AGU PUBLICATIONS

Global Biogeochemical Cycles

RESEARCH ARTICLE
10.1002/2017GB005690

Key Points:
- Revisiting optimization techniques of poorly constrained biogeochemical model parameters in a three-dimensional Earth System Model
- For typical models, noisy observations and evaluation metrics no unique "best" parameter set may exist
- Different parameter sets can be equally consistent with historical observations, while they feature strikingly differing future projections

Correspondence to:
U. Löptien and H. Dietze,
uloeptien@geomar.de;
hdietze@geomar.de

Citation:
Löptien, U., and H. Dietze (2017), Effects of parameter indeterminacy in pelagic biogeochemical modules of Earth System Models on projections into a warming future: The scale of the problem, *Global Biogeochem. Cycles*, *31*, 1155–1172, doi:10.1002/2017GB005690.

Received 11 APR 2017
Accepted 12 JUN 2017
Accepted article online 15 JUN 2017
Published online 21 JUL 2017

Effects of parameter indeterminacy in pelagic biogeochemical modules of Earth System Models on projections into a warming future: The scale of the problem

U. Löptien[1,2] and H. Dietze[1,2]

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany, [2]Institute of Geosciences, Christian-Albrechts-University of Kiel, Kiel, Germany

Abstract Numerical Earth System Models are generic tools used to extrapolate present climate conditions into a warming future and to explore geoengineering options. Most of the current-generation models feature a simple pelagic biogeochemical model component that is embedded into a three-dimensional ocean general circulation model. The dynamics of these biogeochemical model components is essentially controlled by so-called *model parameters* most of which are poorly known. Here we explore the feasibility to estimate these parameters in a full-fledged three-dimensional Earth System Model by minimizing the misfit to noisy observations. The focus is on parameter identifiability. Based on earlier studies, we illustrate problems in determining a unique estimate of those parameters that prescribe the limiting effect of nutrient- and light-depleted conditions on carbon assimilation by autotrophic phytoplankton. Our results showcase that for typical models and evaluation metrics no meaningful "best" unique parameter set exists. We find very different parameter sets which are, on the one hand, equally consistent with our (synthetic) historical observations while, on the other hand, they propose strikingly differing projections into a warming climate.

1. Introduction

In a warming world, our limited ability to reduce emissions of climate-relevant species to the atmosphere becomes a societal concern. To-date several so-called geoengineering options have been suggested to alleviate the effects of ever increasing greenhouse gas emissions [e.g., *Govindasamy and Caldeira*, 2000; *Keller et al.*, 2014; *Tuana et al.*, 2012; *Vaughan and Lenton*, 2011]. Some of these options may, or may not, come along with unanticipated side effects on local to global scales. One way to identify and quantify these side effects is to test the respective options in numerical Earth System Models. These models comprise, as far as we know, the most important components (and processes) impinging on our climate such as the atmosphere, the ocean, ocean biogeochemistry, sea ice, land-ice, and land vegetation. Even so, their projections into a warming future are associated with a considerable degree of uncertainty.

Relatively straightforward is the exploration of the uncertainty that is associated with the anthropogenic emission scenarios. In this context, it is common to pragmatically test through a suite of scenarios that envelope anticipated global economic growth [e.g., *Moss et al.*, 2010]. But apart from the uncertainties associated to these boundary conditions, the models themselves may be inherently deficient. Of particular concern are the biogeochemical processes, because (1) their mathematical description is not derived from first principles and the precise choice of description has a strong impact on the model's sensitivity [e.g., *Jones et al.*, 2003; *Löptien*, 2011; *Taucher and Oschlies*, 2011, and many more to follow] and (2) associated model parameters, which—along with the mathematical equations—determine the sensitivity, are poorly constrained [e.g., *Kriest et al.*, 2010; *Löptien and Dietze*, 2015]. The need to address the latter issue is reflected in an increasing number of studies that set out to estimate or constrain such model parameters, by minimizing a metric that measures the misfit to observational data [e.g., *Fan and Lv*, 2009; *Friedrichs et al.*, 2006; *Rückelt et al.*, 2010; *Schartau et al.*, 2001; *Schartau*, 2005; *Spitz et al.*, 1998; *Tjiputra et al.*, 2007; *Hemmings and Challenor*, 2012; *Matear*, 1995; *Ward et al.*, 2010; *Xiao and Friedrichs*, 2014a, 2014b, among others]. Common to many of these efforts is the failure to identify *the* optimal parameter set [*Ward et al.*, 2010; *Schartau et al.*, 2001; *Rückelt et al.*, 2010]. Among the reasons for failure discussed so far are that the respective equations may not represent underlying processes

adequately [*Fasham et al.*, 1995; *Fennel et al.*, 2001] and overfitting, i.e., the attempt to extract more information than is inherent to the respective data [*Matear*, 1995]. A recent study of *Löptien and Dietze* [2015], based on a slab ocean model, supports the early findings of *Matear* [1995] in that it illustrates that even drastic changes to key model parameters can compensate one another. Further, *Löptien and Dietze* [2015] show that model parameter sets that yield almost identical model solutions (i.e., solutions that are equally consistent with observations) can strongly diverge once the boundary conditions change. Unfortunately, a large fraction of this diverging behavior turned out to be attributed to those parameters which are used to describe light and nutrient limitation of phytoplankton growth. It is unfortunate because the magnitude of these parameters is poorly known while they, at the same time, largely control the sensitivity of the biogeochemical system to ever changing environmental controls.

In summary, a number of studies showed that biogeochemical model parameters are difficult to estimate based on observations in one-dimensional (with the one dimension referring to depth) or slab ocean frameworks. This is problematic, as seemingly equivalent solutions can diverge under anticipated future changes. If this also applies to state-of-the-art three-dimensional model frameworks (with the three dimensions referring to longitude, latitude, and depth), such implications would be disconcerting because the implication would be that a model's forecast skill (or reliability in general) is not necessarily associated with goodness-of-fit to historical observations. The discussion to what extent results from one-dimensional models are indicative for the behavior of full-fledged three-dimensional coupled global ocean general circulation biogeochemical models is ongoing. One may argue that there is no conceptional difference between one-dimensional and three-dimensional frameworks other than that one-dimensional frameworks are based on the assumption of horizontal homogeneity of all relevant environmental variables. Clearly, this assumption does not hold for the real ocean. Currents, the angle of the sunlight penetrating into the surface ocean, temperature, stratification, nutrients, and other variables, which exert control on marine biota and associated biogeochemical cycling, do vary strongly in space. But then, three-dimensional model frameworks do also — intrinsically — depend to a certain degree on horizontal averaging or homogeneity, since ocean models commonly rely on a spatial discretization which assumes the environmental variables to be homogenous within a certain area. For global models the horizontal discretization can be rather coarse — typically of the order of hundreds of kilometers. Hence, one may argue that the conceptional differences between one- and three-dimensional approaches are rather small and inferences can be made from one to another.

In the present study we set out to put the above argument to the test. Based on a predefined and common model-data misfit function, we will determine several parameter sets which fit observational data equally well. In a second step, we test the sensitivity of these equivalent model solutions by projecting the differing model versions into the future. Note that the determination of the respective parameter sets requires an automatized optimization procedure, which just recently has become feasible by advancements in computer technology. So far few pioneering studies have been able to estimate a limited number of parameters in a fully spun-up model [e.g., *Kriest et al.*, 2017]. Our study adds to the emerging field in that our focus is on the feasibility to determine a unique solution and, specifically, on the relation between goodness-of-fit to historical observations and reliability (or robustness) of projections into a warming world. Our approach is based on *twin experiments* where we define an arbitrarily chosen parameter set as the "true" one. From the respective simulation, which we refer to as *genuine truth* in the following, we subsequently sample "synthetic observations." These *synthetic observations* are distorted by noise in order to mimic realistic conditions. Among the advantages of the twin experiment approach (applied also by, e.g., *Friedrichs* [2001], *Gunson et al.* [1999], *Lawson et al.* [1996], *Schartau et al.* [2001], and *Spitz et al.* [1998]) is full control over the "observations." We are not limited by data availability and can specify the noise level at will. Another advantage of our twin experiment approach is that our equivalent model solutions differ only in terms of their underlying parameter values while they apply the same biogeochemical and physical model formulations.

The following section 2 describes the Earth System Model and the experimental setup. In section 3 we present our model results and compare different setups under historical and anticipated future greenhouse gas emissions. A discussion follows in section 4. The paper closes with summary and conclusions (section 5).

2. Methods

Our twin experiment approach is based on the University of Victoria Earth System Climate Model (UVic) [*Weaver et al.*, 2001]. We pick an arbitrary parameter set and define the respective simulation as "the"

Table 1. Model Parameters, Occurring in Equations (1)–(5)[a]

Parameter	Description	Unit
a	Maximum growth rate at 0°C	day^{-1}
T_b	*e*-folding temperature dependence	°C
K^P_{Fe}	Half-saturation constant for Fe uptake	mmol Fe m^{-3}
α	Initial slope of the *P-I* curve	$(W\ m^{-2})^{-1}\ d^{-1}$
K_N	Half-saturation constant for NO_3 uptake	mmol N m^{-3}
K_P	Half-saturation constant for PO_4 uptake	mmol P m^{-3}

[a]The respective notation matches the one in *Keller et al.* [2012].

genuine truth. The genuine truth is distorted by noise. This noisy model output is defined as synthetic observations. Additional experiments, so-called twins, are conducted with the same model that underlies the genuine truth. These experiments are twins because they use the exact same configuration — except for the choice of some of the model parameters of the pelagic biogeochemical component. By using an automized optimization process, as outlined in Appendix A, we determine several parameter sets that lead to equally good fits between the respective simulations and the synthetic observations. Finally, we will use these parameter sets to make model-based projections into a warming future. A comparison between these projections will provide a link between the model's forecast skill (or reliability in general) and goodness-of-fit to historical observations.

In the following section 2.1 we describe our modeling framework, the University of Victoria (UVic) Earth System Climate Model. The synthetic observations that are based on the genuine truth simulation are described in section 2.2. In section 2.3 we define measures of goodness-of-fit to observations (also referred to as metrics or cost) that quantify the difference or consistency between a simulation and the synthetic observations. Section 2.4 describes the spin-up procedure and the boundary conditions used to project into a warming future.

2.1. The Earth System Model

Our study is based on the University of Victoria (UVic) Earth System Climate model (version 2.9) which is of intermediate complexity. The model has been applied to explore geoengeneering options in a number of studies [e.g., *Keller et al.*, 2014; *Oschlies et al.*, 2010; *Matthews et al.*, 2009; *Reith et al.*, 2016; *Weaver et al.*, 2007]. Our configuration is identical to the one in *Keller et al.* [2012] which is also referred to as "the reference simulation" in *Getzlaff and Dietze* [2013] and *Getzlaff et al.* [2016]. Note that porting from one computer hardware to another results in (minor) differences, i.e., bit-precise reproducibility is impeded. This holds especially for properties at the sea ice edges.

All model components use a horizontal resolution of 1.8° in latitude and 3.6° in longitude. The UVic model comprises a single-level atmospheric energy-moisture balance model (with prescribed surface winds), a dynamic-thermodynamic sea ice model, a simple land ice model, and an active terrestrial vegetation component [*Weaver et al.*, 2001]. Further, UVic's ocean is based on a three-dimensional primitive-equation model [*Pacanowski*, 1995]. The vertical discretization comprises 19 levels. The vertical resolution starts with 50 m near the surface and gradually increases to 500 m at depth. A marine pelagic ecosystem model is coupled to the ocean general circulation component. As in the setup of *Keller et al.* [2012] phytoplankton growth (J_D) depends on photosynthetically active radiation (PAR), nitrate (NO_3), phosphate (PO_4), iron (Fe), and temperature (T). An overview of the respective model parameters is provided in Table 1. The specific ecosystem model formulation is as follows:

$$J_D := a\, J_{\mathrm{Fe}} \min(J_{OI}, J_{\mathrm{NO_3}}, J_{\mathrm{PO_4}}). \tag{1}$$

Here a determines the maximum phytoplankton growth and J_{Fe} mimics iron limitation and temperature dependence:

$$J_{\mathrm{Fe}} := \frac{\mathrm{Fe}}{\mathrm{Fe} + K^P_{\mathrm{Fe}}} \cdot \exp(T/T_b). \tag{2}$$

The parameter K^P_{Fe} is the so-called *half-saturation constant* for iron limitation and T_b the *e-folding temperature dependence of biological rates*.

The light limitation is expressed as follows:

$$J_{Ol} := \frac{\alpha \, \mathrm{PAR}}{((\alpha \, \mathrm{PAR})^2 + (J_{\mathrm{Fe}})^2)^{(1/2)}}. \tag{3}$$

The parameter α determines the sensitivity toward the incoming photosynthetically active radiation (*initial slope of the P-I curve*).

Growth limitations due to nitrate (NO_3) and phosphate (PO_4) deficiency are expressed by the so-called *Michaelis-Menten (MM)* formulations. The sensitivity toward the availability of these nutrients is determined by the *half-saturation constants* K_N and K_P, respectively (which are related via the Redfield ratio $R_{N:P}$=1/16; i.e., K_P=$R_{N:P}$ K_N):

$$J_{\mathrm{NO_3}} := \frac{\mathrm{NO_3}}{\mathrm{NO_3} + K_N} \tag{4}$$

and

$$J_{\mathrm{PO_4}} := \frac{\mathrm{PO_4}}{\mathrm{PO_4} + K_P}. \tag{5}$$

In the present study we focus on those parameters that (1) determine the control of light (i.e., parameter α in equation (3)) and nutrients (i.e., the parameters $K^p_{\mathrm{Fe}}, K_N, K_P$ in equations (2), (4), and (5), respectively) on phytoplankton growth and (2) prescribe the maximum growth of phytoplankton (i.e., parameter a in equation (1)).

The ratio behind this choice is that these parameters occur in formulations which are common to most biogeochemical models. Also, it has been suggested earlier that these parameters can be difficult to estimate unambiguously while they can, at the same time, impact the model's sensitivity decisively [*Löptien and Dietze*, 2015].

2.2. The Synthetic Observations

As outlined at the beginning of this section, we use synthetic data instead of real-world observations. We define the equilibrated preindustrial spin-up simulation integrated with the reference configuration of *Keller et al.* [2012], *Getzlaff and Dietze* [2013], and *Getzlaff et al.* [2016] as *the genuine truth*. We assume that surface concentrations of phytoplankton and macronutrients are best suited to estimate the considered parameters, because our focus is on model parameters that determine the growth of phytoplankton in the sunlit surface ocean. We do not consider data from deeper parts of the water column where the absence of sunlight forestalls autotrophic growth.

Based on this genuine truth, we distort 5-daily samples with noise. These noisy data are defined as synthetic observations. The added noise is meant to account for all unresolved processes that can cause deviations between a real-world observation and a model-based estimate. Among these processes are (1) measurement accuracy; (2) misalignment between observations and model estimate in space and/or time; (3) unit conversion; e.g., models typically carry nutrient-contained-by-phytoplankton as a prognostic variable, rather than the often observed chlorophyll *a* content; the conversion from one to another is not straightforward because it is intrinsically dependent on local environmental conditions (that are often unknown); and (4) model biases as a consequence of a model structure that is inconsistent or imprecise compared to real-world dynamics. A problem associated with the construction of the noise is that there is so little information on the magnitude, coherence, and probability density functions of the combined effects of the processes listed above. As for the measurement accuracy it is probably prudent to assume that the noise is uncorrelated or "white" [*Rückelt et al.*, 2010; *Schartau et al.*, 2001]. As for misalignment in space and time the resulting noise is probably correlated, or "red" because ocean dynamics typically feature red rather than white spectra in time [cf. *Hasselmann*, 1976]. As for the conversion from chlorophyll *a* to the nutrient content of phytoplankton there are large biases, strongly correlated in space and time [e.g., *Banse*, 1977]. As for the effects of model biases, the fact that models are typically rather coarsely resolved in space and driven by forcing less detailed than the real-world environment suggests that the resulting noise structure is rather red than white. When it comes to the precise definition of noise, we are, however, pressed to make assumptions.

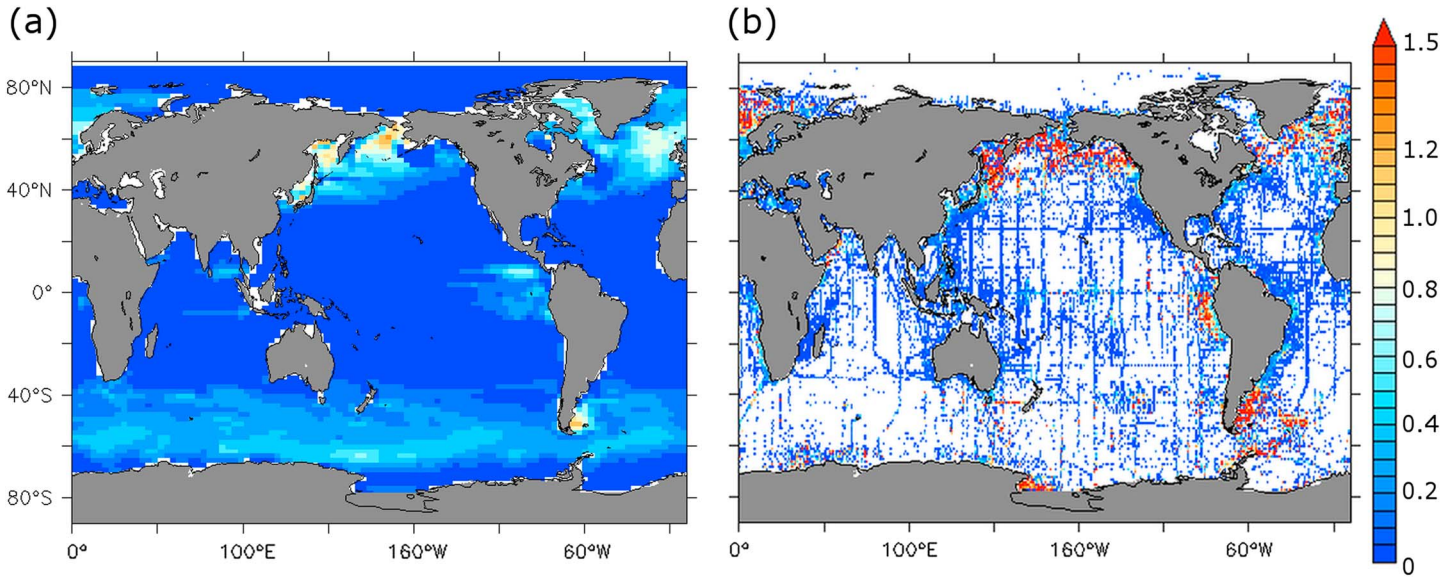

Figure 1. (a) Standard deviation of the noise added to the genuine truth surface nitrate in units mmol N m^{-3}. (b) Annual standard error (i.e., the standard deviation of the mean) of nitrate concentrations in the World Ocean Atlas [*Garcia et al.*, 2009].

In this study, we construct noise based on the assumption that (1) the power spectrum of its temporal variability is rather red than white and (2) the absolute amplitude of the noise is correlated with major patterns of variability in the biogeochemical model variables. Specifically, we produce a low-frequent noise time series by superimposing four autoregressive processes (AR(3)), each featuring a standard deviation of 0.25 in its underlying white noise process [cf. *Löptien and Dietze*, 2015]. The time series are then multiplied by the absolute values of the first four leading empirical orthogonal functions (EOFs) [*Von Storch and Zwiers*, 2001] of modeled surface phytoplankton and nitrate concentrations. (Note that these first four EOFs, based on the genuine truth, explain 73.23 % of the total variance for phytoplankton and 94.32 % for surface nitrate.) By adding the respective noise to our genuine truth we construct our synthetic observations. Resulting negative values in the prognostic variables are truncated and set to zero.

The resulting noise patterns in our synthetic surface nitrate (Figure 1) resembles successfully the "standard errors" (i.e., the standard deviation of the mean) provided by the World Ocean Atlas [*Garcia et al.*, 2009]. The comparison with World Ocean Atlas data further reveals that our synthetic nitrate observations feature less noise (i.e., overall lower noise amplitudes) than typical real-world data. A similar argumentation holds for the noise structure in our synthetic phytoplankton observations (Figure 2). According to a study from *Volpe et al.* [2007] covering the Mediterranean, satellite chlorophyll data may well carry noise amplitudes corresponding to 100% when compared with in situ measurements. For the Argentinean Patagonian continental shelf, *Dogliotti et al.* [2009] report somewhat lower numbers around 20–40%. Note that noise associated to unit conversions, model/data misalignment in space and/or time, and model bias is not included in these numbers but has to be added. We apply noise corresponding to much less than 100% over most of the ocean and suggest that this level is less than actual levels to be expected in real-world applications based on satellite data.

2.3. Metrics of Misfit/Cost

As outlined in the beginning of section 2, we set out to compare and rate model-data misfits of various long-term simulations based on differing parameter sets (cf. Table 2). Such a comparison necessitates the definition of a metric that quantitatively describes the difference (or misfit) between the synthetic observations and a model solution. In the literature, such misfit metrics are also referred to as *objective functions* or *cost functions*, and various approaches exist [e.g., *Stow et al.*, 2009]. Clearly, the result of a parameter optimization depends on the definition of the underlying cost function [cf. *Evans*, 2003]. The present study does not aim to make inferences for all definitions of misfit metrics ever proposed so far, nor for all of those that will ever be proposed. For pragmatic reasons we use two exemplary approaches only. Both are generic in that they are

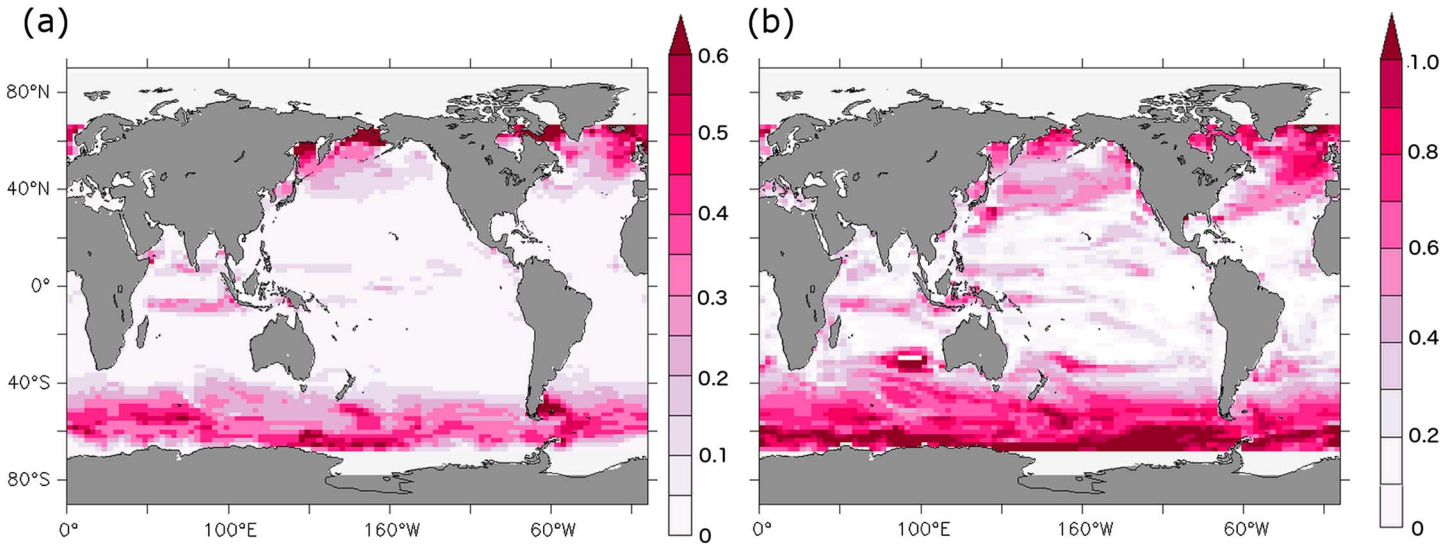

Figure 2. (a) Standard deviation of the noise added to the genuine truth surface phytoplankton in units mmol N m^{-3}. (b) Like in Figure 2a but relative to the local temporal mean phytoplankton concentration in relative units.

based on the sum of the weighted squared differences (RMSE) between model output and observations [cf. *Ward et al.*, 2010; *Schartau et al.*, 2001; *Rückelt et al.*, 2010]:

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{M}\sum_{m=1}^{M} W_m^2 \frac{1}{N_m}\sum_{j=1}^{N_m}\left(a_j - \hat{a}_j\right)^2\right)}, \tag{6}$$

M denotes the number of prognostic variables, N_m the number of (synthetic) observations for each prognostic variable, a_j stands for a observation at time j, and $\hat{a}_j$ the corresponding model result. W_m determines the weight that each data-model pair contributes to the overall cost (RMSE).

Our first metric of misfit, COST_1, is based on the common approach to consider phytoplankton observations only [as in, e.g., *Gregg*, 2008; *Gunson et al.*, 1999]. It is motivated by the fact that satellite observations of chlorophyll *a* provide unrivalled coverage in time on a near-global scale. The COST_1 is calculated as RMSE in equation (6) based on 5-daily synthetic observations throughout one seasonal cycle of surface phytoplankton between 65°S and 65°N. For simplicity we assume equal weights (i.e., $W_m = 1$). The synthetic observations consist of simulated phytoplankton concentrations distorted by noise as described in section 2.2. The sensitivity of COST_1 is biased toward misfits occurring in the higher latitudes, particularly, because there, the phytoplankton variability (in absolute values as considered by equation (6)) is higher than in the more oligotrophic lower latitudes.

Our second metric of misfit, COST_2, serves as an example of more complex cost functions. The point we want to make by defining COST_2 is that misfit metrics are adjustable at will and can be shaped to enhance or reduce their sensitivity toward particular processes. In COST_2, we average 5-daily synthetic surface observations of phytoplankton and nitrate and the respective model results over three regions, the tropics/subtropics (27°S–27°N) and the higher latitudes (27° to 65°) in both hemispheres. The latter feature, in contrast to the tropics, a high seasonality. In total we therewith use six time series, covering a full year, from the model simulations and the synthetic observations (nitrate averaged over the three regions and phytoplankton averaged over the three regions) to calculate the term $\left(a_j - \hat{a}_j\right)^2$ in equation (6). As for the respective weights, all mean

Table 2. Naming of Parameter Sets Leading to Equivalent Model Solutions

Name	Underlying Phytoplankton Growth Parameter Set
Genuine truth	Parameter set underlying the genuine truth simulation from which we draw synthetic observations
OPTI_1	Parameter set obtained by minimizing COST_1 (cf. section 2.3)
OPTI_2	Parameter set obtained by minimizing COST_2 (cf. section 2.3)

Table 3. Misfit to Synthetic Observations (Calculated From the Last Year of the Respective 3000 Yearlong Spin-ups)[a]

Parameter Set of Underlying Simulation	$COST_1$ (mmol N m^{-3})	$COST_2$ (mmol N m^{-3})
Genuine truth	0.3186	0.2824
$OPTI_1$	0.3141	0.2586
$OPTI_2$	0.3319	0.1825

[a]The table reads as follows: a simulation with the parameters underlying the genuine truth yields 0.3186 and 0.2824 in terms of $COST_1$ and $COST_2$, respectively. A simulation with parameters $OPTI_1$ undercuts these misfits (i.e., it is closer to the observations) both in terms of $COST_1$ and $COST_2$.

squares calculated with nitrate (phytoplankton) data are normalized with the overall standard deviation of nitrate (phytoplankton) data. (Note that for the calculation of both standard deviations data from oligotrophic regions defined as hosting nitrate surface concentrations of less than 0.5 mmol N m^{-3} are omitted). The ratio behind this approach is as follows: (1) By using spatial averages of observations and model output in equation (6), problems associated to the potential misalignment between observations and model estimates in space can be reduced in real-world applications. (2) The partitioning into regions that feature a high seasonality and those that are severely depleted in macronutrient or micronutrient concentrations gives control on the strength with which (model) processes affect the misfit. (3) By weighting with the respective standard deviation of the prognostic variables, we assume (and want to ensure) that phytoplankton observation carry as much relevant information into the misfit as the nitrate concentrations.

2.3.1. The Concept of Equivalent Solutions

Key to this study is the concept of *equivalent solutions*. As described in section 2.2, we use a twin experiment approach where we define a model simulation as the genuine truth. From this "truth," we sample our synthetic observations which we distort by noise to mimic real-world conditions. Thus, even the genuine truth deviates from the observations in that it yields a cost higher than zero because of the noise in the synthetic observations. We name in the following all of those model solutions as "equally consistent" *with the observations* that yield the same or a lower model-data misfit than the genuine truth does, relative to the synthetic observations. To test for equivalence, we compare model solutions in equilibrium, after 3000 years of simulation time (starting from steady state for the genuine truth). Note that this definition of *equivalence* depends on the underlying cost function and different cost functions may identify differing solutions as equivalent: cf. Table 3.

2.4. Spin-up Procedure and Future Projections

Using the systematic optimization procedure outlined in Appendix A, we find parameter sets (cf. section 3) that yield model solutions that are, under the given preindustrial boundary conditions, equivalent with

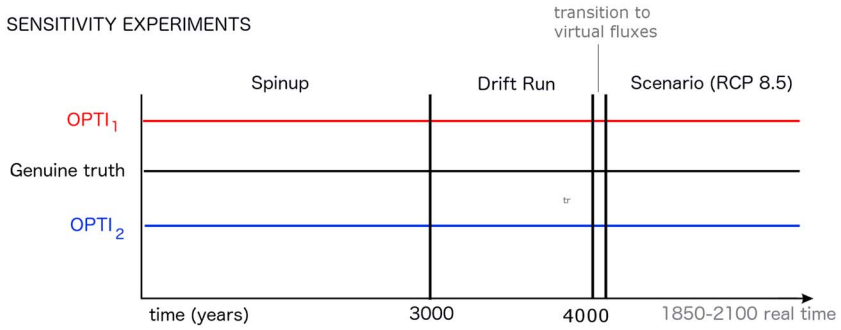

Figure 3. Schematics of the integration procedure. For each of the three parameter sets, a 3000 year long spin-up starting from the equilibrated reference run of *Keller et al.* [2012] is integrated. These spun-up states are compared with synthetic observations, and their equivalence (cf. section 2.3.1) in terms of respective cost (cf. section 2.3) is confirmed. Thereafter (cf. section 2.4), drift runs of 1000 year duration are followed by another 50 year long transition phase (denoted by the perpendicular lines). Subsequently, we run RCP 8.5 emission scenarios with each of the three parameters sets through the period 1850–2100.

respect to the observations (i.e., the respective cost is equal or lower than the cost associated to the parameter set underlying the genuine truth simulation). In a set of scenario runs, we explore how differences in parameter sets underlying these (historically) equivalent model solutions map to differences in model-based projections into a warming future. Following *Keller et al.* [2014], all projections are integrated starting from their respective spin-ups of 3000 years with so-called 1000 year long drift runs where the atmospheric CO_2 content is no longer prescribed but allowed to vary in response to preindustrial emissions. After this switch to *von Neumann* boundary conditions, another transition phase of 50 year duration is annexed. From now on, virtual air-sea fluxes of biogeochemical species are turned on [i.e., changes in DIC due to evaporation, precipitation and runoff - *Weaver et al.*, 2007]. Finally, we run the emission scenario RCP 8.5 [*Riahi et al.*, 2011], covering the period 1850–2100. A schematic of this integration procedure is provided in Figure 3.

3. Results

3.1. Equivalent Model Solutions

As a first step, we examine a set of equivalent model solutions. Table 3 shows that for each of our two cost functions (defined in section 2.3), we find one parameter set that fits the synthetic observations better than parameter set underlying the genuine truth simulation (and the synthetic observations). The two respective parameter sets are named $OPTI_1$ and $OPTI_2$. Remarkably, the long-term simulation based on $OPTI_1$ features less misfit to the observations both in terms of $COST_1$ and $COST_2$. Following our definitions in section 2.3.1, we conclude that long-term model simulations based on the parameter sets $OPTI_1$ and $OPTI_2$ are equally consistent with our synthetic observations because they yield (according to both cost functions in the case of the former, and according to $COST_2$ in the case of the latter) solutions that are equivalent to the genuine truth. This equivalence is also associated to a high visual agreement in, e.g., simulated annual mean surface concentrations of nitrate and phytoplankton (Figures 4 and 5, respectively)—even though some of the underlying parameters differ from each other by more than 100% (Table 4).

3.2. Diverging Model Projections Into a Warming Future

As described in the previous section, both parameter sets $OPTI_1$ and $OPTI_2$ yield model solutions that are equivalent to the historical steady state of the genuine truth model configuration (according to the respective misfit functions). We will now showcase both—robust patterns and high uncertainties, where the supposedly equivalent configurations diverge from one another, when running all three model configurations into a warming future (RCP 8.5 scenario).

Among the robust patterns are the global effects of increasing water column stratification on diffusive vertical nutrient supply. As the surface waters warm and stratification increases, turbulent mixing is suppressed because vertical mixing requires more energy to work against the more stable stratification by mixing dense (heavy) waters upward and push lighter (buoyant) waters to depth. On a global scale the reduced vertical mixing results in less nutrients being transported from depth to the sunlit surface. Less upward transport results in overall decreasing surface nutrient concentrations (Figure 6) that generally sustain a decreasing phytoplankton standing stock at the surface (Figure 7). Although this pattern is rather robust, there are substantial exceptions, with differences occurring both regionally and among the configurations. In this regard, the Arctic and the eastern equatorial Pacific are most prominent given that first, they do not follow the latter rule (less vertical mixing resulting in less nutrients and phytoplankton at the surface) and, second, the responses among the configurations differ considerably. In the following, we will elaborate on the differing model projections in these regions.

1. In the Arctic, the combination of retreating sea ice and increasing stratification (decreasing surface mixed layer depths) results in increasing light levels experienced by phytoplankton. The configurations respond differently to this: close to the North Pole, the genuine truth simulation features none or only weak changes in both surface nutrient and phytoplankton concentrations (Figures 6 and 7). In contrast, both $OPTI_1$ and $OPTI_2$, which both feature a sixfold higher initial slope of the *P-I* curve (Table 4), respond with a substantial increase in phytoplankton (Figure 7) and a related drawdown of surface nutrients (Figure 6).
2. In the eastern equatorial Pacific light-triggered changes, such as found in the Arctic, are of minor importance. Instead, the complex interplay between macronutrient limitation, micronutrient limitation, and nonlinear controls and responses of zooplankton dynamics shape the different sensitivities to environmental changes. Among the main drivers affecting biogeochemical cycles in the eastern equatorial Pacific is the wind-driven upwelling of nutrient-replete waters to the sunlit surface. The associated upward transport of

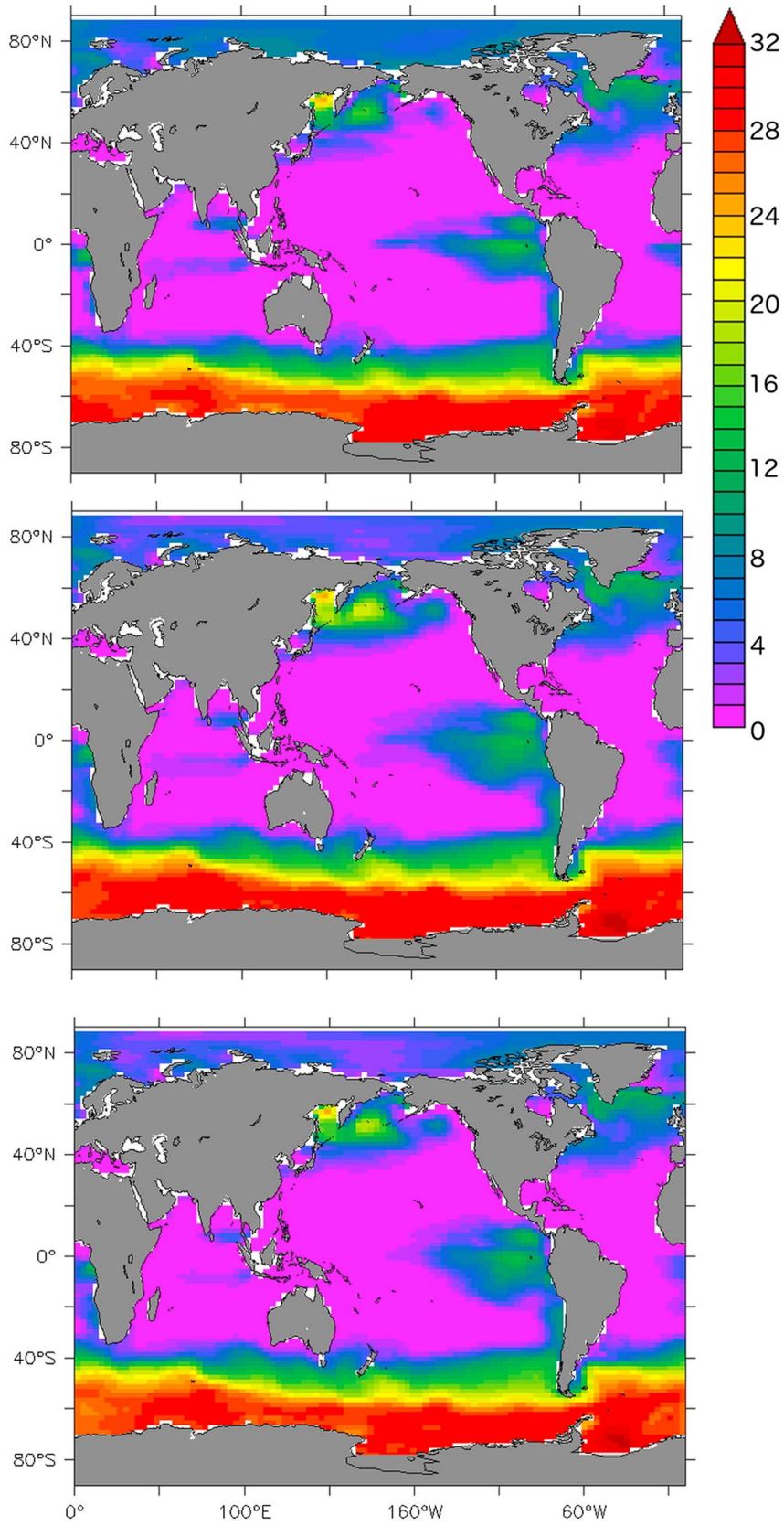

Figure 4. Simulated surface nitrate concentrations in units mmol N m^{-3} at the end of the spin-up (annual mean of year 3000). (top) The genuine truth; (middle and bottom) the simulation based on the parameter sets $OPTI_1$ and $OPTI_2$.

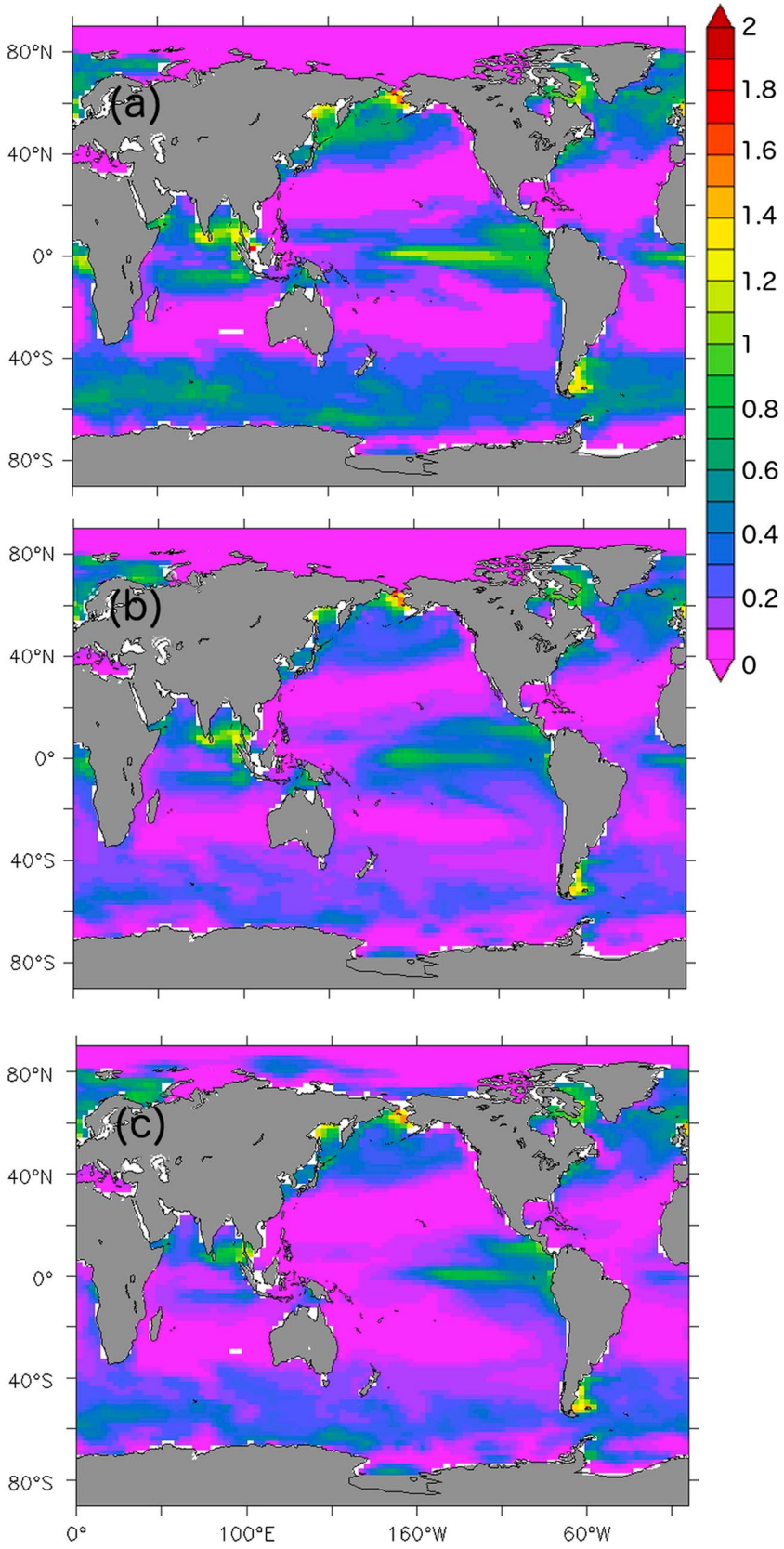

Figure 5. Simulated surface phytoplankton concentrations in units mmol N m^{-3} at the end of the spin-up (annual mean of year 3000). (a) The genuine truth; (b and c) the simulation based on the parameter sets $OPTI_1$ and $OPTI_2$.

Table 4. The Considered Parameters for the Different Model Versions[a]

Parameter	Description	Reference Value (Genuine Truth)	$OPTI_1$	$OPTI_2$	Unit
K_N	Half-saturation constant for NO_3 uptake	0.7	1.7	1.7	mmol N m^{-3}
K_P	Half-saturation constant for PO_4 uptake	0.044	0.11	0.11	mmol P m^{-3}
α	Initial slope of the *P-I* curve	0.1	0.6	0.6	$(W\ m^{-2})^{-1}\ d^{-1}$
a	Maximum phytoplankton growth rate at 0°C	0.6	0.41	0.435	day^{-1}
K^P_{Fe}	Half-saturation constant for Fe uptake	0.1	0.046	0.46	mmol Fe m^{-3}

[a]The respective notation is identical to the one in *Keller et al.* [2012].

nutrients is increasing in a warming world where, as explained above, the diffusive upward transport is hampered and, consequently, deep nutrient concentrations increase such that the upwelling taps into a deep pool of increasing nutrient concentrations. The increased nutrient fluxes fuel an increasing phytoplankton concentration in the vicinity of the major equatorial upwelling sites in all of our simulations (Figure 7). Somewhat counter to intuition is that at the same time, all configurations propose decreasing surface nutrient concentrations (Figure 6). This effect can be explained by the model's representation of zooplankton dynamics where the increased nutrient fluxes trigger an increased phytoplankton growth which, again, feeds a growing zooplankton population. At some stage heavily nonlinear behavior kicks in (partly because the zooplankton grazes also on itself—determined by the underlying Hollinger Type II functional response). This reduces the grazing pressure on phytoplankton in relative terms. As a result, the phytoplankton outgrows the top-down control by zooplankton and can then draw down surface nutrient concentrations in the fertile equatorial upwelling regions. (Please note that, apparently, the zooplankton top-down control used to be so strong in earlier versions of the model, such as used in, e.g., *Schmittner et al.* [2008], that the simulated surface macronutrient concentrations in the HNLC (high-nutrient-low-chlorophyll) region in the equatorial Pacific were realistic—even though the model lacked a representation of the limiting effect of iron.) In other words, the extra pulse of nutrients pushes the model out of its comfort zone and results in massive drawdown of macronutrients in the HNLC region. While the main mechanisms hold for all three configurations, they come in different flavors and have different consequences: In $OPTI_1$ the response is strongest, with larger nutrient drawdown and larger phytoplankton increase than in $OPTI_2$ and the genuine truth. Analyzing the differences between the setups, $OPTI_1$ is the least iron limited (i.e., its K^P_{Fe} is lowest) and has the slowest maximum growth rate of phytoplankton (Table 4). The latter results in a relatively stronger top-down control by zooplankton. As a consequence, $OPTI_1$ features the strongest nonlinear response to the upwelling of waters with increasing macronutrient concentrations: the phytoplankton outgrows the control exerted by the zooplankton, surface macronutrient concentrations are depleted, and more organic material is exported to depth. The associated remineralization of organic material leads to a local increase in oxygen consumption. In the projection based on $OPTI_1$, this triggers an increase in that fraction of the tropical Pacific hosting dissolved oxygen concentrations of less than 5 mmol m^3 (Figure 8a). This increase of the suboxic volume is accompanied with an upward shift (Figure 8b) such that more organic material is remineralized under suboxic conditions (basically because the remineralization profile (or Martin Curve) decays highly nonlinearly with depth) and, in turn, triggers an increase in denitrification rates of 48 Tg Nyr^{-1} by the end of 2100.

Figure 8 also highlights that differences in iron limitation or top-down control can trigger subtle differences in local export of organic matter to depth that amplify to massively different projections of the suboxic volume. In addition to local changes in the export of organic matter, increased stratification (which is locally especially pronounced in the projections based on the genuine truth and $OPTI_1$) is among the factors that can trigger the simulated shallowing oxygen minimum zones depicted in Figure 8b. Consequently, the affected oxygen minimum zones can catch more export. These complex processes that determine the dynamics of oxygen deficits eventually trigger substantial changes to global pelagic denitrification rates, corresponding to a decrease of 4.4 Tg Nyr^{-1} and an increase of 6.4 Tg Nyr^{-1} by the end of 2100 in $OPTI_2$ and the genuine truth, respectively.

In summary, we conclude that the supposedly equivalent (i.e., the biological pump scoring comparable fits to historical observations) model configurations feature substantially different sensitivities to the RCP 8.5

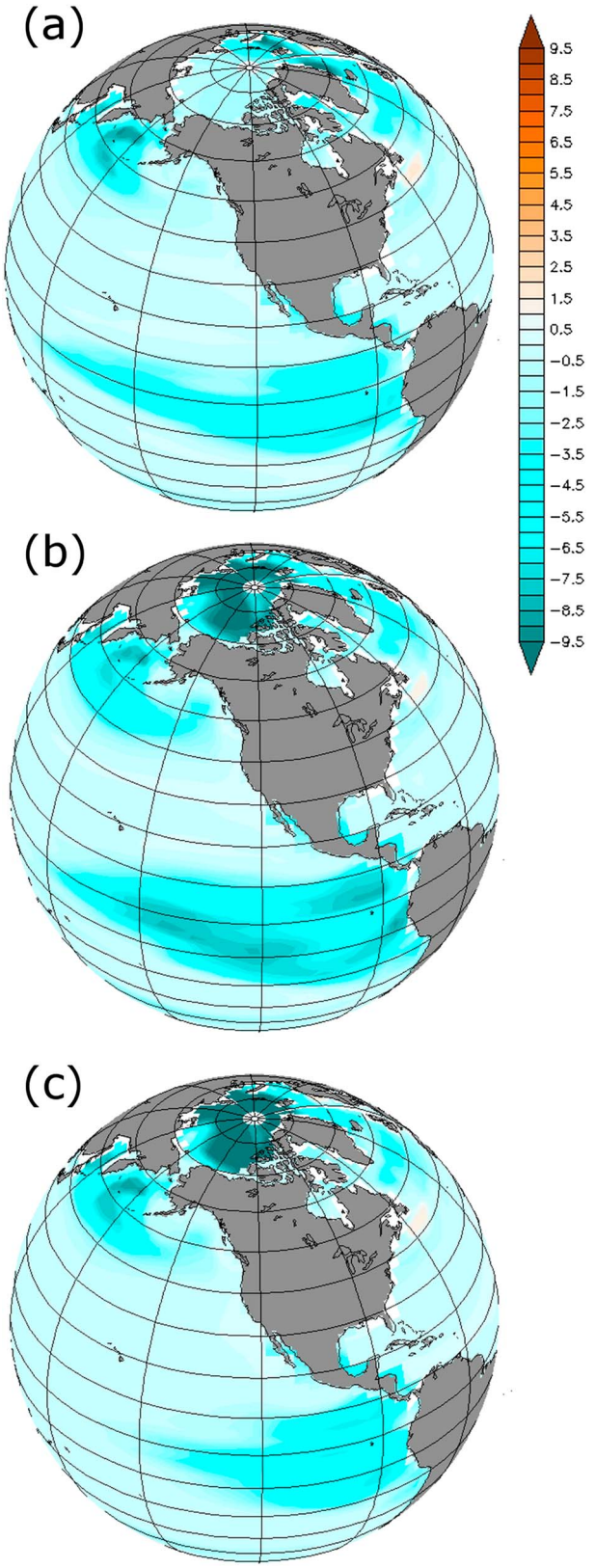

Figure 6. Simulated changes in surface nitrate concentrations as a consequence of rising CO_2 concentrations (emission scenario RCP 8.5) calculated as the annual mean concentration difference between the years 2100–1850; the unit is mmol N m^{-3}. (a) The genuine truth; (b and c) the simulation based on the parameter sets $OPTI_1$ and $OPTI_2$.

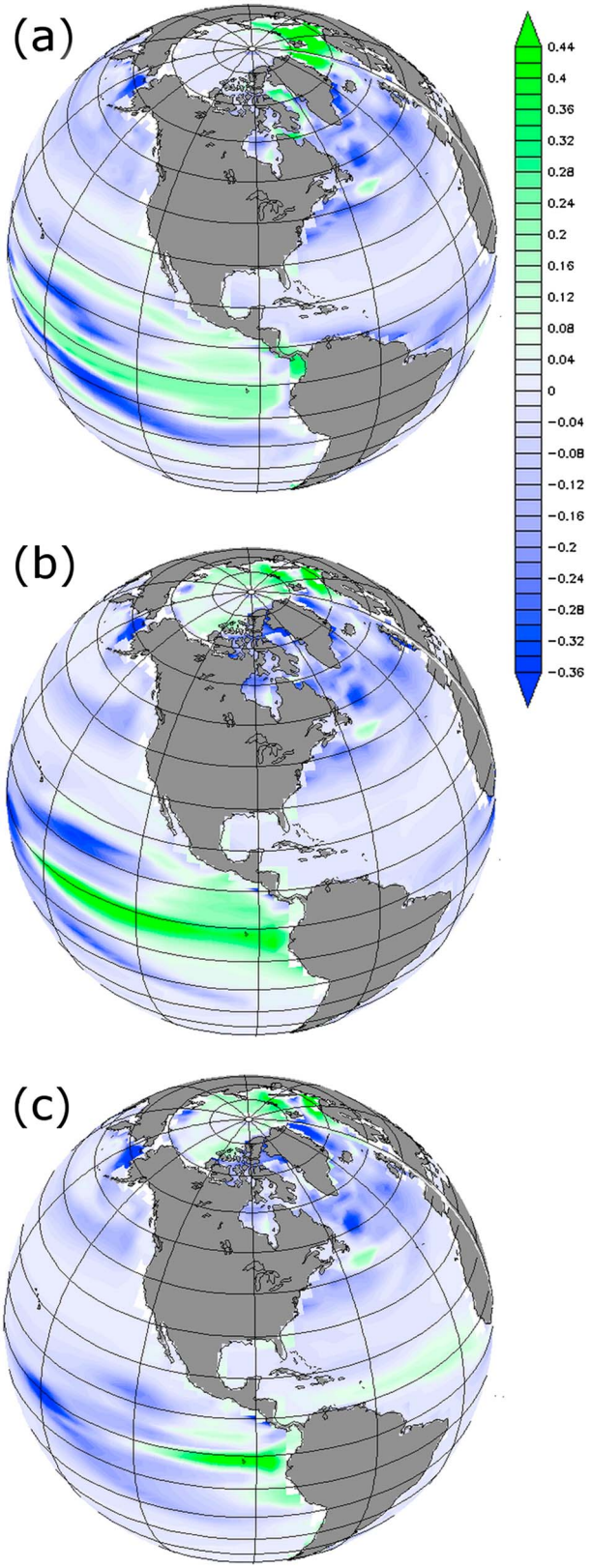

Figure 7. Simulated changes in surface phytoplankton concentrations as a consequence of rising CO_2 concentrations (emission scenario RCP 8.5) calculated as the annual mean concentration difference between the years 2100–1850; the unit is mmol N m^{-3}. (a) The genuine truth; (b and c) the simulation based on the parameter sets $OPTI_1$ and $OPTI_2$.

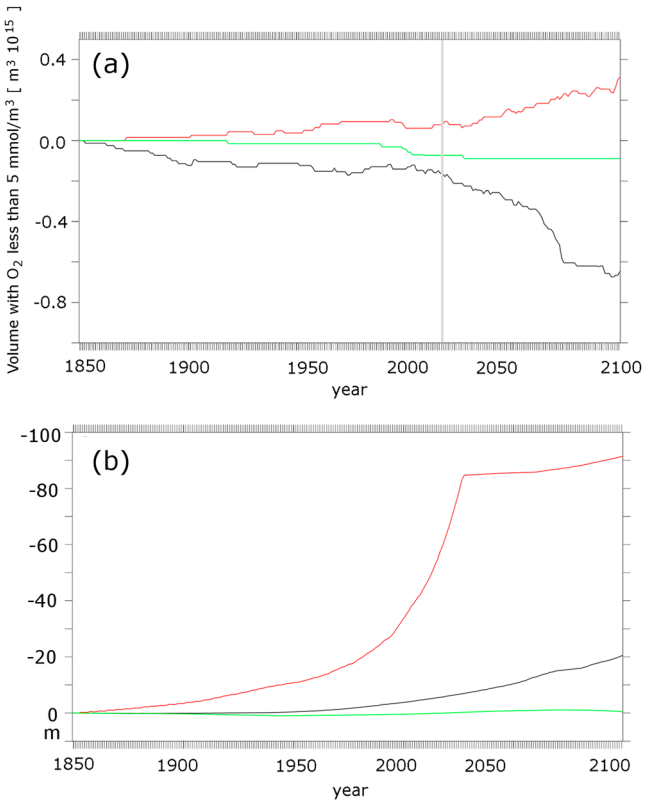

Figure 8. (a) Simulated temporal evolution of the volume of suboxic waters in the tropical Pacific in response to rising CO_2 concentrations (emission scenario RCP 8.5). The black line denotes the genuine truth. The red and green lines refer to simulations based on the parameter sets $OPTI_1$ and $OPTI_2$, respectively. The light gray line marks denote 2016, the date of this study. (b) Changes of the minimum depth of the suboxic volume in the tropical Pacific in meters for the genuine truth (black line), $OPTI_1$ (red line), and $OPTI_2$ (green line). The given values are linearly interpolated from the model grid.

emission scenario, both locally and globally. We find substantial uncertainties associated with projections of surface nutrients, phytoplankton abundance, suboxic volume, denitrification, and, as Figure 9 shows, even oceanic carbon uptake, which is related to differences in the export of organic matter (i.e., the biological pump). The uptake projected by $OPTI_2$ is very similar to the genuine truth projection. In $OPTI_1$, however, the oceanic carbon uptake accelerates faster. Integrated over the period 1850 to 2100 the respective difference adds up to 60 Pg C.

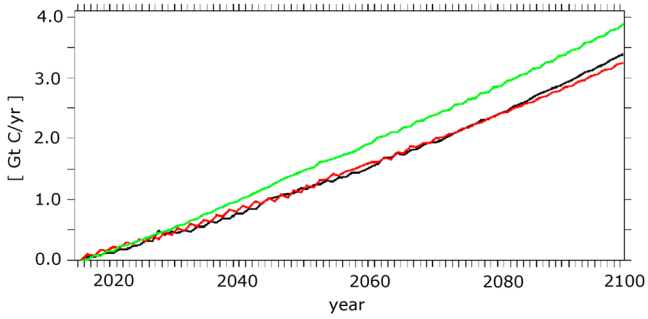

Figure 9. Simulated carbon uptake of the ocean in response to rising CO_2 concentrations (emission scenario RCP 8.5). The black line denotes the genuine truth. The red and green lines refer to simulations based on the parameter sets $OPTI_1$ and $OPTI_2$, respectively.

4. Discussion

Numerical pelagic biogeochemical models rely on parameters that are not known per se but that are typically chosen such that the models fit a set of present-day or historical observations in an "optimal way." This approach is problematic for a number of reasons. For one, the definition of the optimal way inevitably contains a subjective element, i.e., the respective measure of misfit or cost function is adjustable at will. An additional problem is imperfect observations and models: we showed in section 3.1 that for a typical combination of a three-dimensional coupled ocean circulation biogeochemical model, noisy observations, and cost functions, the existence of a meaningful, unique global minimum in the model-data misfits is not guaranteed.

Rather, we find a set of model configurations (i.e., model parameter sets) that fit noisy observations equally well. There have been indications of this indeterminacy early on [*Matear*, 1995], and a detailed analysis and the implications have been illustrated in a slab ocean model framework already [*Löptien and Dietze*, 2015]. The present study, rendered feasible by recent advances in computational efficiency, adds to the discussion in that it illustrates the problem with a full-fledged three-dimensional coupled general ocean circulation biogeochemical model—a module found in all state-of-the-art Earth System Models.

The indeterminacy of a meaningful, unique and globally optimal parameter set would not necessarily be problematic—if the respective equally well fitting parameter sets would always yield model solutions that are similar enough for the purpose at hand. This, however, is not the case because, consistent with the slab ocean results from *Löptien and Dietze* [2015], we find that the indeterminacy does also apply to the so-called Michaelis-Menten parameters. These parameters determine the model's sensitivity to nutrients and light in the sunlit surface ocean and, as such, are paramount as concerns the model's response to changing conditions such as anticipated in a warming climate. Specifically, we find that supposedly equally consistent parameter sets yield substantially different model projections into a warming climate, driven by the emission scenario RCP 8.5. Depending on the projected variable of interest the differences involve both the magnitude and the sign. The oceanic carbon uptake and oxygen minimum zone dynamics provide good examples of this problem (Figures 8 and 9).

As for carbon we find that projections with model configurations that are "equally consistent" with a set of synthetic observations yield differences in oceanic carbon uptake that add up to 60 Pg C until 2100 even though the respective synthetic observations are superior to typical real-world observations in terms of (1) noise, (2) spatial/temporal coverage, and (3) model bias. This 60 Pg C difference is substantial given that it corresponds to 6 years worth of today's anthropogenic carbon emissions.

As for the suboxic volume hosted by the eastern tropical Pacific, we find that model configurations fitting the synthetic observations equally well yield forecasts with differing signs ranging from a decrease of 0.65 $\times 10^{15} m^3$ to an increase of 0.32 $\times 10^{15}$ m^3 until year 2100. This uncertainty is of importance because suboxia triggers denitrification, a process that controls the standing stock of nitrogen in the ocean that is available for marine biota. Accordingly, we find, e.g., for the tropical Pacific, that projections, based on model version that fit present-day observations equally well, do not even agree in the sign of pelagic denitrification rate changes that are to be expected in the decades to come.

5. Summary and Conclusions

We set out to link parameter indeterminacy in pelagic biogeochemical modules of Earth System Models with uncertainties of their projections into the future. We illustrate pitfalls of a standard approach to estimate the parameters of biogeochemical pelagic models with (noisy) historical observations. The illustration is based on UVic 2.9, an Earth System Model of intermediate complexity [*Weaver et al.*, 2001]. Specifically, we focus on model parameters that imprint the limiting effects of nutrients and sunlight on carbon assimilation by autotrophic phytoplankton.

Our approach is based on twin experiments: we construct our own synthetic observations by (1) choosing one arbitrary model parameter set to be the *true* one, (2) run a simulation with this parameter set which we define as the genuine truth, and (3) add noise to the respective model output. Here the noise is meant to cover all kinds of potential errors that are prone to occur when comparing a biogeochemical model to observations (e.g., measurement accuracy, misalignment between observations and model estimate in space and/or time, and unit conversion). The level of noise added is less than in typical real-world applications. In a second step, we run simulations with different sets of model parameters (the so-called twins to the genuine

truth simulation), which fit the synthetic observations equally well. The determination of these parameters relies on automatized optimization procedure outlined in Appendix A. Note that such a systematic optimization procedure just recently has become feasible by advancements in computer technology.

Our results indicate that for typical noise levels inherent to observations, various parameter sets might exist that are equally consistent with the observations and yield strikingly similar solutions for present-day climate—even though respective parameter values may differ by manyfold from one another. Further, our results showcase that the solutions' similarity breaks once the simulations are forced with increasing CO_2 emissions (emission scenario RCP 8.5). This applies especially for near-surface properties, such as projected concentrations of nitrate and phytoplankton at the surface in the tropical Pacific and the Arctic Ocean. These differences also trigger differences farther down in the water column. For instance, projections of the suboxic volume in the eastern tropical Pacific differ substantially, ranging from strongly decreasing to strongly increasing volumes in the decades to come. Consequently, also the projections of that pelagic denitrification that is hosted in the suboxic tropical Pacific do not even agree on the sign of the changes to come. Drastic differences among the supposedly similar model projections were also found in the projections of oceanic carbon. Integrated from the start of the Anthropocene to year 2100 the differences peaked at 60 GT C corresponding to 6 years worth of today's anthropogenic carbon emissions to the atmosphere.

We conclude that the indeterminacy of parameter estimates of pelagic biogeochemical modules of the current generation of Earth System Models, related to the noise inherent to typical observations, may well explain a significant fraction of intermodel differences of projections into a warming future reported by *Friedlingstein et al.* [2006]. Further, we see strong parallels to a recently started discussion in atmospheric and climate science where (1) exercises of *Mauritsen et al.* [2012] revealed that "the tuning process" of a climate model does not necessarily lead to a single, unique parameter set and (2) *Notz* [2015] stated that " … The usefulness of a climate-model simulation cannot be inferred solely from its degree of agreement with observations...."

Forthcoming research will show if and to what extent the definition of new model-observations misfit metrics can reduce the respective uncertainties reported in this study.

Appendix A: Optimization

After a 5000 year spin-up of the genuine truth configuration, we perform two sets of numerical experiments where we assume that the true parameters were unknown and strive to find a parameter set for the maximum phytoplankton growth, the Michaelis-Menten parameters, and the initial slope of the *P-I* curve (a, K^P_{Fe}, K_N, K_P, and α, respectively) that minimizes the misfit between the model and the synthetic observations. The two sets of experiments differ only in terms of the definition of the misfit, as outlined in section 2.3. During the optimization, we use prescribed preindustrial atmospheric CO_2 concentrations.

The major problem with optimization of parameters in full Earth System Models is their high computational demand. Therefore, we optimize only a few parameters rather than the entire set. The goal is to retrieve the parameters which underlie the genuine truth (acting as if these were unknown). The study of *Löptien and Dietze* [2015], however, illustrates in a slab ocean model that it can be impossible to identify a unique, optimal parameter set of those parameters that prescribe the limiting effect of nutrient- and light-depleted conditions on carbon assimilation by autotrophic phytoplankton based on noisy observations. One of the major aims of the present study is to examine whether the same problems associated to parameter identifiability are endemic also to full-fledged Earth System Models. We thus chose some arbitrary extreme values for the half-saturation constant for nitrate (K_N) and phosphate (K_P) and the slope of the *P-I* curve α, which differ strongly from the true ones. In a second step, we search for the model solution that is as close as possible to the synthetic observations by systematically adjusting the remaining phytoplankton growth parameters (i.e., maximum phytoplankton growth rate a and the half-saturation constant for Fe uptake K^P_{Fe}) until the model-data misfit is minimal. Based on *Löptien and Dietze* [2015], we expect that solutions with a lower cost than the genuine truth exist. To emphasize the problems related to the phytoplankton growth parameters, we assume that all other model parameters were known. For the parameter adjustment, or minimization of the misfit, we use an automized numerical optimization algorithm, the heuristic downhill simplex method known as Nelder-Mead Simplex Method [*Lagarias et al.*, 1998]. We start from an arbitrary initial guess of $K^P_{Fe} = 0.07$ and $a = 0.48$.

To reduce the computational costs even further, we follow the ideas presented by *Prieß et al.* [2013]. These authors propose to use a so-called *surrogate* (i.e., a less costly approximation of the full model). In our case

we, utilize very short 5 year simulations, starting from the true state, as *surrogate*. This surrogate is used at the beginning of respective automatized searches for the optimum. By this technique we can limit the amount of necessary full model simulations to a low number. Note that we do not apply the *response correction* suggested by *Prieß et al.* [2013], since we focus on surface parameters only. Another advantage of focusing on the phytoplankton growth parameters is that the surface processes are relatively fast and we are close to steady state already after 100 years of simulation time. By this optimization procedure we find two parameter sets with lower costs relative to the synthetic observations than the genuine truth, when regarding the final year of the 100 year simulations. This relatively short period of 100 years proves sufficient for optimization as the cost functions for the optimal solutions remain lower than the respective costs for the genuine truth plus noise even after 3000 years of simulation time.

Acknowledgments
This work is a contribution of the Sonderforschungsbereich 754 "Climate-Biogeochemistry Interactions in the Tropical Ocean" (www.sfb754.de; D 1570), which is funded by the German Science Foundation (DFG) and of the BMBF-funded project PalMod 4.1 (Paleo Modelling: A national paleo climate modeling initiative; D 1774). Both authors acknowledge essential long-term support by Andreas Oschlies. The authors received very helpful comments from two anonymous reviewers! Both authors were involved in the design of the work, in data analysis, in data interpretation, and in drafting the article. The model output is archived at http://thredds.geomar.de/thredds/catalog/open_access/loeptien_dietze_2016_gbc/catalog.html.

References

Banse, K. (1977), Determining the carbon-to-chlorophyll ratio of natural phytoplankton, *Mar. Biol.*, *41*(3), 199–212, doi:10.1007/BF00394907.
Dogliotti, A. I., I. R. Schloss, G. O. Almandoz, and D. A. Gagliardini (2009), Evaluation of SeaWiFS and MODIS chlorophyll-a products in the Argentinean Patagonian continental shelf (38°S–55°S), *Int. J. Remote Sens.*, *30*(1), 251–273, doi:10.1080/01431160802311133.
Evans, G. T. (2003), Defining misfit between biogeochemical models and data sets, *J. Mar. Syst.*, *40-41*, 49–54, doi:10.1016/S0924-7963(03)00012-5.
Fan, W., and X. Lv (2009), Data assimilation in a simple marine ecosystem model based on spatial biological parameterizations, *Ecol. Modell.*, *220*, 1997–2008, doi:10.1016/j.ecolmodel.2009.04.050.
Fasham, M. J. R., G. T. Evans, D. A. Kiefer, M. Creasey, and H. Leach (1995), The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47°N 20°W, *Philos. R. Soc. B.*, *348*, 203–209, doi:10.1098/rstb.1995.0062.
Fennel, K., K. Losch, J. Schröter, and M. Wenzel (2001), Testing a marine ecosystem model: Sensitivity analysis and parameter optimization, *J. Mar. Syst.*, *28*, 45–63, doi:10.1016/S0924-7963(00)00083-X.
Friedrichs, M. A. M. (2001), A data assimilative marine ecosystem model of the central equatorial Pacific: Numerical twin experiments, *J. Mar. Res.*, *59*, 859–894, doi:10.1357/00222400160497544.
Friedrichs, M. A. M., R. R. Hood, and J. D. Wiggert (2006), Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data, *Deep Sea Res., Part II*, *53*, 576–600, doi:10.1016/j.dsr2.2006.01.026.
Friedlingstein, P., et al. (2006), Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison, *J. Clim.*, *19*(14), 3337–3353, doi:10.1175/JCLI3800.
Garcia, H. E., R. A. Locarnini, T. P. Boyer, J. I. Antonov, M. M. Zweng, O. K. Baranova, and D. R. Johnson (2009), *World Ocean Atlas 2009, Volume 4: Nutrients (Phosphate, Nitrate, Silicate)*, edited by S. Levitus, 398 pp., NOAA Atlas NESDIS 71, U.S. Govern. Print. Office, Washington, D. C.
Getzlaff, J., and H. Dietze (2013), Effects of increased isopycnal diffusivity mimicking the unresolved equatorial intermediate current system in an Earth system climate model, *Geophys. Res. Lett.*, *40*, 2166–2170, doi:10.1002/grl.50419.
Getzlaff, J., H. Dietze, and A. Oschlies (2016), Simulated effects of Southern Hemispheric wind changes on the Pacific oxygen minimum zone, *Geophys. Res. Lett.*, *43*, 728–734, doi:10.1002/2015GL066841.
Govindasamy, B., and K. Caldeira (2000), Geoengineering Earth's radiation balance to mitigate CO_2-induced climate change, *Geophys. Res. Lett.*, *27*(14), 2141–2144, doi:10.1029/1999GL006086.
Gregg, W. (2008), Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, *J. Mar. Syst.*, *69*(3), 205–225, doi:10.1016/j.jmarsys.2006.02.015.
Gunson, J. R., A. Oschlies, and V. Garcon (1999), Sensitivity of ecosystem parameters to simulated satellite ocean color data using a coupled physical-biological model of the North Atlantic, *J. Mar. Res.*, *57*, 613–639, doi:10.1357/002224099321549611.
Hasselmann, K. (1976), Stochastic climate models: Part I. Theory, *Tellus*, *28*, 473–485, doi:10.1111/j.2153-3490.1976.tb00696.x.
Hemmings, J. C. P., and P. G. Challenor (2012), Addressing the impact of environmental uncertainty in plankton model calibration with a dedicated software system: The Marine Model Optimization Testbed (MarMOT 1.1 alpha), *Geosci. Model Dev.*, *5*, 471–498, doi:10.5194/gmd-5-471-2012.
Jones, C. D., P. Cox, and C. Huntingford (2003), Uncertainty in climate-carbon-cycle projections associated with the sensitivity of soil respiration to temperature, *Tellus B*, *55*(2), 642–648, doi:10.1034/j.1600-0889.2003.01440.x.
Kriest, I., V. Sauerland, S. Khatiwala, A. Srivastav, and A. Oschlies (2017), Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0), *Geosci. Model Dev.*, *10*, 127–154, doi:10.5194/gmd-10-127-2017.
Keller, D. P., A. Oschlies, and M. Eby (2012), A new marine ecosystem model for the University of Victoria Earth System Climate Model, *Geosci. Model Dev.*, *5*, 1195–1220, doi:10.5194/gmd-5-1195-2012.
Keller, D. P., E. Y. Feng, and A. Oschlies (2014), Potential climate engineering effectiveness and side effects during a high carbon dioxide-emission scenario, *Nat. Commun.*, *5*, 3304, doi:10.1038/ncomms4304.
Kriest, I., S. Khatiwala, and A. Oschlies (2010), Towards an assessment of simple global marine biogeochemical models of different complexity, *Prog. Oceanogr.*, *86*, 337–360, doi:10.1016/j.pocean.2010.05.002.
Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998), Convergence properties of the Nelder-Mead simplex method in low dimensions, *SIAM J. Optimiz.*, *9*, 112–147, doi:10.1137/S1052623496303470.
Lawson, L. M., E. E. Hofmann, and Y. H. Spitz (1996), Time series sampling and data assimilation in a simple marine ecosystem model, *Deep Sea Res., Part II*, *43*, 625–651, doi:10.1016/0967-0645(95)00096-8.
Löptien, U. (2011), Steady states and sensitivities of commonly used pelagic ecosystem model components, *Ecol. Modell.*, *222*, 1376–1386, doi:10.1016/j.ecolmodel.2011.02.005.
Löptien, U., and H. Dietze (2015), Constraining parameters in state-of-the-art marine pelagic ecosystem models—Is it actually feasible with typical observations of standing stocks?, *Ocean Sci.*, *11*(4), 573–590, doi:10.5194/os-11-573-2015.
Mauritsen, T., et al. (2012), Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, *4*(3), M00A01, doi:10.1029/2012MS000154.
Matear, R. J. (1995), Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at station P, *J. Mar. Res.*, *53*, 571–607, doi:10.1357/0022240953213098.
Matthews, H. D., L. Cao, and K. Caldeira (2009), Sensitivity of ocean acidification to geoengineered climate stabilization, *Geophys. Res. Lett.*, *36*, L10706, doi:10.1029/2009GL037488.

Moss, R. H., et al. (2010), The next generation of scenarios for climate change research and assessment, *Nature*, *463*(7282), 747–756, doi:10.1038/nature08823.
Notz, D. (2015), How well must climate models agree with observations?, *Philos. Trans. R. Soc. A*, *373*(2052), 20140164, doi:10.1098/rsta.2014.0164.
Oschlies, A., M. Pahlow, A. Yool, and R. J. Matear (2010), Climate engineering by artificial ocean upwelling: Channelling the sorcerer's apprentice, *Geophys. Res. Lett.*, *37*, L04701, doi:10.1029/2009GL041961.
Pacanowski, R. C. (1995), MOM 2 documentation, user's guide and reference manual, Tech. Rep. 3, GFDL Ocean Group, Princeton.
Prieß, M., J. Piwonski, S. Koziel, A. Oschlies, and T. Slawig (2013), Accelerated parameter identification in a 3D marine biogeochemical model using surrogate-based optimization, *Ocean Model.*, *68*, 22–36, doi:10.1016/j.ocemod.2013.04.003.
Reith, F., D. P. Keller, and A. Oschlies (2016), Revisiting ocean carbon sequestration by direct injection: A global carbon budget perspective, *Earth Syst. Dyn. Discuss.*, *20*, 1–26, doi:10.5194/esd-2016-20.
Riahi, K., S. Rao, V. Krey, C. Cho, V. Chirkov, G. Fischer, G. Kindermann, N. Nakicenovic, and P. Rafaj (2011), RCP 8.5—A scenario of comparatively high greenhouse gas emissions, *Clim. Change*, *109*(1-2), 33–57, doi:10.1007/s10584-011-0149-y.
Rückelt, J., V. Sauerland, T. Slawig, A. Srivastav, B. Ward, and C. Patvardhan (2010), Parameter optimization and uncertainty analysis in a model of oceanic CO_2 uptake using a hybrid algorithm and algorithmic differentiation, *Nonlinear Anal.*, *11*, 3993–4009, doi:10.1357/002224003322981147.
Schartau, M. (2005), Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic Ocean: Part I—Method and parameter estimates, *J. Mar. Res.*, *62*, 765–793, doi:10.1357/002224003322981156.
Schartau, M., A. Oschlies, and J. Willebrand (2001), Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method, *Deep Sea Res., Part II*, *48*, 1769–1800, doi:10.1016/S0967-0645(00)00161-2.
Schmittner, A., A. Oschlies, H. D. Matthews, and E.D. Galbraith (2008), Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business- as- usual CO2 emission scenario until year 4000 AD, *Global Biogeochem. Cycles*, *22*(1), doi:10.1029/2007GB002953.
Spitz, Y. H., J. R. Moisan, M. R Abbott, and J. G. Richman (1998), Data assimilation and a pelagic ecosystem model: Parameterization using time series observations, *J. Mar. Syst.*, *16*, 51–68, doi:10.1016/S0924-7963(97)00099-7.
Stow, C. A., J. Jolliff, D. J. McGillicuddy, S. C. Doney, J. I. Allen, M. A. M. Friedrichs, K. A. Rose, and P. Wallhead (2009), Skill assessment for coupled biological/physical models of marine systems, *J. Mar. Syst.*, *76*(1), 4–15, doi:10.1016/j.jmarsys.2008.03.011.
Taucher, J., and A. Oschlies (2011), Can we predict the direction of marine primary production change under global warming?, *Geophys. Res. Lett.*, *38*, L02603, doi:10.1029/2010GL045934.
Tjiputra, J. F., D. Polzin, and A. M. Winguth (2007), Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization, *Global Biogeochem. Cycle*, *21*, GB1001, doi:10.1029/2006GB002745.
Tuana, N., R. L. Sriver, T. Svoboda, R. Olson, P. J. Irvine, J. Haqq-Misra, and K. Keller (2012), Towards integrated ethical and scientific analysis of geoengineering: A research agenda, *Ethics Policy Environ.*, *15*(2), 136–157, doi:10.1080/21550085.2012.685557.
Vaughan, N. E., and T. M. Lenton (2011), A review of climate geoengineering proposals, *Clim. Change*, *109*(3-4), 745–790, doi:10.1007/s10584-011-0027-7.
Volpe, G., R. Santoleri, V. Vellucci, M. R. d'Alcala, S. Marullo, and F. d'Ortenzio (2007), The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates, *Remote Sens. Environ.*, *107*(4), 625–638, doi:10.1016/j.rse.2006.10.017.
Von Storch, H., and F. W. Zwiers (2001), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, Cambridge, U. K.
Ward, B. A., M. A. M. Friedrichs, T. A Anderson, and A. Oschlies (2010), Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, *J. Mar. Res.*, *81*, 34–43, doi:10.1016/j.jmarsys.2009.12.005.
Weaver, A. J., et al. (2001), The UVic Earth System Climate Model: Model description, climatology, and applications to past, present and future climates, *Atmos. Ocean*, *39*(4), 361–428, doi:10.1080/07055900.2001.9649686.
Weaver, A. J., K. Zickfeld, A. Montenegro, and M. Eby (2007), Long term climate implications of 2050 emission reduction targets, *Geophys. Res. Lett.*, *34*, L19703, doi:10.1029/2007GL031018.
Xiao, Y., and M. A. M. Friedrichs (2014a), The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model, *J. Geophys. Res. Oceans*, *119*, 2691–2712, doi:10.1002/2013JC009433.
Xiao, Y., and M. A. M. Friedrichs (2014b), Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models in the Mid-Atlantic Bight: Effects of increasing the complexity of the planktonic food web, *Biogeosciences*, *11*, 3015–3030, doi:10.5194/bg-11-3015-2014.