainty: the predicted spread is wider than the observed errors. This rules out the risk of overconfident calibration with artificially tight posteriors described by Hermans et al., (2022) and Yang and Zhu, (2018).

Although not perfectly calibrated, these results confirm that the assimilation has meaningfully constrained the model while maintaining a cautious uncertainty envelope. A conservative estimate of uncertainty is generally preferable to an overconfident one in operational forecasting applications, and the consistently low bias across both floats and all strategies demonstrates that the optimized parameter set produces accurate simulations in bioregions distinct from the training site.

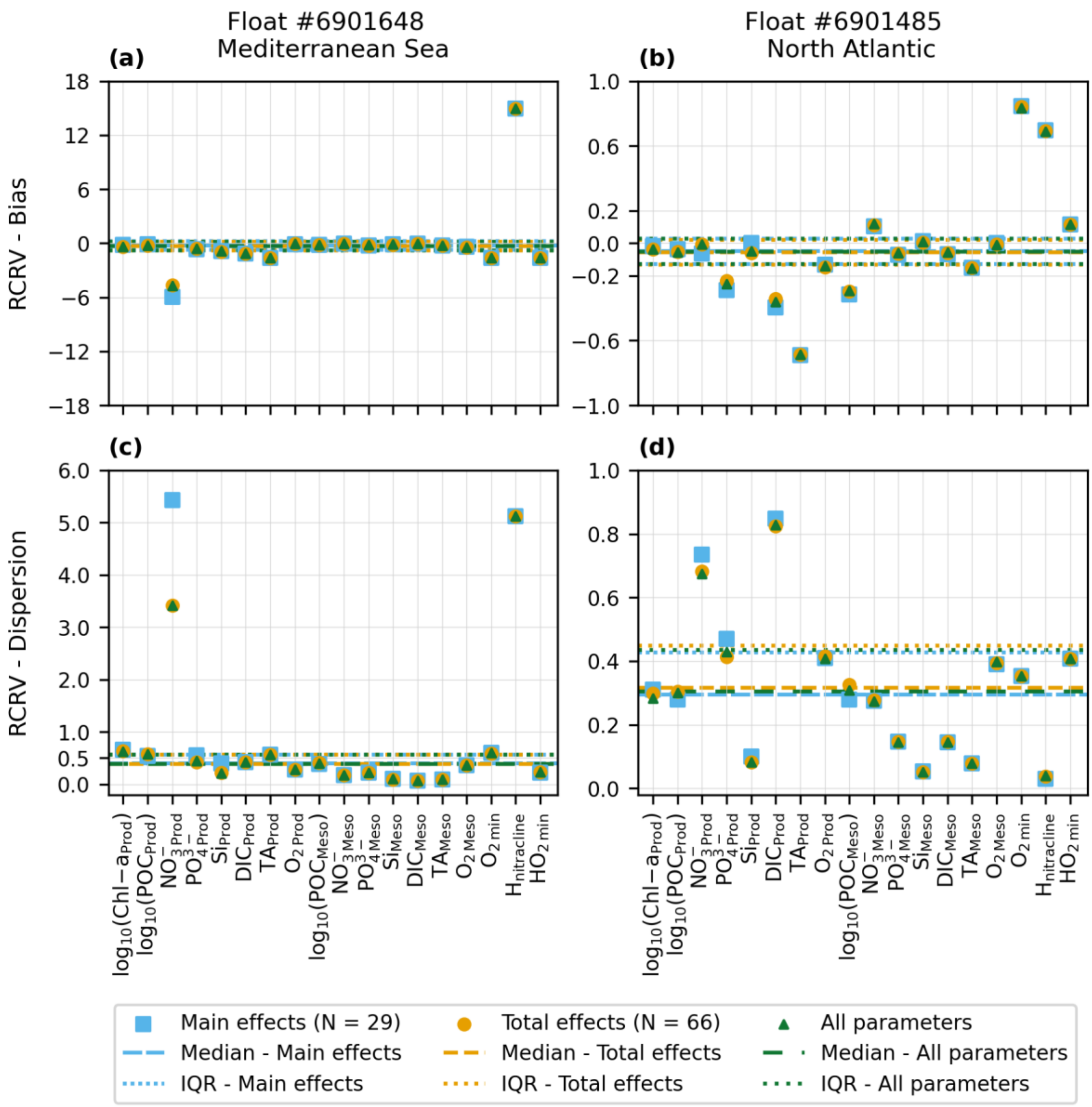


Figure 9. RCRV values across all assimilated metrics. Each column corresponds at one of the two test floats (#6901648 and #6901485). The bias is estimated as the mean RCRV, while the dispersion is quantified using the standard deviation of the RCRV values. The scatter plot shows results obtained using three different parameter selection strategies : 'Main effects' and 'Total effects' correspond to selections based on first-order and total-order Sobol indices, respectively, while 'All Parameters' refers to optimization using the full set of available parameter. The median value across metrics is indicated by the dashed lines. DCM observations did not allow for the reconstruction of the corresponding observational metrics. As a result, RCRV values could not be computed for the DCM depth and DCM intensity metrics.

Table ~~5~~ 4. Comparison of parameter distributions across optimization strategies. The three strategies are represented in the columns. The percentage reduction of the 67% HDI is calculated as the median percentage decrease of the HDI across the indicated

parameters. For each parameter, the HDI reduction is computed as the relative difference between the 67% HDI of the posterior distribution obtained after optimization and the 67% HDI of the initial parameter distribution. The maximum correlation is estimated among the parameter values obtained from the optimized ensemble. The median correlation is the median of the correlations between the parameter values obtained at the end of the optimization.

	Optimization Main effects parameters (N=29)	Optimization of Total effects parameters (N=66)	Optimization of All parameters
HDI reduction on Main effects parameters (%) (± IQR)	-41.3 (± 14.88)	-24.47 (± 11.35)	-27.36 (± 13.07)
HDI reduction on Total effects parameters (%) (± IQR)	-	-20.5 (± 9.18)	-20.4 (± 11.35)
HDI reduction on All parameters (%) (± IQR)	-	-	-16.3 (± 11.02)
Maximum correlation between optimized parameters	0.34	0.29	0.3
Median correlation between optimized parameters (± IQR)	0.04 (± 0.028)	0.033 (± 0.02)	0.032 (± 0.02)

~~Furthermore, the optimization produced posterior parameter ensembles that were effectively decorrelated (Table 5). This is a notable finding, as strong posterior correlations are a common challenge in data assimilation, often indicating that the available data are insufficient to constrain parameters independently. In contrast, our analysis reveals an absence of significant linear dependencies. Across all three optimization strategies, the maximum correlation coefficient observed between any two parameters was low (peaking at 0.34), while the median correlation was statistically indistinguishable from zero (0.032–0.04). These consistently low correlations indicate that the rich, multi-variable BGC-Argo dataset provided sufficient orthogonal constraints to allow the framework to find a solution where parameters were constrained independently of each other.~~

Table ~~6~~ 5. Comparison of the number of 1D simulations and the computational cost required for the parameter screening and optimization steps. The screening step for identifying Main Effects and Total Effects parameters requires the same number of 1D simulations, and therefore the same computational cost. The same parameter optimisation process is applied across all parameter sets. As a result, for the parameter optimisation step the number of simulations and the associated computational cost are identical for each method. The computational cost is expressed in terms of CPU hours required to generate the full ensemble of 1D simulations on a single CPU of an HPC system. Naturally, this value depends on the number of nodes and the specific configuration of the HPC used. The computational cost of other steps, such as metric evaluation, is negligible compared to that of generating the 1D simulation ensemble.

	Selection of parameters by Sobol indices	Parameters optimization
Number of 1D simulations	1 490 944	26 624
CPU-Hours (Hours)	932,4	24

4.7. ~~5~~ Transferability of the Optimized Parameters to the 3D Configuration

The 3D validation dataset comprises approximately 1,430 unique BGC-Argo profiles matched to the IBI model, over the period 2017–2019. After quality control and outlier filtering, this yields between 430,000 and 490,000 observation–model pairs per variable across the 0–500 m depth range. The observation distribution is strongly asymmetric: the Mediterranean basin concentrates 85–90% of the matched pairs, while the North Atlantic (Northern + Iberian Atlantic) contributes the remaining 10–15%, reflecting the denser BGC-Argo float deployment in the Mediterranean during this period.

Table 6. Basin-level NRMSE for the OPTI and REF simulations and relative improvement Δ%NRMSE. Bold entries indicate NRMSE < 1 (performance within observation uncertainty). When both simulations are in bold, a positive Δ%NRMSE does not constitute a meaningful degradation.

Variable	Basin	NRMSE_OPTI	NRMSE_REF	Δ%NRMSE
O_2	Northern	1.359	1.335	1.8%
O_2	Iberian Atlantic	**0.763**	**0.754**	1.2%
O_2	Mediterranean	2.154	2.578	-16.4%
NO_3^-	Northern	1.994	2.274	-12.3%
NO_3^-	Iberian Atlantic	1.781	1.901	-6.3%
NO_3^-	Mediterranean	5.375	5.109	5.2%
PO_4^{3-}	Northern	1.233	1.438	-14.3%
PO_4^{3-}	Iberian Atlantic	1.545	1.612	-4.1%
PO_4^{3-}	Mediterranean	3.969	4.337	-8.5%
Si	Northern	**0.369**	**0.395**	-6.6%
Si	Iberian Atlantic	**0.294**	**0.275**	6.8%
Si	Mediterranean	3.828	3.194	19.8%
DIC	Northern	1.929	2.230	-13.5%
DIC	Iberian Atlantic	1.385	1.370	1.1%
DIC	Mediterranean	8.202	8.265	-0.8%
TA	Northern	**0.489**	1.200	-59.3%
TA	Iberian Atlantic	1.012	1.125	-10.1%
TA	Mediterranean	9.361	10.020	-6.6%

The OPTI simulation generally outperforms REF across the IBI domain (Table 6). Of the 18 variable–basin combinations, 12 show a negative Δ%NRMSE, indicating that the optimized parameters improve model skill. The largest improvements occur for total alkalinity in the Northern basin (Δ%NRMSE = −59.3%), dissolved oxygen in the Mediterranean (−16.4%), phosphate in the Northern basin (−14.3%), and TCO_2 in the Northern basin (−13.5%). Phosphate is the only variable for which OPTI outperforms REF in every basin, with Δ%NRMSE ranging from −4.1% to −14.3%.

The Northern and Iberian Atlantic basins are the regions where the parameter optimization brings the clearest benefits. Nitrate, phosphate, and total alkalinity show consistent improvements in both basins, with Δ%NRMSE ranging from −4.1% (PO_4, Iberian Atlantic) to −59.3% (TALK, Northern basin). TCO_2 improves markedly in the Northern basin (−13.5%) but remains essentially unchanged in the Iberian Atlantic (+1.1%). These improvements are genuine because, for these variables, NRMSE values exceed unity in at least one simulation, confirming that the optimization measurably reduces model errors that are detectable relative to observation uncertainty. Dissolved oxygen is the only variable where OPTI slightly underperforms REF in the Northern basin (+1.8%), though both simulations exceed unity only marginally. Silicate stands apart: both simulations achieve NRMSE well below unity (0.27–0.40) across both basins, indicating that the model reproduces CANYON-B-derived silicate observations within less than half the associated uncertainty regardless of the parameter set.

The Mediterranean presents a more contrasted picture. All NRMSE values exceed unity — often substantially — indicating that model errors consistently surpass observation uncertainty in this basin. Dissolved oxygen and phosphate show genuine improvements (Δ%NRMSE = −16.4% and −8.5%, respectively), confirming that the optimized parameters partially correct the free-running model even in a biogeochemical regime different from the training site. Nitrate and silicate, however, degrade under OPTI (+5.2% and +19.8%), suggesting that the biological rate parameters tuned on a single North Atlantic float do not fully transfer to the oligotrophic Mediterranean, consistent with the nitracline bias identified in the 1D independent validation (Sect. 4.6). The carbonate system variables exhibit very high NRMSE values (8–10 for both TCO_2 and TALK), reflecting the fact that the CONTENT-derived error estimates are much smaller than the actual model–observation discrepancies; within this context, OPTI provides only marginal improvements (-0.8% and -6.6%, respectively).

Surface chlorophyll-a provides a complementary, independent assessment of the optimized parameters (Fig. 10). The Taylor diagram confirms that both simulations reproduce the MOBTAC-observed seasonal cycle across all five boxes with correlations typically exceeding 0.8–0.9. The OPTI simulation systematically outperforms REF: for every box, the green markers sit closer to the MOBTAC reference point than the corresponding red markers, reflecting both a reduced centred

RMSE and a normalised standard deviation brought closer to unity. The improvement is most visible in the Atlantic box 1 where REF tends to underestimate the amplitude of the seasonal cycle — a bias that the optimized parameters correct.

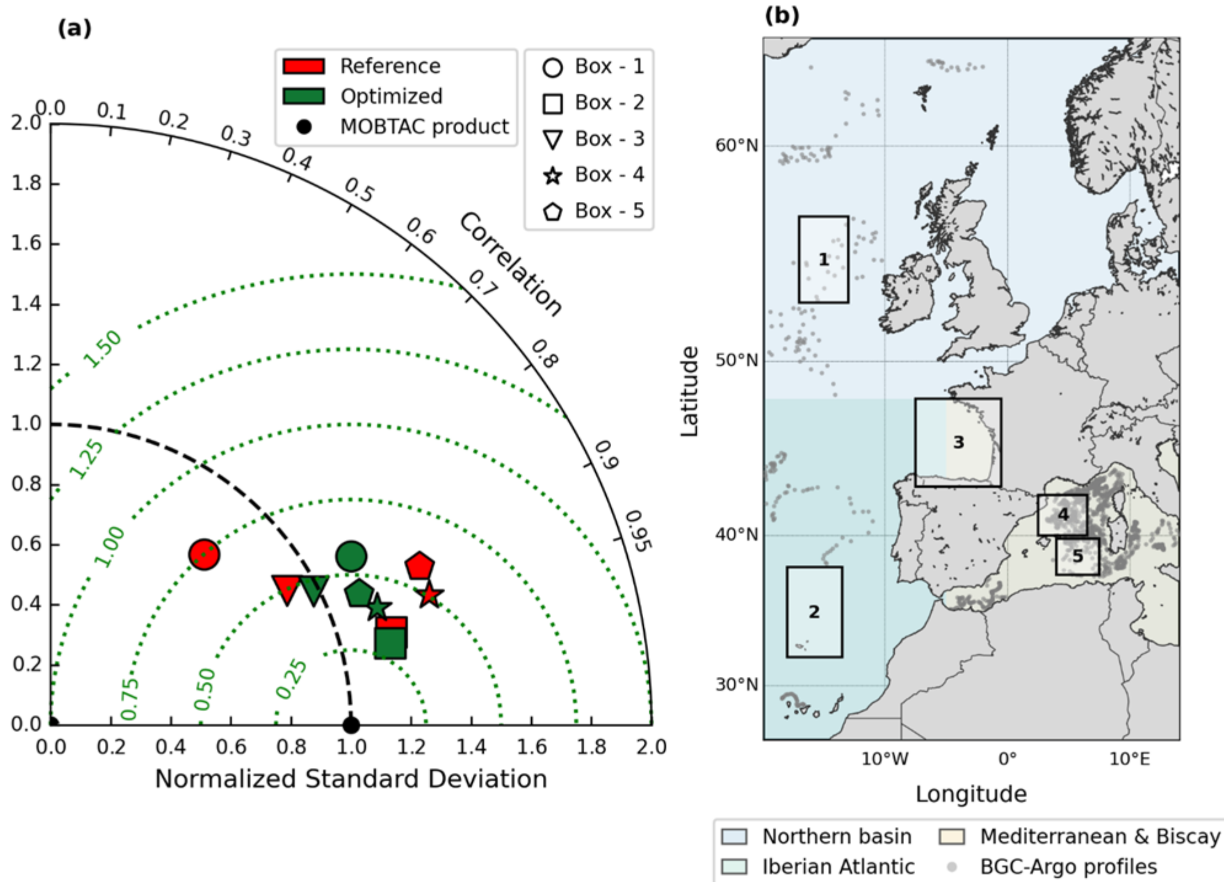


Figure 10. Surface chlorophyll-a validation against the MOBTAC multi-observation reprocessed product over the period 2017–2019. (a) Taylor diagram computed on monthly mean time series in $\log_{10}$ space. The radial distance from the origin represents the normalised standard deviation ($\sigma_{model}/\sigma_{obs}$), the angular position the Pearson correlation coefficient, and the green dotted contours the centred root-mean-square difference. Red markers denote the REF simulation and green markers the OPTI simulation; marker shapes identify the five validation boxes (circle: Box 1; square: Box 2; inverted triangle: Box 3; star: Box 4; pentagon: Box 5). The black filled circle indicates the MOBTAC reference point at unit normalised standard deviation, and the black dashed semicircle marks the locus where model and observed variability are equal ($\sigma_{model}/\sigma_{obs} = 1$). (b) Location of the five validation boxes within the IBI domain: Box 1 (Celtic Sea / Rockall Trough), Box 2 (Canary / Madeira), Box 3 (Bay of Biscay), Box 4 (Gulf of Lion), and Box 5 (Balearic Sea). The three broader sub-regions for subsurface BGC-Argo validation

are delineated by colored polygons: Northern basin (20°W-17°E, 48-65°N), Iberian Atlantic (20-5°W, 26-48°N), and Mediterranean & Bay of Biscay (5°W-17°E, 26-48°N). Individual profile locations from the approximately 1,430 BGC-Argo profiles matched to the IBI model over 2017-2019 are shown as grey dots.

5. Discussion

Our study demonstrates that a comprehensive suite of BGC-Argo observations can be used to robustly constrain all 95 parameters of the PISCES biogeochemical model. The iterative Importance Sampling (iIS) framework successfully reduced the productive-layer model-data misfit by over 50~~%,~~ % relative to the reference simulation with default parameters, yielded ~~portable~~ parameter sets with significantly reduced uncertainty relative to the broad uniform prior distributions, and, crucially, produced posterior parameter distributions with negligible inter-correlation. A key finding of this study is that directly optimizing all 95 model parameters is just as effective as targeting smaller, GSA-informed subsets, with both approaches yielding statistically indistinguishable improvements in ~~model skill.~~goodness of fit. Furthermore, all strategies produced ~~portable parameter sets~~ ensembles with low bias and reliable, albeit cautious, uncertainty envelopes when evaluated against independent BGC-Argo observations. Given that the 'All-parameters' strategy achieves these successful outcomes while avoiding the immense computational cost of the prerequisite GSA and providing a more robust quantification of uncertainty in unassimilated variables, we conclude that it offers the most practical and effective balance between ~~model skill~~ goodness of fit, computational cost, and uncertainty quantification for this type of calibration problem (Table ~~6~~ 5).

In complex biogeochemical models, parameter correlations, identifiability, and overfitting are closely linked. When two or more parameters produce compensating effects on the same model output, their posteriors become correlated, meaning they cannot be estimated independently from the available data — the parameters are said to be unidentifiable. Unidentifiable parameters are a direct pathway to overfitting: the optimisation may find parameter combinations that reproduce the training data well, but for the wrong mechanistic reasons, leading to poor performance on independent observations. Breaking these correlations therefore requires observations that respond differently to each parameter, so that the data can distinguish their individual contributions. This is the rationale for assimilating a comprehensive, multi-variable dataset spanning multiple biogeochemical tracers.

A central achievement of this framework is the advance it represents in addressing parameter equifinality, a longstanding challenge in biogeochemical data assimilation. Historically, studies using sparse datasets have often found strong posterior correlations between parameters, indicating that the available data were insufficient to constrain them independently ~~(Matear, 1995; Fennel et al., 2001; Mamnun et al., 2022).~~ (Fennel et al., 2001; Hurtt and Armstrong, 1996; Mamnun, 2022; Matear, 1995). Our results, however, suggest that this issue was not an inescapable property of the models, but rather a symptom of being under-constrained. The ~~rich~~ comprehensive, multi-variable BGC-Argo dataset provides a diverse and powerful set of orthogonal constraints. Simple trade-offs that cause parameter correlations in data-poor scenarios are no longer possible when twenty distinct metrics are assimilated simultaneously. It is important to note that these twenty metrics do not define separate objective functions optimised via multi-objective methods (Sauerland et al., 2019); rather, they enter a single product likelihood (Eq. S14), so that each ensemble member is evaluated against all constraints at once. For example, while increasing the P-I slope or decreasing phytoplankton mortality might both fit Chl-a data, these changes leave different imprints on nitrate drawdown, oxygen concentrations, etc~~.~~... By leveraging the distinct sensitivities of these multiple tracers, the framework can untangle parameter effects. As a result, our analysis revealed posterior parameter distributions with negligible inter-correlation (median r < 0.04; Table ~~5~~ 4). This does not mean that equifinality has been eliminated; rather, it has changed in character. We still find that a distribution of parameter sets performs well, but crucially, the parameters within this posterior ensemble are not structurally dependent on each other. The framework has thus advanced the state of the problem from one of correlated equifinality to one of uncorrelated equifinality, a critical step towards robust parameter estimation in complex ecosystem models.

~~Despite its successes, this study's primary limitation is its reliance on a 1D vertical model that neglects horizontal advection. The framework produced a validated, regionally tuned parameter set that significantly improves 1D NEMO-PISCES simulations in the North Atlantic. However, the critical next step is to evaluate whether these skill improvements translate to a three-dimensional context. Implementing the optimized parameter set in a fully regional 3D NEMO-PISCES simulation will be essential to test its performance in a more complex and realistic environment.~~

~~The portability tests confirmed that the optimized parameters were robust and not over-fitted, but also revealed their limitations. While the North Atlantic-tuned set performed well when transferred to a different site within the same basin, it failed to reproduce the deep nitracline of the oligotrophic Mediterranean Sea. This finding highlights the regional specificity of certain parameters, a concept supported by studies showing significant parameter variation across different oceanic biomes (Singh et al., 2025). This work therefore paves the way for moving beyond a single global parameterization towards a mosaic of regionally-optimized parameter sets. Systematically applying this framework to the growing global fleet of BGC-Argo floats could enable the creation of maps of PISCES parameters, revealing how parameters vary across the world's oceans. While this would introduce new challenges, such as defining emergent 'parameter bioregions', it represents a significant step towards developing more accurate and regionally-calibrated global biogeochemical models.~~

A recurrent concern in biogeochemical parameter optimization is whether parameters estimated in simplified 1D configurations merely compensate for missing physical processes rather than genuinely improving the biogeochemical formulation (Löptien and Dietze, 2019; Pasquier et al., 2023). The 3D validation presented in Sect. 4.7 directly addresses this question. Implementing the 1D-optimized parameter set from the All paramters strategy in the high-resolution IBI model — which resolves 3D physical processes — shows that skill improvements persist: the OPTI simulation outperforms REF in 12 of 18 variable–basin combinations, with particularly large gains for total alkalinity, nitrate, and phosphate in the North Atlantic.

Surface Chl-a is also better reproduced across the entire domain. These results demonstrate that the 1D optimization has captured genuine improvements in the biogeochemical parameterisation that transfer to a fully three-dimensional, dynamically resolved configuration.

The 3D validation does, however, reveal the regional specificity of the optimized parameter set. In the Mediterranean, OPTI degrades nitrate (+5.2%) and silicate (+19.8%) relative to REF, consistent with the nitracline bias already identified in the 1D independent validation against the Mediterranean float (Sect. 4.6). The North Atlantic-tuned biological rate parameters maintain a shallow nitracline structure that is inappropriate for the oligotrophic Mediterranean. This finding aligns with studies showing significant parameter variation across oceanic biomes (Singh et al., 2025) and motivates a move beyond a single global parameterisation towards regionally-optimized parameter sets. Systematically applying this framework to the growing global fleet of BGC-Argo floats could enable the construction of parameter maps, revealing how PISCES parameters vary across the world's oceans and defining emergent "parameter bioregions."

The use of a 1D vertical model for the optimization presents both advantages and limitations that should be weighed explicitly. The primary advantage is computational efficiency: the 1D configuration reduces the cost of each ensemble member by several orders of magnitude relative to a 3D run, making it feasible to evaluate over 26,000 parameter combinations and to perform a rigorous global sensitivity analysis. The 1D setup also isolates the biogeochemical response from uncertainties in the representation of ocean dynamics, providing a controlled environment in which parameter effects can be cleanly attributed. The optimization is further conditioned on a single training site and a one-year assimilation window, which may not capture the full range of environmental variability encountered in multi-year, basin-scale applications. The fact that the 3D validation confirms the transferability of the optimized parameters largely mitigates the first concern, but extending the assimilation window to multiple years and multiple sites would further strengthen confidence in the generality of the results.

6. Conclusion

We developed and applied a parameter optimization framework that leveraged comprehensive BGC-Argo data from a North Atlantic float in order to constrain the 95 parameters of the PISCES biogeochemical model. This framework was highly effective, reducing the productive-layer model-data misfit by over 50% and yielding decorrelated posterior parameter distributions. This significant advance shifts the long-standing challenge of correlated equifinality to uncorrelated equifinality, where a range of optimal parameter sets can be found independently.

A central finding is that directly optimizing all parameters is the recommended strategy. This approach achieves statistically indistinguishable skill improvements compared to targeting smaller, GSA-informed subsets, without incurring the immense computational cost of a prerequisite GSA. By exploring the full parameter space, this approach provides a more robust quantification of uncertainty in unassimilated variables than optimizing smaller, GSA-informed subsets.

~~Finally, our portability tests show that optimized parameter sets in the North Atlantic are not universally applicable, highlighting the regional specificity of BGC parameters. This work provides a scalable pathway to move beyond a single, compromised global set of parameters. Applying this framework to the international BGC-Argo fleet will enable the creation of a map of regionally tuned parameter sets. This is a crucial step toward a new generation of more accurate, regionally calibrated global biogeochemical models.~~

Crucially, the optimized parameters improve not only the 1D simulations on which they were trained, but also the fully three-dimensional, high-resolution IBI simulation against independent observations, at least, off the continental shelf. The 3D validation against approximately 1430 BGC-Argo vertical profiles and the MOBTAC surface Chl-a product confirms that OPTI outperforms REF in the majority of variable–basin combinations. This demonstrates that a computationally inexpensive 1D optimization can deliver parameter sets that genuinely improve operational 3D biogeochemical forecasting systems.

Two avenues remain to be explored. The generality of the method should be tested by applying it to additional BGC-Argo floats in other oceanic regions, and the resulting parameter sets should be evaluated to determine whether they further improve model performance across biogeochemically distinct basins. A further priority is to enrich the observation vector with zooplankton constraints, given that our GSA identifies zooplankton-related parameters as the dominant source of model sensitivity. The recent development of the Underwater Vision Profiler 6 (UVP6; Picheral et al., 2022), a miniaturised imaging sensor that can be integrated on BGC-Argo floats, may in the future provide zooplankton abundance profiles from the same autonomous platforms, offering a direct pathway to constrain the grazing parameters.

Author contibution

QH, AM, GR, EG conceived the study. QH carried out, sensitivity analysis, optimization experiments with the support of GR, model simulations with the support of EG, and analyses with the support of AM, EG, GR, HC, FD. CP and GS assisted with setup and validation of the ‘3D-Free’ model. OA assisted by providing constraints on parameter values. AM assisted with processing of the BGC-Argo float data. RS assisted in the use of the CANYON-B and CONTENT neural network. AM, QH, EG and GR discussed the results and wrote the paper with contributions from the coauthors.

Supplement link

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors used artificial intelligence tools (ChatGPT) to assist with English language editing of the manuscript

Financial support

This work was supported by the REFINE project. REFINE has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N° 834177).

This work was supported by the Argo-2030 project. Argo-2030 has received the support of the French government within the framework of the "Investissements d'avenir" program integrated in France 2030 and managed by the Agence Nationale de la Recherche (ANR) under the reference "ANR-21-ESRE-0019".

References

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

Bendtsen, J., Vives, C. R., and Richardson, K.: Primary production in the North Atlantic estimated from in situ water column data observed by Argo floats and remote sensing, Front. Mar. Sci., 10, https://doi.org/10.3389/fmars.2023.1062413, 2023.

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian Neural Networks, Front. Mar. Sci., 5, https://doi.org/10.3389/fmars.2018.00328, 2018.

Bock, N., Cornec, M., Claustre, H., and Duhamel, S.: Biogeographical Classification of the Global Ocean From BGC-Argo Floats, Global Biogeochemical Cycles, 36, e2021GB007233, https://doi.org/10.1029/2021GB007233, 2022.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, Journal of Geophysical Research: Oceans, 109, https://doi.org/10.1029/2004JC002378, 2004.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., Garçon, V., Gilbert, D., Gutiérrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., Telszewski, M., Yasuhara, M., and Zhang, J.: Declining oxygen in the global ocean and coastal waters, Science, 359, https://doi.org/10.1126/science.aam7240, 2018.

Brodeau, L., Barnier, B., Treguier, A.-M., Penduff, T., and Gulev, S.: An ERA40-based atmospheric forcing for global ocean circulation models, Ocean Modelling, 31, 88–104, https://doi.org/10.1016/j.ocemod.2009.10.005, 2010.

Campolongo, F., Saltelli, A., and Tarantola, S.: Sensitivity Anaysis as an Ingredient of Modeling, Statistical Science, 15, 377–395, https://doi.org/10.1214/ss/1009213004, 2000.

Cermeño, P., Dutkiewicz, S., Harris, R. P., Follows, M., Schofield, O., and Falkowski, P. G.: The role of nutricline depth in regulating the ocean carbon cycle, Proceedings of the National Academy of Sciences, 105, 20344–20349, https://doi.org/10.1073/pnas.0811302106, 2008.

Chu, P. C., Ivanov, L. M., and Margolina, T. M.: On non-linear sensitivity of marine biological models to parameter variations, Ecological Modelling, 206, 369–382, 2007.

Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the Global Ocean with Biogeochemical-Argo, Annual Review of Marine Science, 12, 23–48, https://doi.org/10.1146/annurev-marine-010419-010956, 2020.

Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., D'Ortenzio, F., Gentili, B., and Schmechtig, C.: Deep Chlorophyll Maxima in the Global Ocean: Occurrences, Drivers and Characteristics, Global Biogeochemical Cycles, 35, e2020GB006759, https://doi.org/10.1029/2020GB006759, 2021.

Denman, K. L.: Modelling planktonic ecosystems: parameterizing complexity, Progress in Oceanography, 57, 429–452, https://doi.org/10.1016/S0079-6611(03)00109-5, 2003.

DeVries, T., Liang, J.-H., and Deutsch, C.: A mechanistic particle flux model applied to the oceanic phosphorus cycle, Biogeosciences, 11, 5381–5398, https://doi.org/10.5194/bg-11-5381-2014, 2014.

Doléac, S., Lévy, M., El Hourany, R., and Bopp, L.: Toward more robust net primary production projections in the North Atlantic Ocean, Biogeosciences, 22, 841–862, https://doi.org/10.5194/bg-22-841-2025, 2025.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO2 Problem, Annual Review of Marine Science, 1, 169–192, https://doi.org/10.1146/annurev.marine.010908.163834, 2009.

Dowd, M., Jones, E., and Parslow, J.: A statistical overview and perspectives on data assimilation for marine biogeochemical models, Environmetrics, 25, 203–213, https://doi.org/10.1002/env.2264, 2014.

Dulvy, N. K., Pacoureau, N., Rigby, C. L., Pollom, R. A., Jabado, R. W., Ebert, D. A., Finucci, B., Pollock, C. M., Cheok, J., Derrick, D. H., Herman, K. B., Sherman, C. S., VanderWright, W. J., Lawson, J. M., Walls, R. H. L., Carlson, J. K., Charvet, P., Bineesh, K. K., Fernando, D., Ralph, G. M., Matsushiba, J. H., Hilton-Taylor, C., Fordham, S. V., and Simpfendorfer, C. A.: Overfishing drives over one-third of all sharks and rays toward a global extinction crisis, Current Biology, 31, 4773-4787.e8, https://doi.org/10.1016/j.cub.2021.08.062, 2021.

Elvira, V. and Martino, L.: Advances in Importance Sampling, in: Wiley StatsRef: Statistics Reference Online, John Wiley & Sons, Ltd, 1–14, https://doi.org/10.1002/9781118445112.stat08284, 2021.

Fennel, K., Losch, M., Schröter, J., and Wenzel, M.: Testing a marine ecosystem model: sensitivity analysis and parameter optimization, Journal of Marine Systems, 28, 45–63, https://doi.org/10.1016/S0924-7963(00)00083-X, 2001.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A. M., Gregoire, M., Jones, E., Kim, H.-C., Lamouroux, J., Murtugudde, R., Perruche, C., and the GODAE OceanView Marine Ecosystem Analysis and Prediction Task Team: Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health, Front. Mar. Sci., 6, 89, https://doi.org/10.3389/fmars.2019.00089, 2019.

Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L.: Ocean biogeochemical modelling, Nat Rev Methods Primers, 2, 76, https://doi.org/10.1038/s43586-022-00154-2, 2022.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., et al.: Global Carbon Budget 2023, Earth System Science Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

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 Jr., D. J., Moore, J. K., Schartau, M., Spitz, Y. H., and Wiggert, J. D.: Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups, Journal of Geophysical Research: Oceans, 112, https://doi.org/10.1029/2006JC003852, 2007.

Galí, M., Falls, M., Claustre, H., Aumont, O., and Bernardello, R.: Bridging the gaps between particulate backscattering measurements and modeled particulate organic carbon in the ocean, Biogeosciences, 19, 1245–1275, https://doi.org/10.5194/bg-19-1245-2022, 2022.

Germineaud, C., Brankart, J.-M., and Brasseur, P.: An Ensemble-Based Probabilistic Score Approach to Compare Observation Scenarios: An Application to Biogeochemical-Argo Deployments, https://doi.org/10.1175/JTECH-D-19-0002.1, 2019.

Gillooly, J. F., Charnov, E. L., West, G. B., Savage, V. M., and Brown, J. H.: Effects of size and temperature on developmental time, Nature, 417, 70–73, https://doi.org/10.1038/417070a, 2002.

Henson, S. A., Laufkötter, C., Leung, S., Giering, S. L. C., Palevsky, H. I., and Cavan, E. L.: Uncertain response of ocean biological carbon export in a changing world, Nat. Geosci., 15, 248–254, https://doi.org/10.1038/s41561-022-00927-0, 2022.

Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., Begy, V., and Louppe, G.: A Crisis In Simulation-Based Inference? Beware, Your Posterior Approximations Can Be Unfaithful, Transactions on Machine Learning Research, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hieronymus, J., Hieronymus, M., Gröger, M., Schwinger, J., Bernadello, R., Tourigny, E., Sicardi, V., Ruvalcaba Baroni, I., and Wyser, K.: Net primary production annual maxima in the North Atlantic projected to shift in the 21st century, Biogeosciences, 21, 2189–2206, https://doi.org/10.5194/bg-21-2189-2024, 2024.

Hoegh-Guldberg, O., Poloczanska, E. S., Skirving, W., and Dove, S.: Coral Reef Ecosystems under Climate Change and Ocean Acidification, Front. Mar. Sci., 4, https://doi.org/10.3389/fmars.2017.00158, 2017.

Homma, T. and Saltelli, A.: Importance measures in global sensitivity analysis of nonlinear models, Reliability Engineering & System Safety, 52, 1–17, https://doi.org/10.1016/0951-8320(96)00002-6, 1996.

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 Science, 14, 371–386, https://doi.org/10.5194/os-14-371-2018, 2018.

Huntley, M. E.: Temperature-Dependent Production of Marine Copepods: A Global Synthesis, The American Naturalist, 140, 201–242, https://doi.org/10.1086/285410, 1992.

Hurtt, G. C. and Armstrong, R. A.: A pelagic ecosystem model calibrated with BATS data, Deep Sea Research Part II: Topical Studies in Oceanography, 43, 653–683, https://doi.org/10.1016/0967-0645(96)00007-0, 1996.

Issan, O., Riley, P., Camporeale, E., and Kramer, B.: Bayesian Inference and Global Sensitivity Analysis for Ambient Solar Wind Prediction, Space Weather, 21, e2023SW003555, https://doi.org/10.1029/2023SW003555, 2023.

Iwanaga, T., Usher, W., and Herman, J.: Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses, Socio-Environmental Systems Modelling, 4, 18155–18155, https://doi.org/10.18174/sesmo.18155, 2022.

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D., Williams, N. L., Boss, E., Haëntjens, N., Talley, L. D., and Sarmiento, J. L.: Biogeochemical sensor performance in the SOCCOM profiling float array, Journal of Geophysical Research: Oceans, 122, 6416–6436, https://doi.org/10.1002/2017JC012838, 2017.

Johnson, K. s, Plant, J. n, Maurer, T. l, Takeshita, Y., Maurer, T. l, and Takeshita, Y.: Processing BGC-Argo pH data at the DAC level, https://doi.org/10.13155/57195, 2025.

Keeling, R. E., Körtzinger, A., and Gruber, N.: Ocean deoxygenation in a warming world, Ann Rev Mar Sci, 2, 199–229, https://doi.org/10.1146/annurev.marine.010908.163855, 2010.

Lacroix, G. and Grégoire, M.: Revisited ecosystem model (MODECOGeL) of the Ligurian Sea: seasonal and interannual variability due to atmospheric forcing, Journal of Marine Systems, 37, 229–258, https://doi.org/10.1016/S0924-7963(02)00190-2, 2002.

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S.: Particle filters for high-dimensional geoscience applications: A review, Quarterly Journal of the Royal Meteorological Society, 145, 2335–2365, https://doi.org/10.1002/qj.3551, 2019.

Limburg, K. E., Breitburg, D., Swaney, D. P., and Jacinto, G.: Ocean Deoxygenation: A Primer, One Earth, 2, 24–29, https://doi.org/10.1016/j.oneear.2020.01.001, 2020.

Liu, T., Qiu, Y., Lin, X., Ni, X., Wang, L., Li, H., and Jing, C.: Dissolved Oxygen Recovery in the Oxygen Minimum Zone of the Arabian Sea in Recent Decade as Observed by BGC-Argo Floats, Geophysical Research Letters, 51, e2024GL108841, https://doi.org/10.1029/2024GL108841, 2024.

Löptien, U., Dietze, H., Preuss, R., and Toussaint, U. V.: Mapping manifestations of parametric uncertainty in projected pelagic oxygen concentrations back to contemporary local model fidelity, Sci Rep, 11, 20949, https://doi.org/10.1038/s41598-021-00334-2, 2021.

Lucazeau, F.: Analysis and Mapping of an Updated Terrestrial Heat Flow Data Set, Geochemistry, Geophysics, Geosystems, 20, 4001–4024, https://doi.org/10.1029/2019GC008389, 2019.

MacLeod, M., Arp, H. P. H., Tekman, M. B., and Jahnke, A.: The global threat from plastic pollution, Science, 373, 61–65, https://doi.org/10.1126/science.abg5433, 2021.

Madec, G., Bell, M., Blaker, A., Bricaud, C., Bruciaferri, D., Castrillo, M., Calvert, D., Chanut, J., Clementi, E., Coward, A., Epicoco, I., Éthé, C., Ganderton, J., Harle, J., Hutchinson, K., Iovino, D., Lea, D., Lovato, T., Martin, M., Martin, N., Mele, F., Martins, D., Masson, S., Mathiot, P., Mele, F., Mocavero, S., Müller, S., Nurser, A. J. G., Paronuzzi, S., Peltier, M., Person, R., Rousset, C., Rynders, S., Samson, G., Téchené, S., Vancoppenolle, M., and Wilson, C.: NEMO Ocean Engine Reference Manual, https://doi.org/10.5281/zenodo.8167700, 2023.

Mamnun, N.: Uncertainties in ocean biogeochemical simulations: Application of ensemble data assimilation to a one-dimensional model, Frontiers in Marine Science, 2022.

Mamnun, N., Völker, C., Vrekoussis, M., and Nerger, L.: Uncertainties in ocean biogeochemical simulations: Application of ensemble data assimilation to a one-dimensional model, Front. Mar. Sci., 9, https://doi.org/10.3389/fmars.2022.984236, 2022.

Martino, L., Elvira, V., and Louzada, F.: Effective sample size for importance sampling based on discrepancy measures, Signal Processing, 131, 386–401, https://doi.org/10.1016/j.sigpro.2016.08.025, 2017.

Matear, R.: Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P, Journal of Marine Research, 53, 1995.

Mattern, J. P., Dowd, M., and Fennel, K.: Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations, Journal of Geophysical Research: Oceans, 118, 2746–2760, https://doi.org/10.1002/jgrc.20213, 2013.

Mayot, N., Buitenhuis, E. T., Wright, R. M., Hauck, J., Bakker, D. C. E., and Le Quéré, C.: Constraining the trend in the ocean CO2 sink during 2000–2022, Nat Commun, 15, 8429, https://doi.org/10.1038/s41467-024-52641-7, 2024.

Mignot, A., Claustre, H., Uitz, J., Poteau, A., D'Ortenzio, F., and Xing, X.: Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation, Global Biogeochemical Cycles, 28, 856–876, https://doi.org/10.1002/2013GB004781, 2014.

Mignot, A., Ferrari, R., and Claustre, H.: Floats with bio-optical sensors reveal what processes trigger the North Atlantic bloom, Nat Commun, 9, 190, https://doi.org/10.1038/s41467-017-02143-6, 2018.

Mignot, A., D'Ortenzio, F., Taillandier, V., Cossarini, G., and Salon, S.: Quantifying Observational Errors in Biogeochemical-Argo Oxygen, Nitrate, and Chlorophyll a Concentrations, Geophysical Research Letters, 46, 4330–4337, https://doi.org/10.1029/2018GL080541, 2019.

Mignot, A., Claustre, H., Cossarini, G., D'Ortenzio, F., Gutknecht, E., Lamouroux, J., Lazzari, P., Perruche, C., Salon, S., Sauzède, R., Taillandier, V., and Teruzzi, A.: Using machine learning and Biogeochemical-Argo (BGC-Argo) floats to assess biogeochemical models and optimize observing system design, Biogeosciences, 20, 1405–1422, https://doi.org/10.5194/bg-20-1405-2023, 2023.

Minamide, M. and Zhang, F.: Adaptive Observation Error Inflation for Assimilating All-Sky Satellite Radiance, https://doi.org/10.1175/MWR-D-16-0257.1, 2017.

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., and Franz, B. A.: Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach, Remote Sensing of Environment, 111, 69–88, https://doi.org/10.1016/j.rse.2007.03.012, 2007.

Ohishi, S., Miyoshi, T., and Kachi, M.: An ensemble Kalman filter-based ocean data assimilation system improved by adaptive observation error inflation (AOEI), Geoscientific Model Development, 15, 9057–9073, https://doi.org/10.5194/gmd-15-9057-2022, 2022.

Olsen, A., Key, R. M., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F. F., and Suzuki, T.: The Global Ocean Data Analysis Project version 2 (GLODAPv2) – an internally consistent data product for the world ocean, Earth System Science Data, 8, 297–323, https://doi.org/10.5194/essd-8-297-2016, 2016.

Orr, J. C.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, Nature, 2005.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, Nature, 437, 681–686, https://doi.org/10.1038/nature04095, 2005.

Owen, A. and Zhou, Y.: Safe and Effective Importance Sampling, Journal of the American Statistical Association, 95, 135–143, https://doi.org/10.2307/2669533, 2000.

Pauly, D., Christensen, V., Guénette, S., Pitcher, T. J., Sumaila, U. R., Walters, C. J., Watson, R., and Zeller, D.: Towards sustainability in world fisheries, Nature, 418, 689–695, https://doi.org/10.1038/nature01017, 2002.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Prieur, C., Viry, L., Blayo, E., and Brankart, J.-M.: A global sensitivity analysis approach for marine biogeochemical modeling, Ocean Modelling, 139, 101402, https://doi.org/10.1016/j.ocemod.2019.101402, 2019.

Raices Cruz, I., Lindström, J., Troffaes, M. C. M., and Sahlin, U.: Iterative importance sampling with Markov chain Monte Carlo sampling in robust Bayesian analysis, Computational Statistics & Data Analysis, 176, 107558, https://doi.org/10.1016/j.csda.2022.107558, 2022.

Reffray, G., Bourdalle-Badie, R., and Calone, C.: Modelling turbulent vertical mixing sensitivity using a 1-D version of NEMO, Geoscientific Model Development, 8, 69–86, https://doi.org/10.5194/gmd-8-69-2015, 2015.

Renardy, M., Joslyn, L. R., Millar, J. A., and Kirschner, D. E.: To Sobol or not to Sobol? The effects of sampling schemes in systems biology applications, Math Biosci, 337, 108593, https://doi.org/10.1016/j.mbs.2021.108593, 2021.

Ristic, B., Arulampalam, S., and Gordon, N.: Beyond the Kalman Filter: Particle Filters for Tracking Applications, Artech House, 299 pp., 2004.

Rodgers, K. B., Schwinger, J., Fassbender, A. J., Landschützer, P., Yamaguchi, R., Frenzel, H., Stein, K., Müller, J. D., Goris, N., Sharma, S., Bushinsky, S., Chau, T.-T.-T., Gehlen, M., Gallego, M. A., Gloege, L., Gregor, L., Gruber, N., Hauck, J., Iida, Y., Ishii, M., Keppler, L., Kim, J.-E., Schlunegger, S., Tjiputra, J., Toyama, K., Vaittinada Ayar, P., and Velo, A.: Seasonal Variability of the Surface Ocean Carbon Cycle: A Synthesis, Global Biogeochemical Cycles, 37, e2023GB007798, https://doi.org/10.1029/2023GB007798, 2023.

Rohr, T., Richardson, A. J., Lenton, A., Chamberlain, M. A., and Shadwick, E. H.: Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models, Commun Earth Environ, 4, 1–22, https://doi.org/10.1038/s43247-023-00871-w, 2023.

Sauerland, V., Kriest, I., Oschlies, A., and Srivastav, A.: Multiobjective Calibration of a Global Biogeochemical Ocean Model Against Nutrients, Oxygen, and Oxygen Minimum Zones, Journal of Advances in Modeling Earth Systems, 11, 1285–1308, https://doi.org/10.1029/2018MS001510, 2019.

Schartau, M. and Oschlies, A.: Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—Method and parameter estimates, J Mar Res, 61, 765–793, https://doi.org/10.1357/002224003322981147, 2003.

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling, Biogeosciences, 14, 1647–1701, https://doi.org/10.5194/bg-14-1647-2017, 2017.

Schmechtig, C.: Processing Bio-Argo chlorophyll-a concentration at the DAC level, 2015.

Schmechtig, C., Claustre, H., Poteau, A., D'Ortenzio, F., Schallenberg, C., Trull, T., and Xing, X.: BGC-Argo quality control manual for the Chlorophyll-A concentration, Ifremer, https://doi.org/10.13155/35385, 2023.

Schmidtko, S., Stramma, L., and Visbeck, M.: Decline in global oceanic oxygen content during the past five decades, Nature, 542, 335–339, https://doi.org/10.1038/nature21399, 2017.

Shu, C., Xiu, P., Xing, X., Qiu, G., Ma, W., Brewin, R. J. W., and Ciavatta, S.: Biogeochemical Model Optimization by Using Satellite-Derived Phytoplankton Functional Type Data and BGC-Argo Observations in the Northern South China Sea, Remote Sensing, 14, 1297, https://doi.org/10.3390/rs14051297, 2022.

Sobol′, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55, 271–280, https://doi.org/10.1016/S0378-4754(00)00270-6, 2001.

Sobol', I. M., Asotsky, D., Kreinin, A., and Kucherenko, S.: Construction and Comparison of High-Dimensional Sobol' Generators, Wilmott, 2011, 64–79, https://doi.org/10.1002/wilm.10056, 2011.

Stramma, L., Johnson, G. C., Sprintall, J., and Mohrholz, V.: Expanding Oxygen-Minimum Zones in the Tropical Oceans, Science, 320, 655–658, https://doi.org/10.1126/science.1153847, 2008.

Tagliabue, A., Kwiatkowski, L., Bopp, L., Butenschön, M., Cheung, W., Lengaigne, M., and Vialard, J.: Persistent Uncertainties in Ocean Net Primary Production Climate Change Projections at Regional Scales Raise Challenges for Assessing Impacts on Ecosystem Services, Front. Clim., 3, https://doi.org/10.3389/fclim.2021.738224, 2021.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and De Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 554–577, https://doi.org/10.1016/j.dsr2.2008.12.009, 2009.

Thierry, V., Bittig, H., and Team, T. A. B.: Argo quality control manual for dissolved oxygen concentration. Version 2.0, 23 October 2018., IFREMER for Argo BGC Group, https://doi.org/10.13155/46542, 2018.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Wang, B. and Fennel, K.: Distinct sources of uncertainty in simulations of the ocean biological carbon pump at different depths, Commun Earth Environ, 5, 1–10, https://doi.org/10.1038/s43247-024-01561-x, 2024.

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

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., and Oschlies, A.: Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, Journal of Marine Systems, 81, 34–43, https://doi.org/10.1016/j.jmarsys.2009.12.005, 2010.

Wikle, C. K., Milliff, R. F., Herbei, R., and Leeds, W. B.: Modern Statistical Methods in Oceanography: A Hierarchical Perspective, Statistical Science, 28, 466–486, 2013.

Wilcox, C., Van Sebille, E., and Hardesty, B. D.: Threat of plastic pollution to seabirds is global, pervasive, and increasing, Proceedings of the National Academy of Sciences, 112, 11899–11904, https://doi.org/10.1073/pnas.1502108112, 2015.

Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., Yashayaev, I., Barker, P. M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, M. O., Schmid, C., Lyman, J. M., McTaggart, K. E., Purkey, S. G., Zilberman, N., Alkire, M. B., Swift, D., Owens, W. B., Jayne, S. R., Hersh, C., Robbins, P., West-Mack, D., Bahr, F., Yoshida, S., Sutton, P. J. H., Cancouët, R., Coatanoan, C., Dobbler, D., Juan, A. G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., Claustre, H., D'Ortenzio, F., Le Reste, S., Le Traon, P.-Y., Rannou, J.-P., Saout-Grit, C., Speich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I. M., Klein, B., Notarstefano, G., Poulain, P.-M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaska, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida, T., Cowley, R., Lovell, J. L., Oke, P. R., van Wijk, E. M., Carse, F., Donnelly, M., Gould, W. J., Gowers, K., King, B. A., Loch, S. G., Mowat, M., Turton, J., Rama Rao, E. P., Ravichandran, M., Freeland, H. J., Gaboury, I., Gilbert, D., Greenan, B. J. W., Ouellet, M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K., Jo, H., et al.: Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats, Front. Mar. Sci., 7, https://doi.org/10.3389/fmars.2020.00700, 2020.

Yang, Z. and Zhu, T.: Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees, Proceedings of the National Academy of Sciences, 115, 1854–1859, https://doi.org/10.1073/pnas.1712673115, 2018.