



RESEARCH ARTICLE

10.1029/2022MS003390

Key Points:

- Model calibration is an important task in the view of parametric uncertainties
- We allow specific uncertain model parameters to be random while optimizing other model parameters, providing an efficient tool for the task
- We apply it to a global biogeochemical circulation model to quantify the impact of zooplankton mortality on the underlying biogeochemistry

Correspondence to:

V. Sauerland,
vsauerland@geomar.de

Citation:

Sauerland, V., von Hallern, C., Kriest, I., & Getzlaff, J. (2023). A CMA-ES algorithm allowing for random parameters in model calibration. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003390. <https://doi.org/10.1029/2022MS003390>

Received 3 SEP 2022

Accepted 4 AUG 2023

Author Contributions:

Conceptualization: Volkmar Sauerland, Julia Getzlaff

Formal analysis: Volkmar Sauerland, Iris Kriest

Investigation: Volkmar Sauerland, Claudine von Hallern, Iris Kriest, Julia Getzlaff

Methodology: Volkmar Sauerland, Claudine von Hallern

Software: Volkmar Sauerland

Validation: Claudine von Hallern, Iris Kriest, Julia Getzlaff

Visualization: Volkmar Sauerland, Iris Kriest

Writing – original draft: Volkmar Sauerland

Writing – review & editing: Claudine von Hallern, Iris Kriest, Julia Getzlaff

© 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

A CMA-ES Algorithm Allowing for Random Parameters in Model Calibration

Volkmar Sauerland^{1,2} , Claudine von Hallern³, Iris Kriest¹ , and Julia Getzlaff¹ 

¹Biogeochemical Modelling, GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Kiel, Germany, ²Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, Kiel, Germany, ³Department Mathematik/CEN, Universität Hamburg, Hamburg, Germany

Abstract In geoscience and other fields, researchers use models as a simplified representation of reality. The models include processes that often rely on uncertain parameters that reduce model performance in reflecting real-world processes. The problem is commonly addressed by adapting parameter values to reach a good match between model simulations and corresponding observations. Different optimization tools have been successfully applied to address this task of model calibration. However, seeking one best value for every single model parameter might not always be optimal. For example, if model equations integrate over multiple real-world processes which cannot be fully resolved, it might be preferable to consider associated model parameters as random parameters. In this paper, a random parameter is drawn from a wide probability distribution for every single model simulation. We developed an optimization approach that allows us to declare certain parameters random while optimizing those that are assumed to take fixed values. We designed a corresponding variant of the well known Covariance Matrix Adaption Evolution Strategy (CMA-ES). The new algorithm was applied to a global biogeochemical circulation model to quantify the impact of zooplankton mortality on the underlying biogeochemistry. Compared to the deterministic CMA-ES, our new method converges to a solution that better suits the credible range of the corresponding random parameter with less computational effort.

Plain Language Summary In geoscience and other fields, researchers use models as a simplified representation of reality. The models include processes that often rely on uncertain parameters that reduce model performance in reflecting real-world processes. The problem is commonly addressed by adapting parameter values to reach a good match between model simulations and corresponding observations. However, seeking one best value for every single model parameter might not always be optimal. For example, if model equations integrate over multiple real-world processes it might be preferable to consider associated model parameters as random parameters that are drawn from a wide probability distribution for every single model simulation. To explore this further, we modified a popular variant of the Covariance Matrix Adaption Evolution Strategy (CMA-ES) in order to tune the other parameters. The new algorithm was applied to a global biogeochemical circulation model to quantify the impact of zooplankton mortality on the underlying biogeochemistry. Compared to the deterministic CMA-ES our new method converges to a solution that better suits the credible range of the corresponding random parameter.

1. Introduction

The aim to understand natural systems is a driving force in science. An important task is the development of associated models, which are used to simulate the (present and future) behavior of a system and to verify concepts about underlying principles and processes. However, complex systems like the Earth system rely on many entangled processes a model can represent neither completely nor precisely. It is therefore often difficult to choose the appropriate model complexity, for example, for a component of an Earth system model (ESM; Eyring et al., 2016).

Decisions about suitable model setups can be difficult if parameterized processes with uncertain parameters have a large influence in the model, and if model simulations are computationally demanding. This is commonplace in global coupled biogeochemical (BGC) ocean models (e.g., Anderson, 2005; Kriest, 2017; Leles et al., 2016; Shimoda & Arhonditsis, 2015). Kriest et al. (2010) showed that the impact of BGC interactions on model representations of key tracers such as oxygen or nutrients can be as large as that of ocean circulation (Najjar et al., 2007). However, BGC parameters like growth rates, consumption rates, and mortality rates are not exactly

known a priori, but are often only roughly constrained by laboratory or in-situ experiments, and integrate over many groups of organisms. Inverse model studies, that adjust parameters to fit the model to more easily observable quantities, such as nutrients or oxygen, can help to define credible intervals (e.g., Kane et al., 2011; Schartau et al., 2017). In practice, BGC parameters of global ocean models are typically adjusted manually in a given circulation, that is, the BGC parameter values are adjusted until simulated tracers yield a good match with observed data. Systematic calibrations of biogeochemical model parameters need to apply an automated optimization procedure in order to minimize an objective model-data misfit measure, but require a large number of model evaluations. However, the high computational demand of global BGC models impedes this approach, which is therefore the exception rather than the rule; moreover, parameters adjusted in a given circulation may not perform well in a different one, thereby requiring a re-calibration whenever the circulation changes (Kriest et al., 2020).

The computational demand may be reduced by either applying computationally cheaper surrogate circulations (Khatiwala et al., 2005; Prieß et al., 2013), or by using efficient optimization algorithms. Several parameter optimization studies invoke gradient information of the misfit function to iteratively approach a locally optimal parameter vector from some initial estimate. The gradient information is mostly obtained by the—computationally efficient—adjoint method. Lawson et al. (1995) introduced it in the context of biogeochemical models. It is used to determine a direction of descent, for example, by a quasi-Newton method as in Friedrichs (2002); Spitz et al. (1998); Tjiputra et al. (2007); Xiao and Friedrichs (2014), and a step size to change the parameter vector. In the face of complex BGC ocean models (including discontinuous functions), the deployment of associated adjoint models is often difficult. Consequently, gradient-based methods are often unstable. Further, the convergence speed of gradient-based methods can suffer from poorly conditioned Hessian matrices, and finding good pre-conditioners can be difficult or time-consuming, nullifying a possible gain in convergence speed. Stochastic derivative-free search algorithms (e.g., used by Hurtt & Armstrong, 1996; Kidston et al., 2011; Kaufman et al., 2018) are known to be more robust. They allow for a thorough yet efficient scan of the parameter space and can also avoid to get stuck in a (first) local optimum (cf. Schartau & Oschlies, 2003; Vallino, 2000). This is of particular importance given the complex topography of the model-data misfit measure typical for complex BGC models (see, e.g., Faugeras et al., 2003; Hurtt & Armstrong, 1996).

The covariance matrix adaption evolution strategy (CMA-ES, Hansen, 2006) is a stochastic search algorithm that is also applicable to poorly conditioned problems. It is popular for its competitiveness concerning effectivity and efficiency (cf., Hansen et al., 2010). Because of its advantages concerning robustness, efficiency and the balance between exploration and exploitation of the search space, Kriest et al. (2017) applied the CMA-ES algorithm in one of the first optimizations of a global ocean BGC model of intermediate complexity in combination with a sufficient model spin-up time (3,000 years) to approach annual tracer equilibria. In this study, we modified the CMA-ES algorithm in order to enhance the concept of parameter adjustment, in the view of uncertainties that arise from insufficiently resolved real-world processes.

When applying parameter optimization, the decision about appropriate model complexity is often accompanied by the question which BGC parameter values should be tuned (i.e., are changed during optimization in order to obtain a better model fit to observations), and which parameters should remain fixed at some reasonable, constant value during optimization. In order to save computational effort, it seems natural to exclude parameters from optimization that are of minor importance for the research question on hand and/or are unimportant with regard to the applied model-data misfit measure. Indeed, Kriest et al. (2017) observed the most insensitive parameter subject to their optimization experiment to converge the slowest. The same behavior was recently observed by Oliver et al. (2021) who applied an even faster (but less explorative) search algorithm to the same coupled BGC model setup, using twin experiment data. However, fixing some parameters from the start might impact the optimization of the other parameters. Actually, it may not always be the most realistic but pragmatic assumption that model parameters are constant at all. Optimal constant parameters can result in unexpected model responses, when the modeling purpose is changed. Our model processes are simplifications of reality. Therefore, a single model process might integrate over several relevant real-world processes, which may change in space and time. An example for unresolved processes is the impact of higher trophic levels on biogeochemical cycling. In general global coupled BGC models resolve only the first two to three trophic levels (of plankton), with zooplankton mortality as upper closure term. Here, zooplankton mortality parameterizes not solely the natural zooplankton mortality but in addition also the feeding pressure of higher trophic levels (HTL), such as fish, on zooplankton. Given that the spatial and temporal distribution of fish populations is variable, a constant mortality parameter of zooplankton (as currently applied in many BGC models) may not be well justified.

Therefore, we target an optimization approach that allows to declare some parameters to be random parameters while optimizing those that are assumed to take fixed values. In this paper, a random parameter is drawn from a probability distribution for every single model simulation (but stays constant within a model simulation). Solutions that are obtained this way might be preferable, since.

1. they are optimal with regard to a large range of (uncertain) parameters, and thereby may render the model applicable in a wider range of contexts (e.g., when coupled to fish models imposing spatially varying feeding pressures on zooplankton), so as to minimize the deterioration of the performance of the biogeochemical model.
2. setting a parameter to the one or the other credible value may bias the optimization of the other parameters toward local optima, with consequences for the biogeochemical cycling (as, e.g., in Kriest et al., 2017).

However, exploring the model-data misfit across a large range of randomly varying parameters would drastically increase the computational demand of the misfit evaluation. Indeed, global ocean BGC models are a paragon for the problem, as single model simulations already have a high computational demand. We provide an efficient way to deal with the given problem by introducing the new R-CMA-ES (a modification of the CMA-ES where R stands for random) in this study. We will examine the effects of the proposed optimization procedure by applying it to the same global BGC model setup as used in Kriest et al. (2017) and Oliver et al. (2021).

The paper is organized as follows. In Section 2 we briefly describe the CMA-ES algorithm as well as our modifications in order to efficiently solve problems with random parameters. While Section 2 is conceptual, the technical descriptions of the algorithms and their pseudo codes can be found in Appendix A and Appendix B, respectively. Further, in Appendix C we test our CMA-ES variant on a set of mathematical benchmark instances and selected random parameters in order to provide evidence that the algorithm serves the intended purpose. In Section 3 we introduce the global ocean biogeochemical model which we choose as our sample application along with an RMSE type model-data misfit measure, which has been used by Kriest et al. (2017) to calibrate the model. Based on an analysis of the calibration experiments of that former study, we declare one parameter a random parameter and derive the associated integrated model-data misfit measure (the expected model-data misfit measure with regard to the random parameter), which we want to optimize, here. In Section 4 the results of our BGC model calibrations with one random parameter are compared to the calibrations carried out by Kriest et al. (2017). We discuss the convergence behavior of both optimization approaches as well as scientific findings by the new calibration experiment and close with some conclusions.

2. Methods

For our approaches to optimization throughout this paper, we distinguish three types of model parameters:

- fixed parameters: parameters stay fixed at a single, scalar value throughout the optimization. These are not further considered in this paper.
- non-random parameters: parameters whose values are to be optimized and change through optimization. The parameter values are drawn from a probability distribution with changing standard deviation (see below, Section 2.1), and are hereafter denoted by p .
- random parameters: parameters whose values are drawn at random from a (wide) probability distribution. In contrast to non-random parameters, the standard deviation of the probability distribution is kept constant throughout the optimization. However, equivalent to non-random parameters, random parameters remain constant during each simulation. Their value can only change for each new simulation.

Random parameters are hereafter denoted as q and characterized in detail in Section 2.2.

A suitable division of the parameters into random parameters and non-random parameters requires some preliminary considerations. For instance, all parameters that are relevant for the research question on hand can be analyzed with regard to their covariances and/or their impact on the model-data misfit function, using multiple model runs (e.g., the simulations of a parameter sensitivity analysis or a deterministic model calibration experiment). Based on this analysis, an appropriate parameter (set) can then be declared random.

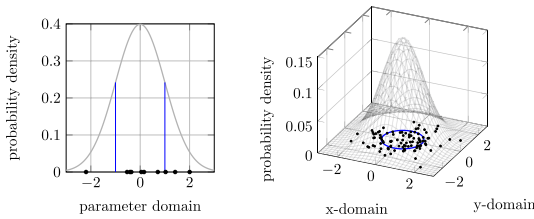


Figure 1. Examples for a uni-variate and a bi-variate normal distribution. Here, it is the standard normal distribution $\mathcal{N}(0, 1)$ (left panel) and $\mathcal{N}(\mathbf{m}, \mathbf{C})$

(right panel) with $\mathbf{m} = (0, 0)^T$ and $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The associated probability

densities are indicated by the gray curve and mesh-grid, respectively. We also highlight the areas of one standard deviation which is an interval on the x -axis (limited by the blue vertical lines) in the uni-variate case, and an ellipse in the plane (blue ellipse) in the bi-variate case. In both cases, we realized some random samples (black dots) from the respective normal distribution.

2.1. The CMA-ES Algorithm

Here, we adapt the Covariance Matrix Adaption Evolution Strategy (CMA-ES) to allow for an efficient model calibration with random parameters. More precisely, we build upon the $(\mu/\mu_w, \lambda)$ -CMA-ES (see Hansen, 2006; Hansen, 2016). From now on, we will simply write CMA-ES to refer to the $(\mu/\mu_w, \lambda)$ -CMA-ES. We start with a brief recapitulation of the CMA-ES. A detailed description and the pseudo code (reduced by some subtleties) of the original algorithm can be found in Appendix A.

The CMA-ES algorithm allows the optimization of n model parameters without the requirement to calculate model derivatives and the associated derivatives of a cost function with respect to the parameters. It is popular for its robustness and efficiency in solving optimization problems that are characterized by a difficult topography of a misfit function $f(\mathbf{p})$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (cf., e.g., Hansen et al., 2010).

The algorithm maintains a multi-variate normal distribution over the n -dimensional parameter space of $\mathbf{p} \in \mathbb{R}^n$. Similar to the definition of a uni-variate normal distribution $\mathcal{N}(m, \sigma)$ by its mean m and its standard deviation σ , a multi-variate normal distribution $\mathcal{N}(\mathbf{m}, \mathbf{C})$ is uniquely defined by a mean vector \mathbf{m} and a positive definite covariance matrix \mathbf{C} . Figure 1 illustrates the situation for the uni-variate and the bi-variate case. The algorithm's normal distribution is initialized such that it covers a sufficient part of the parameter space. The algorithm then iteratively samples a *population*, that is, a set of λ candidate solutions from the normal distribution. From the λ samples the $\mu = \lfloor \frac{\lambda}{2} \rfloor$ best-ranked samples with respect to f are selected. The subset of selected samples is used to calculate an empirical normal distribution (i.e., an empirical mean vector and an empirical matrix of covariances), using rank-dependent weights w which (usually) sum up to 1. The suitable choice of λ , μ , w , and many other operational constants of the algorithm are mainly determined empirically (cf. Hansen, 2016) and details are given in Table A1 in Appendix A. The normal distribution that has been estimated from the sampled parameter vectors is in turn used to update the algorithm's normal distribution to a region with better misfit values. This procedure of sampling and updating the normal distribution is repeated until convergence (or for as many iterations as necessary/desired), that is, until f becomes sufficiently small and all parameter variances vanish. Figure 2 illustrates the convergence behavior of

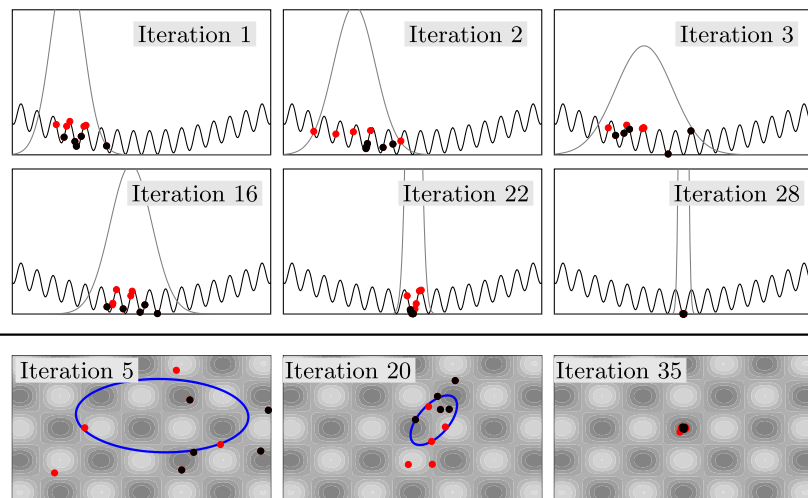


Figure 2. Convergence examples of the CMA-ES algorithm for the uni-variate Griewank function (upper 6 panels) and the bi-variate Griewank function (lower 3 panels). In both cases we draw $\lambda = 10$ samples (indicated as dots) per iteration from the normal distribution. The better half of $\mu = 5$ samples (black dots) is used to update the distribution. In the uni-variate case, the function is represented as black curve and the normal distribution is indicated as gray curve. In the bi-variate case, the function values are represented by the gray-scale color scheme with increasing values from dark to light shades. Here, like in Figure 1, blue ellipses denote the standard deviation of the normal distribution in the two-dimensional parameter space.

CMA-ES for a uni-variate misfit function (upper six panels) and a corresponding bi-variate misfit function (lower three panels), respectively. We see that the variance of the distribution can increase as long as the good samples show a wide spread (e.g., iteration 3 in the one-dimensional case) but decrease if the good samples are close together like in the vicinity of the optimum (iterations 22 and 28 in the one-dimensional case).

The general applicability of CMA-ES to large-scale, real-world problems has been shown by, for example, Kriest et al. (2017, 2020). These studies also benefited from additional features of CMA-ES, that help to accelerate shifts of the distribution's mean into directions of descent and also support reliable updates of the distribution with a small population size λ . These features are given in more detail in Appendix A.

2.2. The R-CMA-ES Algorithm

The CMA-ES provides a unique set of n optimal model parameters \mathbf{p}^* , that minimize a given misfit function. However, as mentioned above, and shown earlier (cf. Kriest et al., 2017), a given misfit function can be rather insensitive to some parameters, thereby slowing down the convergence, and resulting in poorly defined optimal parameter values. Moreover, the parameters may represent a number of unresolved processes, which are only vaguely defined; therefore, these parameters are associated with a large inherent uncertainty. We now aim to account for this uncertainty by introducing a random element into optimization, by extending and modifying CMA-ES as follows:

Let n be the number of non-random parameters \mathbf{p} , $\mathbf{p} \in \mathbb{R}^n$, whose values are to be optimized by a given misfit function, in analogy to CMA-ES described in Section 2.1. We consider these parameters to represent processes that are well-understood and well-defined by the model, and that can be constrained well by the given misfit function. Let further r be the number of random parameters \mathbf{q} , $\mathbf{q} \in \mathbb{R}^r$, that represent unresolved processes, and whose values can not be constrained by the given misfit function.

We may further assume without loss of generality an overall parameter vector $(\mathbf{p}; \mathbf{q})$, whose first n components consist of the non-random parameters and the following r of the random components: $(p_1, \dots, p_n, q_1, \dots, q_r) \in \mathbb{R}^{n+r}$. Our intention behind R-CMA-ES is to efficiently optimize \mathbf{p} given randomly varying \mathbf{q} . In other words, we seek for a parameter vector \mathbf{p}^* that is optimal over the entire range of \mathbf{q} , where the mean value over the range of \mathbf{q} can be generalized by applying a probability density function $\text{pdf}(\mathbf{q})$. Such an optimal solution would then be minimal with respect to $F : \mathbb{R}^n \rightarrow \mathbb{R}$, defined by

$$F(\mathbf{p}) = \int_{\mathbf{q} \in \mathbb{R}^r} \text{pdf}(\mathbf{q}) \cdot f((\mathbf{p}; \mathbf{q})) d\mathbf{q}, \quad (1)$$

which is a weighted integral over a given misfit function $f : \mathbb{R}^{n+r} \rightarrow \mathbb{R}$. For example, if we deal with only one random parameter q ($r = 1$) and allow q to take uniformly distributed values within a credible interval $[a, b]$, then Equation 1 becomes

$$F(\mathbf{p}) = \frac{1}{b-a} \int_{q=a}^b f((\mathbf{p}; q)) dq,$$

that is, the mean f -value over $[a, b]$.

When the evaluation of f is already computationally demanding, its computation over a wide range of \mathbf{q} is prohibitive. Therefore, rather than applying the normal CMA-ES directly to F , we seek to design a modified CMA-ES that operates on the cheaper (point-based) misfit measure $f(\mathbf{p}; \mathbf{q})$, in order to converge to a parameter vector \mathbf{p}^* that minimizes F . The next two subsections describe our corresponding algorithmic modifications including a final sentence of justification at the very end of each subsection.

2.2.1. Distribution Handling for Random Parameters

The basic idea is to modify the update procedure of the maintained multi-variate normal distribution of parameters such that it gets only updated for the non-random parameters but retains its initial distribution property with regard to the random parameters. More precisely, we desire that the standard deviation interval of the marginal distribution w.r.t. the random parameter always remains the entire credible interval of that parameter. Figure 3

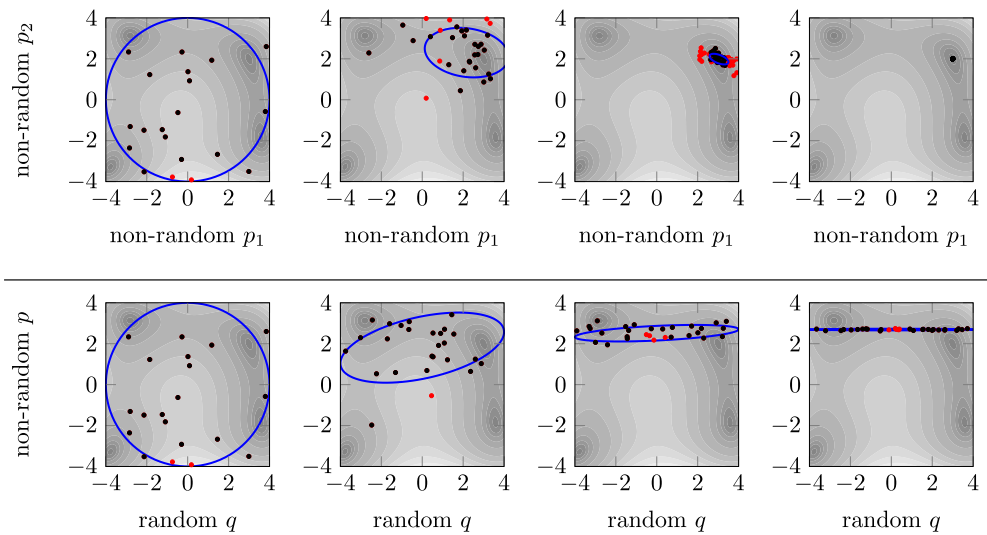


Figure 3. Convergence behavior of CMA-ES variants on a two-dimensional test case with Himmelblau's function. The gray-scale color scheme represents the function values $f(\mathbf{p})$ with increasing values from dark to light shades. We draw $\lambda = 50$ samples per iteration (many samples occur beyond the axis limits of the plot) and select the better $\mu = 25$ samples (black dots) for the distribution update. Blue ellipses denote the standard deviation of the normal distribution in the two-dimensional parameter space. Upper panels: Iterations 1, 7, 10 and 20 of the CMA-ES. Lower panels: Iterations 1, 7, 10 and 20 of R-CMA-ES, where the variance of the first (random) parameter remains unchanged.

sketches the intended behavior for a two-dimensional test function (Himmelblau's function): the top row shows the convergence of the CMA-ES algorithm; the bottom row illustrates the corresponding behavior of R-CMA-ES if the first parameter is random and the second parameter is optimized. In the example, the CMA-ES algorithm converges to a point (parameter vector \mathbf{p}^*) that minimizes Himmelblau's function, that is, the mean of the normal distribution approaches the optimal point while all variances approach zero. In the modified algorithm, the normal distribution is supposed to “converge to a line,” leaving the variance of the first (random) parameter unchanged while forcing the variance of the second parameter to approach zero.

The intended adaption with regard to the normal distribution over random and non-random parameters is justified by the following fact: if the normal distribution converged in the sense that the variances of all non-random parameters approached zero, then the expected $f(\mathbf{p}; \mathbf{q})$ -value of any sample $(\mathbf{p}; \mathbf{q})$ is exactly the integrated value, $F(\mathbf{p})$. However, in order to ensure that R-CMA-ES does actually converge to a parameter vector \mathbf{p} that yields a small integrated misfit value $F(\mathbf{p})$, we also have to adapt the sampling and selection procedure of the algorithm.

2.2.2. Modification of the Sampling and Selection Procedure

In an iteration of the CMA-ES, the good μ samples that are used to update the probability distribution are selected with regard to an f -value ranking. The same kind of selection is, however, not preferable if we want to update the normal distribution toward a solution with a small value of the integrated misfit function F defined by Equation 1. As an example, which is depicted graphically in Figure 4, we consider the two-dimensional linear function $f(p, q) = -p - q$ on the feasible domain $A = [0, 1]^2$ and declare q to be random. The misfit function $f(p, q)$ takes its minimum in the upper right corner $(1, 1)^T$ of A and the integrated misfit function F , defined by

$$F(\mathbf{p}) = \int_{q=0}^1 f(\mathbf{p}, q) dq,$$

is minimized for $q = 1$, that is, on the right boundary of A . With the selection criterion that is based on the ranking according to the f -values, the μ selected samples can often occur quite balanced on both sides of the current distribution mean, implying too small distribution updates toward the right boundary of A , where the actual minimum (line) of A would be found. At the same time, the selected samples might not cover the full credible interval of the random parameter, as f pushes the selection toward its minimum in the upper right corner. Using this kind of selection (sketched in Panel (a) of Figure 4) for the new algorithm, it will likely not converge to the

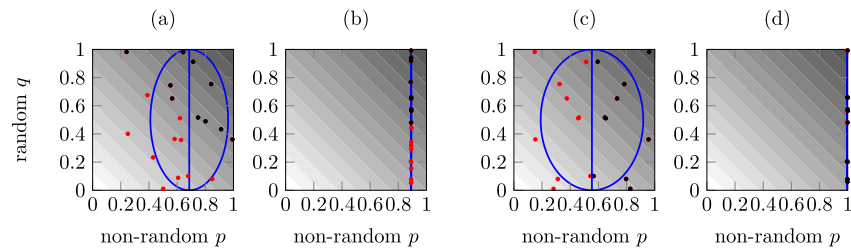


Figure 4. Effects of different selection strategies, illustrated with the linear test function $f(p, q) = -p - q$, where p is a non-random parameter to be optimized, and q is a random parameter. The gray-scale color scheme represents the function values $f(p, q)$ with increasing values from dark to light shades. We draw $\lambda = 20$ samples per iteration (shown as dots) and select $\mu = 10$ samples (black dots) for the distribution update. Blue ellipses denote the standard deviation of the normal distribution. We also show the principal axis of the ellipses that corresponds to the random parameter. Panel (a) shows iteration three of an optimization using independent samples that are ranked and selected with regard to their f -values. Panel (b) shows iteration 50 of the same optimization. Panel (c) and (d) depict iterations 3 and 50 of another optimization, this time using mirrored sampling and pairwise selection (see text).

right boundary of A . For this test, it converges to the solution indicated by the blue (slim, almost line-shaped) standard deviation ellipse in Panel (b) of Figure 4. We therefore prescribe the values of the random parameter to stay independently normally distributed with regard to the subset of μ selected samples (like it is the case with regard to all λ samples).

One possibility to attain the desired independency of selected parameter vectors is mirrored sampling and pairwise selection. Brockhoff et al. (2010) introduced the idea of mirrored sampling as a derandomization technique which can improve the convergence speed of evolution strategies (Auger et al., 2010). Auger et al. (2011) combined mirrored sampling with pairwise selection in order to avoid premature convergence caused by canceling effects on cumulative step-size adaptations. Recently, Wang et al. (2019) proposed mirrored orthogonal sampling which further improved the convergence behavior of CMA-ES. We adapt the idea of mirrored sampling to our situation and generate pairs of samples that are mirrored at the axis (hyper plane) of the random parameter(s). From each pair of mirrored samples we select the one with smaller f -value. An example for this kind of mirrored sampling is sketched in panel (c) of Figure 4. There, the combined sampling and selection procedure forces the distribution to approach the right boundary of the search space as desired (Figure 4, panel (d)).

To summarize, applying the adapted mirrored sampling to the CMA-ES — in conjunction with the afore described distribution handling for random parameters — supports the convergence to a parameter vector \mathbf{p}^* that actually minimizes F instead of f .

2.3. Boundary Handling

Biogeochemical model parameters are usually confined to a reasonable range, for example, negative growth or mortality rates are biologically meaningless. Therefore, when optimizing these parameters, usually credible boundaries are defined.

Yet, during the selection of discrete parameter values from the assumed distribution, the algorithm might select parameters slightly outside this range. Similar to Kriest et al. (2017), we exclude these values by adding a penalty term to the misfit. For the current study we apply a simpler boundary handling, which can be applied, without the need of further adaptations, to the case with random parameters. The procedure is as follows:

We assume that every parameter vector $\mathbf{p} \in \mathbb{R}^n$ must be contained in the set

$$A := [a_1, b_1] \times \dots \times [a_n, b_n],$$

where each interval $[a_i, b_i]$ is the credible interval of parameter p_i . We consider these constraints by adding to the objective function a penalty term, which depends on the distance of \mathbf{p} to its own *best approximation* $\phi(\mathbf{p})$ in A . The best approximation $\phi(\mathbf{p}) \in A$ of \mathbf{p} is the vector in A which has (amongst all vectors in A) the smallest distance to \mathbf{p} . It can be calculated component-wise by setting $(\phi(\mathbf{p}))_i = \min(b_i, \max(a_i, p_i))$ for all $i = 1, \dots, n$. Thus, for $\mathbf{p} \in A$ we have $\phi(\mathbf{p}) = \mathbf{p}$ and for $\mathbf{p} \notin A$ the best approximation $\phi(\mathbf{p})$ lies on the boundary of A . The re-definition of f with regard to the penalty term is given by

$$f(\mathbf{p}) = f(\phi(\mathbf{p})) + c \cdot \sum_{i=1}^n (p_i - (\phi(\mathbf{p}))_i)^2,$$

where c is a (large) constant. We apply this kind of boundary handling with both, the classical CMA-ES and R-CMA-ES.

Since we deal with boundary constraints of the form $p_i \in [a_i, b_i]$, we can restrict CMA-ES to operate on the unit cube $[0,1]^n$. In this case, we obtain a parameter vector $\mathbf{p} \in A$ (with A as defined above) from a sample $\mathbf{x} \in [0,1]^n$ by scaling and shifting every component: $p_i = a_i + (b_i - a_i)x_i$.

We tested R-CMA-ES on a set of mathematical benchmark functions. The results of our test-bed confirmed the intended behavior of the algorithm. Our tests are summarized in Appendix C. They encouraged us to use R-CMA-ES in a real-world application, and compare its effects on algorithm performance, optimal model parameters and potential effects on biogeochemical turnover.

3. The Global Ocean Biogeochemical Model Setup

Kriest et al. (2017) applied the CMA-ES to optimize six parameters of the BGC Model of Oceanic Pelagic Stoichiometry (MOPS; Kriest & Oschlies, 2015). The approach was facilitated by the Transport Matrix Method (TMM; Khatiwala, 2007, 2018; Khatiwala et al., 2005) as a numerically efficient tool to represent the global ocean circulation. The TMM represents advection and mixing in terms of transport matrices which are precomputed from an online ocean circulation model simulation. Here, as in Kriest et al. (2017) and also in Oliver et al. (2021), we apply monthly mean transport matrices from a 2.8° global configuration of the MIT general circulation model (MITgcm), having 15 depth levels (Marshall et al., 1997). MOPS coupled to the TMM simulates globally the concentrations and biogeochemical turnover of seven tracer components, namely phyto- and zooplankton, dissolved and particulate organic matter, phosphate, nitrate and oxygen.

To showcase the impact and performance of R-CMA-ES, we consider the same model setup as in Kriest et al. (2017), and optimize biogeochemical model parameters against a misfit function that includes global observations of nutrients and oxygen. Each biogeochemical model setup is simulated for 3,000 years, after which the model tracers approach a steady annual cycle.

3.1. The Misfit Function f_{rmse}

The misfit function considered in Kriest et al. (2017) is a weighted root mean squared error between simulated annual mean tracer concentrations of oxygen, phosphate, and nitrate and their observed equivalents after being mapped to the model grid with $N = 52,749$ ocean grid boxes. For each grid box and each tracer the corresponding misfit term is weighted with regard to both, the volume V_i of the grid box divided by the total ocean volume V_T , and the global average observed tracer concentration \bar{o}_j ($j = 1$ for phosphate, $j = 2$ for nitrate, and $j = 3$ for oxygen). Denoting the model output for a given parameter vector

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{pmatrix} = \begin{pmatrix} R_{-O_2:P} \\ I_c \\ K_{\text{Phy}} \\ \mu_{\text{Zoo}} \\ b^* \\ \kappa_{\text{Zoo}} \end{pmatrix}$$

as $m_{i,j}$ and the corresponding observation as $o_{i,j}$, $i = 1, \dots, N, j = 1, 2, 3$, the misfit function is

$$f_{\text{rmse}}((p_1, \dots, p_6)^T) = \sum_{j=1}^3 \frac{1}{\bar{o}_j} \left(\sum_{i=1}^N (m_{i,j} - o_{i,j})^2 \frac{V_i}{V_T} \right)^{1/2}. \quad (2)$$

Table 1
Optimal Parameter Values of Calibration Experiment OBS-NARR and Their Corresponding Ranges When Only Simulations Which Deviate From the Best Misfit by Less Than 5% Are Included

Parameter	Bounds	Opt.	5%-misfit ranges	
			Range	Rel. width
Oxygen consumption: phosphate release during aerobic remineralization ($R_{-O_2:P}$)	150–200	167	150–192.5	85.0%
Half-saturation for light (I_c)	4.0–48	9.7	4.0–17.96	31.7%
Half-saturation for phosphate (K_{phy})	0.001–0.5	0.5	0.001–0.530	106%
Maximum grazing rate of zooplankton (μ_{Zoo})	1.0–3.0	1.89	0.603–3.047	122%
Vertical increase in sinking speed of organic matter (b^*)	0.4–1.8	1.34	1.066–1.538	33.7%
Zooplankton mortality parameter (κ_{Zoo})	1.6–4.8	4.57	0.772–5.315	142%

3.2. Zooplankton Mortality as Random Parameter

We here diverge from the setup by Kriest et al. (2017) and apply R-CMA-ES to consider zooplankton mortality κ_{Zoo} (see Kriest & Oschlies, 2015, equations 9–11) as a random parameter. In particular, we aim to optimize the remaining five parameters (the light and nutrient affinity of phytoplankton, maximum zooplankton grazing rate, the oxygen demand of remineralization, and the exponent describing the particle flux curve) such that the model shows a good fit to the misfit function f_{rmse} of Equation 2 over a wide range of zooplankton mortalities. The choice of this parameter can be justified from a modeling point of view. As noted above, zooplankton mortality does not only represent the natural mortality but also the predation by higher trophic levels which is likely to vary, because, for example, of different fish populations preying upon zooplankton. Further, the model assumes a single zooplankton functionality, while in nature many different zooplankton organisms contribute with probably different mortalities. Moreover, the parameter optimization experiment OBS-NARR by Kriest et al. (2017) identified κ_{Zoo} to be most insensitive among the six parameters that are optimized. Table 1 depicts the optimal parameter values of experiment OBS-NARR when only simulations which deviate from the best fit $f_{rmse}(\mathbf{p}^*)$ by less than 5% (according to $(f_{rmse}(\mathbf{p})/f_{rmse}(\mathbf{p}^*) - 1) \cdot 100$) are included. Parameter κ_{Zoo} has the widest range of variability ([0.772,5.315]), exceeding both boundaries of the credible interval [1.6,4.8] used by Kriest et al. (2017). Relative to the credible interval, the range of variability is 142%. Further, together with the half-saturation constant for PO_4 uptake, κ_{Zoo} was the only parameter that showed a strong and long-lasting trend during the optimization experiment OBS-NARR, while the other four parameters approached their optimum value much earlier (cf. Figure 5). Therefore, we decided to optimize the other five parameters considered by Kriest et al. (2017) such that the expected model-data misfit with regard to $\kappa_{Zoo} \in [1.6, 4.8]$ is minimized. More precisely, we aim to obtain five optimal parameters

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \begin{pmatrix} R_{-O_2:P} \\ I_c \\ K_{phy} \\ \mu_{Zoo} \\ b^* \end{pmatrix}$$

such that eventually the integral of f_{rmse} over the random parameter $q = \kappa_{Zoo}$ is minimized. If we assume a uniform distribution of κ_{Zoo} within its credible interval [1.6, 4.8], the definition of the integral is simply the κ_{Zoo} -averaged model-data misfit value

$$F_{rmse}(\mathbf{p}) = \frac{1}{3.2} \int_{\kappa_{Zoo}=1.6}^{4.8} f_{rmse}((p_1, p_2, p_3, p_4, p_5, \kappa_{Zoo})^T) d\kappa_{Zoo}.$$

Table 2

Optimized Parameters and Model-Data Misfit Values of Experiment OBS-NARR and Our New Experiment With Random Zooplankton Mortality Parameter, OBS-RAND

Experiment	OBS-NARR	OBS-RAND
Iterations	182	120
f_{rmse}	0.4499	0.4509
F_{rmse}	0.4566	0.4537
Oxygen consumption: phosphate release during aerobic remineralization ($R_{-O_2:P}$)	167.0	167.9
Half-saturation for light (I_c)	9.66	9.70
Half-saturation for phosphate (K_{phy})	0.5	0.284
Maximum grazing rate of zooplankton (μ_{Zoo})	1.89	1.72
Vertical increase in sinking speed of organic matter (b^*)	1.34	1.35
Zooplankton mortality parameter (κ_{Zoo})	4.57	n.a.

Note. For CMA-ES, the f_{rmse} -value is the misfit value of the algorithm's solution. For R-CMA-ES, we state the empirically best f_{rmse} -value which is obtained a posteriori by fixing the non-random parameters of the solution vector but choosing 100 different κ_{Zoo} -values such that each value covers an area of probability $\frac{1}{100}$. The same 100 κ_{Zoo} -values were also used to approximate the F_{rmse} -values of both algorithms' solution vectors.

More generally, using a probability distribution with density function pdf for the values of κ_{Zoo} , we consider the expected model-data misfit value

$$F_{\text{rmse}}(\mathbf{p}) = \int_{\kappa_{\text{Zoo}}=-\infty}^{\infty} \text{pdf}(\kappa_{\text{Zoo}}) \cdot f_{\text{rmse}}((p_1, p_2, p_3, p_4, p_5, \kappa_{\text{Zoo}})^T) d\kappa_{\text{Zoo}}. \quad (3)$$

In this contribution we assume normally distributed parameter values and define κ_{Zoo} to be $\mathcal{N}(3.2, 1.6)$ distributed, that is, normally distributed with mean 3.2 (the middle of the credible interval) and standard deviation 1.6 (half interval length).

Thus, we compare two model calibration experiments. The first one is the reference experiment OBS-NARR by Kriest et al. (2017). Our new calibration experiment is referred to as OBS-RAND. It deviates from OBS-NARR by defining κ_{Zoo} to be random and $\mathcal{N}(3.2, 1.6)$ distributed, aiming to minimize the integrated misfit function F_{rmse} instead of the parameter point-based misfit function f_{rmse} . The model configuration and the computing facility are the same for both experiments.

4. Results and Discussion

We note that both model calibration approaches have a common set of five non-random parameters to optimize. In order to compare their performance, following optimization we carried out 100 model simulations with the five optimal parameters over the range of zooplankton mortality rates. We then selected the minimum value of Equation 2 as representative for f_{rmse} , and the average over all 100 simulations as representative for F_{rmse} . Note that for OBS-NARR, the former was the target of optimization with CMA-ES, whereas for OBS-RAND F_{rmse} (the integrated misfit) was the implied target. Table 2 shows the results of this analysis, together with the optimal parameter sets. The optimized model-data misfit values of experiment OBS-RAND are quite close to the optimal model-data misfit value of experiment OBS-NARR. Indeed, we can see that the optimizations with R-CMA-ES (experiment OBS-RAND) improved the model-data misfit averaged over the entire range of zooplankton mortalities from 0.4566 to 0.4537 (see Table 2), while the best misfit of 0.4499 for a single (six-)parameter vector has been found by CMA-ES in the reference experiment, OBS-NARR by Kriest et al. (2017).

Compared to experiment OBS-NARR by Kriest et al. (2017), our global ocean biogeochemical model calibrations with random quadratic zooplankton mortality parameter p_{Zoo} hardly affect parameters related to long-term and/or large scale model processes which are affected by circulation and latitude, such as I_c , $R_{-O_2:P}$, and

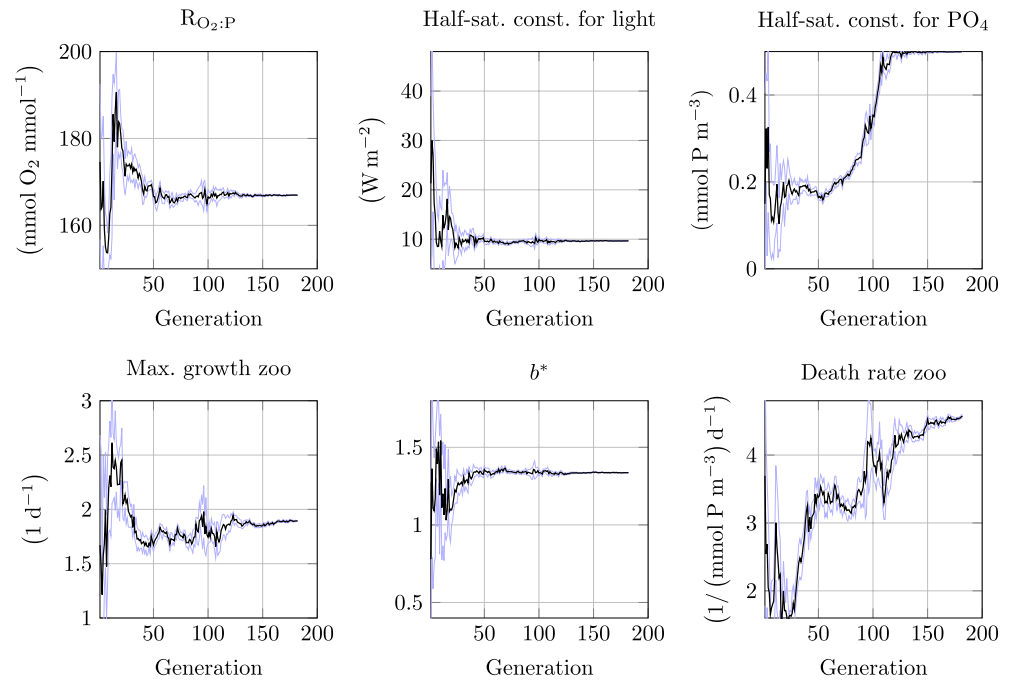


Figure 5. Parameter convergence w.r.t. experiment OBS-NARR.

the sinking speed parameter b^* . On the other hand, the half saturation constant for nutrient uptake K_{phy} is not pushed to the upper bound of its credible interval any more, but is closer to its center. Also, the maximum grazing rate μ_{zoo} of zooplankton drops by 10%. A detailed view of the optimization trajectory is provided in Figure 5, which shows the convergence of the six free model parameters obtained by the reference experiment OBS-NARR while Figure 6 shows the corresponding result of experiment OBS-RAND. Figure 7 illustrates the exploration of the parameter space by R-CMA-ES, as well as its convergence toward five optimal parameters of p^* .

Further we find that R-CMA-ES converged faster compared to CMA-ES, as indicated by the lower number of iterations required for convergence (cf. Table 2). Both of these responses might be a result of the modifications in the optimization procedure, as R-CMA-ES optimizes the expected misfit with regard to the most misfit-insensitive parameter.

Figure 8 shows the model-data misfit measure, f_{rmse} , for 100 parameter sets with varying κ_{zoo} given optimal vector of the non-random parameters p . In OBS-NARR the minimum misfit is obtained at the corresponding κ_{zoo} -value of 4.57 (blue dot). On the contrary, by minimizing the expected misfit $F_{rmse}(p)$ using R-CMA-ES, we obtain a lower minimum misfit over a wider range of κ_{zoo} . When partitioning f_{rmse} with respect to its three components, namely the misfit to observed oxygen, phosphate and nitrate (Figure 9), we see that the improvement is mainly caused by a uniformly better RMSE of oxygen, while the RMSE of nitrate and, to a smaller extend, the RMSE of phosphate increased. Because the oxygen misfit comprises about 45% and both nutrients together only about 55% of the total misfit (cf. the right panel of Figure 9), the increased RMSEs of both nutrients could not compensate the better agreement with oxygen. Verifying the bias of the three tracers we find that the oxygen bias improves by 4 to 8 mmol m^{-3} for all κ_{zoo} -values in the credible interval [1.6,4.8] in experiment OBS-RAND (Figure 10, red lines in lower right panel) compared to the values of the reference experiment OBS-NARR (black lines). Averaging over κ_{zoo} in [1.6,4.8] the oxygen bias improved from +10.1 mmol m^{-3} to +5.6 mmol m^{-3} , that is, we have a smaller overestimation of the observed global oxygen content (172.9 mmol m^{-3}), which is also recognizable in the spatially resolved, depth integrated oxygen bias of Figure 11. Concerning the nutrient bias, there is almost no change for nitrate and phosphate between experiments OBS-RAND and OBS-NARR (also cf. Figure 11), that is, their higher RMSEs appear to be caused by higher pattern errors, only.

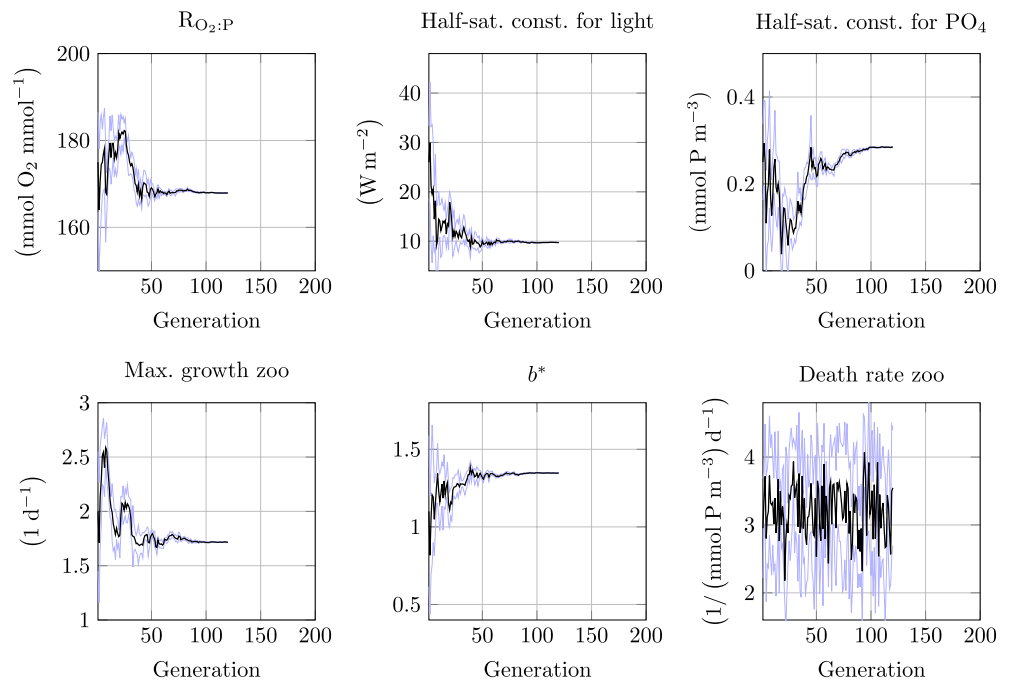


Figure 6. Parameter convergence using R-CMA-ES, keeping the normal distribution of the zooplankton mortality parameter (κ_{Zoo}) as $\mathcal{N}(3.2, 1.6)$, while optimizing the other five free BGC parameters.

4.1. Increasing Zooplankton Mortality Decreases Global Oxygen Inventory

To understand the reasons for the improved oxygen misfit across the range of zooplankton mortalities, we have to investigate the role of those two parameters (μ_{Zoo} and K_{Phy}) that changed significantly between the two optimization approaches. We therefore set up a set of sensitivity simulations applying the optimal values for μ_{Zoo} (hereafter dubbed MU) and K_{Phy} (KP) derived by the two optimization procedures in different combinations (see Table 3). As for OBS-NARR (hereafter called NA) and OBS-RAND (RA) for each sensitivity set-up, we ran a set of 100 simulations for different κ_{Zoo} values. By doing so, we hope to disentangle the effects of μ_{Zoo} , K_{Phy} and κ_{Zoo} on oxygen and biogeochemical cycling.

Figure 10 shows the change of the most important tracer concentrations and fluxes with regard to the four scenarios. We find considerable effects of the mortality rate on plankton concentrations. Increasing zooplankton mortality leads to a strong decline in zooplankton concentrations, less grazing on phytoplankton and thus an increase in phytoplankton concentration. Because phytoplankton mortality is a linear function of its concentration, it increases with increasing zooplankton mortality. Egestion decreases in line with zooplankton concentration (and grazing; not shown here). The decline in zooplankton concentrations is, however, not strong enough to counteract the increase in its mortality rates, leading to an increase in the mortality flux. Zooplankton mortality contributes mostly to the production of sinking detritus, followed by phytoplankton mortality and egestion. The antagonistic responses of detritus production through egestion and the respective mortality fluxes cause a net increase in export production by about 8% for scenario RA. Larger export of organic matter to deep waters, where it is respired and thus consumes oxygen, in turn causes a decline in global average oxygen by about 8% (16.6 Pmol in terms of global inventory or 14.2 mmol m⁻³ global average oxygen) for scenario RA after 3,000 years of simulation. The effects of changes in mortality rate are similar for all four model setups that apply different combinations of μ_{Zoo} and K_{Phy} .

Changes in these two parameters affect the oxygen inventory only by about 1%–2%, but we note that here we investigate a smaller range of variation in the parameters. A smaller K_{Phy} (higher nutrient sensitivity) as in MU causes a more efficient nutrient uptake and larger growth of phytoplankton; yet, this is not reflected in its concentration or mortality, but propagated into zooplankton grazing (not shown), egestion and mortality. The resulting increase in export production in turn causes a decline in oxygen; thus, for MU the decrease in oxygen is caused through zooplankton egestion and mortality. A smaller zooplankton grazing rate in KP causes a decline in its

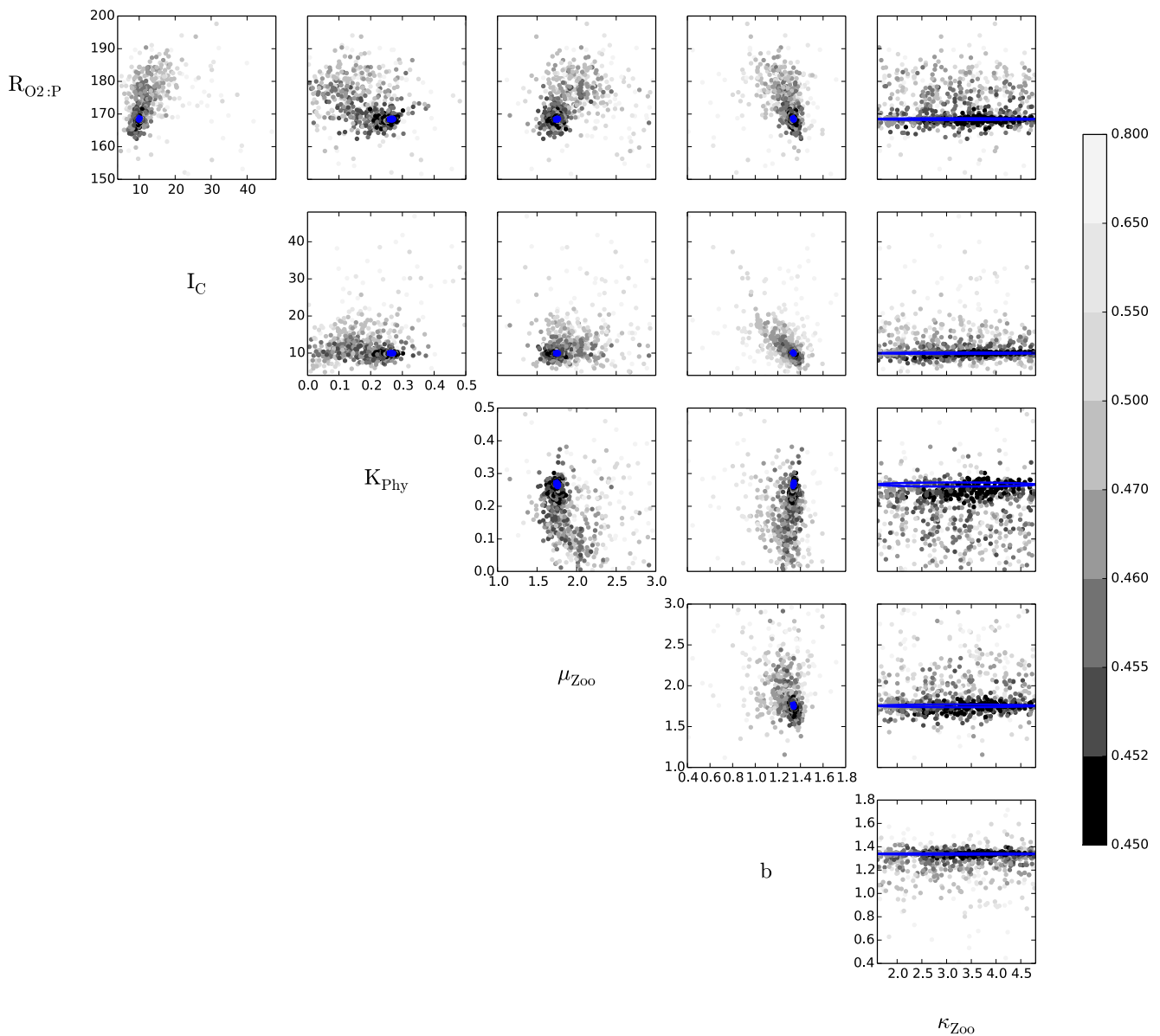


Figure 7. Model simulations of the first 80 iterations of experiment OBS-RAND. The 15 panels represent the projections of the six-dimensional parameter vectors (\mathbf{p} ; \mathbf{q}) (sampled by R-CMA-ES) onto the planes of all possible parameter pairs. The gray-scale color scheme represents the function values $f_{\text{rmse}}(\mathbf{p})$ with increasing values from dark to light shades (see color bar on the right). Blue ellipses denote the projections of the standard deviation ellipsoid of the normal distribution in iteration 80.

concentration, egestion and mortality. However, it also relieves the grazing pressure on phytoplankton, thereby enhancing its concentration and mortality. The resulting increase in export production, which decreases the average oxygen concentration, is thus caused by phytoplankton mortality.

Setup RA combines both parameter changes (small K_{Phy} and μ_{Zoo}) which eventually add up to the highest export production and lowest global average oxygen concentration. Thus, the side effects of introducing a random zooplankton mortality in OBS-RAND lead, to some extent, to a modification in the model's biogeochemical cycling, which eventually results in a global oxygen inventory that is about 6 mmol m^{-3} (about 3%) lower than in OBS-NARR, which relied on a single, fixed value of zooplankton mortality. Nevertheless, the effects of zooplankton mortality on global oxygen content are about two to three times as large.

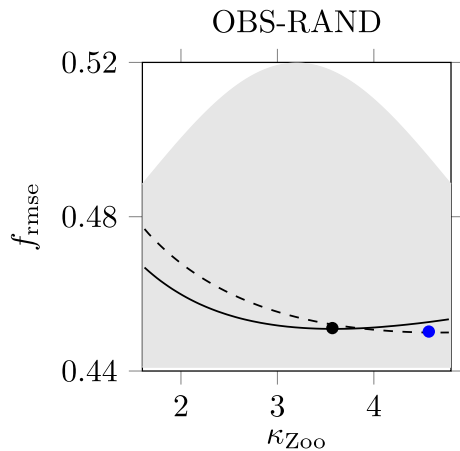


Figure 8. Misfit values f_{rmse} in dependence of the random parameter κ_{Zoo} if all other (optimized) parameters are fixed. The dashed curve corresponds to calibration experiment OBS-NARR. The solid curve corresponds to calibration experiment OBS-RAND. Dots represent the minima along the misfit curves. The area below the probability density function (pdf) of the normal distribution of parameter κ_{Zoo} is colored gray.

5. Conclusions

We introduced R-CMA-ES as a new variant of the Covariance Matrix Adaptation Evolution Strategy. R-CMA-ES allows us to declare one or more parameters to be random, that is, the algorithm seeks to adjust only the other (non-random) parameters in order to minimize a new misfit function, which is the expectation of the former misfit function, integrating over all values of the random parameters. Such a calibration can be more reasonable than searching for single optimal parameter values only. An example are situations where (e.g., due to model simplifications) it is not clear which natural processes are actually covered by a certain parameter and to what extent. Tests with mathematical benchmark functions confirm an efficient convergence behavior of R-CMA-ES as compared to its deterministic counterpart.

We applied R-CMA-ES to a global ocean biogeochemical model setup, which inspired us to develop the algorithm. The model has been considered in a former optimization study by Kriest et al. (2017), who optimized six BGC parameters and observed that two of the parameters—the quadratic loss rate of zooplankton and the phytoplankton half-saturation constant for PO_4 —showed a long lasting drift during optimization. The quadratic loss of zooplankton parameter has high uncertainty because it mimics many different processes (e.g., cannibalism within the highly aggregated zooplankton compartment; predation by fish and higher trophic levels; density-dependent

population control through viral infection), as well as many different zooplankton species. Therefore, this parameter is an ideal candidate to declare random during optimization. Allowing the quadratic loss of zooplankton to vary randomly over a credible interval, R-CMA-ES converged faster than CMA-ES did in the reference experiment of the former study. Moreover, the optimization now also reflects the potential spatio-temporal variability of this parameter, for example, due to higher trophic levels such as fish, which might be of relevance when a BGC model is coupled to a model of higher trophic levels (Getzlaff & Oschlies, 2017; Hill-Cruz et al., 2022). Furthermore, while optimization OBS-NARR seems to cause a bias of the half-saturation constant of phytoplankton for phosphate toward its upper limit, the same parameter is more in the center of its credible interval when we declare the quadratic zooplankton loss random. Another significant change is observed for the zooplankton maximum grazing rate.

Our model results suggest that after 3,000 years of simulation with climatological forcing, the uncertainty in zooplankton mortality causes a variation of oxygen inventory by 8% (16.6 Pmol or 14.2 mmol/m³ average). Smaller changes in maximum zooplankton grazing rate or nutrient affinity of phytoplankton have a smaller effect on biogeochemical fluxes and the long-term global oxygen inventory, but suffice to improve the applied model-data misfit measure over a wide range of zooplankton mortalities. The change in oxygen induced by mortality is of the same order of magnitude as changes induced by anthropogenic climate change (Oschlies, 2021), but it occurs after a considerably longer time span. Also, the oxygen variation across the spectrum of mortality rates is as large as the deviation of many global Earth system models from observations (e.g., Bopp et al., 2013). Our optimizations have shown that it is difficult, if not impossible, to constrain zooplankton mortality with a misfit function that targets at the RMSE of dissolved inorganic tracers (Kriest et al., 2017), which is a common practice when tuning global biogeochemical ocean models. A more flexible tuning strategy as presented here could potentially help to account for this large uncertainty, and may also help to provide a sound and reliable upper closure term and interface for biogeochemical models coupled to higher trophic level (HTL) models.

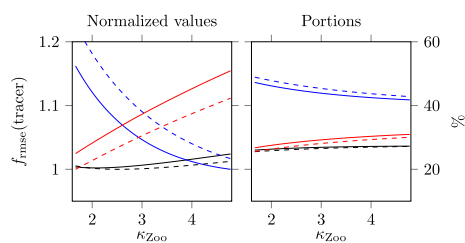


Figure 9. Misfit components of f_{rmse} for the tracers PO_4 (black curves), O_2 (blue curves), and NO_3 (red curves) for experiments OBS-NARR (dashed curves) and OBS-RAND (solid curves) in dependence of the random parameter κ_{Zoo} . The left panel shows values that are normalized w.r.t. the minimum of both corresponding curves. The right panel shows each tracer's portion of the total misfit in percent.

In this study we used an ocean biogeochemical model to showcase the potential advantages of the R-CMA-ES to carry out model calibrations in the view of some uncertain model parameters. Of course, there are many more fields, that face parameter uncertainty, and where a model calibration with random parameters could be useful, such as the personalization cardiac

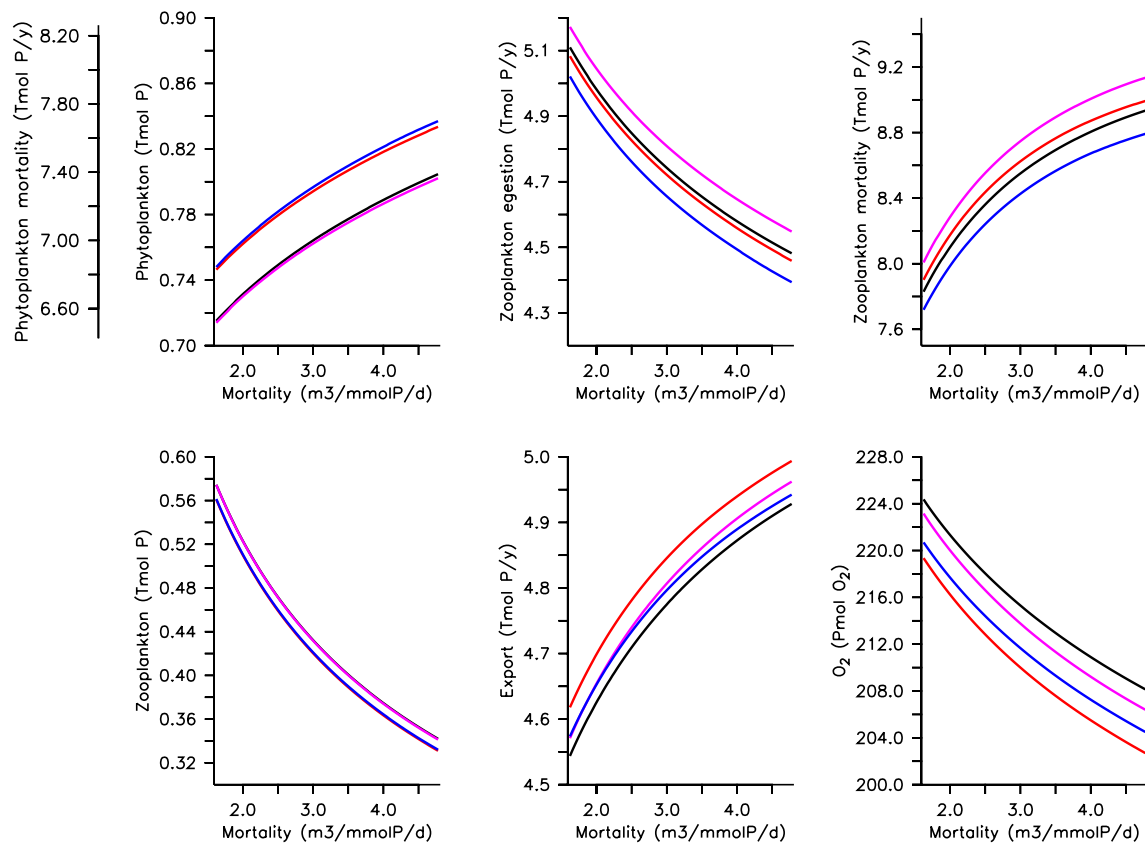


Figure 10. Upper panels: global phytoplankton concentration and its mortality flux to detritus, calculated from phytoplankton $\times 0.85 \times 10.7 \cdot \text{yr}^{-1}$ (left), zooplankton flux to detritus from egestion (middle) and quadratic mortality (right). Lower panels: zooplankton concentration (left), export of detritus sinking through 120 m (middle), and global oxygen inventory (right). The phyto- and zooplankton and biogeochemical fluxes are integrated over the upper 120 m and 1 year. Colors indicate the different scenarios: NA (black), RA (red), MU (pink), and KP (blue).

models (Molléro et al., 2019), modeling of groundwater flow (Elshall et al., 2015; Lykkegaard et al., 2021) or morphodynamic models of a curved channel (Shoarinezhad et al., 2020). All these fields face the issue of parameter uncertainty. In general, a suitable partition of the model parameters into random parameters and non-random parameters needs some problem-dependent pre-considerations. For example, all parameters of interest can be analyzed w.r.t. their covariances and their impact on the model-data misfit function, using multiple model runs (e.g., the simulations of a parameter sensitivity analysis or a deterministic model calibration experiment). Depending on the research question and the observed misfit-sensitivities and covariances of the model parameters, a suitable parameter (set) can be declared random during an optimization, which calibrates the model in the face of (random) parameter uncertainty.

Appendix A: Details of the Classical CMA-ES

As illustrated in Section 2.1 CMA-ES iteratively samples a population of λ candidate solutions from a multi-variate normal distribution $\mathcal{N}(\mathbf{m}, s^2\mathbf{C})$, defined by a mean vector \mathbf{m} , a positive definite covariance matrix \mathbf{C} and an overall scaling factor s . A new normal distribution is empirically re-estimated from the better half of $\mu = \lfloor \frac{\lambda}{2} \rfloor$ samples, and the new probability distribution is used for a smooth update of the former distribution, which in turn is sampled in the next iteration.

A1. Sampling the Normal Distribution

Sampling a parameter vector $\mathbf{p} \in \mathbb{R}^n$ from a multi-variate normal distribution $\mathcal{N}(\mathbf{m}, s^2\mathbf{C})$ is practically realized by choosing n independent samples from the uni-variate standard normal distribution $\mathcal{N}(0, 1)$ (e.g., using the Box-Muller transform) to be the components of a vector $\mathbf{z} \in \mathbb{R}^n$ and defining

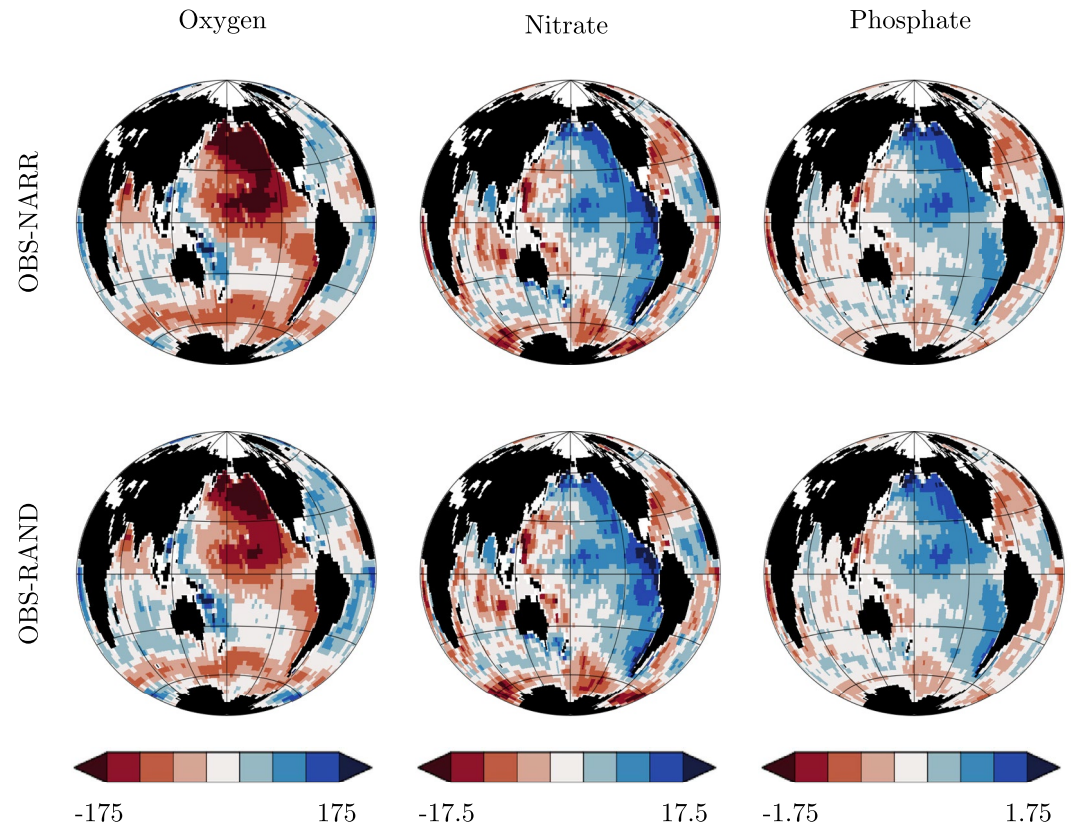


Figure 11. Model deviations from observations of vertically integrated oxygen, nitrate and phosphate (in mol m⁻²) for optimizations OBS-NARR and OBS-RAND, but both in expectation with regard to a random κ_{Zoo} parameter in [1.6, 4.8], that is, we consider the mean result of 100 model simulations with the different κ_{Zoo} values described in the caption of Table 2.

$$p = m + s\mathbf{B}\mathbf{D}z$$

where the matrices **B** and **D** correspond to an eigendecomposition

$$\mathbf{C} = \mathbf{B}\mathbf{D}^2\mathbf{B}^T \tag{A1.}$$

of **C**, that is, the columns $\mathbf{B}^{(i)}$ of **B** are orthogonal eigenvectors of **C**, and \mathbf{D}^2 is a diagonal matrix of corresponding eigenvalues. Geometrically (cf. Hansen, 2016), **B** and **D** can be identified with the so called standard deviation ellipsoid, which is a surface of equal probability density of the normal distribution (see, e.g., the ellipses in Figures 3, 4 and 7). The orientations of the n principal axes of the standard deviation ellipsoid are given by the eigenvectors $\mathbf{B}^{(i)}$ of **C**, and the lengths of its principal axes are given by the roots of the corresponding real and positive eigenvalues $\mathbf{D}_{i,i}^2$. Like in the uni-variate case, the probability that a sample lies within a given area of the search space is obtained by integrating the density function of the probability distribution (pdf) over that area. For a multi-variate normal distribution $\mathcal{N}(\mathbf{m}, \mathbf{C})$, the corresponding probability density function is given as

$$\text{pdf}(\mathbf{p}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{p} - \mathbf{m})^T \mathbf{C}^{-1}(\mathbf{p} - \mathbf{m})\right)}{\sqrt{(2\pi)^n \det \mathbf{C}}}$$

Table 3
Sensitivity Scenarios

Scenario	μ_{Zoo}	K_{Phy}	Remark
NA	1.89	0.5	determined by OBS-NARR
RA	1.72	0.284	determined by OBS-RAND
MU	1.89	0.284	
KP	1.72	0.5	

Note. For each scenario we state the values of parameters μ_{Zoo} and K_{Phy} , respectively. Apart from that, each scenario's set of model simulations is defined by the same 100 different κ_{Zoo} -values between 1.6 and 4.8. The remaining parameter values are constants that have been found by the reference model calibration experiment, OBS-NARR.

A2. Updating the Distribution: Basic Principle

Given any set $S = \{\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(\lambda)}\}$ of λ samples, empirical (re)estimates \mathbf{m}_{emp} and \mathbf{C}_{emp} of the distribution parameters can be calculated such that the expectation of \mathbf{m}_{emp} is \mathbf{m} and the expectation of \mathbf{C}_{emp} is \mathbf{C} . Clearly, the estimates become more reliable the larger λ is. We may assume that the population S is increasingly ordered (ranked) with respect to the considered objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, that is

$$f(\mathbf{p}^{(1)}) \leq f(\mathbf{p}^{(2)}) \dots \leq f(\mathbf{p}^{(\lambda)}).$$

Now, by involving only the better half of $\mu = \lfloor \frac{\lambda}{2} \rfloor$ samples, their distribution estimate $\mathcal{N}(\mathbf{m}_\mu, \mathbf{C}_\mu)$ with corresponding parameters \mathbf{m}_μ and \mathbf{C}_μ will be modified to reproduce that μ samples with higher probability than the other $\lambda - \mu$ samples. CMA-ES uses values $w_1 \geq w_2 \geq \dots \geq w_\mu$ with $\sum_{i=1}^\mu w_i = 1$ to give solutions a rank dependent weight in the updating process of both, \mathbf{m}_μ and \mathbf{C}_μ (a more general version allows to involve all solutions, applying negative weights for the poor ranks). The new mean is, thus, calculated as $\mathbf{m}_\mu = \sum_{i=1}^\mu w_i \mathbf{p}^{(i)}$. A subtlety is the choice of the reference mean value used for estimating \mathbf{C}_μ . Instead of the new empirical mean \mathbf{m}_μ , the mean \mathbf{m} of the former distribution is chosen and yields

$$\mathbf{C}_\mu = \sum_{i=1}^\mu w_i (\mathbf{p}^{(i)} - \mathbf{m})(\mathbf{p}^{(i)} - \mathbf{m})^T.$$

It has the effect that the new distribution is elongated into directions of descend (cf., e.g., iteration 2 in the right example of Figure 2).

A3. Updating the Distribution: Working With Small Populations

As mentioned above, reliable distribution estimates require a sufficiently large number of samples. But, for a competitive computational performance CMA-ES should get along with a rather small number of samples. Therefore the information of former populations is involved by updating the covariance matrix \mathbf{C} to be a (convex) combination of both the current \mathbf{C} and its estimate \mathbf{C}_μ , that is

$$\mathbf{C} \leftarrow (1 - c_\mu)\mathbf{C} + c_\mu\mathbf{C}_\mu. \tag{A2}$$

Using this formula with c_μ as in Table A1, it can be shown that 37% of the current matrix \mathbf{C} 's information dates back at least $\lfloor \frac{1}{c_\mu} \rfloor$ generations, that is, the choice of the smoothing factor c_μ decides about the backward time horizon of the update procedure.

Another feature that facilitates small population sizes λ is to calculate and update a vector \mathbf{p}_c that represents iteration averaged changes of the distribution mean and to use \mathbf{p}_c for a so called rank-one estimate $\mathbf{C}_1 = \mathbf{p}_c \mathbf{p}_c^T$ of the covariance matrix. The idea behind this approach is that, using \mathbf{C}_μ , distribution elongations into directions of descend do not distinguish the sign of the directions. The use of the vector \mathbf{p}_c (called evolution path) mitigates this effect: consecutive changes of the distribution mean into opposite directions would cancel out each other. Similar

Table A1
Operational Constants of the CMA-ES Algorithm (cf. Initialization in Algorithms 1 and 2)

Selection and recombination	Step size control	Covariance matrix adaption
$\lambda = 4 + \lfloor 3 \log n \rfloor$	$\sigma_0 = \frac{1}{3}$	$c_c = \frac{4 + \mu_{\text{eff}}/n}{n + 4 + 2\mu_{\text{eff}}/n}$
$\mu = \lfloor \frac{\lambda}{2} \rfloor$	$\chi = \sqrt{n} \left(1 - \frac{1}{4n} + \frac{1}{21n^2} \right)$	$c_\mu = \min \left(1 - c_1, 2 \frac{\mu_{\text{eff}} + 1/\mu_{\text{eff}} - 2}{(n+2)^2 + \mu_{\text{eff}}} \right)$
$w_i = \frac{\log(\mu + 0.5) - \log(i)}{\sum_{j=1}^\mu \log(\mu + 0.5) - \log(j)}$	$c_\sigma = \frac{\mu_{\text{eff}} + 2}{n + \mu_{\text{eff}} + 5}$	$c_1 = \frac{2}{(n+1.3)^2 + \mu_{\text{eff}}}$
$\mu_{\text{eff}} = \frac{\left(\sum_{i=1}^\mu w_i \right)^2}{\sum_{i=1}^\mu w_i^2} = \frac{1}{\sum_{i=1}^\mu w_i^2}$	$d_\sigma = 1 + c_\sigma$	

Algorithm 1. CMA-ES

```

1: Initialization:
2:   Set  $\lambda, \mu, w, \mu_{\text{eff}}, s_0, \chi, c_\sigma, d_\sigma, c_c, c_\mu, c_1$  according to Table A1
3:   Set  $\mathbf{m} = \frac{1}{2} \cdot \mathbf{1}$ 
4:   Set  $\mathbf{p}_\sigma = \mathbf{p}_c = \mathbf{0}, \mathbf{C} = \mathbf{B} = \mathbf{D} = \mathbf{I}$  and  $\sigma = \sigma_0$ 
5:   while stopping criterion is not met do
6:     Sample probability distribution:
7:       for  $k = 1, \dots, \lambda$  do
8:         Sample  $\mathbf{z}^{(k)}$  from  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ 
9:         Set  $\mathbf{y}^{(k)} = \mathbf{B}\mathbf{D}\mathbf{z}^{(k)}$ 
10:        Set  $\mathbf{p}^{(k)} = \mathbf{m} + s\mathbf{y}^{(k)}$ 
11:       end for
12:       Sort samples by (penalized) objective function values
13:     Update probability distribution:
14:       Update mean:
15:       Set  $\mathbf{m} = \sum_{k=1}^{\mu} w_k \mathbf{p}^{(k)}$ 
16:       Set  $\bar{\mathbf{z}} = \sum_{k=1}^{\mu} w_k \mathbf{z}^{(k)}$ 
17:       Update evolution paths:
18:        $\mathbf{p}_\sigma \leftarrow (1 - c_\sigma)\mathbf{p}_\sigma + \sqrt{c_\sigma(2 - c_\sigma)\mu_{\text{eff}}} \mathbf{B}\bar{\mathbf{z}}$ 
19:        $\mathbf{p}_c \leftarrow (1 - c_c)\mathbf{p}_c + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}} \mathbf{B}\mathbf{D}\bar{\mathbf{z}}$ 
20:       Update covariances and scaling:
21:        $\sigma \leftarrow \sigma \cdot \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma\|}{\chi} - 1\right)\right)$ 
22:       Set  $\mathbf{C}_\mu = \sum_{k=1}^{\mu} w_k \mathbf{y}^{(k)} \mathbf{y}^{(k)T}$  and  $\mathbf{C}_1 = \mathbf{p}_c \mathbf{p}_c^T$ 
23:        $\mathbf{C} \leftarrow (1 - c_\mu - c_1)\mathbf{C} + c_1 \mathbf{C}_1 + c_\mu \mathbf{C}_\mu$ 
24:       Derive  $\mathbf{B}$  and  $\mathbf{D}$  according to (A1)
25:     end while

```

to the smoothing with factor c_μ in the update of \mathbf{C} in Equation A2, the update of \mathbf{p}_c is done with a smoothing factor c_c . With a further smoothing factor c_1 for the rank-one estimate \mathbf{C}_1 , the combined covariance matrix update reads

$$\mathbf{C} \leftarrow (1 - c_\mu - c_1)\mathbf{C} + c_\mu \mathbf{C}_\mu + c_1 \mathbf{C}_1.$$

While \mathbf{C}_μ efficiently involves information from the current population into the update process, \mathbf{C}_1 exploits correlations between generations. The former is important in large populations, the latter is particularly important in small populations.

A4. Step Size Control

Finally, there is an additional explicit adaption of the overall scale (the step size) of the distribution by adapting a scaling factor σ , actually using $\mathcal{N}(\mathbf{m}, \sigma^2 \mathbf{C})$ instead of $\mathcal{N}(\mathbf{m}, \mathbf{C})$. Similar to the evolution path \mathbf{p}_c for the rank-one covariance matrix estimates above, the adaption of the scale σ involves an evolution path \mathbf{p}_σ that mirrors cumulative changes of the mean. The difference between the update formulas of both evolution paths \mathbf{p}_σ and \mathbf{p}_c is that for \mathbf{p}_σ all step sizes are re-scaled with respect to the isotropic normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, where \mathbf{I} is the identity matrix and $\mathbf{0}$ is the zero vector. The expected step size between the mean vectors of two consecutive iterations is therefore the expected length χ of a sample of $\mathcal{N}(\mathbf{0}, \mathbf{I})$, which is

$$\chi := \mathbb{E}(\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|) \approx \sqrt{n} \left(1 - \frac{1}{4n} + \frac{1}{21n^2}\right).$$

Now, a rather small length $\|\mathbf{p}_\sigma\|$ compared to χ indicates that consecutive normalized moves of the mean vector canceled each other out, meaning that the overall scale of the distribution should be reduced by s . Vice versa, an

Algorithm 2. R-CMA-ES: A random coordinates permitting CMA-ES variant

```

1: Initialization:
2:   Set  $\mu, w, \mu_{\text{eff}}, \chi, c_\sigma, d_\sigma, c_c, c_\mu, c_1, s_0$  according to Table A1 (but
    $\lambda = 2\mu$ ) and  $\sigma = \sigma_0$ 
3:   Set  $\mathbf{m} = \frac{1}{2} \cdot \mathbf{1}$   $\parallel$  (dimension  $n + r$ )
4:   Set  $\mathbf{p}_\sigma = \mathbf{p}_c = \mathbf{0}, \mathbf{C} = \mathbf{B} = \mathbf{D} = \mathbf{I}$   $\parallel$  (all with dimension  $n$ )
5: while stopping criterion is not met do
6:   Sample probability distribution:
7:   for  $k = 1, \dots, \mu$  do
8:     Sample  $\mathbf{z}^{(k)}$  from  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ 
9:     Set  $\mathbf{z}^{(\mu+k)} = (-\mathbf{z}_n^{(k)}; \mathbf{z}_r^{(k)})$ 
10:    end for
11:    for  $k = 1, \dots, \lambda$  do
12:      Set  $\mathbf{y}^{(k)} = \mathbf{B}\mathbf{D}\mathbf{z}_n^{(k)}$   $\parallel$   $\mathbf{y}^{(k)}$  has dimension  $n$ 
13:      Set  $\mathbf{p}^{(k)} = \mathbf{m} + \left(\sigma \mathbf{y}^{(k)}; \frac{1}{2} \mathbf{z}_r^{(k)}\right)$ 
14:    end for
15:    Sort samples (see paragraph above) so that selected samples have
    indices in  $[\mu]$ 
16:] Update probability distribution:
17:   Update mean:
18:   Set  $\mathbf{m} = \left(\left(\sum_{k=1}^\mu w_k \mathbf{p}_n^{(k)}\right); \frac{1}{2} \cdot \mathbf{1}_r\right)$ 
19:   Set  $\bar{\mathbf{z}} = \left(\left(\sum_{k=1}^\mu w_k \mathbf{z}_n^{(k)}\right); \mathbf{0}_r\right)$ 
20:   Update evolution paths:
21:    $\mathbf{p}_\sigma \leftarrow (1 - c_\sigma)\mathbf{p}_\sigma + \sqrt{c_\sigma(2 - c_\sigma)\mu_{\text{eff}}}\mathbf{B}\bar{\mathbf{z}}_n$ 
22:    $\mathbf{p}_c \leftarrow (1 - c_c)\mathbf{p}_c + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}}\mathbf{B}\bar{\mathbf{D}}\bar{\mathbf{z}}_n$ 
23:   Update covariances and scaling:
24:    $\sigma \leftarrow \sigma \cdot \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma\|}{\chi} - 1\right)\right)$ 
25:   Set  $\mathbf{C}_\mu = \sum_{k=1}^\mu w_k \mathbf{y}^{(k)} \mathbf{y}^{(k)T}$  and  $\mathbf{C}_1 = \mathbf{p}_c \mathbf{p}_c^T$ 
26:    $\mathbf{C} \leftarrow (1 - c_\mu - c_1)\mathbf{C} + c_1 \mathbf{C}_1 + c_\mu \mathbf{C}_\mu$ 
27:   Derive  $\mathbf{B}$  and  $\mathbf{D}$  from  $\mathbf{C}$  according to (A1)
28: end while

```

evolution path \mathbf{p}_σ longer than χ indicates consecutive distribution drifts into correlated directions which justifies a larger overall scale of the distribution.

A5. Operational Constants and the Pseudo Code

The classical CMA-ES is (reduced by some subtleties) outlined as Algorithm 1 (cf. Hansen, 2016). This pseudo code may serve to verify our adaptations to the new situation, which we will elucidate in detail in Appendix B and summarize as Algorithm 2. Here, we denote the identity matrix with \mathbf{I} , the all-ones vector $(1, \dots, 1)^T$ with $\mathbf{1}$, and the zero vector with $\mathbf{0}$. Note, that in Algorithm 1 the sampled candidate parameter vectors $\mathbf{p}^{(k)}$ are always assumed to be increasingly ordered with respect to the (penalized) objective function, that is, the best μ out of λ parameter vectors are used for updating both, \mathbf{m} and \mathbf{C} .

Appendix B: Details of R-CMA-ES

As already explained in Section 2, we will modify both, the distribution and the sampling and selection procedure of the CMA-ES algorithm in order to efficiently optimize the integrated objective function F defined by Equation 1 instead of its point-based counterpart f .

B1. Modification of the Distribution

We will use the following notations, most of which were already introduced in Section 2.2: we denote by n be the number of non-random parameters which we want to optimize and by r the number of random parameters. We can restrict to the case that the first n components of the parameter vectors are the non-random parameters and the last r components are the random parameters. For $\mathbf{p} \in \mathbb{R}^n$ we write \mathbf{p}_n for the subvector $(p_1, \dots, p_n)^T \in \mathbb{R}^n$ and \mathbf{p}_r for the subvector $(p_{n+1}, \dots, p_{n+r})^T \in \mathbb{R}^r$. Vice versa, if a vector $\mathbf{p} \in \mathbb{R}^n$ and a vector $\mathbf{q} \in \mathbb{R}^r$ are given, then we denote by $(\mathbf{p}; \mathbf{q})$ the vector $(p_1, \dots, p_n, q_1, \dots, q_r)^T \in \mathbb{R}^{n+r}$.

In order to incorporate the variability of a parameter p_i in the procedure, we must fix both, the mean and the variance of the random parameter, that is,

$$p_i \sim \mathcal{N}(0.5, 0.5)$$

in our case, as we operate on $[0,1]^{n+r}$. This can be done by using some suitable modification \mathbf{C}' of the scaled covariance matrix $s^2\mathbf{C}$ (and additionally by keeping $\mathbf{m}_i = 0.5$).

A natural modification is to overwrite the i th column of $s^2\mathbf{C}$ with $0.25 \cdot \mathbf{e}_i$ and the i th row of $s^2\mathbf{C}$ with $0.25 \cdot \mathbf{e}_i^T$, where \mathbf{e}_i is the i th unit vector. It implies that \mathbf{e}_i becomes an eigenvector of the modified covariance matrix \mathbf{C}' with eigenvalue 0.25 and, thus, that one of the principal axes of the standard deviation ellipsoid is parallel to the i th coordinate axis and has length 0.5.

Essentially the same result can be achieved by restricting the maintained multi-variate normal distribution to the set of non-random parameters, that is, $\mathbf{C} \in \mathbb{R}^{n \times n}$, calculating from each sample $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ the corresponding vector of non-random coordinates $\mathbf{y}_n = \mathbf{B}\mathbf{D}\mathbf{z}_n$ (where \mathbf{B} and \mathbf{D} describe the eigendecomposition of \mathbf{C} according to Equation A1), and the corresponding vector of random coordinates $\mathbf{y}_r = \frac{1}{2}\mathbf{z}_r$. Finally, the new scaled and shifted sample (cf. lines 12–13 in Algorithm 2) can be defined by

$$\mathbf{p} = \mathbf{m} + (\sigma \mathbf{y}_n; \mathbf{y}_r).$$

B2. Modification of the Sampling and Selection Procedure

Let $f : \mathbb{R}^{n+r} \rightarrow \mathbb{R}$ be the objective function in the deterministic optimization case and $X_r \subseteq \mathbb{R}^r$ be the subspace of all random components. By pdf we denote the probability density function of the random components' (r -dimensional) normal distribution $\mathcal{N}(0.5 \cdot \mathbf{1}, 0.5 \cdot \mathbf{I})$. We wish to optimize the integrated objective function $F : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by Equation 1:

$$F(\mathbf{p}) = \int_{\mathbf{q} \in X_r} \text{pdf}(\mathbf{q}) \cdot f((\mathbf{p}; \mathbf{q})) d\mathbf{q}.$$

As reasoned in Section 2.2 we want the values of a random component i to stay independently $\mathcal{N}(0.5, 0.5)$ distributed with regard to the μ selected samples (like it is the case with regard to all samples). For this purpose, we adapt the idea of mirrored sampling (Auger et al., 2011) to our situation and generate pairs of samples that are mirrored at the axis (hyper plane) of the random parameter(s), meaning that we sample the vectors $\mathbf{z}^{(k)}$ only for $k \in [\mu]$ and set $\mathbf{z}^{(\mu+k)} := (-\mathbf{z}_n^{(k)}; \mathbf{z}_r^{(k)})$ (cf. lines 8–9 of Algorithm 2). An example for this kind of mirrored sampling is sketched in panel (c) of Figure 4. From each pair of samples we select exactly one sample for the update of the normal distribution, namely the better one w.r.t. f .

B3. Pseude Code

We summarize R-CMA-ES as Algorithm 2. For the update formulas in lines 18–19 as well as for the calculation of \mathbf{C}_μ , we assume that (by sorting) the first μ samples have been selected according to the procedure described above.

Appendix C: R-CMA-ES Test-Bed

We consider an ensemble of mathematical benchmark functions on which we compare CMA-ES to R-CMA-ES. Similar to our real-world application, we restrict R-CMA-ES to deal with a single random component and choose $n = 6$ as problem dimension. Additionally, we use $n = 10$. Our test cases invoke 5 benchmark functions; one linear function, one bowl-shaped (“sphere”) function, one valley-shaped (“Rosenbrock”) function, and two functions (“Griewank” and “Rastrigin”) with many local minima. The details are listed in Table C1.

Table C1
Test Problems

Instance	Test function	Dim.	Feasible domain	Random Param.
Linear1	$f(x) = -\sum_{i=1}^n x_i$	6	$[0, 1]^6$	3
Linear2		10	$[0, 1]^{10}$	1
Sphere1	Sphere	6	$[-1, 1]^6$	3
Sphere2		10	$[-1, 1]^{10}$	1
Rosenbrock1	Rosenbrock	6	$[-1, 3]^2 \times [0.1, 0.3] \times [-1, 3]^3$	3
Rosenbrock2		10	$[0.5, 0.8] \times [-1, 3]^9$	1
Griewank1	Griewank	6	$[-10, 10]^6$	3
Griewank2		10	$[-10, 10]^{10}$	1
Rastrigin1	Rastrigin	6	$[-5.12, 5.12]^6$	3
Rastrigin2		10	$[-5.12, 5.12]^{10}$	1

For problem dimension 6 we set the number of samples per iteration $\lambda = 10$ and pose a maximum iteration number of 200; for problem dimension 10 we use $\lambda = 12$ and pose an iteration limit of 300. As a second stopping criterion a standard deviation of at most 10^{-4} must be satisfied for all parameters. Table C2 represents the results that we obtained by applying both, a Matlab implementation of CMA-ES (Hansen, 2012, cmaes.m, Version 3.61. beta) and our R-CMA-ES algorithm (also implemented in Matlab).

Table C2
Results Obtained With CMA-ES and R-CMA-ES for the Test Instances of Table C1

Instance name	CMA-ES			R-CMA-ES		
	funct. eval.	f -value	F -value	funct. eval.	f -value	F -value
Linear1	537	-6.00	-5.50	1,892	-5.94	-5.44
Linear2	1,054	-10.0	-9.50	3,524	-9.91	-9.41
Sphere1	745	0.00	0.29	1,237	0.00	0.29
Sphere2	1,214	0.00	0.29	1,796	0.00	0.29
Rosenbrock1	1,753	3.43	5.29	1,637	3.38	4.25
Rosenbrock2	3,364	6.00	12.72	3,040	6.79	8.19
Griewank1	1,390	0.05	1.06	1,906	0.02	0.94
Griewank2	1,990	0.01	1.02	3,528	0.00	1.04
Rastrigin1	1,331	6.27	22.97	1,223	4.18	21.67
Rastrigin2	2,287	18.94	33.76	2,107	6.66	24.14

Note. We used 10 optimization trials (with different random numbers) for both algorithms and each instance. In each case we show the mean result over all trials. The f -values and F -values have been calculated like in Table 2 but using 10,000 (instead of 100) values for the random parameter.

Eight out of our ten test instances have the property that the optimal solution to the deterministic optimization task does also belong to the set of optimal solutions to the optimization task with a random parameter. For example, the all-ones vector $\mathbf{1}$ is the global optimum of the benchmark function f of the test instances Linear1 and Linear2, but also a solution to the problem

$$\min_{p \in \{0,1\}^n} F(p).$$

Nevertheless, benchmark functions that exhibit the mentioned property serve well in proving a good convergence behavior of R-CMA-ES in comparison to CMA-ES, if it has similar (or even smaller) f -values and F -values. This is indeed the case for the eight respective test instances. The situation is different for both Rosenbrock instances. Here, the optimal solutions with regard to f and F differ. Indeed, CMA-ES often finds solutions that have a better f -value than the solutions of R-CMA-ES; but vice versa, all solutions found by R-CMA-ES have significantly better F -values.

For the well-shaped test instances Linear1, Linear2, Sphere1, and Sphere2, CMA-ES converged after less function evaluations than R-CMA-ES did; the factor was about 3 for the linear instances and about 1.5 for the sphere instances. However, R-CMA-ES required less function evaluations for both valley-shaped instances Rosenbrock1 and Rosenbrock2 and for the jagged instances Rastrigin1 and Rastrigin2.

Data Availability Statement

The implementation of our optimization algorithm R-CMA-ES and the benchmark function test-bed in MATLAB as well as the C++ implementation for model calibration on HPC platforms can be found on GitHub. The permanent version of the code which we used for the experiments of this article is archived in a public zenodo repository (Sauerland, 2023). For compilation, usage, and further notes, we refer to the README contained in that repository.

The BGC ocean model code and the observational data we used for this study is the same as used by Kriest et al. (2017) and also available on (Sauerland, 2023) (in the folder “MOPS”). The basic TMM and MOPS code as well as required input data for forcing, geometry, and initialization of the model are available to download from (Khaliwala, 2018).

Acknowledgments

This work is a contribution to the “Helmholtz Initiative Climate Adaption and Mitigation (HI-CAM),” the Horizon Europe project “RESCUE” (EU grant agreement ID 101056939), and the BMBF joint project “CO2Meso” (FKZ 03F0876A). Parallel super-computing resources have been provided by the North-German Super-computing Alliance (HLRN). The authors wish to acknowledge use of the Ferret program of NOAA's Pacific Marine Environmental Laboratory for analysis and graphics in this paper. We thank Mariana Hill-Cruz for proofreading the revised version of the manuscript. We thank Christopher J. Somes for reading the manuscript and providing final feedbacks concerning grammar and syntax. Finally, we would like to thank both reviewers for their valuable comments and suggestions. They helped us a lot in improving the structure of the paper.

References

- Anderson, T. R. (2005). Plankton functional type modelling: Running before we can walk? *Journal of Plankton Research*, 27(11), 1073–1081. <https://doi.org/10.1093/plankt/fbi076>
- Auger, A., Brockhoff, D., & Hansen, N. (2010). Mirrored variants of the (1, 2)-CMA-ES compared on the noiseless BBOB-2010 testbed. In M. Pelikan & J. Branke (Eds.), *12th annual genetic and evolutionary computation conference, GECCO 2010, Proceedings, Portland, Oregon, USA, July 7-11, 2010, companion material* (pp. 1551–1558). ACM. <https://doi.org/10.1145/1830761.1830772>
- Auger, A., Brockhoff, D., & Hansen, N. (2011). Mirrored sampling in evolution strategies with weighted recombination. In N. Krasnogor & P. L. Lanzi (Eds.), *13th annual genetic and evolutionary computation conference, GECCO 2011, Proceedings, Dublin, Ireland, July 12-16, 2011* (pp. 861–868). ACM. <https://doi.org/10.1145/2001576.2001694>
- Bopp, L., Resplandy, L., Orr, J., Doney, S., Dunne, J., Gehlen, M., et al. (2013). Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*, 10, 6225–6245. <https://doi.org/10.5194/bg-10-6225-2013>
- Brockhoff, D., Auger, A., Hansen, N., Arnold, D. V., & Hohm, T. (2010). Mirrored sampling and sequential selection for evolution strategies. In R. Schaefer, C. Cotta, J. Kolodziej, & G. Rudolph (Eds.), *Parallel problem solving from nature - PPSN XI, 11th International conference, Kraków, Poland, September 11-15, 2010, Proceedings, Part I* (Vol. 6238, pp. 11–21). Springer. https://doi.org/10.1007/978-3-642-15844-5_2
- Elshall, A. S., Pham, H. V., Tsai, F. T.-C., Yan, L., & Ye, M. (2015). Parallel inverse modeling and uncertainty quantification for computationally demanding groundwater-flow models using covariance matrix adaptation. *Journal of Hydrologic Engineering*, 20(8), 04014087. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001126](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001126)
- Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., et al. (2016). Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, 7(4), 813–830. <https://doi.org/10.5194/esd-7-813-2016>
- Faugeras, B., Lévi, M., Mémyer, L., Verron, J., Blum, J., & Charpentier, I. (2003). Can biogeochemical fluxes be recovered from nitrate and chlorophyll data? A case study assimilating data in the Northwestern Mediterranean Sea at the JGOFS-DYFAMED station. *Journal of Marine Systems*, 40(41), 99–125. [https://doi.org/10.1016/S0924-7963\(03\)00015-0](https://doi.org/10.1016/S0924-7963(03)00015-0)
- Friedrichs, M. A. M. (2002). Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the Central Equatorial Pacific ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(1–3), 289–319. [https://doi.org/10.1016/S0967-0645\(01\)00104-7](https://doi.org/10.1016/S0967-0645(01)00104-7)
- Getzlaff, J., & Oschlies, A. (2017). Pilot study on potential impacts of fisheries-induced changes in zooplankton mortality on marine biogeochemistry. *Global Biogeochemical Cycles*, 31(11), 1656–1673. <https://doi.org/10.1002/2017GB005721>
- Hansen, N. (2006). The CMA evolution strategy: A comparing review. In J. A. Lozano, P. Larranaga, I. Inza, & E. Bengoetxea (Eds.), *Towards a new evolutionary computation. Advances on estimation of distribution algorithms* (pp. 75–102). Springer.
- Hansen, N. (2012). CMA-ES source code. Retrieved from http://www.cmap.polytechnique.fr/~nikolaus.hansen/cmaes_inmatlab.html
- Hansen, N. (2016). The CMA evolution strategy: A tutorial. arXiv:1604.00772. <https://doi.org/10.48550/arXiv.1604.00772>

- Hansen, N., Auger, A., Ros, R., Finck, S., & Posik, P. (2010). Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In *Proceedings of the genetic and evolutionary computation conference, GECCO 2010, Portland, Oregon, USA, July 7-11, 2010, companion material* (pp. 1689–1696).
- Hill-Cruz, M., Frenger, I., Getzlaff, J., Kriest, I., Xue, T., & Shin, Y.-J. (2022). Understanding the drivers of fish variability in an end-to-end model of the Northern Humboldt Current System. *Ecological Modelling*, 472, 110097. (To appear). <https://doi.org/10.1016/j.ecolmodel.2022.110097>
- Hurtt, G. C., & Armstrong, R. A. (1996). A pelagic ecosystem model calibrated with BATS data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2–3), 553–583. [https://doi.org/10.1016/0967-0645\(96\)00007-0](https://doi.org/10.1016/0967-0645(96)00007-0)
- Kane, A., Moulin, C., Thiria, S., Bopp, L., Berrada, M., Tagliabue, A., et al. (2011). Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations. *Journal of Geophysical Research*, 116(C6), C06011. <https://doi.org/10.1029/2009JC006005>
- Kaufman, D. E., Friedrichs, M. A. M., Hemmings, J. C. P., & Smith, W. O., Jr. (2018). Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea. *Biogeosciences*, 15(1), 73–90. <https://doi.org/10.5194/bg-15-73-2018>
- Khatiwal, S. (2007). A computational framework for simulation of biogeochemical tracers in the ocean. *Global Biogeochemical Cycles*, 21(GB3001), 21. <https://doi.org/10.1029/2007GB002923>
- Khatiwal, S. (2018). Transport matrix method software for ocean biogeochemical simulations. [Software]. Zenodo. <https://doi.org/10.5281/zenodo.1246300>
- Khatiwal, 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>
- Kidston, M., Matear, R., & Baird, M. E. (2011). Parameter optimisation of a marine ecosystem model at two contrasting stations in the Sub-Antarctic Zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(21), 2301–2315. <https://doi.org/10.1016/j.dsr2.2011.05.018>
- Kriest, I. (2017). Calibration of a simple and a complex model of global marine biogeochemistry. *Biogeosciences*, 14(21), 4965–4984. <https://doi.org/10.5194/bg-14-4965-2017>
- 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*, 17(12), 3057–3082. <https://doi.org/10.5194/bg-17-3057-2020>
- Kriest, I., Khatiwal, S., & Oschlies, A. (2010). Towards an assessment of simple global marine biogeochemical models of different complexity. *Progress in Oceanography*, 86(3–4), 337–360. <https://doi.org/10.1016/j.pocean.2010.05.002>
- 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., Sauerland, V., Khatiwal, 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>
- Lawson, L. M., Spitz, Y. H., Hofmann, E. E., & Long, R. B. (1995). A data assimilation technique to a predator-prey model. *Bulletin of Mathematical Biology*, 57(4), 593–617. [https://doi.org/10.1016/S0092-8240\(05\)80759-1](https://doi.org/10.1016/S0092-8240(05)80759-1)
- Leles, S. G., Valentin, J. L., & Figueiredo, G. M. (2016). Evaluation of the complexity and performance of marine planktonic trophic models. *Anais da Academia Brasileira de Ciências*, 88(3), 1971–1991. <https://doi.org/10.1590/0001-3765201620150588>
- Lykkegaard, M. B., Dodwell, T. J., & Moxey, D. (2021). Accelerating uncertainty quantification of groundwater flow modelling using a deep neural network proxy. *Computer Methods in Applied Mechanics and Engineering*, 383, 113895. <https://doi.org/10.1016/j.cma.2021.113895>
- 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. *JGR Oceans*, 102(C3), 5733–5752. <https://doi.org/10.1029/96JC02775>
- Molléro, R., Pennec, X., Delingette, H., Ayache, N., & Sermesant, M. (2019). Population-based priors in cardiac model personalisation for consistent parameter estimation in heterogeneous databases. *International Journal for Numerical Methods in Biomedical Engineering*, 35(2), e3158. <https://doi.org/10.1002/cnm.3158>
- Najjar, R. G., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S. C., et al. (2007). Impact of circulation on export production, dissolved organic matter and dissolved oxygen in the ocean: Results from phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2). *Global Biogeochemical Cycles*, 21(GB3007). <https://doi.org/10.1029/2006GB002857>
- Oliver, S. E., Cartis, C., Kriest, I., Tett, S. F. B., & Khatiwal, S. (2021). A derivative-free optimisation method for global ocean biogeochemical models. *Geoscientific Model Development Discussions*, 1–24. <https://doi.org/10.5194/gmd-2021-175>
- Oschlies, A. (2021). A committed fourfold increase in ocean oxygen loss. *Nature Communications*, 12(1), 2307. <https://doi.org/10.1038/s41467-021-22584-4>
- Prieß, M., Koziel, S., & Slawig, T. (2013). Marine ecosystem model calibration with real data using enhanced surrogate-based optimization. *Journal of Computational Science*, 4(5), 423–437. <https://doi.org/10.1016/j.jocs.2013.04.001>
- Sauerland, V. (2023). R-CMA-ES: A random coordinates permitting CMA-ES algorithm. [Software]. Zenodo. <https://doi.org/10.5281/zenodo.8146394>
- 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. <https://doi.org/10.1357/002224003322981147>
- Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., et al. (2017). Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling. *Biogeosciences*, 14(6), 1647–1701. <https://doi.org/10.5194/bg-14-1647-2017>
- Shimoda, Y., & Arhonditsis, G. B. (2015). Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge. *Ecological Modelling*, 320, 29–43. <https://doi.org/10.1016/j.ecolmodel.2015.08.029>
- Shoarinezhad, V., Wieprecht, S., & Haun, S. (2020). Comparison of local and global optimization methods for calibration of a 3D morphodynamic model of a curved channel. *Water*, 12(5), 1333. <https://doi.org/10.3390/w12051333>
- Spitz, Y., Moisan, J. R., Abbott, M. R., & Richman, J. G. (1998). Data assimilation and pelagic ecosystem model: Parameterizations using time series observations. *Journal of Marine Systems*, 16(1–2), 51–68. [https://doi.org/10.1016/S0924-7963\(97\)00099-7](https://doi.org/10.1016/S0924-7963(97)00099-7)
- Tjiputra, J., Polzin, D., & Winguth, A. (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). <https://doi.org/10.1029/2006GB002745>
- Vallino, J. J. (2000). Improving marine ecosystem models: Use of data assimilation and mesocosm experiments. *Journal of Marine Research*, 58(1), 117–164. <https://doi.org/10.1357/002224000321511223>
- Wang, H., Emmerich, M., & Bäck, T. (2019). Mirrored orthogonal sampling for covariance matrix adaptation evolution strategies. *Evolutionary Computation*, 27(4), 699–725. https://doi.org/10.1162/evco_a_00251
- Xiao, Y., & Friedrichs, M. A. M. (2014). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, 119(4), 2691–2712. <https://doi.org/10.1002/2013JC009433>