Ensemble-based parameter estimation for improving ocean biogeochemistry in an Earth system model

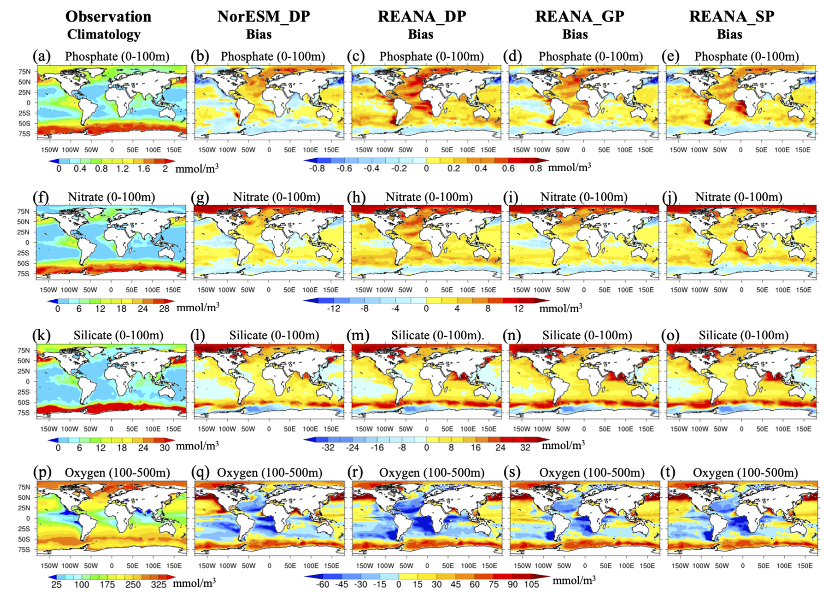
Tarkeshwar Singh[1], Francois Counillon[2], Jerry Fong Tjiputra[3], and Yiguo Wang[1]

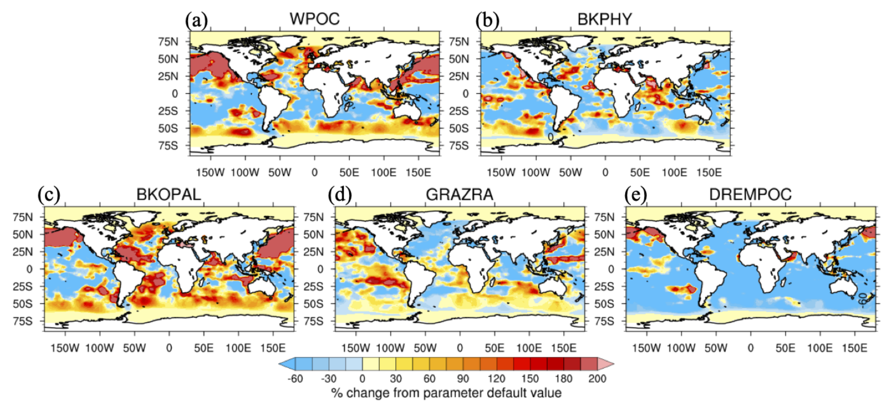
[1]NERSC
[2]Nersc
[3]NORCE

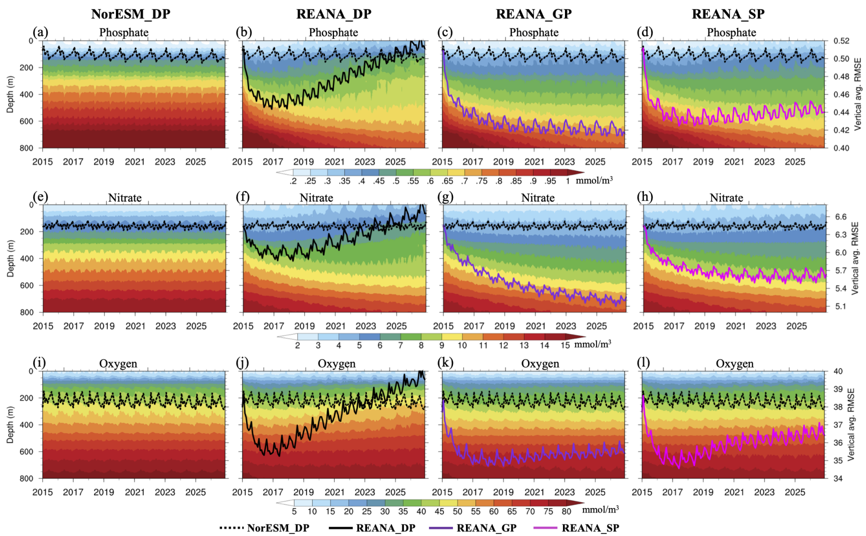
February 26, 2024

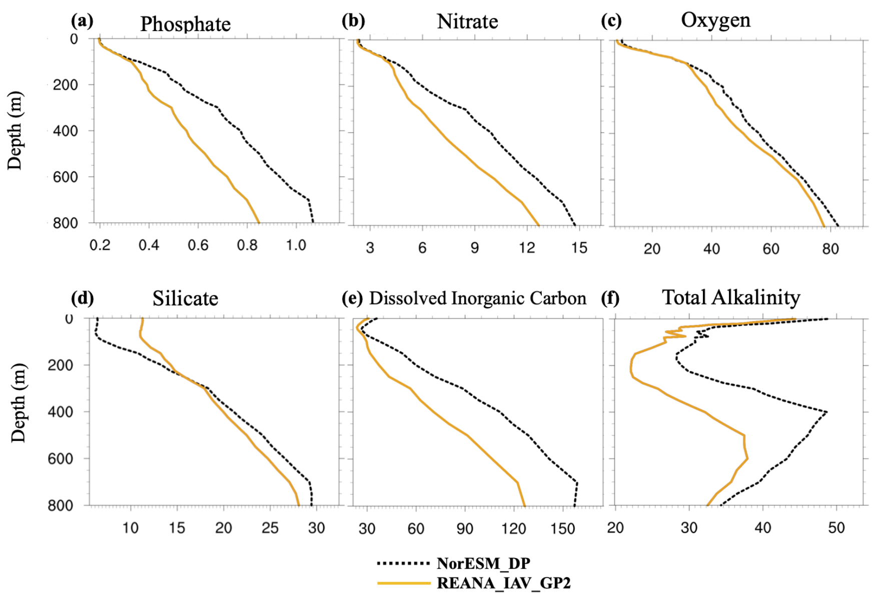
Abstract

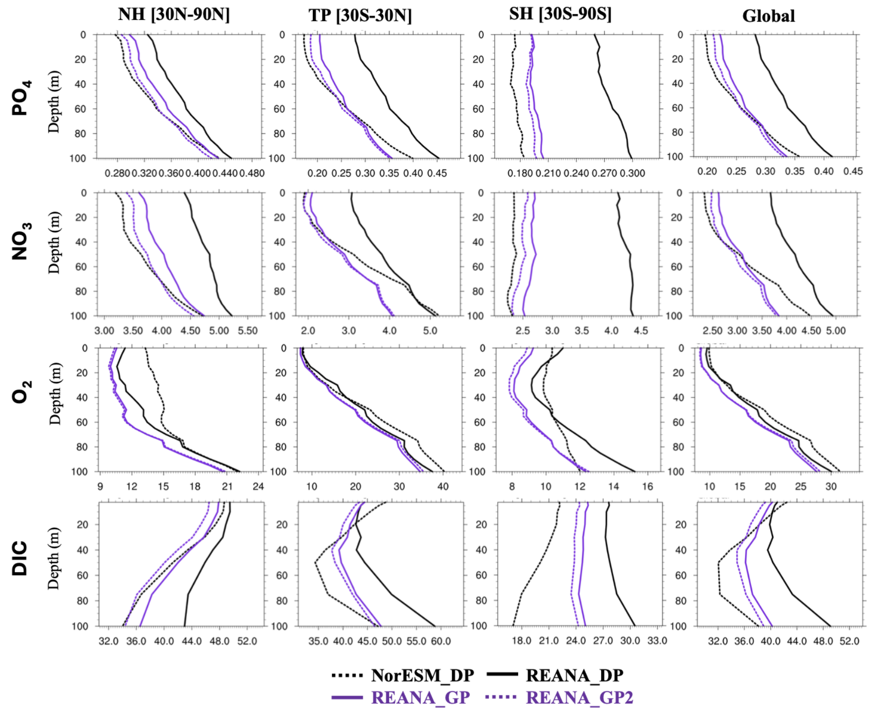
Improved ocean biogeochemistry (BGC) parameters in Earth System Models can enhance the representation of the global carbon cycle. We aim to demonstrate the potential of parameter estimation (PE) using an ensemble data assimilation method to optimise five key BGC parameters within the Norwegian Earth System Model (NorESM). The optimal BGC parameter values are estimated with an iterative ensemble smoother technique, applied a-posteriori to the error of monthly climatological estimates of nitrate, phosphate and oxygen produced by a coupled reanalysis that assimilates monthly ocean physical observed climatology. Reducing the ocean physics biases while keeping the default parameters (DP) initially reduces BGC state bias in the intermediate depth but deteriorates near the surface, suggesting that the DP are tuned to compensate for physical biases. Globally uniform and spatially varying estimated parameters from the first iteration effectively mitigate the deterioration and reduce BGC errors compared to DP, also for variables not used in the PE (such as C0$_2$ fluxes and primary production). While spatial PE performs superior in specific regions, global PE performs best overall. A second iteration can further improve the performance of global PE for near-surface BGC variables. Finally, we assess the performance of the global estimated parameters in a 30-year coupled reanalysis, assimilating time-varying temperature and salinity observations. It reduces error by 20\%, 18\%, 7\%, and 27\% for phosphate, nitrate, oxygen, and dissolved inorganic carbon, respectively, compared to the default version of NorESM. The proposed PE approach is a promising innovative tool to calibrate ESM in the future.


Observation Climatology
NorESM_DP Bias
REANA_DP Bias
REANA_GP Bias
REANA_SP Bias
(a) Phosphate (0-100m)
(b) Phosphate (0-100m)
(c) Phosphate (0-100m)
(d) Phosphate (0-100m)
(e) Phosphate (0-100m)
mmol/m3
(f) Nitrate (0-100m)
(g) Nitrate (0-100m)
(h) Nitrate (0-100m)
(i) Nitrate (0-100m)
(j) Nitrate (0-100m)
mmol/m3
(k) Silicate (0-100m)
(l) Silicate (0-100m)
(m) Silicate (0-100m).
(n) Silicate (0-100m)
(o) Silicate (0-100m)
mmol/m3
(p) Oxygen (100-500m)
(q) Oxygen (100-500m)
(r) Oxygen (100-500m)
(s) Oxygen (100-500m)
(t) Oxygen (100-500m)
mmol/m3

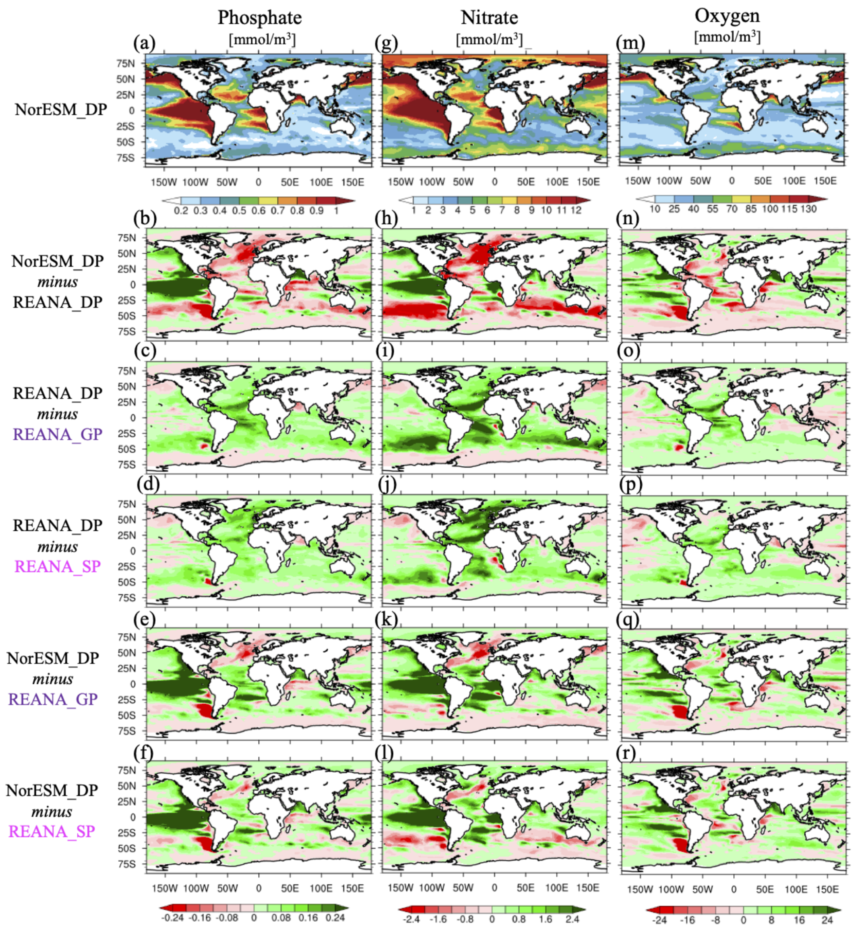

(a) WPOC
(b) BKPHY
(c) BKOPAL
(d) GRAZRA
(e) DREMPOC
-60 -30 0 30 60 90 120 150 180 200
% change from parameter default value


NorESM_DP
REANA_DP
REANA_GP
REANA_SP
(a) Phosphate
(b) Phosphate
(c) Phosphate
(d) Phosphate
(e) Nitrate
(f) Nitrate
(g) Nitrate
(h) Nitrate
(i) Oxygen
(j) Oxygen
(k) Oxygen
(l) Oxygen
Depth (m)
Vertical avg. RMSE
mmol/m³
NorESM_DP
REANA_DP
REANA_GP
REANA_SP

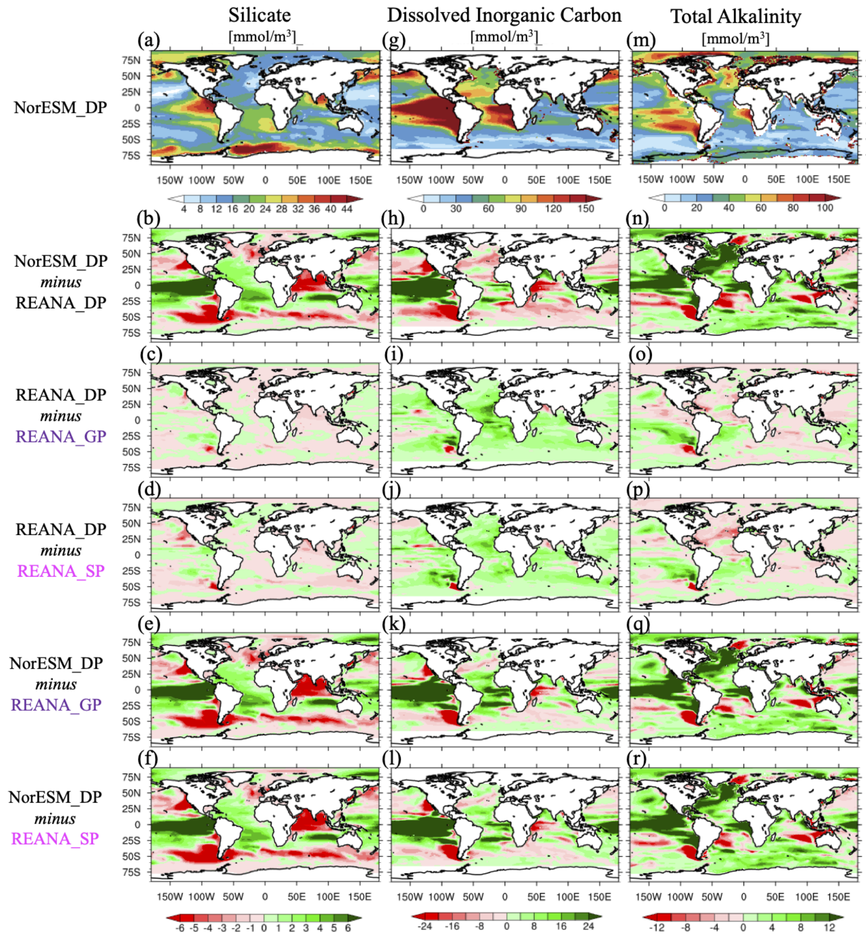

(a) Phosphate
(b) Nitrate
(c) Oxygen
(d) Silicate
(e) Dissolved Inorganic Carbon
(f) Total Alkalinity
Depth (m)
NorESM_DP
REANA_IAV_GP2



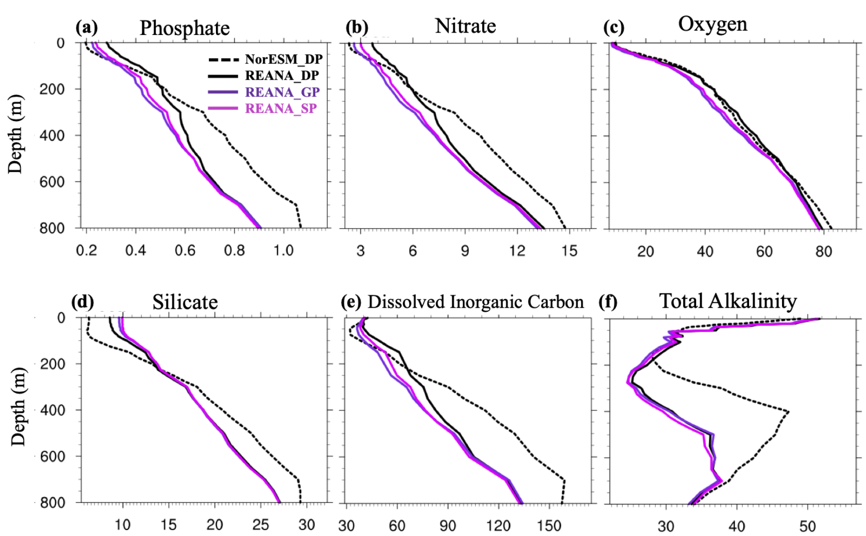
NH [30N-90N]
TP [30S-30N]
SH [30S-90S]
Global
PO4
NO3
O2
DIC
Depth (m)
NorESM_DP
REANA_DP
REANA_GP
REANA_GP2



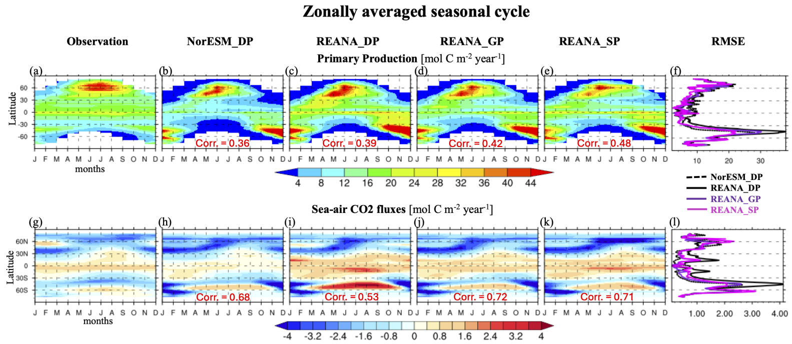
Phosphate
[mmol/m^3]
Nitrate
[mmol/m^3]
Oxygen
[mmol/m^3]
(a)
(g)
(m)
NorESM_DP
(b)
(h)
(n)
NorESM_DP
minus
REANA_DP
(c)
(i)
(o)
REANA_DP
minus
REANA_GP
(d)
(j)
(p)
REANA_DP
minus
REANA_SP
(e)
(k)
(q)
NorESM_DP
minus
REANA_GP
(f)
(l)
(r)
NorESM_DP
minus
REANA_SP
75N 50N 25N 0 25S 50S 75S
150W 100W 50W 0 50E 100E 150E
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
1 2 3 4 5 6 7 8 9 10 11 12
10 25 40 55 70 85 100 115 130
-0.24 -0.16 -0.08 0 0.08 0.16 0.24
-2.4 -1.6 -0.8 0 0.8 1.6 2.4
-24 -16 -8 0 8 16 24

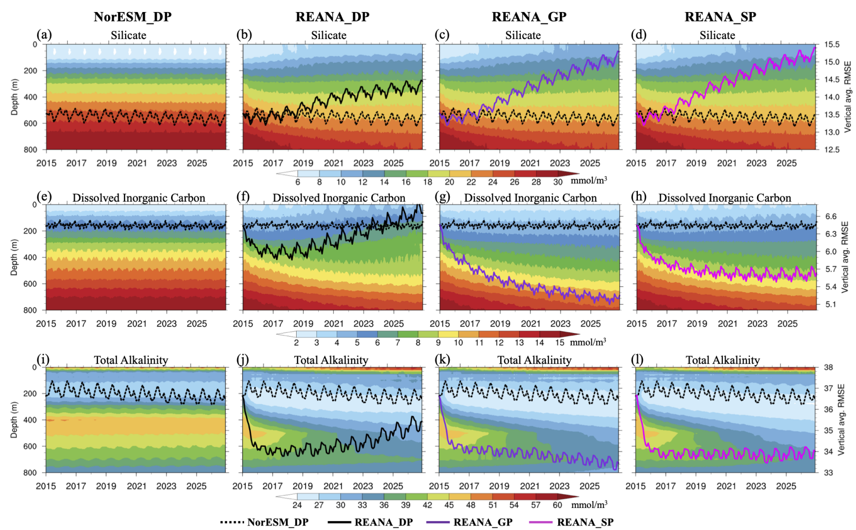


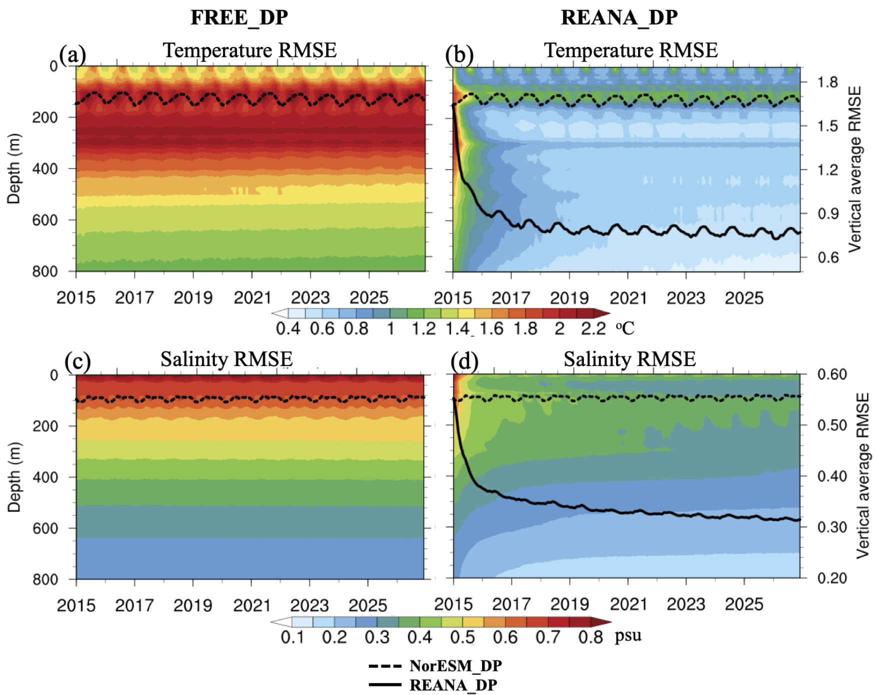
Silicate
[mmol/m³]
Dissolved Inorganic Carbon
[mmol/m³]
Total Alkalinity
[mmol/m³]
NorESM_DP
NorESM_DP
minus
REANA_DP
REANA_DP
minus
REANA_GP
REANA_DP
minus
REANA_SP
NorESM_DP
minus
REANA_GP
NorESM_DP
minus
REANA_SP
(a) (b) (c) (d) (e) (f)
(g) (h) (i) (j) (k) (l)
(m) (n) (o) (p) (q) (r)
75N 50N 25N 0 25S 50S 75S
150W 100W 50W 0 50E 100E 150E
4 8 12 16 20 24 28 32 36 40 44
0 30 60 90 120 150
0 20 40 60 80 100
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
-24 -16 -8 0 8 16 24
-12 -8 -4 0 4 8 12



(a) Phosphate
(b) Nitrate
(c) Oxygen
(d) Silicate
(e) Dissolved Inorganic Carbon
(f) Total Alkalinity
NorESM_DP
REANA_DP
REANA_GP
REANA_SP
Depth (m)
0 200 400 600 800
0.2 0.4 0.6 0.8 1.0
3 6 9 12 15
20 40 60 80
10 15 20 25 30
30 60 90 120 150
30 40 50


Zonally averaged seasonal cycle
Observation
NorESM_DP
REANA_DP
REANA_GP
REANA_SP
RMSE
Primary Production [mol C m-2 year-1]
Sea-air CO2 fluxes [mol C m-2 year-1]
Latitude
months
Corr. = 0.36
Corr. = 0.39
Corr. = 0.42
Corr. = 0.48
Corr. = 0.68
Corr. = 0.53
Corr. = 0.72
Corr. = 0.71
NorESM_DP
REANA_DP
REANA_GP
REANA_SP


NorESM_DP
REANA_DP
REANA_GP
REANA_SP
Silicate
Dissolved Inorganic Carbon
Total Alkalinity
Depth (m)
Vertical avg. RMSE
mmol/m3
NorESM_DP
REANA_DP
REANA_GP
REANA_SP


FREE_DP
REANA_DP
(a)
Temperature RMSE
(b)
Temperature RMSE
Depth (m)
Vertical average RMSE
0
200
400
600
800
1.8
1.5
1.2
0.9
0.6
2015
2017
2019
2021
2023
2025
0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 °C
(c)
Salinity RMSE
(d)
Salinity RMSE
0.60
0.50
0.40
0.30
0.20
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 psu
NorESM_DP
REANA_DP

Ensemble-based parameter estimation for improving ocean biogeochemistry in an Earth system model

Tarkeshwar Singh[1], François Counillon[1,2], Jerry Tjiputra[3], and Yiguo Wang[1]

[1]Nansen Environmental and Remote Sensing Center, and Bjerknes Centre for Climate Research, Bergen, Norway
[2]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[3]NORCE Norwegian Research Centre, and Bjerknes Centre for Climate Research, Bergen, Norway

Key Points:

- Parameters in ocean biogeochemistry model are tuned to compensate for errors in ocean physics
- Ensemble-based data assimilation provides a flexible and computationally efficient framework to train model parameters
- Parameter estimation for ocean biogeochemistry effectively reduces model error also for independent state variables

Corresponding author: Tarkeshwar Singh, `tarkeshwar.singh@nersc.no`

Abstract

Improved ocean biogeochemistry (BGC) parameters in Earth System Models can enhance the representation of the global carbon cycle. We aim to demonstrate the potential of parameter estimation (PE) using an ensemble data assimilation method to optimise five key BGC parameters within the Norwegian Earth System Model (NorESM). The optimal BGC parameter values are estimated with an iterative ensemble smoother technique, applied a-posteriori to the error of monthly climatological estimates of nitrate, phosphate and oxygen produced by a coupled reanalysis that assimilates monthly ocean physical observed climatology. Reducing the ocean physics biases while keeping the default parameters (DP) initially reduces BGC state bias in the intermediate depth but deteriorates near the surface, suggesting that the DP are tuned to compensate for physical biases. Globally uniform and spatially varying estimated parameters from the first iteration effectively mitigate the deterioration and reduce BGC errors compared to DP, also for variables not used in the PE (such as $C0_2$ fluxes and primary production). While spatial PE performs superior in specific regions, global PE performs best overall. A second iteration can further improve the performance of global PE for near-surface BGC variables. Finally, we assess the performance of the global estimated parameters in a 30-year coupled reanalysis, assimilating time-varying temperature and salinity observations. It reduces error by 20%, 18%, 7%, and 27% for phosphate, nitrate, oxygen, and dissolved inorganic carbon, respectively, compared to the default version of NorESM. The proposed PE approach is a promising innovative tool to calibrate ESM in the future.

Plain Language Summary

Earth System Models heavily rely on parametrisation that accounts for unresolved processes. Fine-tuning these numerous parameters is challenging because there are multiple sources of error, and the parameter's sensitivity is interlinked. The ocean biogeochemistry models are particularly challenging as they are heavily parameterised, and observations are sparse. We show that ocean biogeochemistry parameters in an Earth System Model that contributed to the Coupled Model Intercomparison Project have been tuned to compensate for bias in ocean physics. Reducing these biases yields suboptimal performance in ocean biogeochemistry. Here, we demonstrate that data assimilation can provide a successful framework for tuning such parameters within the Norwegian Earth System Model. The method can effectively reduce error, which is also true for variables not used in training. Performance with calibrated parameters and constrained bias in ocean physics achieve superior performance than the default version. The new calibration method will be instrumental in enhancing the performance of our future model.

1 Introduction

Ocean biogeochemical (BGC) models are crucial for estimating the ocean's major chemical elements and biomass, including carbon, oxygen, nitrogen, and phytoplankton. The ocean BGC component is essential to Earth System Models (ESM) and simulating and predicting our future climate (Flato, 2011; Orr et al., 2017). It plays a pivotal role in regulating the atmospheric carbon dioxide concentration and its feedback to the climate system (e.g., J. Tjiputra et al., 2010). The inclusion of ocean BGC in ESM is also crucial in understanding the interactions between the ocean ecosystem and other components of the Earth system and their impact, such as ocean deoxygenation and acidification (Kwiatkowski et al., 2020; J. F. Tjiputra et al., 2023) and their predictability (Fransner et al., 2020; Doney et al., 2012). However, BGC simulations in ESMs are inherently associated with significant uncertainty, emphasizing these models' continuous need for improvement.

The accuracy of BGC models is limited by various sources of errors, which are dominated by imperfect descriptions of the physical environment that drives the biology and the sub-optimally tuned empirical parameterisations of the biogeochemical dynamics (Doney et al., 2004). The BGC model uses numerous poorly known parameters to describe the complex biogeochemical process, such as the growth of phytoplankton and their grazing rate by zooplankton. These parameters are often obtained from small-scale laboratory experiments conducted on individual species, but in models, they are employed in a broader context to describe entire categories of organisms. For instance, parameter values are assumed to be constant globally and manually adjusted to capture the observed large-scale BGC variability within observational uncertainty. However, such a manual tuning process is very complicated and time-consuming, specifically with large models when the number of parameters increases with the complexity of biogeochemical models. Furthermore, many non-linear physical, chemical, and biological processes influence the marine ocean system, making it challenging to isolate the impact of individual biogeochemical parameters on the overall system. Finally, biogeochemical parameters are often not completely independent of each other and should not be tuned in isolation. For example, changes in one nutrient may affect the uptake rate of another nutrient by phytoplankton, making it difficult to isolate the effects of each nutrient individually. We explore the potential of advanced computational tools and statistical methods to provide an efficient framework to calibrate models and improve their accuracy.

Data assimilation (DA) provides a mathematical framework to estimate model states and parameters based on observational data. For state estimation, the model state variables are updated after a model integration and are used to produce reanalyses and forecasts. For parameter estimation with ensemble methods, the ensemble system is run forward with perturbed parameter values, and DA finds the optimal parameter likelihood that minimizes the misfit of the model's state variables with observations (, e.g., Jazwinski, 2007; Spitz et al., 1998; Fennel et al., 2001; Anderson, 2001; Annan et al., 2005; Friedrichs et al., 2006; J. F. Tjiputra et al., 2007; Bagniewski et al., 2011). The Ensemble Kalman Filter (EnKF; Evensen, 1994) is a sequential DA method that uses a Monte Carlo model forward integration to estimate the forecast error covariance. The EnKF has been successfully applied in several biological models for the state and the parameter estimation (Eknes & Evensen, 2002; Allen et al., 2003; Nerger & Gregg, 2008; Simon et al., 2012; Gharamti, Samuelsen, et al., 2017; Gharamti, Tjiputra, et al., 2017a; Natvik & Evensen, 2003). Typically BGC parameters are estimated with a single-column model and at the few locations where the spatial-temporal multivariate observations are good enough (e.g., see Gharamti, Tjiputra, et al., 2017b; Mamnun et al., 2022). However, the optimal parameters retrieved from the different stations differ substantially (Gharamti, Tjiputra, et al., 2017b), which exhibits the complex interactions between multiple parameters and dimensions. It is thus unclear whether one can use parameters from a single location in ESMs and achieve optimal performance globally (e.g., McDonald et al., 2012; Hoshiba et al., 2018; B. Wang et al., 2020). (Simon et al., 2015) attempt to retrieve spatially and time-varying BGC parameters in a forced ocean and BGC regional model of the North Atlantic. While the BGC state error was reduced in some places, the system diverged due to unrealistic parameter values in some regions. Singh et al. (2022) attempted to retrieve spatially varying parameters but constant in-time parameters in an idealised framework within an Earth System model. Synthetic observations – mimicking the shortage of ocean biogeochemistry observations– were generated from the same model run with (spatially) varying perturbed parameter values. The dual-one-step-ahead-smoother (DOSA) technique can recover the spatially varying parameter values and perform nearly optimally – i.e., produce errors comparable to the model with perfect parameter values. In this study, we follow on Singh et al. (2022) and test the parameter estimation technique with a real framework.

We investigate whether parameter estimation using the ensemble DA method can improve the representation of ocean biogeochemistry within the Norwegian Earth System Model (NorESM). However, our first attempt repeating the method from Singh et al. (2022) failed. Hence, the state error inherited from the other components (ocean and atmosphere) grows faster than the parametric error from the BGC model, which causes the parameter estimation to overshoot realistic ranges. Therefore, we revised our framework and decided to train the BGC parameters a-posteriori based on the reanalysis performance with state assimilation in the other components to sustain their error to a low level. The optimisation is iterative (iterative ensemble smoother, IES Evensen, 2018), and we optimize five BGC parameters that play a key role in the carbon cycle. Many studies have demonstrated that resolving local or region-based features of the parameters can be beneficial for the biogeochemical modelling (Schartau & Oschlies, 2003; Hemmings et al., 2004; Losa et al., 2004; J. F. Tjiputra et al., 2007; Roy et al., 2012; Doron et al., 2013). Therefore, we compare the performance of globally uniform and spatially varying parameter values. To our knowledge, this study is the first to perform the BGC parameter estimation in a state-of-the-art fully-coupled ESM with real observations, and we show that a global parameter estimation can achieve a substantial error reduction.

The subsequent sections of this paper are structured as follows. Section 2 provides an overview of the parameter estimation framework, encompassing details related to the model, assimilation algorithms, observations, and experimental design. Section 3 presents and discusses the numerical results. The concluding remarks of this work are outlined in Section 4.

2 Parameter estimation framework and Experimental Design

This section provides an overview of the Norwegian Climate Prediction Model (NorCPM1, Bethke et al., 2021), and outlines the different data assimilation methods used for constraining the state of the ocean component and estimating the biogeochemical parameters. We then describe the experimental design and the observations used for this study.

2.1 The Norwegian Climate Prediction Model

NorCPM1 is a system developed to provide coupled reanalysis (Counillon et al., 2016) and seasonal-to-decadal climate prediction (Kimmritz et al., 2019; Y. Wang et al., 2019; Bethke et al., 2021). It is based on the Norwegian Earth System Model (NorESM1; Bentsen et al., 2013) and provides a suite of data assimilation solutions based on ensemble Kalman Filter methods. NorCPM1 contributed to the Decadal Climate Prediction Project (DCPP) of the sixth Coupled Model Intercomparison Project (CMIP6; Bethke et al., 2021) and to the World Meteorological Organisation "lead centre for operational annual-to-decadal prediction" (Hermanson et al., 2022).

2.1.1 The Norwegian Earth System Model (NorESM)

NorCPM1 was built on the NorESM version 1 configured at a medium resolution (NorESM1-ME; Bentsen et al., 2013; J. Tjiputra et al., 2013), which uses CMIP6 forcing, and includes a bug fix in the atmospheric chemistry and was further re-calibrated (Bethke et al., 2021). The model is based on the Community Earth System Model (CESM1.0.4; Hurrell et al., 2013), but with modified aerosol chemistry in the atmosphere and the ocean component is replaced with an isopycnal coordinate ocean general circulation model. The ocean carbon cycle model is based on the Hamburg Ocean Carbon Cycle (HAMOCC5.1; Maier-Reimer et al., 2005) model. The marine ecosystem module, which is the focus of this study, is based on an NPZD-type (Nutrients-Phytoplankton-Zooplankton-Detritus) ecosystem model (Six & Maier-Reimer, 1996; J. Tjiputra et al., 2010).

At the euphotic layer (i.e., top 100 m), the model simulates primary productivity driven by phytoplankton growth, which is limited by temperature, light, and multi-nutrients (phosphate, nitrate, and dissolved iron) availability. Consumed surface nutrients are assimilated into phytoplankton soft tissue and eventually turned into dissolved and particulate organic matters (DOM and POM) through a chain of processes such as zooplankton grazing, phytoplankton exudation, zooplankton excretion, and mortality. A single species of DOM is implemented and advected by the ocean circulation, while POM primarily sinks into the ocean interior. POM and DOM are remineralized to inorganic nutrients, given sufficient oxygen concentration, at constant remineralization rates. In this version of NorESM, the BGC component does not provide any feedback to the ocean physics component.

For more details of other components of NorCPM1, including their overall performance, the reader is referred to Bethke et al. (2021).

2.1.2 DOSA-EnKF DA algorithm for ocean reanalysis

This study employs the dual one-step ahead iterative smoothing ensemble scheme (DOSA-EnKF, Gharamti et al., 2015; Gharamti, Tjiputra, et al., 2017b) for the online assimilation of ocean physics observations. DOSA-EnKF is an ensemble data assimilation technique that uses a dual iteration step, where the state variables undergo both a smoothing and an analysis step. Here, we implement DOSA-EnKF based on the deterministic EnKF DA algorithm (DEnKF; Sakov & Oke, 2008), a square-root version of the EnKF. The DEnKF algorithm performs the DA in two sub-steps: in the first step, it estimates the ensemble mean (Equation 1) that minimises the distance from the truth based on its distance from observations, while in the second sub-step, the ensemble perturbation is updated to adjust the ensemble anomaly (Equation 2). This method yields an approximate but deterministic form of the traditional stochastic EnKF and outperforms the latter, particularly with small ensembles (Sakov & Oke, 2008). It also inflates the errors by construction and is intended to perform well for operational applications. Below, we present the mathematical formulations of the DEnKF method in the DOSA framework:

Consider an ensemble of model state $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, ...\mathbf{x}^m] \in \mathbf{R}^{n \times m}$, with each column $\mathbf{x}^i \in \mathbf{R}^n$ representing an individual ensemble member. Here, n denotes the size of the state vector, and m is the ensemble size. Let $\overline{\mathbf{X}}$ denote the ensemble mean, which is a column vector in $\in \mathcal{R}^{n\times 1}$. The ensemble anomaly $\mathbf{A} \in \mathcal{R}^{n\times m}$ of the model state can be computed as $\mathbf{A} = \mathbf{X} - \overline{\mathbf{X}}\mathbf{1}_m^{\mathrm{T}}$, with $\mathbf{1}_m = [1, 1, ..., 1] \in \mathcal{R}^{1\times n}$, where the superscript T denotes a matrix transpose.

In the first step of the DOSA scheme, a given ensemble of the analysed model state $\mathbf{X}_{k-1}^a$ at time $k-1$ is integrated forward at time k using the dynamical model ($\mathcal{M}$): $\mathbf{X}_k^f$=$\mathcal{M}(\mathbf{X}_{k-1}^a)$; with the superscript f denoting the forecast and a the analysis. Observations $\mathbf{y}_k$ at time k are used to produce a smoothed estimate of the ensemble mean and anomaly at the previous analysis step $k-1$ as follows:

$$\overline{\mathbf{X}_{k-1}^s} = \overline{\mathbf{X}_{k-1}^a} + \mathbf{K}_{k-1,k}(\mathbf{y}_k - \mathbf{H}\overline{\mathbf{X}_k^f}). \tag{1}$$

$$\mathbf{A}_{k-1}^s = \mathbf{A}_{k-1}^a - \frac{1}{2}\mathbf{K}_{k-1,k}\mathbf{H}\mathbf{A}_k^f. \tag{2}$$

where the superscript s denotes the smooth state, $\mathbf{H}$ is the observations operator which maps the model state to the observations space, and $\mathbf{K}_{k-1,k}$ is Kalman gain formulated as :

$$\mathbf{K}_{k-1,k} = \mathbf{A}_{k-1}^a(\mathbf{A}_k^f)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{A}_k^f(\mathbf{A}_k^f)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{3}$$

where $\mathbf{R}$ is the observation error covariance matrix.

In the second step, the model is integrated forward to time k again but from the smoothed ensemble of state $\mathbf{X}_{k-1}^s$; i.e. $\mathbf{X}_k^{f2}$=$\mathcal{M}(\mathbf{X}_{k-1}^s)$.

The observations at time k, $\mathbf{y}^k$, are then used again to produce an analysis of state $\mathbf{X}_k^a$ and anomaly $\mathbf{A}_k^a$ ensemble at time k. Unlike the standard DOSA filter, we have inflated $\mathbf{R}$ by a factor of 2. Hence, if the model is persistent, the standard formulation would assimilate observations twice. Multiplying the error variance by a factor of 2, the scheme becomes equivalent to the ES–MDA (Ensemble Smoother with Multiple Data Assimilation, Raanes et al., 2019) and results will coincide with the standard EnKF for a linear model.

$$\overline{\mathbf{X}_k^a} = \overline{\mathbf{X}_k^{f2}} + \mathbf{K}_{k,k}(\mathbf{y}_k - \mathbf{H}\overline{\mathbf{X}_k^{f2}}). \tag{4}$$

$$\mathbf{A}_k^a = \mathbf{A}_k^{f2} - \frac{1}{2}\mathbf{K}_{k,k}\mathbf{H}\mathbf{A}^{f2}. \tag{5}$$

Here, $\mathbf{A}_k^{f2}$ is the ensemble anomaly metrics constructed from second time model forecast $\mathbf{X}_k^{f2}$. $\mathbf{K}_{k,k}$ is the standard Kalman gain and represented similar to Equation 3 (with k instead of k-1).

In this study, the assimilation cycle is monthly, and we update the full ocean state vector (Counillon et al., 2016). The adjustment of the other compartments (e.g., atmosphere and sea ice) occurs dynamically during the system integration. The model state vector ($\mathbf{X}$) is constructed in isopycnal coordinates and includes temperature, salinity, layer thickness and velocities of the entire water column (53 layers, see Counillon et al., 2014; Y. Wang et al., 2017). We use the upscaling super layer algorithm developed by Y. Wang et al. (2016) to update layer thickness, which prevents the analysis from returning negative quantities and preserves heat, mass and salt.

The ocean data assimilation (ODA) yields updates of the BGC mass component because of the update of the layer thickness. Bethke et al. (2021) demonstrated that this approach conserves well BGC properties and does not introduce spurious upwelling at the Equator, an artefact often notified by ODA (While et al., 2010; Park et al., 2018).

The remaining assimilation configurations are set as in Bethke et al. (2021). Assimilation is done in a local analysis framework (Sakov et al., 2012). A latitude-varying quasi-Gaussian localization function (Gaspari & Cohn, 1999) is used to smooth the assimilation impact at the boundary of the localisation radius (Y. Wang et al., 2017). We use the moderation and a pre-screening technique (Sakov et al., 2012) to sustain the ensemble spread during the assimilation.

2.1.3 Iterative ensemble smoother for offline parameter estimation

The online parameter estimation from Singh et al. (2022) was initially tested but did not work successfully. Parameter values reached unrealistic values, and performance was poorer than the default parameter values. Unlike in the perfect twin experiment where the model errors are, by construction, only coming from the model parameter, the model errors now also originate from the other components (e.g., ocean) or from BGC parameters not considered in the parameter estimation. Here, the error inherited from the ocean grows faster than that of the parameter considered, which confuses the relation between the parameter values and the model-data misfits. Furthermore, as the model error is much larger than internal variability, sustaining a consistent ensemble spread level, even with an adaptive inflation scheme, is challenging without reaching an unrealistic inflation factor (not shown). This is particularly challenging for ocean biogeochemistry because error is primarily dominated by sporadic bloom events that occur at different times of the year in different locations.

This led to parameter estimation converging to different solutions, depending on when the experiment was started (e.g., in winter or summer), where the first bloom occurred.

We propose an offline approach to estimate parameters using an iterative ensemble smoother (Evensen, 2018) based on the DEnKF scheme. First, each ensemble member runs with a random parameter value over the entire simulation period, and one can then relate the parameter value with the misfit between the state variables and the observations. The update is applied a-posteriori on the aggregated innovation vector of the whole period in a single analysis step. It differs from the online assimilation method, where analysis is performed sequentially with a joint monthly state and parameter update (Singh et al., 2022).

The iterative ensemble smoother scheme nicely handles the challenges mentioned above. First, as we continuously assimilate the ocean state with the DOSA scheme, the errors inherited from the ocean physic are sustained at a low error level, and errors related to the BGC parameter can grow and dominate the overall error term. Second, as the assimilation considers all calendar months jointly, there is no longer preferential influence related to the start of the cycle.

Let the ensemble of model parameters be $\theta \in \mathcal{R}^{p\times m}$, with the ensemble mean $\overline{\theta} \in \mathcal{R}^{p}$ and the ensemble anomaly $\mathbf{A}_\theta \in \mathcal{R}^{p\times m}$. Here, p denotes the number of parameters, and m is the size of the ensemble. The offline parameter estimation using the iterative ensemble smoother is described below.

The ensemble of the analyzed model state $\mathbf{X}^a_{k-1} \in \mathcal{R}^{n\times m}$ at time $k-1$, are integrated forward sequentially T times using the parameter ensemble θ^i with the dynamical model ($\mathbf{X}^f_k$=$\mathcal{M}(\mathbf{X}^a_{k-1}, \theta^i)$) and with ocean assimilation in between each model integration step. We denote the aggregated model forecast ensemble members over time; denoted as $\mathbf{X}^f_{1:T} = [\mathbf{X}^f_{1_{1:T}}, \mathbf{X}^f_{2_{1:T}} ....\mathbf{X}^f_{m_{1:T}}] \in \mathcal{R}^{nT\times m}$. Here, m is the number of ensemble members, and nT denotes the number of model states n times the number of time steps T. For a monthly cycle and a yearly optimisation window, T=12. Let the $\overline{\mathbf{X}^f}_{1:T} \in \mathcal{R}^{nT}$ is the ensemble means, $\mathbf{A}^f_{1:T}$ is the ensemble anamoly and observation is $\mathbf{y}_{1:T} \in \mathcal{R}^o$, aggregated over the corresponding time of $\mathbf{X}^f_{1:T}$. Here, o is the total number of observations aggregated over time $k = 1$ to T. The offline parameter estimation using the DEnKF algorithm is as follows:

$$\overline{\theta^{i+1}} = \overline{\theta^i} + \mathbf{K}(\mathbf{y}_{1:T} - H\overline{\mathbf{X}^f}_{1:T}) \tag{6}$$

$$\mathbf{A}^{i+1}_\theta = \mathbf{A}^i_\theta - \frac{1}{2}\mathbf{KHA}^f_{1:T} \tag{7}$$

$$\mathbf{K} = \mathbf{A}_\theta^i (\mathbf{A}_{1:T}^f)^\mathrm{T} \mathbf{H}^\mathrm{T} \left(\mathbf{H} \mathbf{A}_{1:T}^f (\mathbf{A}_{1:T}^f)^\mathrm{T} \mathbf{H}^\mathrm{T} + \mathbf{R} \right)^{-1}. \tag{8}$$

The ensemble of estimated parameters θ^{i+1} can be reconstructed as: $\theta^{i+1} = \overline{\theta^{i+1}} + \mathbf{A}_\theta^{i+1}$.

The approach can be repeated iteratively, where the forward model integration $\mathcal{M}$ is rerun with the estimated parameter values from the previous iteration. For instance, the model integration for a second iteration can be expressed as $\mathbf{X}_k^f = \mathcal{M}(\mathbf{X}_{k-1}^a, \theta^{i+1})$. Consequently, we can iterate equations 6 and 7, utilizing the updated ensemble mean $\overline{\mathbf{X}^f}_{1:T}$ and anomaly $\mathbf{A}_{1:T}^f$, both of which are derived from the second model integration. The observation error should be multiplied by the number of iterations one intends to perform (Evensen, 2018). Consequently, this approach enhances the accuracy of model simulations by iteratively refining the non-linear response to the model parameters (in the linear case, the solution would be identical). The model starts with the same initial state $\mathbf{X}_{k-1}^a$ for each iteration.

The approach can estimate global and spatially varying parameters. In global parameter estimation, the model parameters are a vector ($\in \mathcal{R}^p$) and are estimated based on the innovation from the entire domain. The spatially varying parameter estimates ($\in \mathcal{R}^p \times m$) are based on the innovation from the local domain, similar to the state estimation of the ocean physics state (see section 2.1.2).

2.2 Observations

We use two distinct sets of global observations to estimate model parameters: (1) ocean physics monthly climatology for state constraints and (2) BGC monthly climatology for parameter estimation, both from the World Ocean Atlas 2018 release (WOA18) datasets (Locarnini et al., 2018; Zweng et al., 2019; Garcia et al., 2019a, 2019b). The ocean physics estimate consists of temperature and salinity (TS), and the BGC includes nutrients (PO_4; phosphate and NO_3; nitrate) and oxygen (O_2). Nutrients extend down to 800 m depth, whereas temperature, salinity and oxygen profiles are available down to 1500 m deep. To estimate the observation error needed for the assimilation, we use the quadratic sum of the error estimate from the WOA18 dataset and add one deseasoned (with the mean seasonal cycle removed) time standard deviation from our model computed over 1980-2010 to account for representation error (Janjić et al., 2018).

The WOA18 climatological estimates are available at a regular horizontal grid with a spatial resolution of 1°x1°. These gridded datasets are generated through objective analysis, which involves interpolating and extrapolating data from individual measurement points to create a continuous gridded product. While the ocean physics climatological estimate is quite accurate overall, the situation differs for BGC data. BGC measurements are sparse and heterogeneously distributed in space and time. Therefore, for BGC, we only use the estimate where at least one measurement is available within the corresponding grid-square area. Ocean physics observations in ice-covered regions are excluded. The sea ice mask is estimated using the 30-year climatological mean sea ice data from the NorESM historical model simulation (1980-2010). BGC measurements that are located within 4 model grid points away from the sea ice point are excluded.

The validation of parameter estimation results involves including additional observations not used during the parameter estimation process. It includes silicate (SI), dissolved inorganic carbon (DIC), total alkalinity (TA), primary production (PP) and sea-air CO_2 fluxes. These observations, referred to as independent data, play a crucial role in assessing the impact of parameter estimation methods.

This study further utilised the time-varying observations of sea surface temperature (SST) data from the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation SST version2 (Reynolds et al., 2002), and subsurface ocean temperature and salinity hydrographic profile observations from the EN4 dataset (EN4.2.2; Gouretski & Reseghetti, 2010) to create a 30 years reanalysis. The aim was to assess the performance of estimated parameters within the context of interannually varying ocean physics and with ocean variability from an independent period.

A summary of observations utilized for parameter estimation and validation is provided in Table 1.

2.3 Experimental setup for BGC parameter estimation and verification

The HAMOCC model in NorESM1 incorporates a range of parameters to effectively simulate the biogeochemical characteristics of the ocean, with a comprehensive list provided by Maier-Reimer et al. (2005). In this study, our primary objective is to optimize a selection of five parameters chosen explicitly for their influences on the carbon cycle and with connections to the assimilated BGC observations. These parameters include: 1) the half-saturation constant for nutrient uptake during phytoplankton growth (BKPHY), 2) the maximum zooplankton grazing rate (GRAZRA), 3) the sinking speed for particulate organic carbon (WPOC), 4) the half-saturation constant for silicate uptake during biogenic opal production (BKOPAL), and 5) the remineralization rate of particulate organic carbon (DREMPOC). The parameter's acronyms (e.g., BKPHY) are the same as in the model source codes to facilitate straightforward tracking.

Table 1. List of observations utilized (Column-1), with the second column detailing the data type; whether time-varying or climatology. The reference period for the latter type is provided in the third Column. The second-to-last column signifies if data is assimilated (in ODA or PE), while the last column contains the observational reference.

Observation variables	data type	Clim. Period	Remark	References
Temperature	climatology	2005–2017	Assimilated (ODA)	WOA 2018 (Locarnini et al., 2018)
Salinity	climatology	2005–2017	Assimilated (ODA)	WOA 2018 (Zweng et al., 2019)
Oxygen	climatology	1960–2018	Assimilated (PE)	WOA 2018 (Garcia et al., 2019a)
Phosphate	climatology	1960–2018	Assimilated (PE)	WOA 2018 (Garcia et al., 2019b)
Nitrate	climatology	1960–2018	Assimilated (PE)	WOA 2018 (Garcia et al., 2019b)
Silicate	climatology	1960–2018	Independent	WOA 2018 (Garcia et al., 2019b)
Dissolved Inorganic Carbon	climatology	2004-–2017	Independent	Keppler et al. (2020)
Total Alkalinity	climatology	1972-2017	Independent	Broullón et al. (2019)
Sea–air CO2 fluxes	climatology	1982-–2015	Independent	Landschützer et al. (2017)
Net primary production	climatology	2003-–2012	Independent	Average of the three remote sensing products (VGPM, Eppley-VGPM, and CbPM) from the Moderate Resolution Imaging Spectroradiometer (Behrenfeld & Falkowski, 1997; Westberry et al., 2008)
Sea Surface Temperature	Time-varying	–	Assimilated (ODA)	NOAA OI SST version2 (Reynolds et al., 2002)
Temperature	Time-varying	–	Assimilated (ODA)	EN4.2.2 (Gouretski & Reseghetti, 2010)
Salinity	Time-varying	–	Assimilated (ODA)	EN4.2.2 (Gouretski & Reseghetti, 2010)

The BGC parameter estimation method uses monthly ocean BGC outputs produced by a coupled reanalysis that assimilates ocean physics climatological observations. As such, it becomes easier to relate the BGC model error to the BGC parameter values. The monthly temperature and salinity climatology observations are repeatably assimilated every year. We perform the offline parameter estimation from a yearly cycle of the ocean reanalysis after it has reached stable performance.

We now describe the series of experiments that we have used to estimate parameters and assess their performance. Each experiment uses a 30-member model ensemble run and is outlined as follows:

- **NorESM_DP** – an ensemble of historical runs with default parameters (DP). It serves as a benchmark to assess the baseline performance of the default version of NorESM that contributed to CMIP6. The ensemble was initialised from a random pre-industrial state in 1850, run until 2014 with CMIP6 historical forcing, and extended with the Shared Socioeconomic Pathway (SSP) 2-4.5 scenario forcing from 2015 to 2026.
- **REANA_DP** – a reanalysis using DP and with ocean constrained to follow monthly climatology of temperature and salinity repeated every year (section 2.1.2). The reanalysis is branched from NorESM_DP and runs from 2015 until 2026. It helps to estimate how the BGC model responds to reducing bias in ocean physics.

- **NorESM_PP** – an ensemble of simulations (as NorESM_DP) utilizing perturbed BGC parameters (PP). The simulation was branched from NorESM_DP in January 2005 and run until January 2015. The PP is generated by adding Gaussian perturbation to the default values of five chosen parameters, with the standard deviation set at 30% of the default value. A 30-member PP ensemble is generated for every chosen parameter, and spatially constant values are assigned to each member. The valid range of these parameters is unknown due to our limited understanding and observations (J. F. Tjiputra et al., 2007). We employ lower and upper thresholds during the ensemble generation process to ensure that the parameter values remain within a reasonable range. Lower and upper bounds are defined as percentage changes from their respective default values, ranging from 60% to 200% of the default value. This experiment serves as a spin-up run to build sensitivity to the perturbed parameters.
- **REANA_PP** – a reanalysis same as REANA_DP, but with perturbed BGC parameters and initialized from NorESM_PP. The mismatch of the REANA_PP reanalysis data with observed BGC climatology is used to estimate the parameters in the first iteration (Iteration-1, Section 2.1.3). The estimated parameter values are discussed and presented in Section 3.2.
- **REANA_GP** – a reanalysis (as REANA_PP) with global estimated parameters obtained from Iteration-1. The performance of the REANA_GP is assessed to investigate the benefit of global parameter estimation (Section 3.3). The output data from this experiment is also used to estimate the global parameters in the second iteration (Iteration-2) and presented in Section 3.2.
- **REANA_SP** – a reanalysis (as REANA_GP), but with spatial estimated parameters (Section 3.2). This experiment is conducted to evaluate the performance of spatially varying estimated parameter values and to investigate if there is any benefit over the global one (Section 3.3).
- **REANA_GP2** – a reanalysis same as REANA_GP, but with global estimated parameters taken from Iteration-2. This experiment helps to investigate the benefit of the multi-iteration approach for parameter estimation (Section 3.4).
- **REANA_IAV_GP2** – an extended reanalysis using time-varying ocean physics observations-i.e., considering internal variability (IAV). It runs from 1985 to 2022 with the same global estimated parameter used in REANA_GP2. It assimilates time-varying observation of SST and TS profiles in a monthly cycle. The initial state was branched from NorESM_DP in 1982. For REANA_IAV_GP2, the first eight years of the reanalysis (1985-1992) are discarded (considered as a spinup to adjust to the new parameter), and the remaining 30 years (1993-2022) are used for validation.

A summary of all experiments is given in Table 2.

Table 2. List of experiments.

Experiment Name	Observations (if assimilated)	Starting initial ensemble	Parameters used	Time period	Description
NorESM_DP	-	Pre-industrial	Default parameters	1850 - 2026	Baseline
REANA_DP	TEM+SAL clim.	NorESM_DP	Default parameters	2015 - 2026	Reanalysis to evaluate the impact of reduced ocean physics bias on BGC
NorESM_PP	-	NorESM_DP	Perturbed parameters	2005 - 2015	Spin-up run for PP.
REANA_PP	TEM+SAL clim.	NorESM_PP	Perturbed parameters	2015 - 2026	Input data to perform first iteration PE
REANA_GP	TEM+SAL clim.	NorESM_PP	Global estimated parameters from Iteration-1	2015 - 2026	Reanalysis to evaluate the first iteration global PE and also input data to perform the second iteration
REANA_SP	TEM+SAL clim.	NorESM_PP	Spatial estimated parameters from Iteration-1	2015 - 2026	Reanalysis to evaluate if any benefit of spatial PE over global one
REANA_GP2	TEM+SAL clim.	NorESM_PP	Global estimated parameters from Iteration-2	2015 - 2026	Reanalysis to evaluate benefit of second iteration global PE
REANA_IAV_GP2	TEM+SAL time-varying	NorESM_DP	Global estimated parameters from Iteration-2	1985 - 2022	Reanalysis to evaluate second iteration global PE with interannually varying ocean forcings.

2.4 Statistical Metrics

The performance of the numerical experiments is analyzed based on the following measures -

$$\mathrm{RMSE} = \sqrt{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{W_i}(\overline{\mathrm{x}}_{\mathrm{i}}^{\mathrm{f}} - \mathrm{y_i})^2} \tag{9}$$

$$\mathrm{Bias} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{W_i}(\overline{\mathrm{x}}_{\mathrm{i}}^{\mathrm{f}} - \mathrm{y_i}), \tag{10}$$

where RMSE is the area-weighted root mean square error, W_i is the area of i^{th} model grid cell, and N is the total number of data points. $\overline{x}^f$ represents the model ensemble mean, and y corresponds to the observed values. We perform bi-linear interpolation of the observations to the model grid.

3 Results and Discussions

3.1 Impact of reduced ocean physics bias on BGC

The accuracy of the underlying ocean physics strongly influences the accuracy of BGC simulations. Utilizing data assimilation techniques to reduce ocean physics bias is expected to improve the performance of BGC simulations. However, the performance of BGC simulations can also be negatively impacted by inaccurate parameters, which may have been tuned to account for bias in the physics component. We compare the performance of REANA_DP and NorESM_DP, which both utilise default model parameters but with REANA_DP constraining the bias in the ocean's physical state. The monthly climatological observations are used for verification. Error in NorESM_DP temperature and salinity is large and similar every year (Figure 1). REANA_DP rapidly reduces error in temperature and salinity and sustains it at a low level (to about 44% for temperature and 50% for salinity).

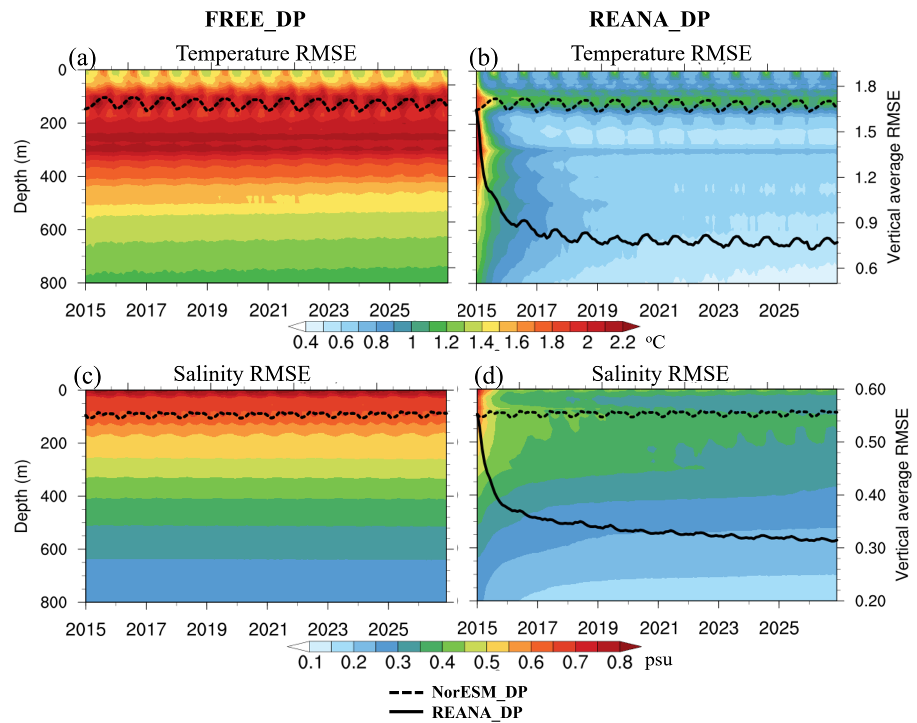


Figure 1. Hovmöller diagram of global monthly RMSE in NorESM_DP (left column) and REANA_DP (right column) computed against WOA18 climatological temperature (top) and salinity (bottom). The black dotted and solid lines represent the monthly vertically-averaged RMSE (right y-axis) in NorESM_DP and REANA_DP, respectively.

REANA_DP yields a pronounced improvement initially for phosphate, nitrate and oxygen (see Figure 2) within the first two years – particularly evident below 400 m. However,

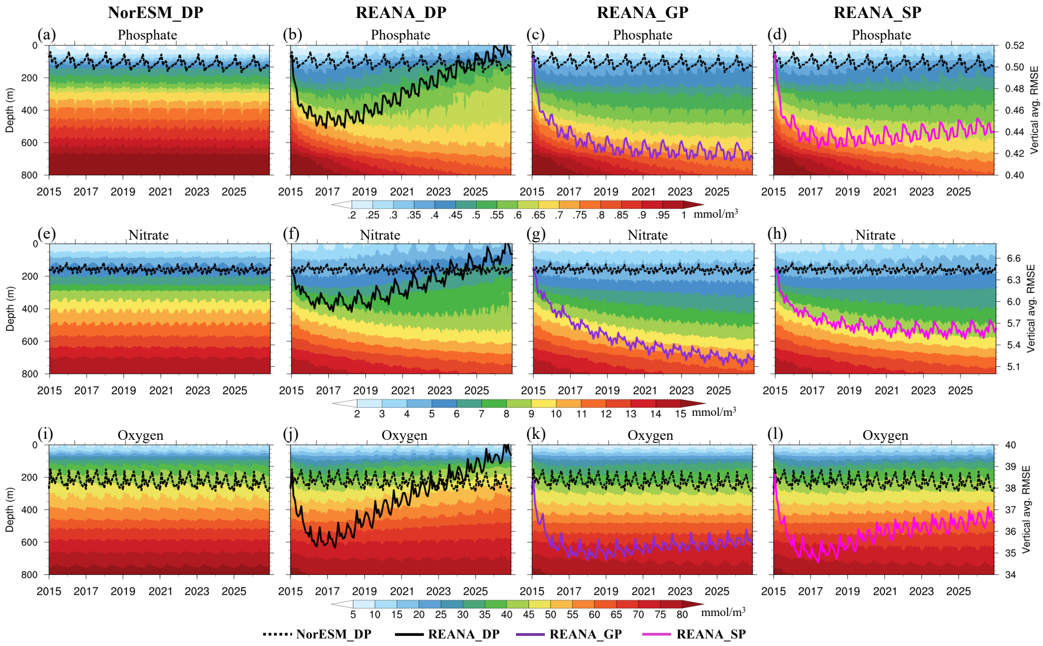


Figure 2. Global Hovmöller diagram of RMSE in NorESM_DP (column-1), REANA_DP (column-2), REANA_GP (column-3) and REANA_SP (column-4) computed again WOA18 monthly climatology of phosphate, nitrate and oxygen (shown in rows 1 to 3, respectively). The lines are the vertically monthly averaged RMSE (right y-axis).

after that period, the error starts to grow in the top 500 meters, and the vertically and globally averaged RMSE is degraded beyond 10 years. A similar behaviour is also evident for dissolved inorganic carbon (DIC), silicate and total alkalinity (TA) (Figure 3). The degradation is quickest for silicate, for which the overall error is already degraded after a couple of years, while there are still some improvements for alkalinity after ten years. The errors below ≈500 m are consistently reduced for DIC, TA and silicate in the REANA_DP experiment.

To investigate the degradation issue, we further examine the REANA_DP nutrients spatial distribution in the euphotic zone (0-100 m). REANA_DP brings an excessive concentration of nutrients (phosphate and nitrate) to the euphotic zone (see Figure 4), which is too high roughly everywhere in the globe. Phosphate and nitrate concentrations are higher over the Atlantic than in other regions (Figure 4c,h). Consequently, the REANA_DP exhibits higher phytoplankton growth and primary production rate than the NorESM_DP (not shown). Regions experiencing elevated phytoplankton growth accumulate larger amounts of particulate organic matter, which subsequently sinks to deeper ocean layers. This increases the input of organic matter and leads to enhanced remineralization, which consumes oxygen and releases nutrients. The enhanced remineralization in REANA_DP leads to oxygen

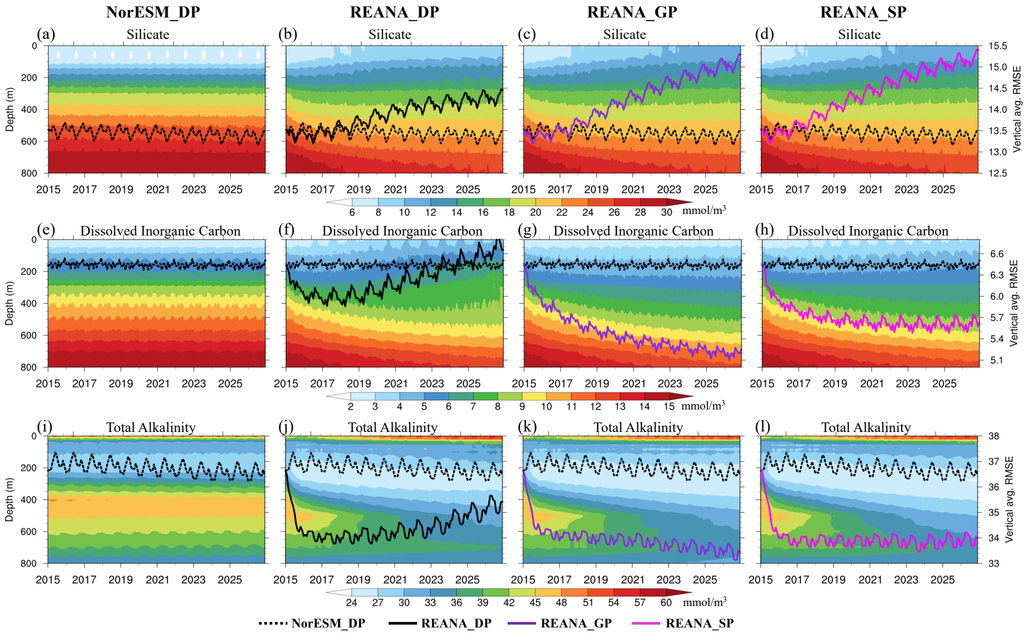


Figure 3. Same as Figure 2 but for silicate, dissolved inorganic carbon and total alkalinity.

depletion below the euphotic layer, which further increases the negative oxygen bias at intermediate ocean depths (approximately 100-500 meters) in biological active regions (Figure 4r). In summary, the improved ocean physics leads to increased nutrient transport to the ocean surface, improving the subsurface nutrient distribution (Figure 2b,f). However, the accumulation of nutrients in surface waters stimulates drift at near-surface layers over time.

In addition to the mean states, as presented so far, we also evaluate the performance of the upper-ocean process seasonal cycle (biological production and air-sea CO_2 fluxes). The biases in the seasonal cycles of biological production have been identified as one of the key factors contributing to the uncertainty in projected carbon sinks and storage in the ESMs participating in CMIP5/6 (Kessler & Tjiputra, 2016; Goris et al., 2018; Rodgers et al., 2023), and their improvements have been prioritized in recent model development (J. F. Tjiputra et al., 2020).

For primary production, the NorESM_DP simulates considerably lower winter production (January–March in the Northern Hemisphere and July–September in the Southern Hemisphere) within the extratropical oceans (between 30° and 65°N and south of 30°S) compared to observational data (Figure 5a,b). Further, NorESM_DP depicts an excessively strong spring bloom over the Southern Ocean and registers lower production in the tropical region relative to the observed estimates. REANA_DP enhances primary production near the equator but tends to overestimate the seasonal blooms in temperate regions of both hemispheres (Figure 5c). Notably, in the Northern Hemisphere, the bloom initiates too early, in February, as opposed to April in observation. The production is also too strong in the Southern Hemisphere, extending from the equator to the extratropical region, a behaviour that significantly deviates from NorESM_DP and observational data.

We also analyse the sea-air CO_2 fluxes (Figure 5). NorESM_DP performs well in representing the seasonal cycle of sea-air CO_2 fluxes. However, REANA_DP significantly degrades its performance compared to NorESM_DP in the tropical and Southern Hemisphere regions due to an overestimation of outgassing. This issue, which is consistent with the higher upwelling rate (i.e., of carbon-rich deep water to surface), is further evident from the high RMSE values (Figure 5l).

The above results suggest that while reducing the ocean physical bias yields some benefit initially, the performance is overall degraded as the default BGC parameters were tuned to compensate for the biases in the ocean physics. Tuning BGC parameters to account for bias in the physical components is challenging when developing community code such as an Earth System Model. If the performance of the ocean physics were to be improved, the system would result in degraded performance of the BGC component unless a new calibration was to be repeated again, which may slow down the model upgrade. Furthermore, large biases in ocean physics can lead to inaccuracies in the representation of BGC processes and their interactions. Thus, we can expect to achieve superior performance by re-tuning the BGC parameters while reducing the bias in the ocean physics.

3.2 Offline global and spatial BGC parameter estimation

We present the globally uniform and spatially varying estimated parameters resulting from the offline iterative ensemble smoother technique. The output of REANA_PP (with perturbed parameters) is used to estimate the parameters (Section 2.3).

Table 3 displays the default and ensemble mean of global estimated values for the five chosen BGC parameters. In Iteration-1, two parameters, the half-saturation constant for nutrient uptake (BKPHY) and the remineralization rate (DREMPOC), exhibit a reduction (of 49% and 60%, respectively) from their default values. The remaining three parameters, the maximum grazing rate (GRAZRA), the sinking speed (WPOC), and the half-saturation constant for silicate uptake (BKOPAL), are increased (by 13%, 20% and 68%, respectively). The estimated values from Iteration-2 demonstrate changes in a similar direction as Iteration-1, except for DREMPOC, which shows nearly no changes between Iteration-1 and -2.

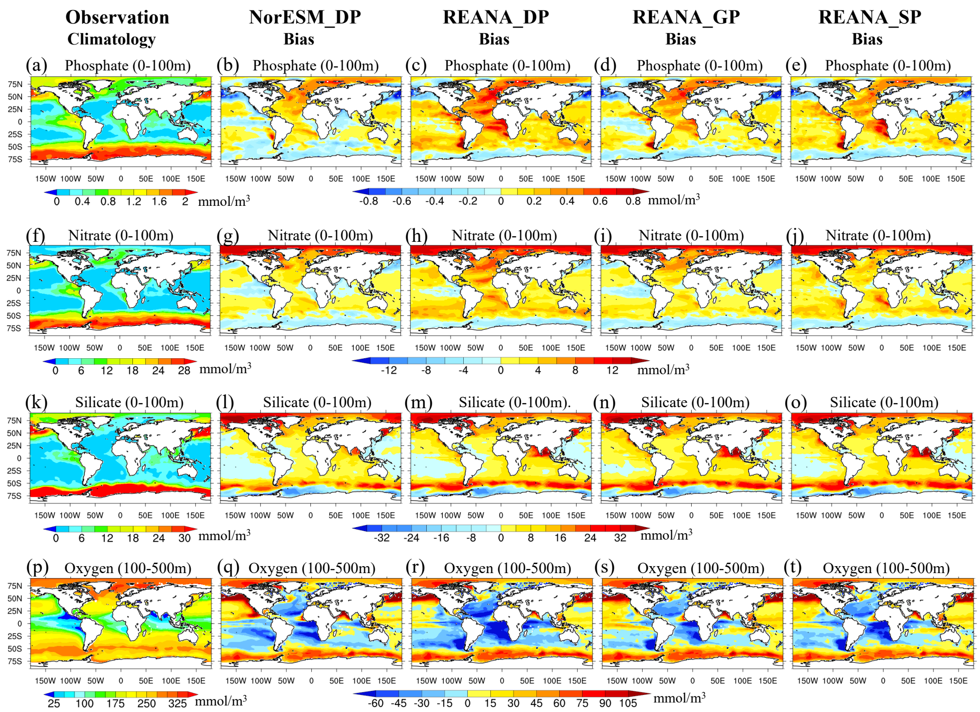


Figure 4. WOA18 climatological estimate of (a) phosphate, (f) nitrate and (k) silicate averaged over euphotic zone (0-100 m depth) and over 100-500 m depth for oxygen. The second column depicts the corresponding climatological biases [2017-2026] in NorESM_DP. The third, fourth and fifth columns are for REANA_DP, REANA_GP and REANA_SP, respectively.

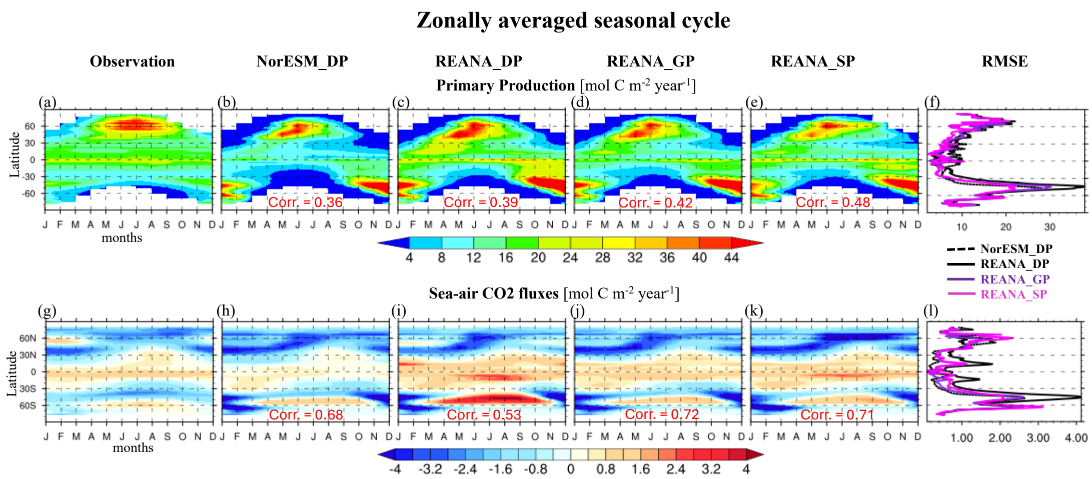


Figure 5. Hovmöller plots of the zonally averaged seasonal cycle of (a) observed climatology of primary production, and corresponding simulated climatology based on 2017-2026 period for (b-e) NorESM_DP, REANA_DP, REANA_GP and REANA_SP, respectively, together with (f) latitudinal varying RMSE from all experiments. The RMSE is computed using longitudinal and monthly climatology data. (g-l) same as for primary production but for sea-air CO_2 fluxes. Negative values show a sink to sea, whereas positive values indicated net outgassing of CO_2 fluxes from the ocean.

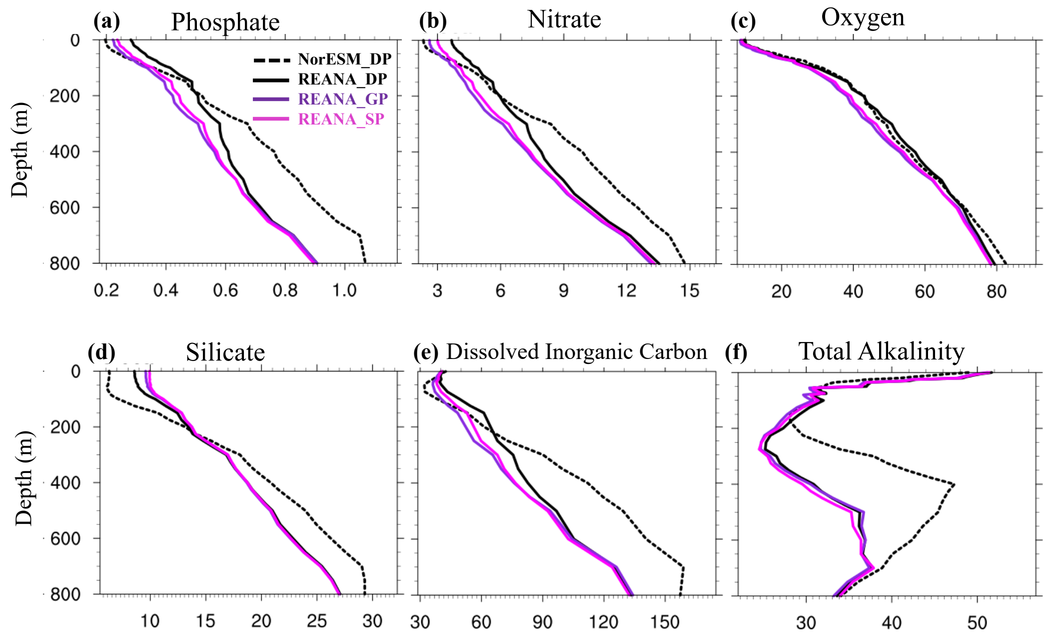


Figure 6. Global RMSE profile for monthly climatological [2017-2026] value of (a) phosphate (b) Nitrate (c) Oxygen (d) Silicate (e) dissolved inorganic carbon and (f) total alkalinity in all four experiments NorESM_DP (black dashed lines), REANA_DP (black solid lines), REANA_GP (purple lines) and REANA_SP (magenta lines).

The spatial distributions of the parameter values obtained from the offline spatial parameter estimation method in Iteration-1 are shown in Figure 7. Changes are significant for all five parameters compared to their default ones. Some parameters exceed the specified upper or lower boundaries, which were enforced to remain within the prescribed bounds. Some regional patterns clearly emerge from the map of all parameters, and estimates from the biologically active region are in agreement with the global estimation. Singh et al. (2022) found that the NorESM model is merely insensitive to parameter values in the biologically less active region, and we should thus be cautious in over-interpreting parameter values there. Increased values of the WPOC parameter can be seen over many biologically active regions, which agrees well with the global estimation. Similarly, lower values for DREMPOC are found over most regions, which is also in line with the global estimate.

3.3 Performances of reanalysis with estimated BGC parameters

We evaluate the performance of ocean reanalysis rerun with the spatially varying and global estimated BGC parameters from Iteration-1 (REANA_SP, and REANA_GP). Performances are compared to the free run with the default parameter (NorESM_DP) and the reanalysis with the default parameter (REANA_DP). We start by analysing the RMSE of

Table 3. Default and global estimated BGC parameters values.

Parameter Name	*unit*	Default values	Global estimated values (ensemble mean)	
			Iteration-1	Iteration-2
BKPHY	$mmol\,P\,m^{-3}$	2.0x10^{-7}	1.02x10^{-7}	0.60x10^{-7}
GRAZRA	day^{-1}	1.0	1.13	1.27
BKOPAL	$mmol\,Si\,m^{-3}$	1.5x10^{-6}	2.52x10^{-6}	3.30x10^{-6}
WPOC	day^{-1}	5.0	6.0	6.5
DREMPOC	day^{-1}	3.0x10^{-2}	0.90x10^{-2}	0.91x10^{-2}

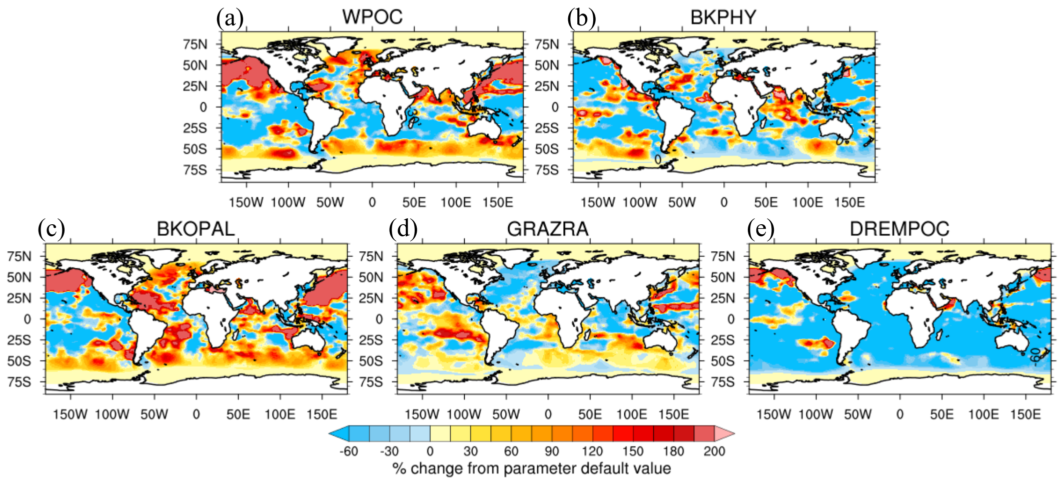


Figure 7. Parameters values obtained from the spatial parameter estimate in the first iteration. The background colour represents the percentage change from their default values for each parameter [100*(estimated_value - default_value)/default_value].

the monthly climatological estimate. The observations of phosphate, nitrate, and oxygen were used for the parameter estimation and, therefore, are not independent variables for evaluating the performance of estimated parameters. On the contrary, silicate, DIC, TA, primary production and CO_2 flux are fully independent.

REANA_GP improves the performance of all state variables used in the parameter estimation and effectively mitigates the drifting issues observed with default parameters (Figures 2, 6, and Table 4). The drift in performance that led to a degradation in the surface down to approximately 500 m in REANA_DP is also effectively reduced. Errors continue to decrease for phosphate and nitrate, but there is still a slight increase for oxygen. The latter seems to stabilize, and the vertically integrated error settles at a lower error level than in NorESM_DP. Similar improvements are also verified for independent variables DIC and TA (Figures 3, 6). The global error in REANA_GP is reduced by 15.6%, 16.4%, 7.9%, 7.7% and 1.9% for phosphate, nitrate, oxygen, dissolved inorganic carbon and total alkalinity, respectively, than NorESM_DP (Table 4). However, REANA_GP degrades performance compared to NorESM_DP and even REANA_DP for silicate. It is unexpected to see improvements in nitrate and phosphate and degradation in silicate, which suggests an internal inconsistency between the formulation of the silicate cycle and other nutrients in the ocean BGC model. This is analysed in more detail below.

REANA_GP reduces error nearly everywhere compared to REANA_DP, and also reduces the error compared to NorESM_DP (Figures 8, 9). This is important because it implies that the model with reduced bias in ocean physics and re-tuned parameters can perform better than the default model. We previously observed that REANA_DP accumulates excessive nutrients in the upper ocean layers. In REANA_GP, this issue is effectively mitigated, and phosphate and nitrate concentrations are reduced (Figure 4c,d,h,i). The phosphate and nitrate concentration dynamics are primarily influenced by three parameters considered in this study: the half-saturation constant for nutrient uptake (BKPHY), sinking speed (WPOC), and remineralization rate (DREMPOC). The estimated value of BKPHY is notably lower than the default values in the REANA_GP (as listed in Table 3), facilitating the higher consumption of near-surface phosphate and nitrate concentrations for phytoplankton growth compared to the default values. Additionally, the increase in the sinking speed (WPOC) accelerates the export of organic matter (containing an excess of nutrients) from the surface into depth. Similarly, the third parameter, DREMPOC, is reduced substantially, slowing the pace of nutrients released into the deeper ocean. It reduces the nutrient availability in the ocean interior, leading to reduced nutrient transport to the surface. These three estimated parameter values jointly act to balance phosphate and nitrate concentrations at the surface, preventing the excessive accumulation observed in REANA_DP. Consequently, this leads to improved phosphate and nitrate concentrations in the euphotic zones, showing the adequate impact of the parameter adjustment.

Excess surface nutrients in REANA_DP induce high phytoplankton growth, which leads to excessive oxygen depletion below the mixed layer through organic matter remineralization. One parameter that indirectly regulates phytoplankton growth is the maximum zooplankton grazing rate (GRAZRA), which is increased in REANA_GP. This, combined with reduced surface levels of nutrients, reduces the primary production in upper oceans. As such, REANA_GP compares more favourably with observation than REANA_DP (not shown). Furthermore, this reduces the flux of particulate organic carbon and oxygen consumption below the mixed layer (in REANA_GP) and further alleviates the anomalously low oxygen simulated in the tropical upwelling system (REANA_DP; Figure 4r,s).

For silicate, a slight degradation is seen in REANA_GP compared to REANA_DP (Figure 4n) and a strong degradation is found compared to NorESM_DP. It suggests that the degradation caused by correcting the ocean's physical bias could not be counteracted by adjusting the parameters selected. Hence, our selection of BGC parameters was tailored to focus on the carbon cycle and have negligible impact on silicate. Only BKOPAL, among the parameters selected, can influence the surface silicate. Our global parameter estimation resulted in an increase of BKOPAL, which led to a reduction of the silicate uptake during biogenic opal production and resulted in increasing surface silicate compared to REANA_GP. Another factor contributing to the silicate degradation is the reduction of phytoplankton concentrations, which implicitly reduces diatom production in REANA_GP compared to REANA_DP. This leads to an excess of surface silicate concentrations as diatom consumes silicate. There are several approaches in which we can improve the silicate simulation. One would be to include the silicate observations in the parameter estimation training. Another possibility is to extend the list of selected parameters to include those influencing the silicate cycle. For instance, the vertical sinking speed of biogenic opal (WOPAL) and deep remineralization constant for the opal (DREMOPAL) in the HAMOCC model will have a comparable impact on silicate than what WPOC and DREMPOC (used in this study) has on the nitrate cycle at depth.

The simulation REANA_SP, which uses spatial varying parameters, overall demonstrates very comparable performance to REANA_GP (Figures 2, 3, 4, 8, and 9). The North Atlantic stands out as a region where REANA_SP improved performance compared to REANA_GP (for example, nitrate, Figures 8i,j)). Conversely, performances are degraded compared to REANA_GP in the Southern Ocean. This coincides well with the density of the observation network, which is much higher in the North Atlantic than in the rest of the domain and that is particularly poor in the Southern Ocean. We suspect that the training period is too short and that the estimation fails in regions where the monthly climatological estimates are too inaccurate due to a lack of data.

Table 4. The overall global RMS error reduction (in %) for the BGC simulated climatology [2017-2026] by global and spatially estimated parameters w.r.t REANA_DP (column-2,3, respectively) and NorESM_DP (column-4,5, respectively). The green colour represents improvement, and red represents degradation.

Variables (0-800 m)	**% error reduction w.r.t REANA_DP**		**% error reduction w.r.t. NorESM_DP**	
	REANA_GP	**REANA_SP**	**REANA_GP**	**REANA_SP**
Phosphate	12.2%	9.3%	15.6%	12.9%
Nitrate	14.4%	10.3%	16.4%	12.3%
Oxygen	6.7%	4.9%	7.9%	6.1%
Silicate	-3.1%	-4.3%	-7.3%	-8.4%
Dissolved Inorganic Carbon	17.5%	15.8%	7.7%	5.8%
Total Alkalinity	7.8%	7.6%	1.9	1.6%

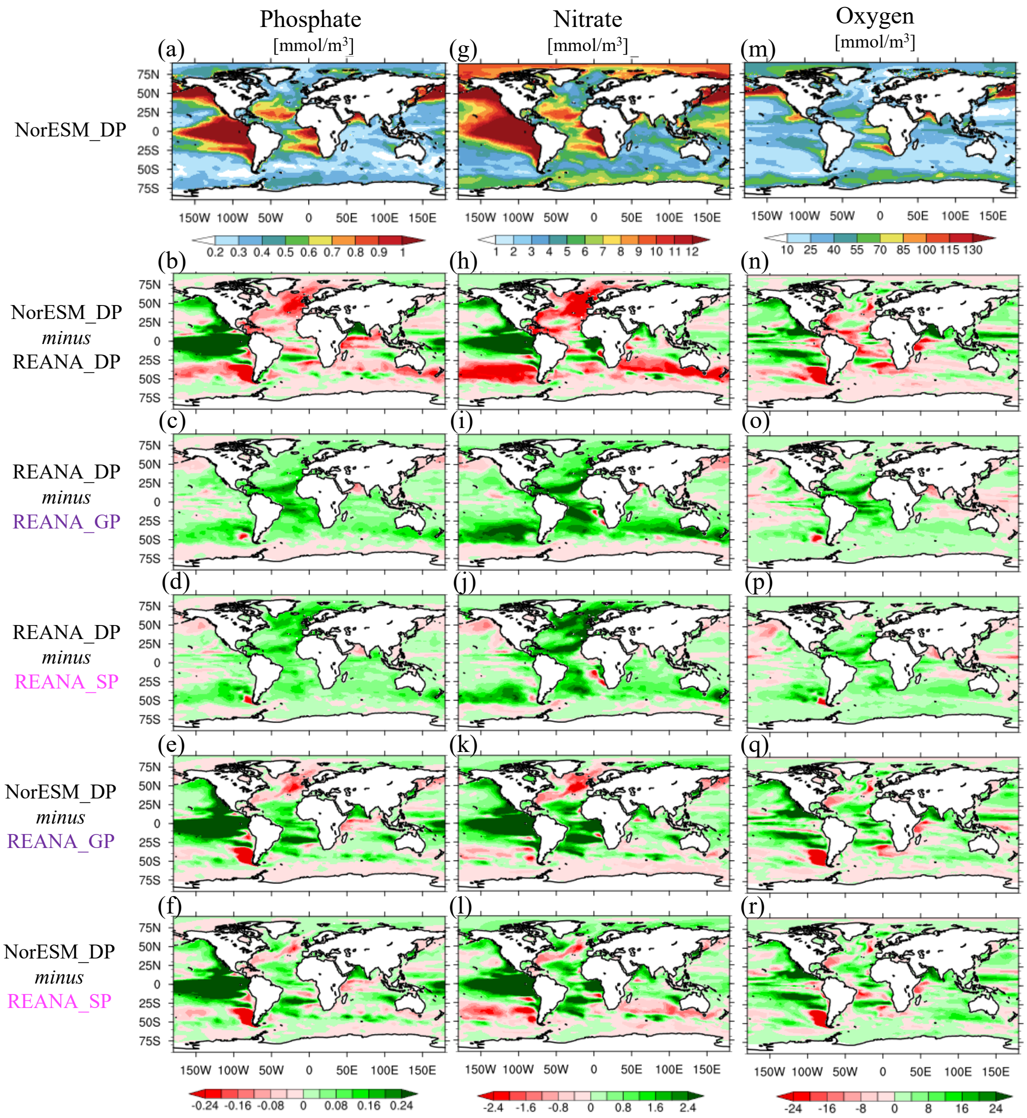


Figure 8. Average of RMSE in the top 800 m, (Row-1) for NorESM_DP monthly climatology [2017-2026] of (a) phosphate, (g) nitrate, and (m) oxygen. the second row shows the RMSE difference between NorESM_DP and REANA_DP (green colour indicates that REANA_DP outperforms NorESM_DP). Similarly, the third and fourth rows display the RMSE difference of NorESM_DP with that of REANA_GP and REANA_SP. Finally, rows five and six depict the RMSE difference of NorESM_DP with that of REANA_GP and REANA_SP.

For primary production, REANA_GP and REANA_SP improve the performance compared to REANA_DP and NorESM_DP (Figure 5d,e), particularly in the tropical and the Southern Ocean. The pattern correlation is also increased, and the RMSE is lower in REANA_GP and REANA_SP than in REANA_DP (Figure 5f). Both REANA_GP and REANA_SP enhance winter production (particularly within the Southern Hemisphere) and exhibit improvements during all seasons in the tropics. Moreover, spatially varying parame-

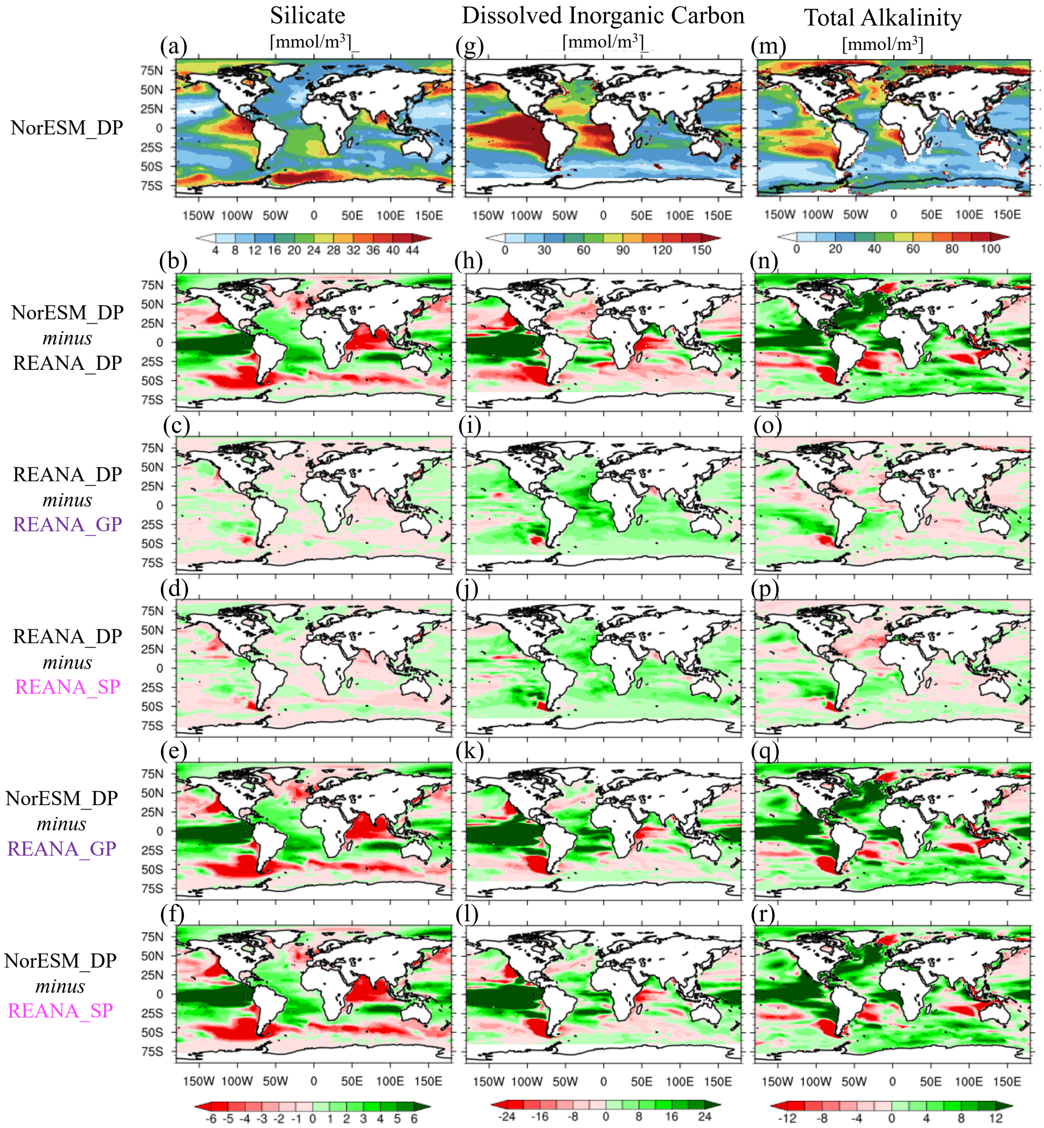


Figure 9. Same as Figure 8 but for silicate, dissolved inorganic carbon and total alkalinity.

582 ter estimation (REANA_SP) demonstrates a better Southern Hemisphere spring bloom than the global estimates (REANA_GP). In fact, REANA_SP shows the lowest RMSE compared to the other three experiments for the southern bloom (Figure 5f).

585 The seasonal cycle of sea-air CO_2 fluxes is also improved in REANA_GP and REANA_SP, and they correct the degradation seen in REANA_DP in the tropical and Southern Hemisphere (Figure 5). While a slight degradation is observed in the high latitudes of the Northern Hemisphere, both estimated parameter simulations slightly improve the pattern correlation compared to NorESM_DP (0.68, 0.53, 0.72, and 0.71 for NorESM_DP, REANA_DP, REANA_GP, and REANA_SP, respectively). REANA_GP performs slightly

better than REANA_SP in the tropics. This underscores the improvement achieved through the estimated parameters in simulating the mean seasonal variations in CO_2 fluxes.

To conclude, we can see that the parameter estimation clearly improves performance compared to both simulations with default parameters (NorESM_DP and REANA_DP) already at the first iteration. A small degradation remains near the surface (Figure 6). The global parameter estimation provides overall better and more stable performance. We will, therefore, continue with that scheme and assess whether further improvements can be achieved with more iterations.

3.4 Benefit of a second iteration with global estimation

We now analyse the performance of REANA_GP2, which is a new reanalysis produced with the parameter estimated from REANA_GP output in the second iteration (Section 2.1.3 and 2.3). It should be noted that we have not increased the observation error for the two iterations as we would from the DA theory with the iterative ensemble smoother (Section 2.1.3). The motivation for that choice was that the BGC monthly climatology observation error is very uncertain. We saw the multi-iteration as a way to iterate until performance degrades for independent data (criteria for stopping iteration). However, as a consequence, we cannot directly compare the performance of the second iteration with the first one because they do not use the same observation error.

We focus on the top 100 m where REANA_GP shows a degradation compared to NorESM_DP. The performance below that depth is nearly identical between REANA_GP and REANA_GP2 (not shown). The REANA_GP2 RMSE profile in the 0-100 m depth range shows some improvements over REANA_GP, particularly for phosphate, nitrate and DIC (Figure 10). The reduction is most pronounced for phosphate in the surface and subsurface layers of the Northern Hemisphere and tropical regions. For nitrate, REANA_GP2 matches the performance of NorESM_DP near the surface in the tropical regions. The improvement of REANA_GP over NorESM_DP for oxygen is further enhanced, particularly for the southern region. However, there is no improvement in silicate and total alkalinity (not shown). Nevertheless, We can conclude that the second iteration has further reduced the error. The improvement of parameter estimation with the iterative ensemble method may relate to the non-linear response of the model error to the parameter values (Evensen, 2018) or to the effective reduction of observation error caused by the second iteration. While we could have continued with more iterations, we stopped here as the error reduction is already much smaller than the first iteration.

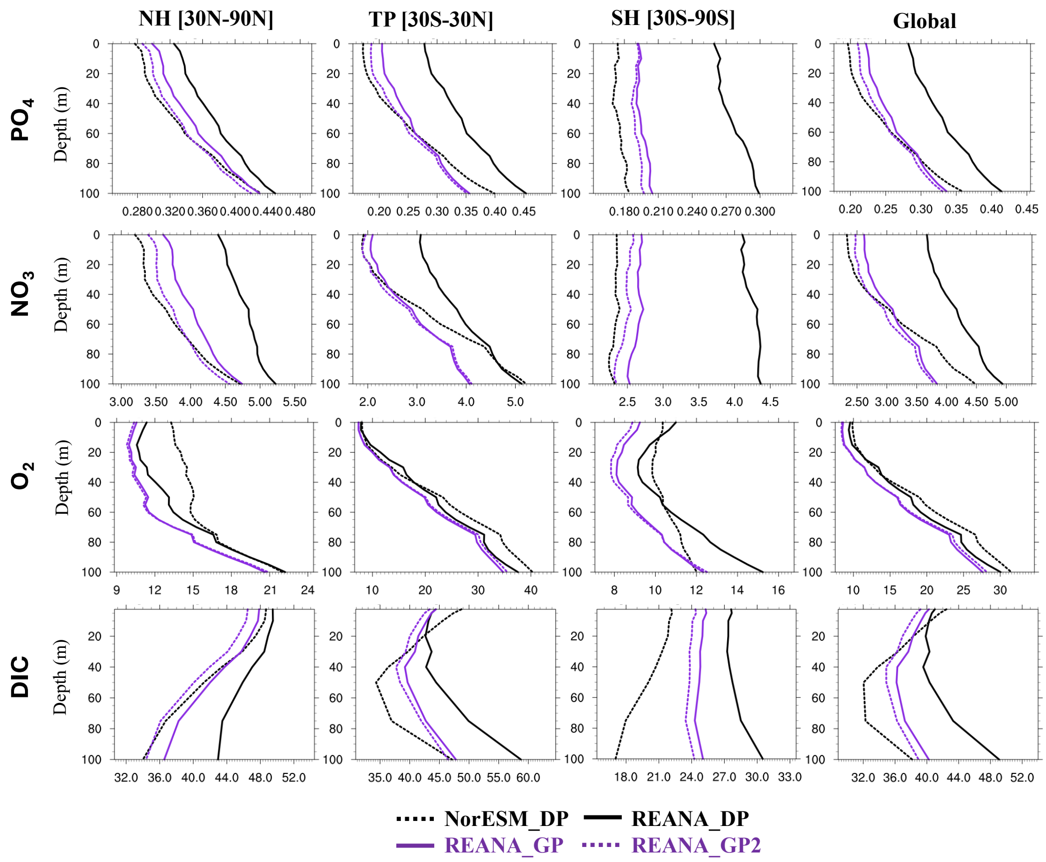


Figure 10. Northen Hemisphere (NH; column-1), Tropics (TP; column-2), Southern Hemishare (SH; column-3) and global (column-4) mean vertical RMSE profile for climatology [2017-2026] phosphate (row-1), nitrate (row-2), oxygen (row-3) and dissolved inorganic carbon (row-4). The profiles are extended from the surface to 100 m depths (euphoric zone) for NorESM_DP (black dashed lines), REANA_DP (black solid lines), REANA_GP (purple lines) and REANA_GP2 (dashed purple lines).

3.5 Performance of global estimated parameters with time-varying ocean forcing

The training and verification of the estimated parameter were performed until now with ocean conditions constrained to observed climatology. We verify whether the improvements are sustained in a reanalysis that assimilates time-varying ocean observations – i.e., in a system that depicts realistic internal variability but where the ocean state error is still constrained to a low error level. The reanalysis REANA_IAV_GP2 experiments are conducted for 1985-2020 with the BGC parameters obtained from Iteration-2 (Section 2.3). As BGC observations are lacking, we will still perform our validation towards BGC climatology, and the reanalysis is validated for the period 1993–2020 to leave the system time to adjust to the new parameters (Section 2.3).

The results are overall in very good agreement with the previous analysis. The errors in the REANA_IAV_GP2 climatological BGC mean state profiles are reduced well for all variables except for silicate that is still degraded near the surface compared to the NorESM_DP (as previously). Interestingly, the degradation near the surface compared to NorESM_DP for phosphate and nitrate is no longer noticeable. We suspect that the assimilation of climatological temperature and salinity may have caused a spurious effect near the surface that is not present when assimilating joint SST and high-resolution vertical profile data. Compared to NorESM_DP, REANA_IAV_GP2 shows a reduction in the overall RMS errors for 0-800 m depth by 20%, 18%, 7%, 27%, and 17% for phosphate, nitrate, oxygen, dissolved inorganic carbon and total alkalinity, respectively.

This comparison is highly promising, considering that the model with default parameters was carefully tuned to get the best possible fit with observations and contributed to CMIP6.

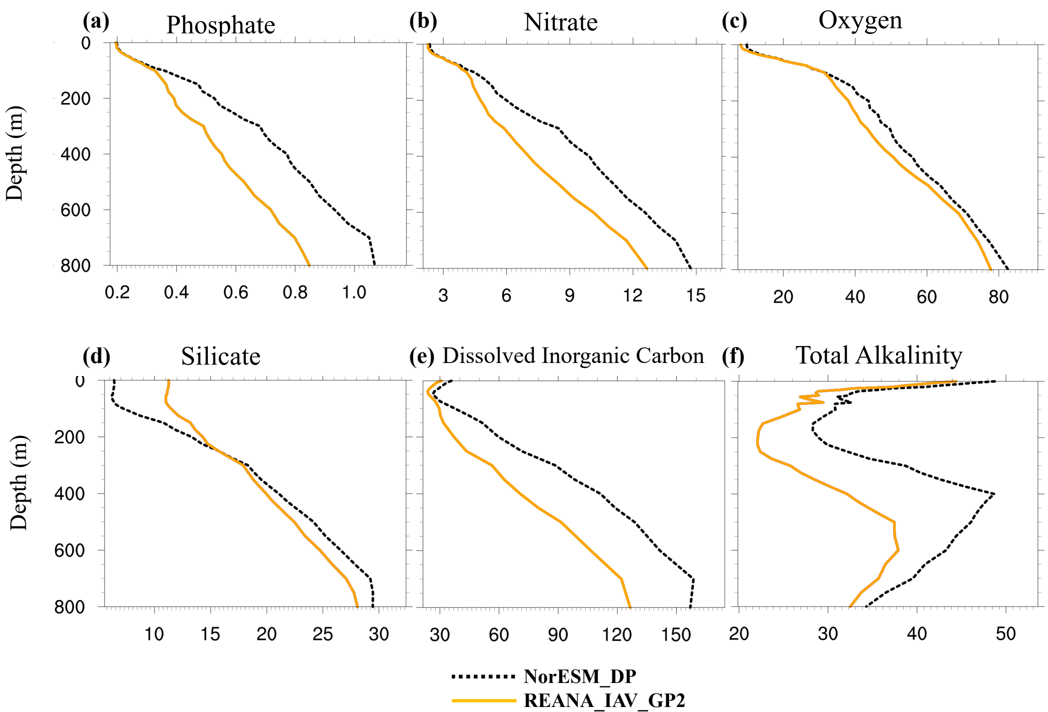


Figure 11. Global mean vertical RMSE profile for climatology [averaged over 1993-2022] (a) phosphate (b) Nitrate (c) Oxygen (d) Silicate (e) dissolved inorganic carbon and (f) total alkalinity in NorESM_DP (black dashed lines) and REANA_IAV_GP2 (orange solid lines).

4 Summary and conclusions

Enhancing the representation of ocean biogeochemistry in a fully coupled Earth System Model (ESM) is challenging due to its dependence on various factors, such as the accuracy of underlying ocean physics and numerous poorly constrained biogeochemical (BGC) parameters governing the intricate biogeochemical processes. Data assimilation methods such as the Ensemble Kalman Filter have recently been shown to offer an automatic and efficient framework for optimizing these parameters (Eknes & Evensen, 2002; Allen et al., 2003; Nerger & Gregg, 2008; Simon et al., 2012; Gharamti, Samuelsen, et al., 2017; Gharamti, Tjiputra, et al., 2017a), and tested within a full ESM in idealised twin experiment framework (Singh et al., 2022). The methods can estimate multiple parameters jointly based on several observation data sets, considering their respective uncertainty.

Here, we follow on Singh et al. (2022) and demonstrate the potential of ensemble data assimilation parameter estimation for full ESM with real observations for the first time. This framework is based on an iterative ensemble smoother, and we assimilate multivariate BGC climatological observations in offline mode to estimate jointly a set of BGC parameters. This study uses the NorCPM system, which combines the fully coupled NorESM model with ensemble data assimilation methods. We focus on five key BGC parameters influencing the ocean carbon cycle. The chosen parameters characterize the major surface biological processes such as phytoplankton growth, zooplankton grazing, sinking and remineralization of organic matter and nutrient uptake.

A key challenge with tuning the model is that these parameters may have been tuned to compensate for errors not intrinsic to the component to which the parameter belongs. This can be challenging with community codes such as ESMs because if errors are reduced in another component, the performance of the other components will degrade without returning and slow down the model development. In our application, ocean physics plays a significant role in driving the ocean BGC variability. Here, we show that if we sustain error to a low level in ocean physics, constraining the ocean physical bias errors initially yields a substantial reduction of error; there is first a strong reduction of error followed by subsequent growth of error near the surface because of an excessive nutrient up-welling at the ocean surface. This suggests that BGC parameters in the default version of NorESM have been tuned to compensate for oceanic physical biases.

We have tested two versions of parameter estimation, one with globally uniform parameters and one where parameters can vary spatially. The parameters are trained based on the error of monthly climatology estimates of nitrate, phosphate and oxygen from a coupled reanalysis, which assimilates temperature and salinity monthly climatology. Rerunning the reanalysis with updated parameters shows that both versions yield substantial improvements for quantities used in the training but also for independent quantities (Total alkalinity, DIC, sea-air CO_2 fluxes and primary production). A degradation is found for silicate, which primarily relates to our selection of BGC parameters that cannot fully control the silicate cycle. Overall, reanalysis with global parameter estimates performed better than the one with spatial parameters. This is likely a consequence of inaccuracy in the monthly climatological observations of BGC in regions where data is very sparse (such as in the Southern Ocean). Using several iterations to estimate the parameters can improve the result. A final long reanalysis was performed from 1982-2022 with updated parameters, and it is shown that improvements are sustained with transient historical forcing and with ocean physics having realistic time variability. The parameter estimation reduces the error of BGC in the coupled reanalysis by 20%, 18%, 7%, 27%, and 17% for phosphate, nitrate, oxygen, dissolved inorganic carbon and total alkalinity, respectively, for the period 1993-2020.

This study highlights the potential of our parameter estimation framework to facilitate the calibration of ESM. The method can ingest as many observations as available and optimally account for the respective uncertainty of the observation and model. It can also converge with multiple parameters simultaneously and with the coupled model that will be used at the end. This practice differs from the conventional approach, where uncertain parameters from each model component are often trained in forced configuration first and are often tested in isolation. Some of these parameters need to be re-tuned in the coupled configuration, requiring substantial human and computational resources. Our proposed approach is fully automated, requires fewer human decisions than standard calibration techniques and, in principle, can be applied for different model components simultaneously and, therefore, is computationally more efficient. The iterative ensemble smoother simplifies the portability of the method to other models. One only needs to implement the input/output routines (reading and writing of the model variables and parameters) within the data assimilation code to apply the method.

We show that parameter estimation can substantially reduce errors in the system when we correct the ocean physics bias. Another test (not presented here) found that a comparable improvement can be achieved without correcting for the ocean physics bias –i.e. adjusting model parameters in the ESMs free ensemble run. We still advocate that a preferable pathway would be to train each component of the ESM with the bias of the other components sustained to a low level, e.g. using data assimilation. Following up on the training of the BGC parameter, here we have only constrained errors in the ocean physics, but depending on the objectives, e.g., improving the representation of air-sea CO_2 fluxes, including land and atmospheric constraints, could be valuable. This is now possible with the recent development of NorCPM (Nair et al., 2024; Garcia-Oliva et al., 2023).

Unlike in Singh et al. (2022), we had to reduce the complexity of our ensemble data assimilation method (with the iterative ensemble smoother rather than the dual one step ahead smoother), and we pursued global estimates rather than spatially varying ones. These were motivated by the challenges of our application – which are characterized by sparse and inaccurate observations in some locations, variability and error dominated by sporadic events and slower sensitivity to model parameters compared to error growth from other independent sources. We are still confident that with an improved observation network, one should be able to achieve superior performance with spatially varying parameters.

Acknowledgments

This study was partly funded by Horizon Europe European TRIATLAS (no. 817578), OceanICU (no. 101083922) projects and Trond Mohn Foundation project (no. BFS2018TMT01), the Bjerknes Centre for Climate Research (BCCR) - Centre of Climate Dynamics (SKD) Strategic Project PARCIM (Proxy Assimilation for Reconstructing Climate and Improving Model) and by the Norwegian Research Council project (grant nos. 270061 and 301396). This work has also received a grant for computer time from the Norwegian Program for supercomputing (NOTUR2, project NN9039K) and a storage Grant (NORSTORE, NS9039K). We greatly acknowledge Nicholas Williams (NERSC, Bergen) for reviewing and providing English corrections.

Data Availability Statement

Model simulations presented in this article have been organised and made available. It contains a separate section for each of the experiments presented (NorESM_DP, NorESM_PP, REANA_DP, REANA_PP, REANA_GP, REANA_SP, REANA_GP2 and REANA_IAV_GP2,). Each folder contains the ensemble mean outputs from the ensemble simulations in NetCDF format. The full simulations will be made available on https://archive.sigma2.no with a specific doi upon acceptance of the manuscript. To retrieve the simulations, one can use the following link (700 GB):

wget -c `https://ns9039k.web.sigma2.no/Parameter_Estimation.tar.gz`

References

Allen, J., Eknes, M., & Evensen, G. (2003). An ensemble kalman filter with a complex marine ecosystem model: hindcasting phytoplankton in the cretan sea. In *Annales geophysicae* (Vol. 21, pp. 399–411).

Anderson, J. L. (2001). An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, *129*(12), 2884–2903.

Annan, J., Hargreaves, J., Edwards, N., & Marsh, R. (2005). Parameter estimation in an intermediate complexity earth system model using an ensemble kalman filter. *Ocean modelling*, *8*(1-2), 135–154.

Bagniewski, W., Fennel, K., Perry, M. J., & D'asaro, E. (2011). Optimizing models of the north atlantic spring bloom using physical, chemical and bio-optical observations from a lagrangian float. *Biogeosciences*, *8*(5), 1291–1307.

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography*, *42*(1), 1–20.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., ... others (2013). The norwegian earth system model, noresm1-m–part 1: description and basic evaluation of the physical climate. *Geoscientific Model Development*, *6*(3), 687–720.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., ... others (2021). Norcpm1 and its contribution to cmip6 dcpp. *Geoscientific Model Development Discussions*, 1–84.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., ... others (2019). A global monthly climatology of total alkalinity: a neural network approach. *Earth System Science Data*, *11*(3), 1109–1127.

Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., & Zheng, F. (2014). Seasonal-to-decadal predictions with the ensemble kalman filter and the norwegian earth system model: a twin experiment. *Tellus A: Dynamic Meteorology and Oceanography*, *66*(1), 21074.

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the norwegian climate prediction model. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 32437.

Doney, S. C., Lindsay, K., Caldeira, K., Campin, J.-M., Drange, H., Dutay, J.-C., ... Yool, A. (2004). Evaluating global ocean carbon models: The importance of realistic physics. *Global Biogeochemical Cycles*, *18*(3). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GB002150` doi: https://doi.org/10.1029/2003GB002150

Doney, S. C., Ruckelshaus, M., Emmett Duffy, J., Barry, J. P., Chan, F., English, C. A., ... Talley, L. D. (2012). Climate change impacts on marine ecosystems. *Annual Review of Marine Science*, *4*(1), 11-37. Retrieved from `https://doi.org/10.1146/annurev-marine-041911-111611` (PMID: 22457967) doi: 10.1146/annurev-marine-041911-111611

Doron, M., Brasseur, P., Brankart, J.-M., Losa, S. N., & Melet, A. (2013). Stochastic estimation of biogeochemical parameters from globcolour ocean colour satellite data in a north atlantic 3d ocean coupled physical–biogeochemical model. *Journal of Marine Systems*, *117*, 81–95.

Eknes, M., & Evensen, G. (2002). An ensemble kalman filter with a 1-d marine ecosystem model. *Journal of Marine Systems*, *36*(1-2), 75–100.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, *99*(C5), 10143–10162.

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, *22*(3), 885–908.

Fennel, K., Losch, M., Schröter, J., & Wenzel, M. (2001). Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, *28*(1-2), 45–63.

Flato, G. M. (2011). Earth system models: an overview. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 783–800.

Fransner, F., Counillon, F., Bethke, I., Tjiputra, J., Samuelsen, A., Nummelin, A., & Olsen, A. (2020). Ocean biogeochemical predictions—initialization and limits of predictability. *Frontiers in Marine Science*, *7*, 386.

Friedrichs, M. A., Hood, R. R., & Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated arabian sea data. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(5-7), 576–600.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., … others (2019a). World ocean atlas 2018. vol. 4: Dissolved inorganic nutrients (phosphate, nitrate and nitrate+ nitrite, silicate).

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., … others (2019b). World ocean atlas 2018, volume 3: Dissolved oxygen, apparent oxygen utilization, and dissolved oxygen saturation.

Garcia-Oliva, L., Counillon, F., Bethke, I., & Keenlyside, N. S. (2023). Intercomparison of initialization methods for seasonal-to-decadal climate predictions with the norcpm. *Authorea Preprints*.

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757.

Gharamti, M., Ait-El-Fquih, B., & Hoteit, I. (2015). An iterative ensemble kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. *Journal of Hydrology*, *527*, 442–457.

Gharamti, M., Samuelsen, A., Bertino, L., Simon, E., Korosov, A., & Daewel, U. (2017). Online tuning of ocean biogeochemical model parameters using ensemble estimation techniques: Application to a one-dimensional model in the north atlantic. *Journal of Marine Systems*, *168*, 1–16.

Gharamti, M., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., & Bertino, L. (2017a). Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling*, *112*, 65–89.

Gharamti, M., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., & Bertino, L. (2017b). Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling*, *112*, 65-89. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1463500317300215` doi: https://doi.org/10.1016/j.ocemod.2017.02.006

Goris, N., Tjiputra, J. F., Olsen, A., Schwinger, J., Lauvset, S. K., & Jeansson, E. (2018). Constraining projection-based estimates of the future north atlantic carbon uptake. *Journal of Climate*, *31*(10), 3959–3978.

Gouretski, V., & Reseghetti, F. (2010). On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep Sea Research Part I: Oceanographic Research Papers*, *57*(6), 812–833.

Hemmings, J. C., Srokosz, M. A., Challenor, P., & Fasham, M. J. (2004). Split-domain calibration of an ecosystem model using satellite ocean colour data. *Journal of Marine Systems*, *50*(3-4), 141–179.

Hermanson, L., Smith, D., Seabrook, M., Bilbao, R., Doblas-Reyes, F., Tourigny, E., ... others (2022). Wmo global annual to decadal climate update: A prediction for 2021–25. *Bulletin of the American Meteorological Society*, *103*(4), E1117–E1129.

Hoshiba, Y., Hirata, T., Shigemitsu, M., Nakano, H., Hashioka, T., Masuda, Y., & Yamanaka, Y. (2018). Biological data assimilation for parameter estimation of a phytoplankton functional type model for the western north pacific. *Ocean science*, *14*(3), 371–386.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., ... others (2013). The community earth system model: a framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360.

Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S. E., Dance, S. L., ... others (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278.

Jazwinski, A. H. (2007). *Stochastic processes and filtering theory.* Courier Corporation.

Keppler, L., Landschützer, P., Gruber, N., Lauvset, S., & Stemmler, I. (2020). Mapped observation-based oceanic dissolved inorganic carbon (dic), monthly climatology from january to december (based on observations between 2004 and 2017), from the max-planck-institute for meteorology (mobo-dic_mpim)(ncei accession 0221526), noaa national centers for environmental information [data set]. *NOAA National Centers for Environmental Information. Dataset. doi. org/10.25921/yvzj-zx46*.

Kessler, A., & Tjiputra, J. (2016). The southern ocean as a constraint to reduce uncertainty in future ocean carbon sinks. *Earth System Dynamics*, *7*(2), 295–312.

Kimmritz, M., Counillon, F., Smedsrud, L. H., Bethke, I., Keenlyside, N., Ogawa, F., & Wang, Y. (2019). Impact of ocean and sea ice initialisation on seasonal prediction skill in the arctic. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4147–4166.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., ... others (2020). Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from cmip6 model projections. *Biogeosciences*, *17*(13), 3439–3470.

Landschützer, P., Gruber, N., Bakker, D., Landschützer, P., et al. (2017). An updated observation-based global monthly gridded sea surface pco2 and air-sea co2 flux product from 1982 through 2015 and its monthly climatology. *Asheville, NC: National Centers for Environmental Information*.

Locarnini, M., Mishonov, A., Baranova, O., Boyer, T., Zweng, M., Garcia, H., ... others (2018). World ocean atlas 2018, volume 1: Temperature.

Losa, S. N., Kivman, G. A., & Ryabchenko, V. A. (2004). Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data? *Journal of marine systems*, *45*(1-2), 1–20.

Maier-Reimer, E., Kriest, I., Segschneider, J., & Wetzel, P. (2005). The hamburg ocean carbon cycle model hamocc5. 1-technical description release 1.1.

Mamnun, N., Völker, C., Vrekoussis, M., & Nerger, L. (2022). Uncertainties in ocean biogeochemical simulations: Application of ensemble data assimilation to a one-dimensional model. *Frontiers in Marine Science*, *9*. Retrieved from `https://www.frontiersin.org/articles/10.3389/fmars.2022.984236` doi: 10.3389/fmars.2022.984236

McDonald, C., Bennington, V., Urban, N., & McKinley, G. (2012). 1-d test-bed calibration of a 3-d lake superior biogeochemical model. *Ecological Modelling*, *225*, 115–126.

Nair, A. S., Counillon, F., & Keenlyside, N. (2024). Improving subseasonal forecast skill in the norwegian climate prediction model using soil moisture data assimilation. *Geoscientific Model Development Discussions*, *2024*, 1–29.

Natvik, L.-J., & Evensen, G. (2003). Assimilation of ocean colour data into a biochemical model of the north atlantic: Part 1. data assimilation experiments. *Journal of marine systems*, *40*, 127–153.

Nerger, L., & Gregg, W. W. (2008). Improving assimilation of seawifs data by the application of bias correction with a local seik filter. *Journal of marine systems*, *73*(1-2), 87–102.

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., ... others (2017). Biogeochemical protocols and diagnostics for the cmip6 ocean model intercomparison project (omip). *Geoscientific Model Development*, *10*(6), 2169–2199.

Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., & Zhang, S. (2018). Modeling global ocean biogeochemistry with physical data assimilation: a pragmatic solution to the equatorial instability. *Journal of Advances in modeling earth systems*, *10*(3), 891–906.

Raanes, P. N., Stordal, A. S., & Evensen, G. (2019). Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, *26*(3), 325–338.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite sst analysis for climate. *Journal of climate*, *15*(13), 1609–1625.

Rodgers, K. B., Schwinger, J., Fassbender, A. J., Landschützer, P., Yamaguchi, R., Frenzel, H., ... others (2023). Seasonal variability of the surface ocean carbon cycle: A synthesis. *Global Biogeochemical Cycles*, e2023GB007798.

Roy, S., Broomhead, D. S., Platt, T., Sathyendranath, S., & Ciavatta, S. (2012). Sequential variations of phytoplankton growth and mortality in an npz model: A remote-sensing-based assessment. *Journal of Marine Systems*, *92*(1), 16–29.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., & Korablev, A. (2012). Topaz4: an ocean-sea ice data assimilation system for the north atlantic and arctic. *Ocean Science*, *8*(4), 633–656.

Sakov, P., & Oke, P. R. (2008). A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters. *Tellus A: Dynamic Meteorology and Oceanography*, *60*(2), 361–371.

Schartau, M., & Oschlies, A. (2003). Simultaneous data-based optimization of a 1d-ecosystem model at three locations in the north atlantic: Part i—method and parameter estimates. *Journal of Marine Research*, *61*(6), 765–793.

Simon, E., Samuelsen, A., Bertino, L., & Dumont, D. (2012). Estimation of positive sum-to-one constrained zooplankton grazing preferences with the denkf: a twin experiment. *Ocean Science*, *8*(4), 587–602.

Simon, E., Samuelsen, A., Bertino, L., & Mouysset, S. (2015). Experiences in multiyear combined state–parameter estimation with an ecosystem model of the north atlantic and arctic oceans using the ensemble kalman filter. *Journal of Marine Systems*, *152*, 1–17.

Singh, T., Counillon, F., Tjiputra, J., Wang, Y., & Gharamti, M. E. (2022). Estimation of ocean biogeochemical parameters in an earth system model using the dual one step ahead smoother: A twin experiment. *Frontiers in Marine Science*, *9*, 775394.

Six, K. D., & Maier-Reimer, E. (1996). Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles*, *10*(4), 559–583.

Spitz, Y., Moisan, J., Abbott, M., & Richman, J. (1998). Data assimilation and a pelagic ecosystem model: parameterization using time series observations. *Journal of Marine Systems*, *16*(1-2), 51–68.

Tjiputra, J., Assmann, K., Bentsen, M., Bethke, I., Otterå, O., Sturm, C., & Heinze, C. (2010). Bergen earth system model (bcm-c): model description and regional climate-carbon cycle feedbacks assessment. *Geoscientific Model Development*, *3*(1), 123–141.

Tjiputra, J., Roelandt, C., Bentsen, M., Lawrence, D., Lorentzen, T., Schwinger, J., ... Heinze, C. (2013). Evaluation of the carbon cycle components in the norwegian earth system model (noresm). *Geoscientific Model Development*, *6*(2), 301–325.

Tjiputra, J. F., Negrel, J., & Olsen, A. (2023). Early detection of anthropogenic climate change signals in the ocean interior. *Scientific Reports*, *13*(3006). Retrieved from https://doi.org/10.1038/s41598-023-30159-0 doi: 10.1038/s41598-023-30159-0

Tjiputra, J. F., Polzin, D., & Winguth, A. M. (2007). Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization. *Global biogeochemical cycles*, *21*(GB1001). doi: 10.1029/2006GB002745

Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., ... others (2020). Ocean biogeochemistry in the norwegian earth system model version 2 (noresm2). *Geoscientific Model Development*, *13*(5), 2393–2431.

Wang, B., Fennel, K., Yu, L., & Gordon, C. (2020). Assessing the value of biogeochemical argo profiles versus ocean color observations for biogeochemical model optimization in the gulf of mexico. *Biogeosciences*, *17*(15), 4059–4074.

Wang, Y., Counillon, F., & Bertino, L. (2016). Alleviating the bias induced by the linear analysis update with an isopycnal ocean model. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 1064–1074.

Wang, Y., Counillon, F., Bethke, I., Keenlyside, N., Bocquet, M., & Shen, M.-l. (2017). Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. *Ocean Modelling*, *114*, 33–44.

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., ... Gao, Y. (2019). Seasonal predictions initialised by assimilating sea surface temperature observations with the enkf. *Climate Dynamics*, *53*(9), 5777–5797.

Westberry, T., Behrenfeld, M., Siegel, D., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*(2).

While, J., Haines, K., & Smith, G. (2010). A nutrient increment method for reducing bias in global biogeochemical models. *Journal of Geophysical Research: Oceans*, *115*(C10).

Zweng, M., Seidov, D., Boyer, T., Locarnini, M., Garcia, H., Mishonov, A., ... others (2019). World ocean atlas 2018, volume 2: Salinity.