JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2023MS003941

Using Shortened Spin-Ups to Speed Up Ocean Biogeochemical Model Optimization

S. Oliver[1], S. Khatiwala[2], C. Cartis[3], Ben Ward[4], and Iris Kriest[5]

[1]National Oceanography Centre, Southampton, UK, [2]Department of Earth Sciences, University of Oxford, Oxford, UK, [3]Mathematical Institute, University of Oxford, Oxford, UK, [4]School of Ocean and Earth Science, University of Southampton, Southampton, UK, [5]GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Kiel, Germany

Abstract The performance of global ocean biogeochemical models can be quantified as the misfit between modeled tracer distributions and observations, which is sought to be minimized during parameter optimization. These models are computationally expensive due to the long spin-up time required to reach equilibrium, and therefore optimization is often laborious. To reduce the required computational time, we investigate whether optimization of a biogeochemical model with shorter spin-ups provides the same optimized parameters as one with a full-length, equilibrated spin-up over several millennia. We use the global ocean biogeochemical model MOPS with a range of lengths of model spin-up and calibrate the model against synthetic observations derived from previous model runs using a derivative-free optimization algorithm (DFO-LS). When initiating the biogeochemical model with tracer distributions that differ from the synthetic observations used for calibration, a minimum spin-up length of 2,000 years was required for successful optimization due to certain parameters which influence the transport of matter from the surface to the deeper ocean, where timescales are longer. However, preliminary results indicate that successful optimization may occur with an even shorter spin-up by a judicious choice of initial condition, here the synthetic observations used for calibration, suggesting a fruitful avenue for future research.

Plain Language Summary Global ocean biogeochemical models allow us to simulate ocean biological and chemical variables throughout the global ocean, and are necessary for climate change projections. They are very computationally expensive due to the required spin-up needed for the model to reach a steady state, and so any improvements to the model need to be sought in an efficient way. One way we investigate here is to first make the model less computationally expensive by using a shorter spin-up, and then we apply an efficient optimization algorithm to tune the model to observations. By shortening the spin-up from 3,000 to 2,000 years we show that we can still reach a successfully optimized model. Preliminary results indicate that successful optimization may occur with an even shorter spin-up when the biogeochemical model is initialized from a more accurate state, which highlights a future avenue for research that may encourage more systematic tuning of computationally expensive ocean biogeochemical models.

Key Points:

- Global ocean biogeochemical models are computationally expensive due to the long spin-up time required to reach equilibrium
- A shortened spin up of 2,000 years during parameter optimization can be successfully optimized
- How short a spin-up one can successfully optimize is influenced by the parameters being calibrated and the initial conditions of the model

Correspondence to:
S. Oliver,
sophy.oliver@noc.ac.uk

Citation:
Oliver, S., Khatiwala, S., Cartis, C., Ward, B., & Kriest, I. (2024). Using shortened spin-ups to speed up ocean biogeochemical model optimization. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS003941. https://doi.org/10.1029/2023MS003941

Received 21 JUL 2023
Accepted 12 AUG 2024

Author Contributions:
Conceptualization: S. Khatiwala
Formal analysis: S. Oliver
Methodology: S. Oliver, S. Khatiwala, C. Cartis
Supervision: S. Khatiwala, C. Cartis
Visualization: S. Oliver, Iris Kriest
Writing – original draft: S. Oliver
Writing – review & editing: S. Oliver, S. Khatiwala, C. Cartis, Ben Ward, Iris Kriest

1. Introduction

Ocean biogeochemical models simulate biogeochemical processes within the ocean, including interactions between dissolved nutrients, plankton, and sinking particles, and are used to understand how the ocean draws down carbon via the biological carbon pump (Christina & Passow, 2007). While no ocean biogeochemical model will represent reality perfectly, the goal is to do so as well as possible. A crucial aspect toward this goal is the process of parameter optimization, which involves calibrating the parameters (biogeochemical model constants) until a minimum discrepancy is found between the model and observations (Evans, 2003; Hourdin et al., 2017; Ward et al., 2010).

Global ocean biogeochemical models are particularly computationally expensive due to the long turnover time taken for waters sinking out of the ocean surface to flush through the ocean interior, which is of order 1,000 years, taking longest for the Pacific Ocean (Khatiwala et al., 2012; Shah et al., 2017). Global ocean biogeochemical models therefore need to "spin-up" to equilibrate ocean biogeochemistry through the balance of circulation and biogeochemical fluxes, which can take 3,000–10,000 years (Primeau & Deleersnijder, 2009; Wunsch & Heimbach, 2008).

The length of time global models are spun-up in practice varies greatly, from decades to millennia (Séférian et al., 2016, 2020). In addition, the models are also initialized differently, as some models start from observed (inorganic) biogeochemical tracer concentrations, while others start from constant values or values derived from previous model simulations (e.g., Séférian et al., 2016). Hence, some models may start near a "desired" (observed) state, which may, nevertheless, not be the ultimate model state arising from model dynamics; these models are often spun up only for a few decades. Models starting from a previous simulation may possibly be close to their intrinsic equilibrium state, and only require short spin-ups. Models starting from spatially constant values are typically spun up on long time scales, which allows enough time for the tracers to adjust to their dynamical forcing and parameterizations.

The large computational expense of these models has encouraged the use of more efficient simulation of biogeochemical tracers, such as the Transport Matrix Method (TMM, Khatiwala, 2007; Khatiwala et al., 2005) or similar methods (e.g., DeVries & Primeau, 2011; Primeau, 2005), and Newton-Krylov methods (e.g., Bardin et al., 2014; Fu & Primeau, 2017; Khatiwala, 2008; Li & Primeau, 2008; Lindsay, 2017). It has also encouraged more efficient optimization, such as by applying reduced model complexity and surrogate optimization (e.g., Kuhn & Fennel, 2019; Priess et al., 2013), and efficient derivative-free optimization algorithms (e.g., Kriest et al., 2017, 2020; Oliver et al., 2022; Sauerland et al., 2019). Here, we explore another method to further reduce the computational expense of global ocean biogeochemical model optimization, namely the calibration of a model after a shorter spin-up. If successful, the computational expense saved may be enough to allow systematic parameter calibration of computationally expensive global ocean biogeochemical models which in many cases were previously deemed only suitable for manual calibration. It will also allow developers to run more optimization iterations, providing more scope to converge closer to an optimal parameter configuration, resulting in better model performance.

We thereby posed the null hypothesis that with shorter equilibration times, an optimization algorithm does not have sufficient information to recover parameter values that simulate the true equilibrated fields. We calibrated a global ocean biogeochemical model with various spin-up lengths to synthetic observations (generated by a previous model run), where the optimized parameters were known, and therefore it was possible to determine if optimization was successful. We then go on to reject the null hypothesis by successfully optimizing with a shortened equilibration time. However, we also delved into the relationship between misfits after different spin-up lengths across the entire parameter space to try and understand why some optimizations were unsuccessful. Section 2 describes the methodology, including the ocean biogeochemical model, general circulation framework, parameters for optimization, chosen misfit function, and the design of our experiments. Section 3 shows the results of the interrogation and optimization of a shortened model spin-up, which are then discussed in Section 4, and final conclusions made in Section 5.

2. Methods

2.1. The Choice of Model and Parameters to Optimize

The chosen global ocean biogeochemical model, the Model of Oceanic Pelagic Stoichiometry (MOPS-2.0: Kriest & Oschlies, 2015), simulates the cycling of nine biogeochemical tracers within the water column. The tracers include dissolved inorganic phosphate, nitrate and carbon, dissolved organic phosphate and nitrate, alkalinity, dissolved oxygen, phytoplankton, zooplankton, and detritus. These tracers allow MOPS to track the cycling of nutrients and oxygen between plankton and the marine environment. In this study MOPS was coupled to an "offline" version of the MITgcm (2.8° horizontal resolution with 15 vertical levels) ocean circulation model (Marshall et al., 1997) using the TMM (Khatiwala, 2007, 2018; Khatiwala et al., 2005). This configuration of MITgcm has been widely used in previous studies (e.g., Dutkiewicz et al., 2005; Kriest et al., 2017; Lauderdale et al., 2013, 2017; Parekh et al., 2005), and has been shown to capture the large-scale global ocean circulation. The TMM involves coupling MOPS to an annually repeating ocean circulation. There is no interannual variability in the circulation or the biogeochemical forcing.

The six biogeochemical parameters selected for optimization are described in Table 1, as in Kriest et al. (2017) and Oliver et al. (2022). These chosen biogeochemical parameters were originally deemed by Kriest et al. (2017) to be particularly influential to the global distributions of dissolved oxygen and nutrients. Several of them are poorly constrained by direct observations, hence they would likely benefit from parameter optimization. We note that four of the parameters to be optimized (I_C, K_{PHY}, μ_{ZOO}, and k_{ZOO}) are closely related to the fast

Table 1
Parameters Chosen for Calibration

Name	Description	Units
$R_{O_2:P}$	Ratio of oxygen consumption to phosphate release during remineralization when oxygen is available	$mmolO_2$:mmolP
I_C	Phytoplankton half-saturation for light	Wm^{-2}
K_{PHY}	Phytoplankton half-saturation for phosphate	$mmolPm^{-3}$
μ_{ZOO}	Zooplankton maximum grazing rate	d^{-1}
k_{ZOO}	Zooplankton quadratic mortality rate	$(mmol\ Pm^{-3})^{-1}\ d^{-1}$
$b*$	Exponent of the remineralization Martin Curve (Martin et al., 1987) where the vertical flux of sinking organic matter is related to a constant raised to the power of $-b*$	

biogeochemical turnover in the euphotic layer, while especially $b*$ and $R_{O_2:P}$ can affect the deep and large-scale nutrient and oxygen distribution.

2.2. Model Spin-Up

The spin up time required for MOPS to reach equilibrium is taken to be 3,000 years, which is consistent with Kriest et al. (2017) and Oliver et al. (2022). By this time most of the tracers within MOPS have approached equilibrium (see Figure 2 in Kriest and Oschlies (2015)). Figure 1 displays the work flow of model runs, parameter configurations and respective spin-up lengths required for the optimization experiments, and Table 2 states the parameter values for the different configurations.

Starting from spatially uniform initial conditions based on average modern ocean concentrations, the model was first spun up with a parameter vector $\mathbf{p}_{ini}$ over 3,000 years (see Section 2.4 for how $\mathbf{p}_{ini}$ was chosen). The resulting vectors of global tracer concentrations $\mathbf{x}_{ini}$ for each of the model tracers provides the initial conditions for the further experiments. Starting from $\mathbf{x}_{ini}$, the model was then simulated for 3,000 years with parameter vector $\mathbf{p}_{target}$ to provide the reference solution $\mathbf{x}_{target}$ (synthetic observations). Likewise, starting from $\mathbf{x}_{ini}$ the model parameters

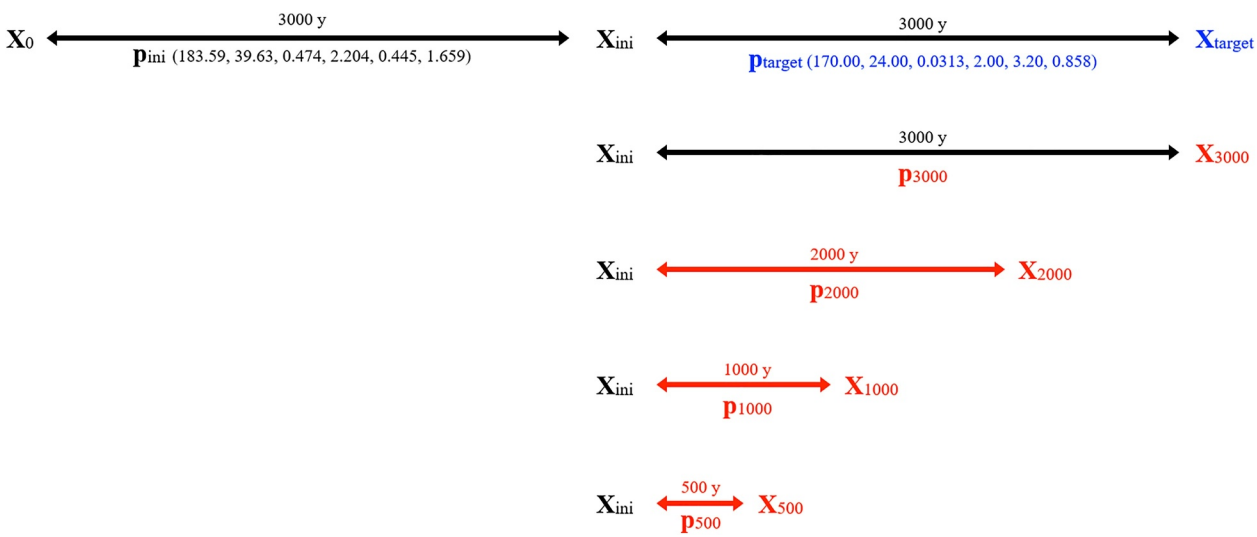

Figure 1. Model runs, parameter configurations and respective spin-up lengths required for the optimization experiments. $\mathbf{x}_0$ are uniform global tracer concentrations and $\mathbf{x}_{ini}$, $\mathbf{x}_{target}$, and $\mathbf{x}_{3000,2000,\ldots}$ are resulting vectors of global tracer concentrations modeled by MOPS. $\mathbf{p}_{ini}$ and $\mathbf{p}_{target}$ are model parameter vectors specified in Table 2 and $\mathbf{p}_{3000,2000,\ldots}$ are parameter vectors chosen by the optimization algorithm during an optimization experiment. At the start of an optimization experiment, $\mathbf{p}_{3000,2000,\ldots}$ are the same as $\mathbf{p}_{start}$ (specified in Table 2), then they vary throughout the optimization experiment (see Figure 3), until the optimization algorithm converges to the optimal parameter vectors (also specified in Table 2) by reducing the misfit between the model and synthetic observations to below a defined threshold (see Section 2.3). The model spin-up length is either 500, 1,000, 2,000 or 3,000 years, which are indicted as horizontal lines and colored black for a 3,000 years spin up and in red for all spin ups shorter than 3,000 years. $\mathbf{x}$ and $\mathbf{p}$ are written in blue for the model run providing the synthetic observations, and in red for the model runs to be calibrated during the optimization experiments.

Table 2
Table of Results for Optimization Experiments With MOPS Initiated From $\boldsymbol{x}_{ini}$

Parameters	$R_{O_2:P}$	I_C	K_{PHY}	μ_{ZOO}	k_{ZOO}	$b*$	Misfit	Success?
Upper bound	200.000	48.000	0.5000	4.000	10.000	1.800		
Lower bound	150.000	4.000	0.0001	0.100	0.000	0.400		
Optimal ($\mathbf{p}_{target}$)	170.000	24.000	0.0313	2.000	3.200	0.858		
Initial conditions ($\mathbf{p}_{ini}$)	183.590	39.630	0.4740	2.204	0.445	1.659		
Start ($\mathbf{p}_{start}$)	197.000	18.901	0.0549	2.925	3.155	0.543		
MOPS optimization results								
500 years	200.000	35.087	0.0044	1.800	3.270	0.400	42.562	No
1,000 years	199.488	28.887	0.0001	2.442	1.790	0.622	9.639	No
2,000 years	177.477	25.278	0.0341	2.310	3.585	0.814	0.623	Yes
3,000 years	170.589	24.128	0.0045	2.102	3.886	0.867	0.041	Yes

Note. The upper section of the table states (row 1) the upper and (row 2) lower parameter bounds, (row 3) the optimal parameter values to be recovered, (row 4) the parameter configuration used to create $\mathbf{x}_{ini}$, the equilibrated tracer distributions from which MOPS was initialized, and (row 5) the location in parameter space from which DFO-LS starts each optimization. The lower section of the table show the optimized parameter values resulting from the optimization experiments with various spin-up lengths. The penultimate column shows the lowest misfit reached by the optimizer. The final column states whether the optimization was successful or not, which was determined by the misfit being minimized to within realistic observational noise.

listed in Table 1 were then optimized against $\mathbf{x}_{target}$ after spin-up times of 500, 1,000, 2,000, and 3,000 years. To avoid a premature convergence of optimization due to initial parameter values that are already close to the target parameters, the initial parameter guess $\mathbf{p}_{start}$ was chosen to be different from the target (see Section 2.4 for how this was chosen). We note that in an ideal case, the resulting optimal parameter sets ($\mathbf{p}_{500}$, $\mathbf{p}_{1000}$,…) correspond exactly to $\mathbf{p}_{target}$. The difference of tracer distributions resulting from these optimizations ($\mathbf{x}_{500}$, $\mathbf{x}_{1000}$,…) to the synthetic observations provided by $\mathbf{x}_{target}$ form the basis for the formulation of the misfit function (see Section 2.3).

Given the variety of methods to initialize the models mentioned previously (e.g., observed biogeochemical tracer concentrations, constant values, etc.) and the potential consequences on the time scale of their temporal adjustment, in a second set of experiments we repeated the optimizations, this time starting the model from $\mathbf{x}_{target}$. These experiments aimed to investigate the effect of a good prior knowledge of the model state on optimization and the effect on spin-up times. Any good guess by the optimization algorithm of an optimal parameter vector is likely to require only minor adjustment by the global tracer distributions. Success or failure may thus become visible after a short time.

2.3. Minimizing the Misfit Function

Fifty seven misfits were calculated for the three well observed oceanic tracers (oxygen, phosphate, and nitrate) for 19 previously established biome regions of similar ocean biogeochemical properties (see Figure 2 in Oliver et al. (2022), and references therein), as in Equation 1. These 57 regional, tracer specific misfits are

$$r_{qj}(\boldsymbol{p}) = \frac{1}{\left(\sum_{i=1}^{N} \mathbf{x}_{\text{target, iq}} \frac{V_i^{i\in j}}{V_j}\right)\frac{V_j}{V_T}} \sqrt{\left(\sum_{i=1}^{N} (\mathbf{x}_{iq}(\boldsymbol{p}) - \mathbf{x}_{\text{target, iq}})^2 \frac{V_i^{i\in j}}{V_j}\right)\frac{V_j}{V_T}}, \quad (1)$$

where $\boldsymbol{p}$ is the parameter values (e.g., $\mathbf{p}_{3000}$ or $\mathbf{p}_{2000}$ from Figure 1), q the tracer index from 1 to 3, j the region index from 1 to 19, and i the index of the grid box within the model domain, ranging from 1 to N, which is 52,749. $\mathbf{x}_{iq}(\boldsymbol{p})$ is the model solution with parameters $\boldsymbol{p}$ at grid point i for tracer q (e.g., $\mathbf{x}_{3000}$ or $\mathbf{x}_{2000}$ from Figure 1), and $\mathbf{x}_{\text{target,iq}}$ the corresponding target synthetic observations. The misfit is normalized by the volume-weighted mean tracer concentration for that region and weighted, first, by individual grid point volumes V_i relative to the volume V_j of region j and, second, by the region's total volume relative to the global ocean volume V_T, where $V_T = \sum_{j=1}^{19} V_j$.

Essentially, each of the regional, tracer specific misfits r is a root sum square error between the modeled tracer and the target tracer within one region, weighted by volume, and also normalized by the volume-weighted mean of the target tracer for the same region.

These 57 regional, tracer specific misfits were provided to the chosen optimization algorithm for minimization, but for illustrative purposes they were also combined into a single global misfit value in Equation 2. A plot of this global misfit as a function of time for two different choices of model parameters is shown in Figure A1.

$$f(\boldsymbol{p}) = \sum_{q=1}^{3} \sum_{j=1}^{19} r_{qj}(\boldsymbol{p})^2. \tag{2}$$

A baseline misfit was also defined, where the term $(\mathbf{x}_{iq}(\boldsymbol{p}) - \mathbf{x}_{\text{target,iq}})$ of Equation 1 is replaced by the uncertainty associated with the equivalent real observations of oxygen, phosphate and nitrate (Garcia et al., 2018a, 2018b). Since our misfit is defined with respect to annual mean data we require an annual mean standard deviation without the variability of the seasonal cycle. To create this, we take the numerical mean (weighted by the number of observations) of the monthly standard deviations provided in the World Ocean Atlas database (WOA18, Garcia et al., 2018a, 2018b) for the upper 800 m (phosphate and nitrate) or 1,500 m (oxygen). We then concatenate this with the annual standard deviation below those depths, and finally linearly interpolate onto the model grid. The baseline misfit indicates the threshold for which, if the observations used for calibration were real and not synthetic, any further misfit reduction by an optimization algorithm is within observational uncertainty (Oliver et al., 2022). In the misfit function we do not weight the synthetic observations to these uncertainties, as our synthetic observations have no uncertainties and therefore overfitting is not an issue and convergence to a misfit of zero is possible. Instead we use these uncertainties to indicate when overfitting might become an issue, if the observations were real and not synthetic.

The chosen algorithm to optimize the six biogeochemical parameters was the derivative-free optimization using least squares (DFO-LS) algorithm (Cartis et al., 2019). This is a deterministic, iterative, local optimization algorithm which minimizes a function $f(\boldsymbol{p})$ with bounded variables $\boldsymbol{p}$, and was used as in Oliver et al. (2022). The parameter bounds were previously chosen by Kriest et al. (2017) and are specified in Table 2. They are wide to allow the optimization algorithm to explore a wide range of potential parameters. An optimization run was deemed successful if DFO-LS managed to reduce the global misfit to below the baseline threshold.

2.4. The Influence of Spin-Up Length on the Misfit Landscape

In addition to the optimization experiments, we investigated how the relationship between assumed equilibrated misfits (calculated after a 3,000 years spin-up) and non-equilibrated misfits (calculated after various shorter spin-ups) vary throughout the parameter space. MOPS was evaluated multiple times across the parameter space and the misfits calculated after 100, 500, 1,000, 2,000, and 3,000 years. We used the stratified-random strategy Latin hypercube Sampling (LHS: McKay et al., 1979) to efficiently sample the parameter space $n \times 15$ times (where n is the number of parameters). This resulted in MOPS evaluations at 90 different locations across the 6-dimensional parameter space, leaving no large gaps of unsampled parameter space. One of these MOPS evaluations, which yielded an above-average misfit value, provided the parameter values for $\mathbf{p}_{\text{start}}$ (see Section 2.2). The MOPS evaluation which yielded the largest misfit provided the parameter values for $\mathbf{p}_{\text{ini}}$.

3. Results

Figure 2 shows the misfit reduction during the four optimization experiments where MOPS was initiated from $\mathbf{x}_{\text{ini}}$, and Figure 3 shows how well the target parameters were recovered (which is particularly what optimization strives to do). To successfully recover the optimal parameter values close enough that the misfit was reduced to within realistic observational noise, the length of spin-up needed to be at least 2,000 years. As expected, calibrating a 3,000 years spin up resulted in the lowest misfit (see Table 2) and convergence closest to the target parameters ($\mathbf{p}_{\text{target}}$), with the exception of K_{PHY}. The misfit is least sensitive to K_{PHY} (see Figure 3 and Section 4 in Oliver et al. (2022)) and therefore convergence to this optimal parameter is not necessary to achieve a low misfit. When applying a spin-up shorter than 2,000 years the optimization failed to recover the target parameters and the resulting final misfit remained at least 10 times the observational noise. With a 500 years spin-up, DFO-LS is unable to improve the misfit at all throughout the optimization, and even triggers several restarts to try and move

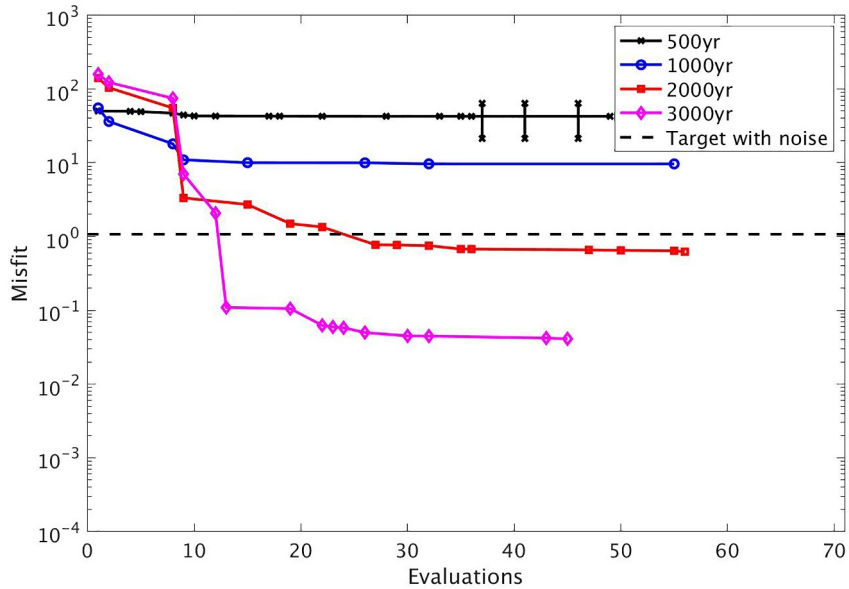

Figure 2. Misfit reduction during each optimization experiment with various lengths of model spin-up for MOPS initiated from $\mathbf{x}_{ini}$. Experiments are shown with a model spin-up of 500 (black line with crosses), 1,000 (blue line with circles), 2,000 (red line with squares), and 3,000 (magenta line with diamonds) years. The baseline misfit (horizontal black dashed line) indicates the threshold below which any reduction in misfit is within realistic observational noise. The misfits have been plotted on the *y*-axis, with the number of evaluations of MOPS that were required by DFO-LS to reach that misfit on the *x*-axis. Only evaluations of MOPS which had a lower global misfit than previously achieved in each experiment have been plotted with markers, and vertical lines indicate a DFO-LS restart, which DFO-LS triggers when optimization progress is particularly poor. All four experiments start from the same point in parameter space, but the first evaluations yield different misfit values due to their various spin-up lengths.

out of local minima it may be trapped in, with little success. It seems DFO-LS attempts to push the parameters toward a minimum in the misfit outside of their set bounds (see Figures 3a and 3f), such as one with a very low b^* value of less than 0.4.

There may be several causes for the effects of spin-up times on the misfit function, the convergence to model parameters and optimization performance. By simulating an ensemble of 90 different parametric model setups, generated through LHS, and examining the misfit in relation to spin-up length and parameter value, we aim to further disentangle the effects of parameter values and spin-up length. The relationships between misfits calculated after different spin-ups are shown in Figure 4. For a short spin-up of 100 years, as in subplot (a), the relationship between a 3,000 years misfit and a 100 years misfit does not fall on a single linear line, but instead is split into two branches, the lower of which has a negative slope. This lower branch shows no correlation when comparing a 500 years misfit to a 3,000 years misfit (Figure 4c), and a slightly positive correlation for a 1,000 years misfit (Figure 4e). By a 2,000-year spin-up, the relationship between the misfit after 2,000 and 3,000 years fits onto a single linear line, with no second lower branch (Figure 4g).

When the relationship between the misfit after 3,000 years and 500 years is separated out into the contribution by individual tracers (see Figure 4, subplots (b, d, and f) for phosphate, oxygen, and nitrate respectively), the prominence of the second lower branch varies slightly. The overall relationship is most strongly linear for phosphate. The stronger sensitivity of nitrate and oxygen to spin-up time is possibly related to their flexible global inventory, which ultimately adjusts through the coupling of deeper processes (i.e., remineralization and denitrification) and surface boundary conditions (i.e., air-sea gas exchange, nitrogen fixation; see also Kriest and Oschlies (2015)). In addition, because this adjustment ultimately affects the global bias, both tracers typically dominate the misfit function (Kriest et al., 2017).

Figure 4 further illustrates the relationships between misfits in their dependence on the Martin b exponent (b^*), and highlights the second lower branch in Figures 4a and 4c corresponded to low values of b^*. A high b^* corresponds to shallow remineralization, and therefore only affects regions with a dynamic physical turnover. Hence,

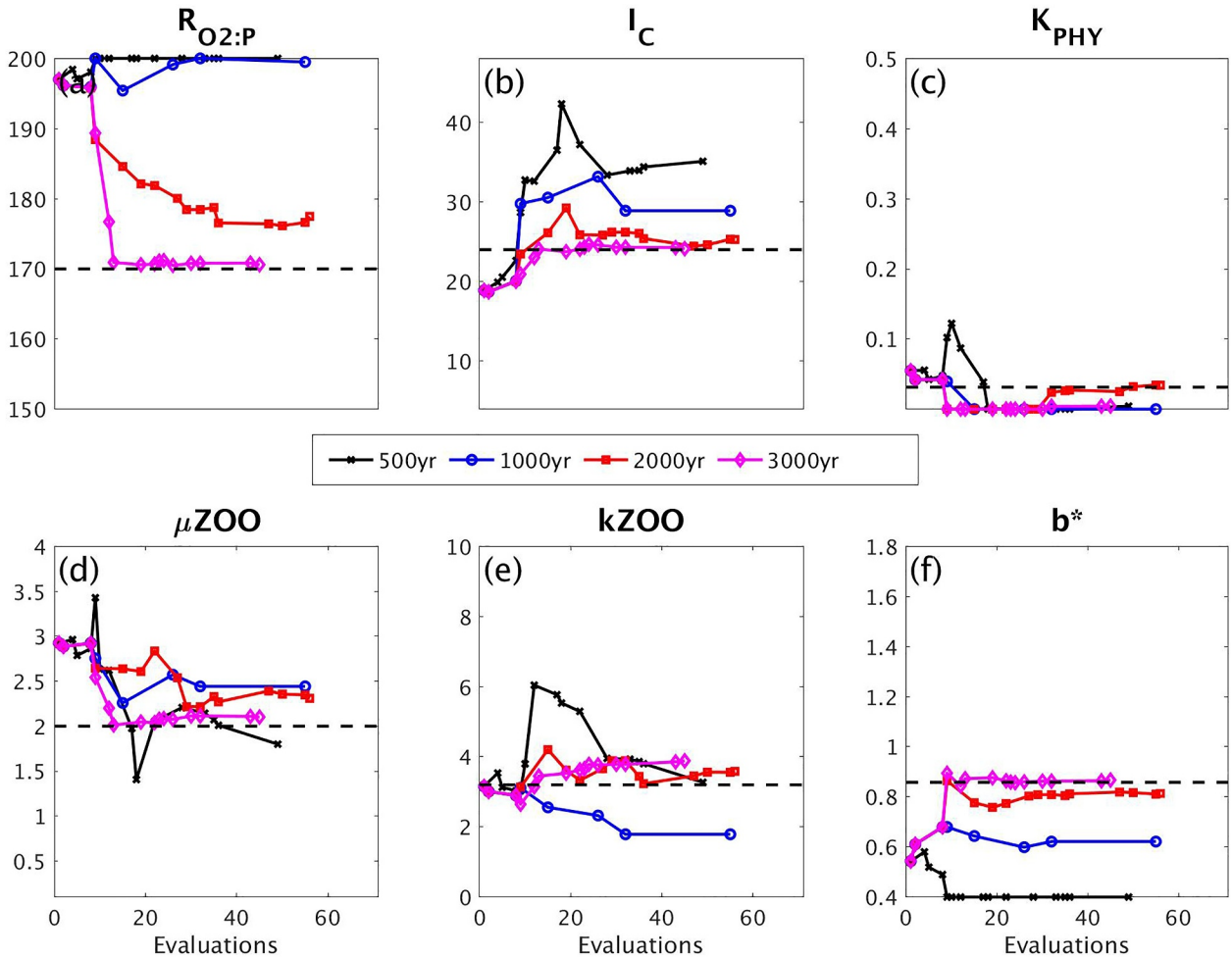

Figure 3. Parameter optimization of MOPS initiated from $\mathbf{x}_{ini}$ with a 500 (black line with crosses), 1,000 (blue line with circles), 2,000 (red line with squares) and 3,000 (magenta line with diamonds) year spin-up, for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO}, and (f) b^*. The parameter values have been plotted on the y-axis, with the number of evaluations of MOPS that were required by DFO-LS to reach that parameter value on the x-axis. Also shown are the known optimal parameter values (horizontal black dashed line). Only parameter values which corresponded to evaluations of MOPS which resulted in a lower global misfit than previously achieved in each experiment (see Figure 2) have been plotted with markers.

the misfit of model simulations with a high b^* is similar after 500 and 3,000 years. On the other hand, a low b^* affects the deep ocean, which is affected by more sluggish circulation acting on millennial time scales. As a consequence, the total misfit after 500 years of spin-up is less variable in these runs with a low b^* than the response of the misfit after 3,000 years, which is highly variable. This variation is caused by large scale circulation which may resupply nutrient-rich or oxygen depleted waters to the surface after 3,000 years. The horizontal offset between the two branches is due to the difference between a less adjusted ocean (which will average around the misfit between the initial conditions and the target state) and a more adjusted ocean, which will fall along a distribution of misfits depending on the parameter values. To put the influence of b^* into context of the other parameters, Figure A2 illustrates the relationship between misfits after a 3,000 years and a 500 years spin-up, with the color of each subplot's data points corresponding to each of the six parameter values the 90 LHS MOPS runs were configured with. This shows that in addition to b^*, I_C also has a similar but lesser influence on the branching in these relationships, where a particularly high I_C is enough to cause an outlier of a low b^* to lie on the upper branch.

Figure 5 displays the additional set of simulations in which the initial conditions were the same as the target state. Promisingly, we achieve a successful optimization with a spin-up length as short as 100 years. In these experiments we initiate MOPS from $\mathbf{x}_{target}$ (created using $\mathbf{p}_{target}$) then run MOPS with some parameter vector different from $\mathbf{p}_{target}$. Therefore, with time the ocean will only diverge further and further from $\mathbf{x}_{target}$, which is why DFO-

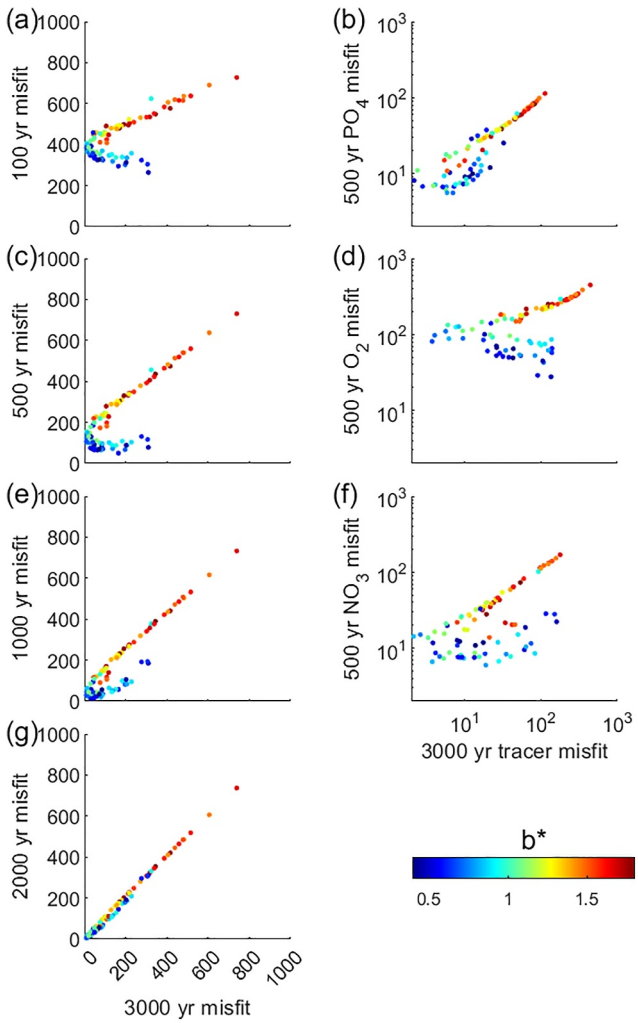

Figure 4. The left-hand column of subplots show the relationships between equilibrated misfits calculated after a 3,000 years spin-up on the *x*-axes and non-equilibrated misfits calculated after a (a) 100, (c) 500, (e) 1,000 and (g) 2,000 years spin-up length on the *y*-axes, created by 90 MOPS runs sampled on a 6-dimensional Latin Hypercube, all initiated from $\mathbf{x}_{ini}$. The colors correspond to the value of the Martin b exponent (b^*) in each MOPS run (though note the other parameter values also vary—see Figure A2). The right-hand column of subplots break down subplot (c) further by showing the relationship between the misfit associated with (b) phosphate, (d) oxygen, and (f) nitrate after a 3,000 years spin-up and a 500 years spin-up. Note the logarithmic scale for panels (b, d, and f).

LS could achieve the lowest misfit for the experiment using the shortest spin-up length. The fact that these optimization experiments were successful regardless of spin up length suggests that a judicious choice of initial conditions may allow for very short spin-ups, but further work is needed to determine exactly how close one needs to be, while taking into account the sparsity and uncertainty of real observations.

4. Discussion

4.1. The Success of This Computational "Short Cut"

For MOPS, with this specific general circulation model and selected biogeochemical parameters for optimization, we showed one can successfully calibrate MOPS to observations while shortening the spin-up length from 3,000 years to 2,000 years. This achieves a computational saving of one third for each evaluation of MOPS, however the optimization algorithm required more evaluations to successfully optimize the model, thereby negating this computational saving. The number of evaluations required by an optimization algorithm may be dependent on certain choices, such as the model, the optimization algorithm, the parameters being optimized and the model's initial conditions. Therefore, this additional computational expense of further evaluations required to optimize a shorter spin-up is difficult to generalize and may not always be the case. Here we further investigated the influence of the parameters being optimized and the model's initial conditions on this computational saving, to better aid future studies attempting a similar route. This is a worthwhile endeavor because during an optimization study the model is evaluated multiple times, and dependent on the computational resources available sometimes every evaluation needs to be run sequentially, even if the chosen optimization algorithm allows for parallelism. Therefore, any computational saving is non-trivial and may allow model developers to achieve closer convergence within the same computational expense budget allocated to a certain model development task.

4.2. The Influence of Certain Parameters

Successful recovery of model parameters from the synthetic observations was not possible with a spin-up length shorter than 2,000 years, as DFO-LS could not converge to the optimal parameters. Successful optimization occurs for a shortened spin-up if the lowest misfits after a shortened spin-up correspond to the lowest misfits after a 3,000 years spin-up. However, this was not the case for spin-up lengths shorter than 2,000 years, due to parameters influencing the transport of material to the deeper ocean, namely the Martin b exponent (b^*). This exponent influences the remineralization depth of sinking organic matter (Martin et al., 1987), where a high b^* causes shallower remineralization. Therefore, the influence of a parameter perturbation in this part of parameter space will be more restricted to the surface ocean, where oceanic timescales are shorter. Conversely, a low b^* results in more remineralization at depth, and hence a greater transport of organic matter and nutrients to the deep ocean conveyor belt where oceanic timescales are longer (Khatiwala et al., 2012; Primeau & Deleersnijder, 2009; Wunsch & Heimbach, 2008). On long time scales, b^* in concert with large scale circulation is very effective on distributing nutrients and oxygen along the conveyor belt (Kriest et al., 2012; Kwon & Primeau, 2006; Kwon et al., 2009). Any misfit signal after a parameter perturbation in this area of parameter space is far-reaching and takes longer than 2,000 years to reach equilibrium (Kriest & Oschlies, 2015). Therefore, how short a spin up we can use during parameter optimization (and hence the computational performance of parameter optimization) highly depends on the regime of parameter space one is targeting. This is particularly the case for nitrate and oxygen, as these tracers are influenced by the process of denitrification in MOPS, which increases the time required for the model to equilibrate (Kriest & Oschlies, 2015).

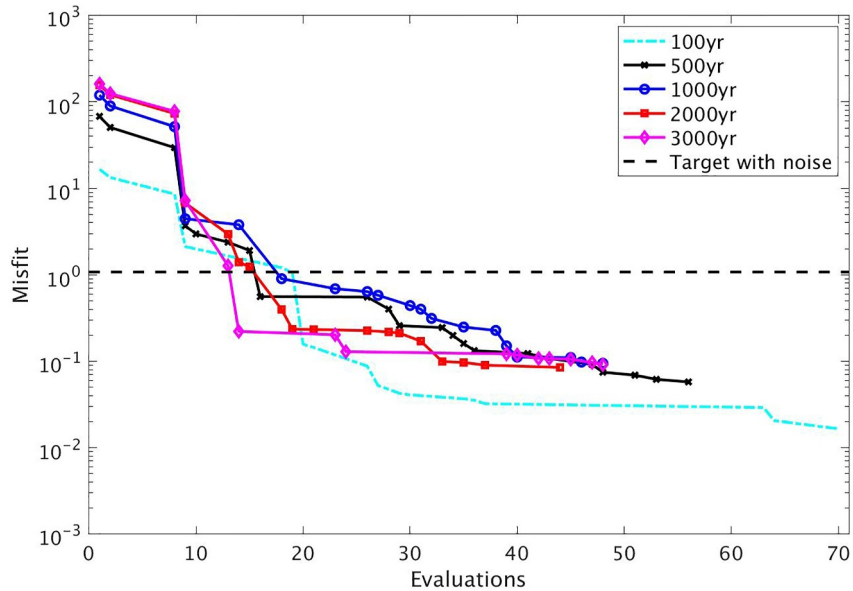

Figure 5. Same as Figure 2, but with MOPS initiated from $\mathbf{x}_{\text{target}}$, and with an additional optimization experiment with a 100 years spin-up length (cyan dotted dashed line).

4.3. The Influence of the Initial Conditions

Our optimizations that start from the synthetic (target) tracer distributions show a fast convergence of optimization for all spin-up lengths and the optimization algorithm does not require many additional evaluations of MOPS to do so. This indicates that the system can be adequately constrained only by looking at upper ocean processes, of which a 100 years spin up is limited to influencing, particularly if organic tracers are also included in the misfit function (Kriest et al., 2023). This also indicates a significant computational saving can be achieved when optimizing a shortened spin-up, if good initial tracer distributions are chosen. However, in this identical twin experiment we assume perfect knowledge and reproduction of the "real" world (the model). Translating this into calibration against observations, a perfect model started from observed tracers will only exhibit a small drift, and the effects of any change in model parameters should become evident early on. However, in practice, global model setups will diverge from the real world both physically and in terms of biogeochemical parameterizations. Hence, when initializing these models from observations (a common practice; see Séférian et al. (2016)) they may be far from the intrinsic final model state, and the effects of tuning will only become evident after millennia.

4.4. Caveats

This study calibrates to synthetic observations that we know the model can reproduce. Therefore we have not investigated the consequences of the model being unable to exactly recreate reality, as would be the case when calibrating to real observations. These could be investigated in future work by adding noise to the synthetic calibration data, whereby the model could be improved to a state closer to the synthetic observations but not exactly to it, which is more realistic. The ability to reach this closer state may depend on the length of the model spin-up, especially when the model is started from an initial state which is far from its intrinsic equilibrium.

5. Conclusions

The specific question addressed in this study is how short a spin up can still yield correct optimal parameters during parameter optimization of a global ocean biogeochemical model. It is important to emphasize that this is different from the broader question of how long one needs to run a model to reach equilibrium or how (and how quickly) one can achieve it, which is on the order of at least 3,000 years (Primeau & Deleersnijder, 2009; Wunsch & Heimbach, 2008). It is common practice for biogeochemical modelers to tune parameters by running the model for several decades to a few hundred years. While this is done for entirely practical reasons, what our study shows,

and in a much more systematic way than previously, is that one cannot simply use such a shortened spin up during parameter optimization, as the converged upon parameter configuration is not guaranteed to be the correct one if the model were fully spun up to equilibrium.

Our results show that the minimum length for successful parameter optimization is 2,000 years, which is far longer than the decades to centuries models are typically integrated for tuning purposes. To optimize the model's representation of meso- and bathypelagic tracers a longer spin-up is required, which provides sufficient time for the modified parameters to influence the deeper ocean. The need for a longer spin-up is exacerbated if the parameters chosen for optimization influence vertical transport of organic matter, and if the model includes processes which increase the time required to reach equilibrium, such as denitrification and nitrogen fixation, which occur in distinct regions.

Despite potentially reducing the computational cost of each iteration of the optimization process through a reduction in spin-up length down to 2,000 years, our chosen optimization algorithm required more evaluations to achieve success, thereby negating this computational saving. However, other optimization algorithms may behave differently, and our further investigation into varying the initial conditions of the biogeochemical model shows that it may still be possible to use a shortened spin up under certain circumstances to reliably yield correct optimal parameters. While this result is preliminary, it does suggest this is a fruitful avenue for further research. If the saving is found to be substantial it could open up the possibility of carrying out at least some parameter tuning on a wider range of models than hitherto feasible. While existing models differ in structure, complexity and resolution, they all respond on similar timescales that are determined by the ocean circulation and underlying biogeochemical processes. Therefore, the results of this study should be applicable to beyond just the particular combination of circulation and biogeochemical models used here.

Appendix A: Supplementary Figures

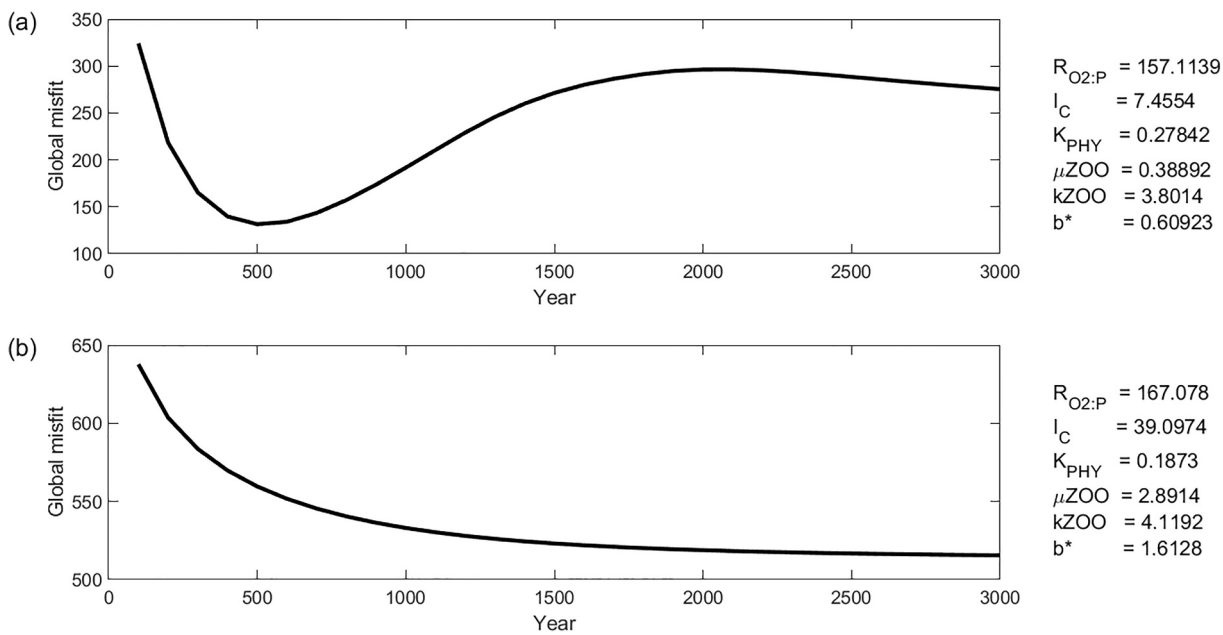

Figure A1. The global misfit during a 3,000 years spin up initiated from $\mathbf{x}_{ini}$ for two different MOPS parameter configurations, with a particularly (a) low b^* and I_C, and (b) high b^* and I_C. Parameter values are stated on the figure.

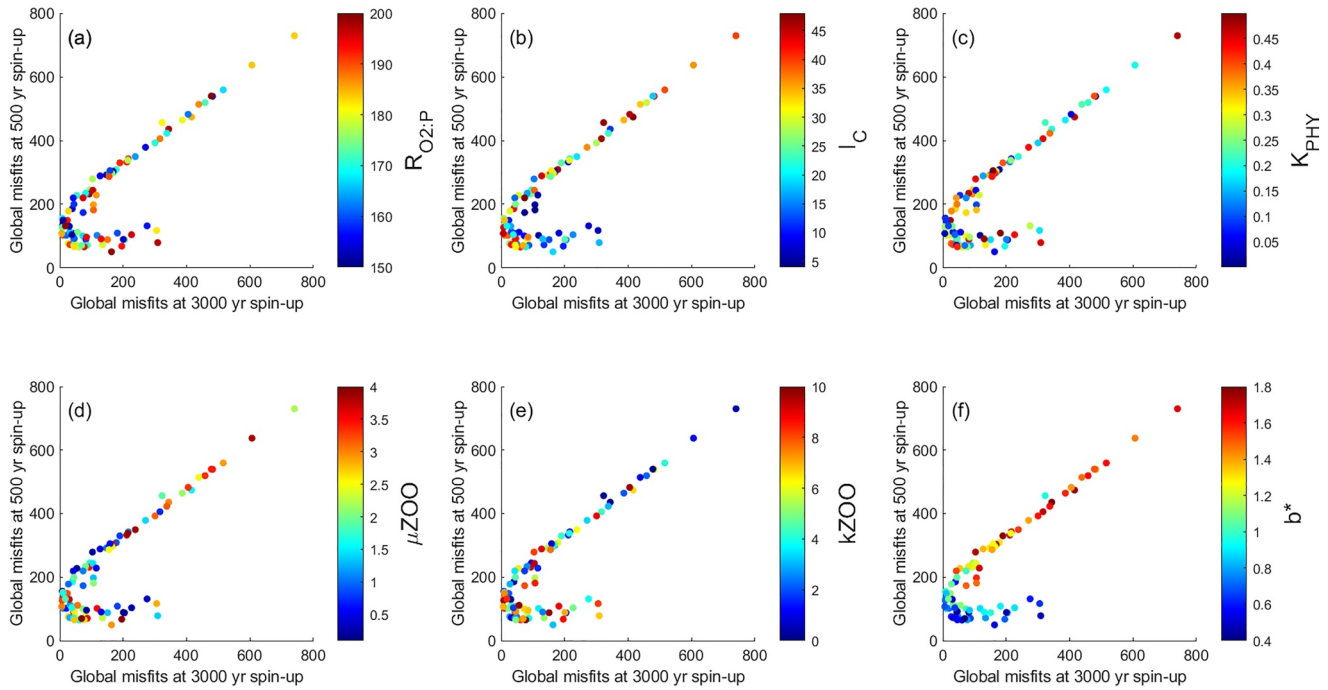

Figure A2. Misfit values for each of the 90 MOPS runs, calculated after a spin-up length of 3,000 years on the *x*-axes and 500 years on the *y*-axes. The 90 misfits are colored according to parameter values of (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO}, and (f) b*.

Data Availability Statement

The base TMM and MOPS code used here can be downloaded from https://github.com/samarkhatiwala/tmm (Khatiwala, 2018), and the transport matrices and forcing fields from https://doi.org/10.5281/zenodo.5517238 (Khatiwala, 2021). The most recent DFO-LS source code is available at https://github.com/numericalalgorithmsgroup/dfols and the optimization framework used to couple this to MOPS is available at https://doi.org/10.5281/zenodo.5517610 (Oliver & Tett, 2021). All data and code required to recreate the figures shown in this study can be downloaded from https://doi.org/10.5281/zenodo.10794545 (Oliver et al., 2024).

Acknowledgments

Computing resources were provided by the University of Oxford Advanced Research Computing (ARC) facility (https://doi.org/10.5281/zenodo.22558, Richards (2015)). Also thanks to Adrian Martin at the National Oceanography Centre for comments on the manuscript. Studentship and funding was provided to Sophy Oliver by the Natural Environmental Research Council (NERC) (NE/L002612/1), the Oxford Doctoral Training Partnership in Environmental Research, and the Met Office, with further support provided by NERC funding for the CLASS (NE/R015953/1) and CUSTARD (NE/P021247/2) projects. Samar Khatiwala was supported by UK NERC (NE/M020835/1 and NE/W007258/1).

References

Bardin, A., Primeau, F., & Lindsay, K. (2014). An offline implicit solver for simulating prebomb radiocarbon. *Ocean Modelling*, *73*, 45–58. https://doi.org/10.1016/j.ocemod.2013.09.008

Cartis, C., Fiala, J., Marteau, B., & Roberts, L. (2019). Improving the flexibility and robustness of model-based derivative-free optimization solvers. *ACM Transactions on Mathematical Software*, *45*(3), 1–35. https://doi.org/10.1145/3338517

Christina, L., & Passow, U. (2007). Factors influencing the sinking of POC and the efficiency of the biological carbon pump. *Deep Sea Research Part II: Topical Studies in Oceanography*, *54*(5–7), 639–658. https://doi.org/10.1016/j.dsr2.2007.01.004

DeVries, T., & Primeau, F. W. (2011). Dynamically and observationally constrained estimates of water-mass distributions and ages in the global ocean. *Journal of Physical Oceanography*, *41*(12), 2381–2401. https://doi.org/10.1175/jpo-d-10-05011.1

Dutkiewicz, S., Follows, M. J., & Parekh, P. (2005). Interactions of the iron and phosphorus cycles: A three-dimensional model study. *Global Biogeochemical Cycles*, *19*(1), 1–22. https://doi.org/10.1029/2004GB002342

Evans, G. T. (2003). Defining misfit between biogeochemical models and data sets. *Journal of Marine Systems*, *40–41*, 49–54. https://doi.org/10.1016/S0924-7963(03)00012-5

Fu, W., & Primeau, F. (2017). Application of a fast Newton-Krylov solver for equilibrium simulations of phosphorus and oxygen. *Ocean Modelling*, *119*, 35–44. https://doi.org/10.1016/j.ocemod.2017.09.005

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, R., et al. (2018a). *World Ocean Atlas 2018. Vol. 4: Dissolved inorganic nutrients (phosphate, nitrate and nitrate+nitrite, silicate)*. A. Mishonov (Ed.). (Vol. 35), NOAA Atlas NESDIS 84.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, R., et al. (2018b). *World Ocean Atlas 2018, volume 3: Dissolved oxygen, apparent oxygen utilization, and dissolved oxygen saturation*. A. Mishonov (Ed.). (Vol. 38). NOAA Atlas NESDIS 83.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. https://doi.org/10.1175/BAMS-D-15-00135.1

Khatiwala, S. (2007). A computational framework for simulation of biogeochemical tracers in the ocean. *Global Biogeochemical Cycles*, *21*(3), 1–14. https://doi.org/10.1029/2007GB002923

Khatiwala, S. (2008). Fast spin up of Ocean biogeochemical models using matrix-free Newton-Krylov. *Ocean Modelling*, *23*(3–4), 121–129. https://doi.org/10.1016/j.ocemod.2008.05.002

Khatiwala, S. (2018). Transport matrix method software for ocean biogeochemical simulations [Software]. *Zenodo*. https://doi.org/10.5281/ZENODO.1246300

Khatiwala, S. (2021). MITgcm 2.8deg transport matrix configuration [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.5517237

Khatiwala, S., Primeau, F., & Holzer, M. (2012). Ventilation of the deep ocean constrained with tracer observations and implications for radiocarbon estimates of ideal mean age. *Earth and Planetary Science Letters*, *325–326*, 116–125. https://doi.org/10.1016/j.epsl.2012.01.038

Khatiwala, S., Visbeck, M., & Cane, M. A. (2005). Accelerated simulation of passive tracers in ocean circulation models. *Ocean Modelling*, *9*(1), 51–69. https://doi.org/10.1016/j.ocemod.2004.04.002

Kriest, I., Getzlaff, J., Landolfi, A., Sauerland, V., Schartau, M., & Oschlies, A. (2023). Exploring the role of different data types and timescales for the quality of marine biogeochemical model calibration. *Biogeosciences Discussions*, *2023*, 1–33. https://doi.org/10.5194/bg-20-2645-2023

Kriest, I., Kähler, P., Koeve, W., Kvale, K., Sauerland, V., & Oschlies, A. (2020). One size fits all? - Calibrating an ocean biogeochemistry model for different circulations. *Biogeosciences Discussions*, *17*(12), 3057–3082. https://doi.org/10.5194/bg-2020-9

Kriest, I., & Oschlies, A. (2015). MOPS-1.0: Towards a model for the regulation of the global oceanic nitrogen budget by marine biogeochemical processes. *Geoscientific Model Development*, *8*(9), 2929–2957. https://doi.org/10.5194/gmd-8-2929-2015

Kriest, I., Oschlies, A., & Khatiwala, S. (2012). Sensitivity analysis of simple global marine biogeochemical models. *Global Biogeochemical Cycles*, *26*(2), 1–15. https://doi.org/10.1029/2011GB004072

Kriest, I., Sauerland, V., Khatiwala, S., Srivastav, A., & Oschlies, A. (2017). Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0). *Geoscientific Model Development*, *10*(1), 127–154. https://doi.org/10.5194/gmd-10-127-2017

Kuhn, A. M., & Fennel, K. (2019). Evaluating ecosystem model complexity for the northwest North Atlantic through surrogate-based optimization. *Ocean Modelling*, *142*, 101437. https://doi.org/10.1016/j.ocemod.2019.101437

Kwon, E. Y., & Primeau, F. (2006). Optimization and sensitivity study of a biogeochemistry ocean model using an implicit solver and in situ phosphate data. *Global Biogeochemical Cycles*, *20*(4), 1–13. https://doi.org/10.1029/2005GB002631

Kwon, E. Y., Primeau, F., & Sarmiento, J. L. (2009). The impact of remineralization depth on the air-sea carbon balance. *Nature Geoscience*, *2*(9), 630–635. https://doi.org/10.1038/ngeo612

Lauderdale, J. M., Garabato, A. C. N., Oliver, K. I., Follows, M. J., & Williams, R. G. (2013). Wind-driven changes in southern ocean residual circulation, ocean carbon reservoirs and atmospheric CO_2. *Climate Dynamics*, *41*(7–8), 2145–2164. https://doi.org/10.1007/s00382-012-1650-3

Lauderdale, J. M., Williams, R. G., Munday, D. R., & Marshall, D. P. (2017). The impact of southern ocean residual upwelling on atmospheric CO_2 on centennial and millennial timescales. *Climate Dynamics*, *48*(5–6), 1611–1631. https://doi.org/10.1007/s00382-016-3163-y

Li, X., & Primeau, F. W. (2008). A fast Newton-Krylov solver for seasonally varying global ocean biogeochemistry models. *Ocean Modelling*, *23*(1–2), 13–20. https://doi.org/10.1016/j.ocemod.2008.03.001

Lindsay, K. (2017). A Newton-Krylov solver for fast spin-up of online ocean tracers. *Ocean Modelling*, *109*, 33–43. https://doi.org/10.1016/j.ocemod.2016.12.001

Marshall, J., Adcroft, A., Hill, C., Perelman, L., & Heisey, C. (1997). A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research*, *102*(C3), 5753–5766. https://doi.org/10.1029/96JC02775

Martin, J. H., Knauer, G. A., Karl, D. M., & Broenkow, W. W. (1987). VERTEX: Carbon cycling in the northeast Pacific. *Deep-Sea Research, Part A: Oceanographic Research Papers*, *34*(2), 267–285. https://doi.org/10.1016/0198-0149(87)90086-0

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245. https://doi.org/10.1080/00401706.2000.10485979

Oliver, S., Cartis, C., Kriest, I., Tett, S., & Khatiwala, S. (2022). A derivative-free optimisation method for global ocean biogeochemical models. *Geoscientific Model Development*, *15*(9), 3537–3554. https://doi.org/10.5194/gmd-15-3537-2022

Oliver, S., Khatiwala, S., Cartis, C., Ward, B., & Kriest, I. (2024). Data supplement for Oliver et al. (2024) Using shortened spin-ups to speed up ocean biogeochemical model optimisation. *Zenodo*. https://doi.org/10.5281/zenodo.10794545

Oliver, S., & Tett, S. (2021). OPTCLIMSO optimisation framework (version 1) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.5517610

Parekh, P., Follows, M. J., & Boyle, E. A. (2005). Decoupling of iron and phosphate in the global ocean. *Global Biogeochemical Cycles*, *19*(2), 1–16. https://doi.org/10.1029/2004GB002280

Priess, M., Koziel, S., & Slawigs, T. (2013). Marine ecosystem model calibration through enhanced surrogate-based optimization. *Advances in Intelligent Systems and Computing*, *197*(5), 193–208. https://doi.org/10.1016/j.jocs.2013.04.001

Primeau, F. (2005). Characterizing transport between the surface mixed layer and the ocean interior with a forward and adjoint global ocean transport model. *Journal of Physical Oceanography*, *35*(4), 545–564. https://doi.org/10.1175/JPO2699.1

Primeau, F., & Deleersnijder, E. (2009). On the time to tracer equilibrium in the global ocean. *Ocean Science*, *5*(1), 13–28. https://doi.org/10.5194/os-5-13-2009

Richards, A. (2015). University of Oxford advanced research computing. *Zenodo*. https://doi.org/10.5281/zenodo.22558

Sauerland, V., Kriest, I., Oschlies, A., & Srivastav, A. (2019). Multiobjective calibration of a global biogeochemical ocean model against nutrients, oxygen, and oxygen minimum zones. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1285–1308. https://doi.org/10.1029/2018MS001510

Séférian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., et al. (2020). Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. *Current Climate Change Reports*, *6*(3), 95–119. https://doi.org/10.1007/s40641-020-00160-0

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., et al. (2016). Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827–1851. https://doi.org/10.5194/gmd-9-1827-2016

Shah, S. H. A. M., Primeau, F. W., Deleersnijder, E., & Heemink, A. W. (2017). Tracing the ventilation pathways of the deep North Pacific Ocean using Lagrangian particles and Eulerian tracers. *Journal of Physical Oceanography*, *47*(6), 1261–1280. https://doi.org/10.1175/JPO-D-16-0098.1

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., & Oschlies, A. (2010). Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems*, *81*(1–2), 34–43. https://doi.org/10.1016/j.jmarsys.2009.12.005

Wunsch, C., & Heimbach, P. (2008). How long to oceanic tracer and proxy equilibrium? *Quaternary Science Reviews*, *27*(7–8), 637–651. https://doi.org/10.1016/j.quascirev.2008.01.006