One size fits all? - Calibrating an ocean biogeochemistry model for different circulations

Iris Kriest[1], Paul Kähler[1], Wolfgang Koeve[1], Karin Kvale[1], Volkmar Sauerland[1,2], and Andreas Oschlies[1]

[1]GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany
[2]Department of Mathematics, Kiel University, Christian-Albrechts-Platz 4, D-24118 Kiel, Germany

Correspondence: Iris Kriest (ikriest@geomar.de)

Abstract.

Global biogeochemical ocean models are often tuned to match the observed distributions and fluxes of inorganic and organic quantities. This tuning is typically carried out "by hand". However, this rather subjective approach might not yield the best fit to observations, is closely linked to the circulation employed, and thus influenced by its specific features and even its faults. We here investigate the effect of model tuning, via objective optimisation, of one biogeochemical model of intermediate complexity when simulated in five different offline circulations. For each circulation, three of six model parameters have adjusted to characteristic features of the respective circulation. The values of these three parameters – namely, the oxygen utilisation of remineralisation, the particle flux parameter and potential nitrogen fixation rate — correlate significantly with deep mixing and ideal age of NADW and the outcrop area of AAIW and SAMW in the Southern Ocean. The clear relationship between these parameters and circulation characteristics, which can be easily diagnosed from global models, can provide guidance when tuning global biogeochemistry within any new circulation model. The results from 20 global cross-validation experiments show that parameter sets optimised for a specific circulation can be transferred between similar circulations without losing too much of the model's fit to observed quantities. When compared to model intercomparisons of subjectively tuned, global coupled biogeochemistry-circulation models, each with different circulation and/or biogeochemistry, our results show a much lower range of oxygen inventory, OMZ volume and global biogeochemical fluxes. Export production depends to a large extent on the circulation applied, while deep particle flux is mostly determined by the particle flux parameter. Oxygen inventory, OMZ volume, primary production and fixed nitrogen turnover depend more or less equally on both factors, with OMZ volume showing the highest sensitivity, and residual variability. These results show a beneficial effect of optimisation, even when a biogeochemical model is first optimised in a relatively coarse circulation, and then transferred to a different, finer resolution circulation model.

1   Introduction

Global models of marine biogeochemistry are applied to prognostic problems, such as the future exchange of CO_2 between ocean and atmosphere, the evolution of oxygen minimum zones (OMZs) under a changing climate, or future primary production, which is the ultimate food source for fish. Unfortunately, in steady state, these models vary greatly in their representation

of, for example, ocean oxygen inventory (up to 50% of the current value; Bopp et al., 2013) or primary production (varying by more than 90% of present day global production; Bopp et al., 2013). The largest uncertainty is related to OMZ volume – here Bopp et al. (2013) report a range of variation that is several times the observed volume, depending on the criterion (maximum oxygen concentration) used for OMZ definition.

Because these coupled models differ both in their physical and biogeochemical setups, to date the contribution of the different model components to this large variation is not clear (e.g., Cabre et al., 2015). Studies indicate a strong impact of physics on deep oxygen levels, leading to a divergence of up to 150 mmol O_2 m^{-3} (Najjar et al., 2007; Seferian et al., 2013). On the other hand, Kriest et al. (2012) and Kriest and Oschlies (2015) showed that the impact of biogeochemical model structure and parameters on deep oxygen profiles can be equally large; also, in the latter study OMZ volume varied among different biogeochemical model setups up to three times the observed value (for an OMZ criterion of 8 mmol O_2 m^{-3}). Thus, so far neither oxygen content nor OMZ volume seem well constrained, possibly because of both circulation and biogeochemistry.

In practice, biogeochemical models are often tuned to the corresponding circulation, in order to make the model results (nutrients, oxygen, organic components) to agree better with observations (e.g., Schwinger et al., 2016). To keep the number of computationally expensive, global simulations low, usually this model calibration is carried out "by hand", i.e. by subjectively tuning some biogeochemical model parameters until the models show a "good" fit to the observed tracer fields. The criterion for "good" is not absolute - it may consist of a sufficient visual match, good indices of some core statistics, (e.g., in Taylor plots; Taylor, 2001), the root-mean-squared error of model results vs. observed quantities, or non-parametric methods such as the Bhattacharya-distance (e.g., Ilyina et al., 2013). Ideally these converge, i.e., they result in a well-defined set of parameters, which provide an optimal fit for all metrics.

However, biogeochemical models include a high-dimensional parameter space with respect to the biogeochemical constants, many of which are not well known. Carrying out a sensitivity study helps to explore a model's sensitivity to its constants (e.g., Kriest et al., 2012), but this is a time-consuming task, both in terms of work-hours and computational time. The computational demand is amplified by the fact that global biogeochemical ocean models require a long time to equilibrate (many millennia), owing to the sluggish circulation (e.g., Wunsch and Heimbach, 2008; Primeau and Deleersnijder, 2009), and because biogeochemical processes act in concert with it (e.g., Kriest and Oschlies, 2015). Short spinup times, on the other hand, will produce model results that still depend on initial conditions, and can hamper a thorough assessment of model skill. Because of these difficulties, there is no common recipe for model spinup (and calibration). This complicates model inter-comparison (Seferian et al., 2016).

Recently, tools have become available to speed up model equilibration, either by efficient offline methods (Khatiwala, 2007), or by root-finding algorithms that solve for the model's steady state (Li and Primeau, 2008; Khatiwala, 2008). Using these tools, automatic calibration of global biogeochemical ocean models becomes more feasible (e.g., DeVries et al., 2014; Holzer et al., 2014; Letscher et al., 2015; Kriest et al., 2017; Kriest, 2017). However, these approaches have so far mostly been applied to biogeochemical models of low complexity, or to circulation models of rather coarse resolution. As physical processes and resolution can play a large role for the representation of biogeochemical tracer distributions (e.g., Najjar et al., 2007; Duteil

et al., 2014), it would be desirable to apply optimisation directly to the more highly resolved models applied in prognostic simulations. Yet, to date this approach has proven to be prohibitive, due to the large computational demand mentioned above.

Therefore, we currently must accept a trade-off between finely resolved representation of physical transport processes, and well-tested and objectively optimised biogeochemical models. The calibration of model biogeochemistry in one computationally cheap circulation may elucidate the model's behaviour in its (biogeochemical) parameter space, and indicate a best set of parameters consistent with observed tracer fields. If the mean transport simulated by models were independent from model resolution, one could then transfer these parameters into the more expensive, high-resolution model. However, like biogeochemical models, physical parameterisations also reflect an idealised system, which can introduce errors in small- and large-scale patterns and processes. Biogeochemical model calibration is affected by these physical errors - the resulting optimal parameters can thus strongly depend on the circulation applied (Löptien and Dietze, 2019). So far it is not clear how model dynamics and performance will change, once these calibrated parameters are transferred to a different circulation, that resolves physical processes in more detail.

To investigate the mutual effects of circulation, biogeochemical model parameters, and model performance we have tested the effect of five different circulations on the objectively optimised parameters of a biogeochemical model. The ocean models differ in resolution as well as physical forcing and dynamics. Biogeochemical model calibration against nutrients and oxygen was carried out using a quasi-evolutionary algorithm, which carries out a dense scan of a six-dimensional, biogeochemical parameter space. Differences in optimal parameters are discussed before the background of large-scale physical properties. In portability experiments we examine model performance when parameters optimal for one circulation are transferred into another circulation. We finally quantify the effects of changes in parameter sets vs. those of circulation on global quantities such as oxygen inventory, OMZ volume, and global biogeochemical fluxes.

2 Models, experiments, and optimisations

2.1 Circulation and physical transport

All model simulations and optimisations apply the Transport Matrix Method (TMM; Khatiwala, 2007, 2018) as an efficient "offline" method for ocean passive tracer transport. The TMM represents advection and mixing in the form of transport matrices that have been calculated from an ocean circulation model simulation prior to the biogeochemical simulations performed here. For our model simulations we apply monthly mean transport matrices (TMs), as well as monthly wind speed, temperature and salinity for air-sea gas exchange. One set of TMs and forcing have been derived from a 2.8° global configuration of the MIT ocean model with 15 vertical levels (Marshall et al., 1997, see also Table 1). Using this rather coarse spatial grid and a time step length of 1/2 day for tracer transport, a model setup with seven tracers can be integrated for 3000 years in ≈ 1-1.5 hours on 4 nodes of Intel Xeon Ivybridge at the North-German Supercomputing Alliance (www.hlrn.de). This circulation is hereafter referred to as MIT28. For the second physical model configuration (hereafter referred to as ECCO) we apply TMs derived from a circulation of the Estimating the Circulation and Climate of the Ocean (ECCO) project, which provides circulation fields that yield a best fit to hydrographic and remote sensing observations over the 10-year period 1992 through 2001 with a horizontal

resolution of $1° \times 1°$ and 23 levels in the vertical (Stammer et al., 2004). A full spinup (3000 years) of the coupled model requires about 9 hours on 16 nodes of Intel Xeon Ivybridge.

Finally, three sets of transport matrices have been derived from version 2.9 of the University of Victoria Earth System Climate
95   Model (UVic ESCM; hereafter called "UVic"; Weaver et al., 2001), a coarse-resolution ($1.8° \times 3.6° \times 19$ vertical layers) ocean-atmosphere-biosphere-cryosphere-geosphere model. TM extraction was carried out as described by Kvale et al. (2017). One set of TMs is identical to that described in Kvale et al. (2017), and includes tidal mixing and a high-mixing scheme in the Southern Ocean, as well as an increased low latitude isopycnal diffusivity (configuration "UHigh"). This configuration utilizes a vertical diffusion coefficient of $0.43 \, \mathrm{cm}^2 \, \mathrm{s}^{-1}$ to stabilise meridional overturning in its linear, 3rd order upwind-biased advection scheme
100 (UW3, Holland et al., 1998; Griffies et al., 2008). This vertical diffusion coefficient is more than double the "standard" value of $0.15 \, \mathrm{cm}^2 \, \mathrm{s}^{-1}$, typically used with the UVic ESCM configured with the default 1st order flux corrected transport (FCT, Weaver and Eby, 1997) advection scheme. Reasons for the change in advection scheme for the application of the TMM to UVic are given in Kvale et al. (2017), but it is important to note here that the UW3 configuration has not benefitted from the more than two decades of careful parameter adjustments that users of the FCT configuration appreciate. The annual maximum global
105 meridional overturning strength in the UHigh configuration is 18.5 Sv, but other physical features of the circulation have not been previously assessed in detail. Two further sets of UVic ESCM TMs do not include regional adjustments to mixing. They have been tuned to a maximum annual average global overturning circulation of either 20 Sv (named U20) or 17.5Sv (named U17.5). This tuning was achieved by adjusting the vertical diffusion coefficient ($0.409 \, \mathrm{cm}^2 \, \mathrm{s}^{-1}$ in U17.5, and $0.4179 \, \mathrm{cm}^2 \, \mathrm{s}^{-1}$ in U20). The physical circulation parameterisation is otherwise identical to UHigh; utilising UW3 advection and the same
110 tidal mixing scheme. Differences arising in the calibrations between UVic ESCM TMs therefore reflect both differences in the application of regional mixing "corrections" (UHigh versus U17.5 and U20), as well as in global overturning strengths and secondary effects from changed values of the vertical diffusion coefficient. We note that none of the circulation configurations, aside from UHigh, have been previously evaluated against the most commonly used UVic ESCM FCT configuration (e.g., Weaver et al., 2001; Schmittner et al., 2005; Somes et al., 2013).

115 ## 2.2   Properties of circulation models

The five circulations differ in many aspects. First, being supported by observational data, ECCO's spatial salinity and density distribution agrees very well with observations, while the MIT28 and UVic circulations show, for example, a too shallow depth of the $\sigma = 27.5$-isopycnal in the Atlantic Ocean, and too saline waters in the deep northern North Atlantic (Figure 1 and S1). In addition, the three UVic circulations all suffer from a too weak formation and northward propagation of Antarctic Intermediate
120 Waters (AAIW), as identified from water of low salinity at ≈ 1000m depth in the southern hemisphere. MIT28, U20 and U17.5 also show a too large outcrop area of dense waters ($\sigma_\theta \geq 27.5$) in the Southern Ocean, which does not agree with the observed pattern. Here ECCO and UHigh better match observations.

Striking differences also occur with respect to the annual maximum mixed-layer depth, as derived from a potential density difference to the surface of $\Delta\sigma_\theta \geq 0.03$ (Figure 2), calculated from the models' monthly mean temperature and salinities.
125 Obviously, MIT28 shows a too large area of deep mixing around $60°$S in the Southern Ocean, which does not agree with

mixed-layer depths derived from observed temperature and salinity. On the other hand, only this circulation exhibits deep mixing in the Labrador Sea, which is in agreement with observations. All configurations of UVic exhibit a too large area of deep mixing in the Southern Ocean, while ECCO tends to underestimate mixing in this area.

Differences between circulations are also reflected in the global distributions of ages diagnosed from the models. MIT28, U20 and U17.5 show very old (> 1400 years) waters in the deep northern North Pacific (Figure 3). Here, ECCO and UHigh contain much younger waters, mostly below 1400 years. In MIT28 the age increases rapidly with depth in the Southern Ocean (up to more than 800 years). Especially ECCO, but also UVic exhibit much younger waters below 2000 m in this region. Finally, the U20 and U17.5 configurations result in too old deep waters in the northern North Atlantic (Khatiwala et al., 2012, their Fig. 4). In general, ECCO and UHigh agree much better with mean age constrained with transient (radiocarbon, CFCs) and hydrographic (temperature, salinity, nutrients, oxygen) tracer observations (Khatiwala et al., 2012).

To summarise, our applied offline circulations for MIT28 and ECCO differ strongly with respect to resolution and many global physical properties, with the UVic configurations in between these two. As a data-constrained circulation, ECCO shows the best overall agreement with observations of all five circulations.

2.3 Derived indicators of circulation

The underlying circulation models, from which the TMs and forcing were extracted, differ in many aspects, such as parameterisation of mixing, forcing, sea-ice, etc., all of which can affect their dynamic behaviour and the quantities and diagnostics described above. For example, in the Southern Ocean the eastward transport of waters through the Drake passage can affect the properties and formation of SAMW (e.g., Sallee et al., 2010, 2013), while parameterisation of sea ice in the models might affect the formation and ventilation of AABW (Dutay et al., 2002), with consequences for water mass age. It is beyond the scope of this paper to compare and discuss the details of the circulation models. Instead, to examine the potential impact of their characteristic features on optimal biogeochemical model parameters (section 3.2), we will focus on three diagnostics that can be easily derived from most circulation models (see Table S1 for simulated and observed values).

Area of deep mixing: Deep mixing in the North Atlantic supplies oxygen to the ocean (Khatiwala et al., 2012). In the Southern Ocean mode and intermediate waters acquire their biogeochemical signatures north of the Antarctic Circumpolar Current (ACC), before being subducted into the interior ocean. These waters then ventilate the thermocline of the subtropics in the southern hemisphere (Sallee et al., 2013). However, as also shown in other studies (for example, Sallee et al., 2013, who found a large variability in Southern Ocean mixed-layer depths simulated by 21 ocean circulation models) the circulations applied in our study differ strongly in the extent and location of deep mixing in this region. To account for the potential effects of this variability on optimal parameter choice, we evaluated the area of annual maximum deep mixing in the two regions. Mixed-layer depth was defined by a density difference of $\Delta\sigma_\theta \geq 0.03$ (in line with de Boyer Montégut et al., 2004; Dong et al., 2008; Sallee et al., 2013), calculated from monthly mean potential temperature and salinity. For the Southern Ocean (south of $40°$S) and the North Atlantic (north of $40°$N) we then calculated the area, where the annual maximum mixed-layer exceeds either 200 or 400 m (the range of mixed-layer depths simulated and observed in the Southern Ocean; Sallee et al., 2013). Altogether, we thus obtain four different indicators for ocean ventilation through deep ocean mixing

Outcrop area of mode and intermediate water masses: On centennial timescales the Antarctic Intermediate Water (AAIW) and Subantarctic Mode Water (SAMW) formed in the Southern Ocean determine nutrient concentrations in subtropical areas. Their nitrate deficit and isotopic composition carries signatures of denitrification and nitrogen fixation (Rafter et al., 2013; Tuerena et al., 2015). Given that the models differ so strongly with respect to the surface density in the Southern Ocean (Figure 1), we evaluated the outcrop area of waters defined by a density of $26.5 \leq \sigma_\theta < 27.5$ and $27.5 \leq \sigma_\theta$ in both the Southern Ocean and the northern North Atlantic (defined as above). In the Southern Ocean the first criterion approximately reflect SAMW and AAIW combined, and the second criterion Circumpolar Deep Water (CDW) (similar to the definitions used by Palter et al., 2010; Iudicone et al., 2011; Rafter et al., 2012, 2013). North Atlantic waters defined by densities of $26.5 \leq \sigma_\theta < 27.5$ and $\sigma_\theta \geq 27.5$ coincide mainly with the region between 40°N-60°N and the Greenland Sea, respectively (Figure 1).

Age of water masses: We use the concept of water mass (or ventilation) age as a diagnostic for the combined effects of ocean circulation, mixing and ventilation on the time elapsed since a water parcel has been isolated from the atmosphere. We distinguish between average ideal age in three different water masses by applying the criteria of Matsumoto et al. (2004). According to their water mass definitions, North Atlantic Deep Water (NADW) comprises all waters in the North Atlantic between 0° and 60°N and 1500-2500 m depth. North Pacific Deep Water (NPDW) is defined for a region between 0° and 60°N and 1500-5000 m depth. Finally, Circumpolar Deep Water (CDW) consists of all waters south of 45°S, for a depth between 1500-5000 m (see also Figure 3). We note that, using these region definitions, the average age of NADW is influenced by waters of the Eastern Tropical Atlantic (ETA), which are quite old in ECCO circulation ($> 400y$), and young in UHigh (Figure 3, left panels). One reason for this could be different rates of overturning, which is between 13 to 14 Sv in ECCO (together with a weak western boundary current; Wunsch and Heimbach, 2006), while the UVic configurations are characterised by higher overturning around 17.5 to 20 Sv. Finally, because the eastern tropical Pacific (ETP) is the main region of fixed nitrogen loss in the models (Kriest and Oschlies, 2015), but the water age in this region varies strongly among the circulations applied in our study (Figure S2), we also calculated average age in the Eastern Equatorial Pacific (ETP) between ±20° latitude, east of 160°W, and within 150-500 m depth as a fourth potential indicator.

2.4 The biogeochemical model

For all model simulations we apply the Model of Oceanic Pelagic Stoichiometry ("MOPS"), which simulates the biogeochemical cycling among phosphate, phytoplankton, zooplankton, dissolved organic phosphorus (DOP) and detritus. The model is described in detail in Kriest and Oschlies (2015), and we here only give a brief overview on model structure, with focus on parameters (and processes) affected by optimisation (see below). All components are calculated in units of mmol P m^{-3}. We assume a constant nitrogen-to-phosphorus ratio of organic matter of 16 [mol N:mol P]. The oxygen demand of aerobic remineralisation is given by $R_{-O2:P}$ [mol O$_2$:mol P], following the stoichiometry by Paulmier et al. (2009). In the model oxygen-dependent aerobic remineralisation of organic matter follows a saturation curve with half-saturation constant K_{O2} [mmol m^{-3}]. With declining oxygen, denitrification takes over as long as nitrate is available above a defined threshold $DIN_{\min}$ [mmol m^{-3}]. Suboxic remineralisation (denitrification) also follows a saturation curve for the oxidant nitrate, defined by the half-saturation

constant for nitrate, K_{DIN} [mmol m^{-3}]. The model assumes immediate coupling of the different processes involved in ni-
trate reduction to dinitrogen, following the stoichiometry derived by Paulmier et al. (2009). Loss of fixed nitrogen (through
denitrification) is balanced by a temperature-dependent parameterisation of nitrogen fixation, which relaxes the nitrate-to-
phosphate ratio to d with a maximum rate μ_{NFix} [μmol m^{-3} d^{-1}]. Detritus sinks with a vertically increasing sinking speed:
$w = a\,z$ [m d^{-1}]. With a constant degradation rate $r = 0.05$ d^{-1}, in equilibrium this is equivalent to a depth-dependent particle
flux curve, corresponding to a power law: $F(z) = (z/z_0)^{-b}$, with $b = r/a$ (see Kriest and Oschlies, 2008). Depending on the
rain rate to the sea floor, a fraction of detritus deposited at the bottom of the deepest vertical box is buried in some hypothetical
sediment. Non-buried detritus is resuspended into the deepest box of the water column, where it is treated as regular detritus.
The global annual burial of organic phosphorus and nitrogen is resupplied in the next year via river runoff (when simulated
with MIT28 or ECCO TMs) or at the ocean surface (TMs derived from UVic). We note that in uncalibrated models (e.g., Kriest
and Oschlies, 2015) this creates differences of a few percent in the global inventory of nitrate and oxygen. Differences in the
regional distribution of nutrients and oxygen are largely comparable in magnitude to those caused by the numerical sinking
scheme of detritus (Kriest and Oschlies, 2011). The effect of this process is subject to further research.

2.5 Optimisation algorithm

Optimisation of $n = 6$ biogeochemical model parameters (see section 2.7 for choice of parameters to be optimised) is carried
out using an Estimation of Distribution Algorithm, namely the Covariance Matrix Adaption Evolution Strategy (CMA-ES;
Hansen and Ostermeier, 2001; Hansen, 2006). The application of this algorithm to the coupled biogeochemistry-TMM frame-
work has been presented in detail in Kriest et al. (2017), and we here only give a brief overview: In each iteration ("generation")
the algorithm defines a population of 10 individuals (biogeochemical parameter vectors of length n), sampled from a multi-
variate normal-distribution in $\mathbb{R}^n$. Following the simulation of these 10 model setups over 3000 years to near-steady state,
after which, for example, global oxygen and nitrate inventory change only by a small amount (see Figure 2 by Kriest and
Oschlies, 2015), the misfit (cost) function (presented in section 2.6) is evaluated, and information of the current as well as
previous generations is used to update the probability distribution in $\mathbb{R}^n$ such that the likelihood to sample solutions resulting
in a good fit to observations increases. Therefore, the population (the number of model simulations per generation) in CMA-ES
is smaller, and of lower computational demand than in classical evolutionary algorithms, making the algorithm applicable to
this computationally expensive problem. On the other hand, with its quasi-stochastic sampling CMA-ES can still to a certain
degree perform well with misfit functions characterised by a rough topography (i.e., many local optima; Kriest et al., 2017;
Kriest, 2017).

2.6 Misfit (cost) function

As in Kriest et al. (2017) and Kriest (2017) the standard misfit to observations J is defined as the root-mean-square error
(RMSE) between simulated and observed annual mean phosphate, nitrate, and oxygen concentrations (Garcia et al., 2006a, b),
mapped onto the respective three-dimensional model geometry. Deviations between model and observations are weighted by

7

the volume of each individual grid box, V_i, expressed as the fraction of total ocean volume, V_T. The resulting sum of weighted deviations is then normalised to the global mean concentration of the respective observed tracer:

$$J_{\text{RMSE}} = \sum_{j=1}^{3} J(j) = \sum_{j=1}^{3} \frac{1}{\bar{o}_j} \sqrt{\sum_{i=1}^{N} (m_{i,j} - o_{i,j})^2 \frac{V_i}{V_T}} \tag{1}$$

$j = 1, 2, 3$ indicates the tracer type (phosphate, nitrate, oxygen) and $i = 1, ..., N$ are the model locations of N model grid
230 boxes. $\bar{o}_j$ is the global average observed concentration of the respective tracer. $m_{i,j}$ and $o_{i,j}$ are simulated and observed concentrations, respectively. By weighting each individual misfit with volume, J_{RMSE} serves as a long time-scale geochemical estimator in contrast to a misfit function focusing, e.g., on (rather fast) turnover in the surface layer, or resolving the seasonal cycle.

2.7 Parameter optimisations and cross-validation experiments

235 The optimisations presented here are based upon two successive optimisations presented by Kriest et al. (2017) and Kriest (2017). Both studies applied model MOPS coupled to TMs derived from MIT28. The cost function, as presented in Equation 1 was calculated after a spin up of 3000 years, as also used in the present study.

In the first optimisation, Kriest et al. (2017) optimised four parameters related to plankton growth and loss terms, together with b and $R_{-\text{O2:P}}$. The optimal parameters of this first calibration led to a better agreement of simulated global biogeochemical
240 fluxes to observations of primary and export production, zooplankton grazing, particle flux at 2000 m, and organic matter burial at the sea floor (Kriest et al., 2017). In a subsequent optimisation Kriest (2017) kept the four optimal plankton parameters fixed, and calibrated four parameters related to remineralisation and nitrogen fixation (namely K_{O2}, K_{DIN}, DIN_{min} and μ_{NFix} described in section 2.4), together with b and $R_{-\text{O2:P}}$ (see Table 1). This second optimisation by Kriest (2017) led to a better match to independent estimates of pelagic denitrification, without deteriorating the matches to observed primary and export
245 production, zooplankton grazing, particle flux at 2000m and organic matter burial at the sea floor. It is hereafter referred to as MIT28* and serves as the starting point for the four additional optimisations presented in this paper.

In particular, we repeated optimisation MIT28* against Equation 1 in circulations ECCO, UHigh, U20 and U17.5 described above. In the following we refer to these four additional optimisations and optimal parameter sets as ECCO*, UHigh*, U20* and U17.5*. To test the portability of optimised parameters to different physical settings, we then transferred the parameter
250 sets of MIT28*, ECCO*, UHigh*, U20* and U17.5* to the other four circulations, again simulating each coupled model for 3000 years. Thus, we present results from 25 different model simulations with five different parameter sets and five different circulations.

3 Results

3.1 Performance of optimised models

When optimised for different circulations the coupled models show similar values of the misfit function J^* (Table 1). The misfit decreases with more realistic circulation and physics (according to the criteria described in section 2.2), and is lowest for ECCO*. Global mean phosphate profiles are quite similar, and vary less then 10% of the observed concentrations at depths below 500 m (Figure 4, panel (A) and Figure S3, black lines). The low variation of phosphate might be explained by the fixed phosphorus inventory of the model. Only at the surface, where nutrient concentrations become low, relative variation is larger. However, despite optimisation some regional mismatches remain: for example, the model when optimised for the three UVic circulations shows a considerable overestimate of deep (> 3000 m) phosphate in the Atlantic (Figure S4). All optimal models further underestimate phosphate in the mesopelagic of the northern North Pacific.

Vertical nitrate profiles are also quite similar to each other (Figure 4, panel (B) and Figure S3, red lines), although the nitrate inventory is allowed to adjust dynamically to the loss of fixed nitrogen during denitrification, and its balance through nitrogen fixation at the sea surface. Because optimisation also attempts to match nitrate observations, the differences between the models are nevertheless quite small, and result in deviations of 1-2% of observed global mean nitrate (see Table 1). The global pattern of nitrate residuals (Figure S5) generally corresponds to that of phosphate. However, the spatial distribution of fixed nitrogen gain and loss causes variations in the global distribution of the nitrate deficit in relation to phosphate, as expressed as $N^* = NO_3 - 16 \times PO_4$ (Figure 5).

In agreement with observations, all optimised models show a large nitrate deficit in the Pacific Ocean, manifest in strongly negative N^* (Figure 5). This lack of nitrate is caused by denitrification in the ETP, and is balanced by nitrogen fixation in tropics and subtropics (see Figure S6). In the Atlantic Ocean, where denitrification is largely absent, simulated N^* is far less negative than in other areas, but it is never positive as suggested by the observations. This mismatch can be explained by the prerequisites of the biogeochemical model applied: In MOPS, nitrogen fixation relaxes nitrate to $16\times$phosphate with a time constant defined by μ_{NFix} whenever N^* is negative; otherwise it does not occur. Because of this process parameterisation, N^* is restricted to values ≤ 0. Finally, the Southern Ocean has moderate values of N^*, owing to the mixing of water masses of different origin. Thus, although global average nutrient profiles match observations well, with little differences among optimal models, some regional biases remain, which differ among the models. Also, because of the different processes involved in nutrient turnover, phosphate and nitrate distributions are not exactly the same, with consequences for the nitrate deficit in different oceanic domains.

Global mean oxygen profiles of optimised models vary more strongly than those of nutrients, up to ≈ 40 mmol O_2 m^{-3} in the deep ocean, and thus more than 20% the observed value (Figure 4, panel (C) and Figure S3, blue lines). Overall, a finer resolution and a more realistic circulation improve the representation of this tracer, reducing the global oxygen bias, which ranges from 5.3 mmol O_2 m^{-3} (U17.5*) to almost zero (ECCO*; Table 1). On a regional scale all models show some common biases: South of $40°$S all optimised models overestimate zonal mean oxygen in subsurface waters above ≈ 1500 m (MIT28*) to ≈ 2000 m (ECCO*), or even further downward (UVic simulations; Figure S7). In the Pacific Ocean these too high oxygen

9

concentrations propagate northward. Finally, in mesopelagic waters of the northern North Pacific all models overestimate oxygen, especially above 1000 m. Common to all models is further an underestimate of oxygen in subsurface waters (down to ≈ 1000 m) in the subtropics and tropics of the southern hemisphere. The Atlantic Ocean is generally characterised by too low oxygen concentrations at greater depths. Here, the five optimal models differ: ECCO* exhibits too low mesopelagic oxygen in the tropical and subtropical Atlantic. MIT28* underestimates oxygen particularly in deep waters of the southern hemisphere, and north of 60°N. The optimal UVic configurations are biased low in deep waters of the northern hemisphere. These low oxygen concentrations of the UVic configurations are accompanied by too high phosphate and nitrate in the deep North Atlantic (Figure S5), and indicate too high remineralisation of organic matter in these depths. Together with circulation these result in a too strong accumulated signal of remineralisation.

Thus, while there are common features among the five optimal models, there are also some striking differences especially in the Atlantic. These differences can be explained with the large impact of the Pacific on the misfit function. Owing to its large volume, and the comparatively large impact of oxygen on the misfit function (see Kriest et al., 2017), optimisation will attempt to minimise especially oxygen misfits in this region, and tune the biogeochemical model parameters to compensate for potential errors of the respective circulation. These different parameters affect the oxygen distribution in the Atlantic, which does not contribute so much to the global misfit. Different physical properties of the circulations then cause divergent patterns in the oxygen distribution of this region.

3.2 Best parameters of optimisations in different circulations differ

As presented and discussed by Kriest (2017), optimisation of MOPS in the circulation of MIT28 reduces the particle flux length scale by increasing b from a "default" value of ≈ 1.1 (Kriest and Oschlies, 2015) to 1.39. It further results in a high nitrate threshold for the onset of denitrification ($DIN_{\min} = 15.8$ mmol m^{-3}, compared to the default value of 4 mmol m^{-3} applied by Kriest and Oschlies, 2015), a low affinity of denitrification to nitrate ($K_{DIN} = 32$ mmol m^{-3}), and a low maximum nitrogen fixation rate ($\mu_{NFix} = 1.19 \, \mu$mol m^{-3} d^{-1}; Table 1), which is only about half of the default value applied by Kriest and Oschlies (2015). The optimised oxygen affinity of remineralisation is very high, as indicated by a low value of K_{O2} (the half-saturation constant for oxygen). We note that K_{O2} also regulates the inhibition of denitrification by oxygen; when this parameter becomes very low, denitrification is more strongly inhibited by oxygen. Hence, the optimal model configuration MIT28* induces only moderate denitrification, and prevents a decline of the global nitrate inventory through this process (see also Kriest and Oschlies, 2015). The oxygen demand of remineralisation, $R_{-O2:P}$, remains close to the value derived from observations (170 ± 10 mmol O$_2$:mmol P; Anderson and Sarmiento, 1994). Finally, as noted by Kriest (2017), some parameters are only weakly constrained by the misfit function: for example, an almost ten-fold increase in K_{O2} results in a misfit function not larger than 1% of the optimal misfit (see also Table 1). One reason for this low sensitivity of the misfit to a variation in K_{O2}, K_{DIN} or $DIN_{\min}$ is the small volume occupied by suboxic zones, where these parameters can play a role for dissolved inorganic tracer concentrations (Kriest, 2017).

Optimising the same set of parameters in either the three different UVic circulations or the ECCO circulation also results in a high threshold for the onset of denitrification, $DIN_{\min}$ (Table 1). As for MIT28* the dependencies of oxic and suboxic

remineralisation on oxygen or nitrate, expressed through K_{O2} and K_{DIN}, may vary largely within the parameter space without having a large impact on the misfit function.

In contrast, $R_{-O2:P}$ and b are constrained very well by the misfit function, as indicated by the narrow range of parameters that result in a good fit to observations. For example, all solutions of the ECCO*, which result in a misfit within 1% of the optimal fit require a b value between 1.4 and 1.5, and a stoichiometric demand for oxygen between 150 and 154 mol O_2: mol P (Table 1). Optimal $R_{-O2:P}$ decreases from 170 mol O_2: mol P (MIT28*) over 162 mol O_2: mol P (UHigh*) to 151 mol O_2: mol P (ECCO*). The exponent determining the shape of the particle flux curve, b, also varies among the five optimisations, between 1.27 (UHigh*) and 1.46 (ECCO*). The range of good (within 1% of the optimal fit) values for b differs between ECCO* and UHigh* (b between 1.2-1.3). Also, the range for good values for $R_{-O2:P}$ of ECCO* does not overlap with that of the other optimisations.

Optimal μ_{NFix} also varies considerably among the different optimisations, from 1 to 3 μmol m^{-3} d^{-1}. Here, the range of good parameter values for U20* (1.0-1.4 μmol m^{-3} d^{-1}) does not overlap with that of ECCO*, UHigh* and U17.5* (all between 1.5-3 μmol m^{-3} d^{-1}). Overall, MIT28* and U20* benefit from a low maximum nitrogen fixation rate around 1 μmol m^{-3} d^{-1}, while the other models require a larger rate between 2-3 μmol m^{-3} d^{-1}.

Therefore, to achieve a good fit to observations different circulations seem to require markedly different parameters for the oxygen utilisation by remineralisation, ($R_{-O2:P}$), the exponent determining the particle flux curve (b) and the potential rate of nitrogen fixation (μ_{NFix}). Other parameters vary little (DIN_{min}), or (as indicated by overlapping ranges of good parameters) the differences among them might not be relevant (K_{O2} and K_{DIN}) for a misfit function targeting on the global scale. It is important to note that the relevant parameters do not seem to be correlated with each other (Figure S8). Apparently, different characteristics of each circulation influence the choice of the optimisation algorithm for the optimal values of $R_{-O2:P}$, b and μ_{NFix}.

To investigate the potential dependence of optimal parameters on circulation, we examined the area of dense-water outcrop and deep mixing in two different regions (see section 2.3 for definition). Together with average age of four different regions or water masses, we investigate 12 different diagnostics for each circulation for their influence on optimal biogeochemical parameter estimates. In most cases the optimal parameters are not correlated with physical properties (Table 2). However, the oxygen demand of remineralisation, $R_{-O2:P}$ is significantly correlated with the area of deep mixing in the northern North Atlantic for maximum mixed-layer depths of 200 and 400 m ($p < 0.05$). For both criteria $R_{-O2:P}$ increases with increasing area of deep mixing (Figure 6). In addition, $R_{-O2:P}$ also correlates with a deep mixing area defined by > 200 m in the Southern Ocean, albeit not significantly (Table 2). In other words, more vigorous mixing in areas of deep, intermediate or mode water formation allows for a higher oxygen utilisation by remineralisation. Parameter b describing the particle flux curve correlates significantly with the ideal age of NADW (Table 2), and increases with increasing age of this water mass (Figure 6). Finally, the maximum potential rate of nitrogen fixation μ_{NFix} correlates with the outcrop area of waters with moderate density ($26.5 \leq \sigma_\theta < 27.5$) waters in the Southern Ocean. An increase in the outcrop area of these waters - which correspond roughly to the sum of AAIW and SAMW - results in an increase of optimal maximum nitrogen fixation rate (Figure 6). The age of mesopelagic waters in the ETP seems to play a small role, despite the fact that it is an important area in the global nitrogen

budget (Figure S6). Possible reasons for the dependence of these three parameters on physical diagnostics will be discussed in section 4.1.

3.3 Cross-validation experiments: Can we transfer parameters optimal for one circulation to another circulation?

Given that three optimal parameters differ among the model circulations, we investigate model performance and dynamics when these parameter sets are swapped among circulations. Of course, every coupled model performs best (with respect to J^* of Equation 1) when simulated with parameters optimal for the respective circulation, as indicated by the lowest relative misfit along the main diagonal in panel (A) of Figure 7. When exchanging the biogeochemical parameters optimal for ECCO or MIT28 circulation with parameters optimised in any other circulation, the model performance with respect to respect to J^* deteriorates. The misfit function changes less when the optimal parameters are swapped among the different UVic circulations.

In the model the global oxygen inventory adjusts dynamically to the combined effects of circulation and biogeochemical parameters, causing a large impact of this tracer on the misfit function (Kriest et al., 2017). Therefore, optimisation attempts to reduce the global oxygen bias, which is low for each optimal model configuration, indicated by the low values along the main diagonal of Figure 7, panel (B). The oxygen bias induced by the changes in parameter set and circulation depends on the combination of these two: For example, the low value for $R_{-O2:P}$ and the high value for b of ECCO* causes a large overestimate of the oxygen inventory in any other circulation (indicated by warm colours in the second column of Figure 7, panel (B)). Vice versa, applying optimal parameter sets from MIT28*, UHigh*, U20* or U17.5* to the ECCO circulation results in a too low oxygen inventory (indicated by cold colours in the second row from the bottom of Figure 7, panel (B)). When swapping parameter sets among the different configurations of the UVic circulation, the effect is much less pronounced. Apparently, despite their different overturning and mixing, the UVic circulations are more similar to each other than those of MIT28 and ECCO. We note that these large differences in oxygen inventory arise mainly from deeper ($27.5 \leq \sigma_\theta$) layers, while the oxygen inventories of waters lighter than $\sigma_\theta = 27.5$ are quite similar (Figure S9).

OMZ volume is biased low for the parameter set of ECCO*, and in the MIT28 circulation (Figure 7, panel (C)). This underestimate is likely caused by the very low $R_{-O2:P}$ and high b of ECCO*, or the vigorous mixing in the MIT28 circulation, which both tend to increase subsurface oxygen concentrations. Otherwise, OMZ volume does not seem to be closely related to the parameter set or circulation, likely because this diagnostic is largely independent of the applied misfit function, and depends on the local circulation pattern (see discussion by Sauerland et al., 2019).

Thus, because of different physical model properties, the biogeochemical model MOPS, when coupled to the MIT28 and ECCO circulation requires unique and different sets of parameters for optimal model performance. In the UVic circulations the model is more flexible with regard to parameters; yet, when aiming for independent diagnostics such as OMZ volume, there is no clear relationship between OMZ volume and changes in parameter set or circulation.

3.4 Effect of parameters and circulation on phosphate and oxygen concentrations in different water masses

Because of the regional biases of nutrients and oxygen in the North Atlantic, North Pacific and Southern Ocean (see section 3.1), and because the values of $R_{-O2:P}$ and b selected in the calibration process correlate significantly on water mass properties of the

12

NADW, NPDW and CDW (section 3.2), we here examine more closely how these parameters affect the large scale distribution of phosphate (as a conserved nutrient) and oxygen, which can adjust dynamically at the model's air-sea interface, and is thus a non-conservative tracer.

Kriest et al. (2012) showed that b, the parameter determining the particle flux length scale, has a large influence on the distribution of phosphate along the "conveyor belt" (i.e., along waters of different age), in agreement with the results obtained by Bacastow and Maier-Reimer (1991) and Kwon and Primeau (2006). We here carry out an analysis similar to that by Kriest et al. (2012) and evaluate average phosphate and oxygen within the NADW, NPDW and CDW, with region definitions as described above for water mass age (section 2.3).

Within each circulation a smaller value of b (corresponding to faster sinking particles) increases phosphate in the NPDW, and decreases it in the NADW (Figure 8, panel (A)), confirming the pattern found by Kriest et al. (2012). When plotting average phosphate in the NPDW against average phosphate in the CDW there is no such relationship, but the average value in the CDW varies only little (Figure 8, panel (B)), and is near the observed value of 2.26 mmol m^{-3}. The spread of phosphate concentrations caused by different parameter sets (same symbol with different colours) is about the same (between 0.1-0.15 mmol m^{-3}) as the spread caused by different circulations (different symbols of the same colour). Thus, both biogeochemistry and circulation seem to play an equally large role for the distribution of phosphate between NADW and NPDW. The variation caused by biogeochemical parameters is smaller for CDW, indicating that in this region physical processes play a larger role.

In contrast to phosphorus, the oxygen inventory is not fixed, but regulated by the interplay of circulation, air-sea exchange, and biogeochemical turnover. Because of this we find a pattern that is very different from that of phosphate when examining the distribution of average oxygen in different water masses. Now, within each circulation average oxygen in the NPDW increases almost linearly with average oxygen in the CDW and NADW (Figure 9), highlighting the role of these waters for the ventilation of the deep North Pacific. In most circulations, a large value of $R_{-O2:P}$ results in a low oxygen content in all water masses. All optimised coupled model configurations suggest average oxygen between 220-250 mmol m^{-3} in the NADW. For the NPDW all optimised models simulate average oxygen concentrations around 150 mmol m^{-3}, and thereby overestimate the observed value of 125 mmol m^{-3} by one fifth. Average oxygen in the CDW of optimal model simulations varies between $\approx 210 - 240$ mmol m^{-3}, encompassing the observed value of 215 mmol m^{-3}. Thus, given the quite wide range of potential parameter values, optimisation improves the global oxygen bias (see Table 1), but some residual regional bias especially in the NPDW remains.

3.5 Effect of parameters and circulation on global oxygen inventory and OMZ volume

A declining trend of global average oxygen with increasing $R_{-O2:P}$ is also reflected in panel (A) of Fig. 10, but circulation also plays a role for the oxygen inventory, with the ECCO circulation showing the lowest values. To have a closer look at the individual contributions of circulation and biogeochemistry to the overall variability of oxygen inventory and OMZ volume we have calculated their maximum spread caused by varying only the circulation (keeping the biogeochemical parameter set constant; ΔCirc) and by varying only the biogeochemical parameters (keeping the circulation constant; ΔPar). For example, to determine ΔCirc for global average oxygen or OMZ volume (here denoted as X), for each parameter set

13

i simulated with the five different circulations $j = 1...5$ we compute the difference between the maximum and minimum value $\Delta X_i = \max(X_{i,j=1...5}) - \min(X_{i,j=1...5})$, and then determine the maximum of these differences: $\Delta\text{Circ} = \max(\Delta X_i)$. The computation of ΔPar is done analogously. We also compute the maximum across all optimal model configurations $\Delta\text{Opt} = \max(X_{i=j}) - \min(X_{i=j})$, and across all 25 experiments presented in this study: $\Delta\text{All} = \max(X_{i,j}) - \min(X_{i,j})$. Using this approach, a value for ΔPar close to ΔAll indicates that the model variability is mainly induced by the biogeochemical parameter set, whereas a relatively large value for ΔCirc indicates a major impact of circulation. Table 3 and Figure 11 show the results of this comparison, and Figure 12 illustrates the variability for each circulation or parameter set, normalised by the average over all optimal models. Note that the longest horizontal or vertical line in Figure 12 corresponds to ΔPar and ΔCirc in Table 3, while the width of the grey shaded square corresponds to ΔAll.

Firstly, biogeochemical parameters (ΔPar) as well as circulation (ΔCirc) play an about equally large role for the global average oxygen, which varies by ≈ 24 mmol m^{-3} (Table 3 and Fig. 11), or about $\pm 15\%$ of the average optimal value (see also Figure 12). The variation decreases to less than one fourth of this value if we restrict our analysis to only optimal models (ΔOpt); as noted above, this strong decrease arises because all optimal models have adjusted $R_{-\text{O2:P}}$ to account for different ventilation in the high latitudes. Considering all 25 experiments, i.e., the accounting for variation induced by both biogeochemical and physical configurations (ΔAll), leads to a variation six times as large as for the optimal configurations (ΔOpt).

The variability is much more pronounced when considering the OMZ volume as defined by two criteria, 50 mmol m^{-3} and 80 mmol m^{-3}. Again, both circulation and parameter set play an about equally large role; but the impact of changes in parameters or circulation varies across the different models (Fig. 12). For example, applying the parameter set of MIT28* to a different circulation causes a very strong increase in OMZ volume (vertical black lines in Fig. 12), while model MOPS coupled to the circulation of MIT28 is quite robust with respect to different parameters (horizontal black lines in Fig. 12). On the other hand, when coupled to the ECCO circulation the biogeochemical model is quite sensitive to the biogeochemical parameter set (horizontal red lines in Fig. 12), but its optimal parameter set ECCO* has a smaller effect when applied to other circulations (vertical red lines in Fig. 12). This diverging effect of parameters and circulation among the different models eventually causes a large spread of 67.2×10^{15} m^3 of global OMZ volume across all model experiments (ΔAll in Table 3 and Fig. 11), i.e., more than 100% the observed volume. The effects are even more pronounced when considering a criterion of 80 mmol m^{-3} for OMZ definition (Table 3). The OMZ volume does not show any consistent trend with b, $R_{-\text{O2:P}}$, or circulation (Fig. 10, panel (B)), although models with high b and low $R_{-\text{O2:P}}$ tend to result in a smaller OMZ volume. The circulation of MIT28 shows the lowest OMZ volume.

To summarise, oxygen inventory and OMZ volume are almost equally influenced by physics and biogeochemistry. Optimisation reduces the spread induced by either biogeochemistry or physics to about 30% percent for average oxygen, but less for OMZ volume, which varies strongly across all model experiments. We note that the individual contributions of ΔPar and ΔCirc for both diagnostics do not add up to ΔAll (Table 3), indicating that the effects of biogeochemical parameters and circulation are not linear and additive.

3.6 Effect of parameters and circulation on global biogeochemical fluxes

Oxygen and nutrient distributions are influenced by the production of organic matter in the euphotic zone, and its subsequent transport to the ocean interior by physical and biogeochemical processes. In addition, denitrification in combination with nitrogen fixation can affect the global nitrogen inventory, and the spatial distribution of the nitrate deficit (see section 3.1). We finally here investigate how these fluxes are affected by the two parameters $R_{-O2:P}$ and b.

In our model experiments simulated global primary production depends slightly more on circulation (ΔCirc) than on biogeochemical parameters (ΔPar; Table 3, Figures 11 and 12). An increase in b (corresponding to slowly sinking particles) causes primary production to increase (Figure 10, panel (D)), likely because of the higher nutrient retention in the euphotic zone, shallow remineralisation and enhanced entrainment of nutrients into the surface layers. Because the latter process depends on physical dynamics, we also find an influence of the circulation model on global primary production. Further, our optimisations did not include parameters relevant for plankton dynamics at the surface, which can also explain the comparatively large impact of circulation. The variation across all optimal models of our study (ΔOpt) is much smaller (about one third) than the variation across all model experiments (ΔAll).

Circulation also plays a large role for export production (particle flux through 100-130 m, depending on model grid), as it supplies new nutrients to the well-lit upper ocean which will, under steady state conditions, be exported again. Somehow surprisingly, export production is not strongly determined by b (Fig. 10, panel (E)). This parameter affects directly the sinking of organic matter out of the euphotic zone. On the other hand, it determines the subsurface concentration of nutrients, as a source for upwelling and entrainment of nutrients. A large b, corresponding to slow sinking and shallow remineralisation, increases nutrients within and below the euphotic zone, and thus primary production as the ultimate source of export production; on the other hand, it prevents fast settling of organic particles out of the euphotic zone. Because the plankton parameters were not changed during optimisation, global grazing follows almost linearly primary production ($r = 0.95$), and the statements and conclusions made with respect to the former flux largely apply to grazing (no figure). Therefore, the combined antagonistic effects of b on surface (and subsurface) nutrient turnover, subsurface nutrient concentrations (as a source of nutrient entrainment and mixing) and direct organic particle flux in the upper few hundred meters explain the relatively small variation caused by biogeochemical parameters on export production (ΔPar; Table 3 and Figures 11 and 12), which is only about half as much the variation due to circulation (ΔCirc).

Deep particle flux, on the other hand, is almost entirely determined by b, and circulation plays a negligible role for this flux (Fig. 10, panel (F)). The large influence of this parameter is also reflected in its range over all model simulations, which is only slightly larger than the range of flux in optimally configured models (Table 3). Therefore, simulated organic matter supply to the deep ocean and deep nutrient concentrations will, to a large extent, depend on the prescribed particle flux profile, with potential effects on the long-term storage of carbon dioxide.

The loss of fixed nitrogen through pelagic denitrification is tightly related to the extent of OMZs, and thus varies quite strongly among the different experiments, with no clear trend for either b, $R_{-O2:P}$ (Figure 10, panel (C)) or μ_{NFix} (no figure). The range of variation due to parameters and circulation is about equally large (about 50% of the average optimal global flux;

490   Table 3 and Fig. 11). Overall, this global flux is affected by both circulation and changes in biogeochemical parameters, which induce changes of about $\pm 30\%$ around the mean optimal flux for each model circulation or parameter set (Fig. 12). Global fixed nitrogen loss of the optimised models varies much less, likely because optimisation adjusts the parameters to match observed nitrate profiles.

To summarise, at least for this particular biogeochemical model circulation and biogeochemistry affect global biogeochem-
495   ical fluxes in different ways. While primary production and fixed nitrogen loss are almost equally influenced by physics and biogeochemistry, export production depends mainly on physics. Deep particle flux, on the other hand, is affected to a large extent by b. Optimisation reduces the spread induced by changing either biogeochemistry or physics to about 50% percent for fixed nitrogen loss and for primary production (compare Δ Opt with Δ Circ or Δ Par). In contrast, there is no such reduction in model variability for export production (which is mainly determined by circulation) or deep particle flux (which is mainly
500   determined by b).

4   Discussion

4.1   Why do different circulations require different parameters?

As we have seen in section 3.2, three optimal parameters depend significantly on three unique diagnostics that result from different features of the circulation model. These diagnostics are related to the northern North Atlantic and the Southern
505   Ocean; the ETP seems to play a lesser role.

The strong correlation of $R_{-O2:P}$ with the area of deep mixing clearly confirms that these two model properties (physics and biogeochemistry) regulate global ocean oxygen distribution and inventory in concert. The larger the area of deep mixing, the more oxygen can – or should – be respired in the model, in order to match observed oxygen concentrations. Despite optimisation, the optimal models show an average oxygen concentration of ≈ 145 mmol m^{-3} in the NPDW (Figure 9), which
510   is higher than the observed value of 125 mmol m^{-3}. Given that the models differ strongly in their physical properties, this residual mismatch of all optimal models especially in the NPDW may point towards a deficiency of the biogeochemical model. For example, the spatially homogeneous, and thus inflexible, particle flux profile may not be adequate to simulate the very dynamic response of ecosystem dynamics and particle size structure to regionally variable mixing and nutrient supply (e.g., Guidi et al., 2015; Marsay et al., 2015). Here, a more flexible model resulting in variable sinking speeds of particles (e.g.,
515   Gehlen et al., 2006; Niemeyer et al., 2019), or, more generally, spatially flexible remineralisation length scales (e.g., Weber et al., 2016), might be of advantage.

The parameter determining particle flux, b, correlates with the ideal age of NADW (Figure 6). This physical diagnostic comprises several aspects of circulation: a large area of deep mixing in the northern North Atlantic supplies this region with "young" waters. At the same time, a strong Atlantic Meridional Overturning circulation (AMOC) and/or confined Deep West-
520   ern Boundary current (DWBC) can more quickly export the preformed properties to the southern parts of the basin. Depending on the parameterisation of mixing and other physical processes, biogeochemical tracers are distributed more efficiently in the

16

west-east direction, or mixed with deeper waters. When these combined properties of a model cause a long residence time of waters in the NADW, the resulting age of this water mass will be quite high, and vice versa.

The circulations applied in our study vary with regard to several aspects in this region: in contrast to all other models the circulation of MIT28 has a large area of deep mixing in the north, including the Labrador Sea (see Figure 2). There is also a quite strong and wide transport of young waters in the western part of the North Atlantic via the DWBC, as indicated by the southward propagation of relatively young waters between 1500-2500 m in the western part of the basin (Figure 3). At the same time there is a strong lateral spreading of these young waters from the western part (see also Dutay et al., 2002). All processes combined lead to relatively young average age of NADW in MIT28. ECCO, in contrast, shows only comparatively shallow mixing in the northern North Atlantic (Figure 2), little southward transport of these waters in the DWBC (Wunsch and Heimbach, 2006), and a large extent of older waters in the Eastern Tropical Atlantic (Figure 3). This circulation is characterised by the oldest average age of NADW.

Our optimisations suggest that models with old NADW adjust to a large b, or slow particle sinking (for example, $b = 1.46$ of ECCO*). As ideal age becomes younger, optimal b decreases. Why is this the case? So far, we can not attribute this exclusively to the area of deep mixing (Table 2), or, for example, to the overturning circulation, which is quite low in ECCO (between 13-14 Sv; Wunsch and Heimbach, 2006), moderate (17.5 and 18.5 Sv) in U17.5 and UHigh, and high (20 Sv) in U20. Instead, the average age of NADW, and resulting optimal b, likely reflects the combined effects of various model physical and biogeochemical parameterisations: the adjustment of b to smaller values decreases shallow production and remineralisation (see Figure 10). It also increases export of phosphorus to deep waters, and finally to the NPDW (see Figure 8). Circulation models with high physical turnover in the NADW (e.g., UHigh), as indicated by young NADW, can more easily resupply nutrients to surface waters, and therefore balance the loss due to particle sinking in this region.

As shown in Figure 6, the maximum potential rate of nitrogen fixation μ_{NFix} increases with area of surface waters defined $26.5 \leq \sigma_\Theta < 27.5$ in the Southern Ocean, i.e., waters reflecting the formation and ventilation of AAIW and SAMW. A broad view of large scale circulation and the spatial separation of fixed nitrogen loss and gain helps to understand the adjustment of maximum nitrogen fixation rate to physical processes in the Southern Ocean. Denitrification is a very localised process, occurring mainly in the Eastern Tropical Pacific (ETP) (Figure S6). On the other hand, simulated nitrogen fixation takes place throughout large parts of the tropical and subtropical regions of the Pacific, Atlantic and Indian Ocean. Even though nitrogen fixation in the Atlantic accounts only for a fraction of global fixed N gain (see also Marconi et al., 2017, for evidence from observations), in our models it nevertheless contributes to the stabilisation of the global fixed-nitrogen budget. A very negative N*, as arises from denitrification in the ETP, has to arrive in the Atlantic for nitrogen fixation to trigger a competitive advantage of nitrogen fixation. These two regions in the Atlantic and Pacific Ocean are connected through large scale circulation, which transports N* on centennial to millennial time scales from areas of fixed nitrogen loss to areas of fixed nitrogen gain. When passing the CDW of the Southern Ocean, these waters can act as a "mixer of deep waters with distinct isotopic signatures and nutrient stoichiometry" (Tuerena et al., 2015); the resulting mixed properties provide the source of AAIW and SAMW. The subsequent transport via AAIW and SAMW then can trigger nitrogen fixation, e.g., in the Atlantic, and balance the nitrate deficit arising mainly in the Pacific Ocean.

17

As shown in Figure 5, the nitrate deficit N^* differs among the different models. For example, MIT28* exports water with an N-deficit of ≈ 3 mmol m^{-3} from the Southern Ocean to the low latitudes (promoting nitrogen fixation). This model adjusts to a low rate of maximum potential nitrogen fixation of 1.19 μmol m^{-3} d^{-1}. On the other hand, UHigh* simulates SAMW and AAIW that contain a lower N-deficit of ≈ 2 mmol m^{-3}, which – depending on phosphate availability – will result in lower nitrogen fixation. The optimal high parameter of UHigh* of almost 3 μmol m^{-3} d^{-1} can partially compensate for this. The effect of N^* is, however, not consistent across all optimal models: U17.5* also shows a small nitrate deficit in this region, but has a still relatively low maximum nitrogen fixation rate. Here, other effects might play a role, such as a stronger ventilation and consequently younger waters in the ETP (Figure S2), which induce a smaller OMZ (Table 1), less denitrification in this region (Figure S6), and thus a lower nitrate deficit in this area, that is to be eventually balanced by nitrogen fixation.

In our analysis we have combined outcrop area of two water masses, SAMW and AAIW into one single diagnostic. Separating the impact of the two water masses on this parameter, we find that the correlation of μ_{NFix} with SAMW outcrop area (when defined as by $26.5 \leq \sigma_\Theta < 27.0$) is less significant ($r = 0.81$) than that with AAIW ($27.0 \leq \sigma_\Theta < 27.5; r = 0.88$), which is somehow in contrast to the findings by Palter et al. (2010). Their model experiments showed that the largest fraction (between 45 to 68%, depending on model configuration) of water volume at the surface between 30°S and 30°N stems from SAMW, highlighting the role of this water mass for nutrient supply in the tropics and subtropics. A possible explanation for this difference between our results and the results by Palter et al. (2010) could be the slightly different definition of water masses. Further, in our models waters denser than $\sigma_\Theta = 26.5$ are influenced by the small nitrate deficit of surface waters in the subtropical southern hemisphere (Figure 5), which moderates the signal arising from denitrification in the Pacific.

4.2   Can optimisation help to improve model performance?

As shown in section 3.2, each circulation requires its own set of parameters for an optimal fit to dissolved inorganic tracers. Optimisation facilitates the identification of these constants; on the other hand, it requires many model evaluations, so this approach is prohibitive for models of high resolution because of computational constraints. It would be desirable to optimise a biogeochemical model in a computationally cheaper circulation and then transfer the optimal parameters to a different model that includes more physical details, but is computationally more expensive. However, as shown in Sections 3.3, 3.5 and 3.6, model performance can deteriorate when simulated with non-optimal parameters, and result in a considerable spread of the independent diagnostics. Is there any advantage of calibrating biogeochemical models in these rather coarse scale, simplified circulations, if the parameters are to be transferred to a different circulation? To answer this question, we here discuss the model variability before the background of earlier model studies and observed estimates.

The mean diagnostic across all 25 model experiments (Mean(All) of Table 3) differs only slightly from the mean across only the optimal model configurations (Mean(Opt)), and is close to observed quantities (Table 3), in agreement with Kriest et al. (2017) and Kriest (2017), who found that optimisation against global nutrients and oxygen can help to improve global model performance. Further, the overall maximum variation across all 25 experiments (ΔAll of Table 3) is usually less than 50% of that found by Bopp et al. (2013), who examined seven global models of different biogeochemical structure and circulation

590 for their global average oxygen, OMZ volume, primary and export production. This indicates that optimisation can help to improve model performance and reduce its uncertainty, even if parameters were optimised in a different circulation.

4.2.1 Model uncertainty, oxygen inventory and OMZ volume

As shown in section 3.5 changes in circulation and biogeochemical parameters affect model performance with respect to global average oxygen about equally, resulting in an overall variation that is less than 25% of the observed value (ΔAll of Table 3). In

595 contrast, the global OMZ volume shows a large response to variations in circulation and model parameters, and varies by more than 100% (ΔAll of Table 3) to 200% (Bopp et al., 2013) of the observed volume. To our knowledge, no global model study exists so far that systematically distinguishes between the effects of circulation and biogeochemistry on global OMZ volume; our study suggests that both are equally important. Even the range across optimal models (ΔOpt) is still quite large, which can be explained by the fact that the target of optimisation (the RMSE to nutrients and oxygen, as of Equation 1) is only weakly

600 correlated to the fit to OMZs (Sauerland et al., 2019). Application of a revised misfit function, or multi-objective optimisation as presented by Sauerland et al. (2019), can help to better constrain the relevant model parameters, and better represent OMZs. Nevertheless, the mean of all models in our study deviates by less than 5% from the observed mean. Obviously, a good representation of nutrients and oxygen can improve the fit to OMZs to some extent.

However, many global circulation models suffer from a deficient representation of physical processes in the tropics and

605 subtropics, for example in their representation of the equatorial undercurrent (Dietze and Loeptien, 2013), or from inadequate ventilation from the Southern Ocean and North Pacific (see Cabre et al., 2015, and citations therein). The first problem can possibly be cured by a higher resolution, which leads to a more realistic OMZ ventilation by equatorial and off-equatorial undercurrents (Duteil et al., 2014). Parameterisation of intermediate jets (Getzlaff and Dietze, 2013) can also lead to a better agreement of the models. Tuning of biogeochemistry before the background of inadequate physics could compensate for the

610 physical errors, but also result in misleading model parameterisations ('overtuning'), with potential consequences for future projections (Löptien and Dietze, 2019). Given the yet unexplored structural and parameter sensitivity of models employed in global assessments, and the large error with respect to OMZ volume and expansion (Table 3; Cocco et al., 2013; Bopp et al., 2013; Cabre et al., 2015), a careful analysis of different error sources (physical and biogeochemical) can help to determine the reasons for model divergence. The study presented here can serve as a first step towards this.

615 ### 4.2.2 Model uncertainty and global biogeochemical fluxes

The loss of fixed nitrogen through pelagic denitrification is tightly linked to OMZs, and therefore also almost equally influenced by circulation and biogeochemical parameters. Our average optimal model estimates are in agreement with recent estimates by Eugster and Gruber (2012) and Somes et al. (2013). The variation due to biogeochemical parameters is lower than found by Somes et al. (2013), who varied the nitrate threshold for denitrification from 20 to 32 mmol m^{-3}, i.e. a wider range than

620 identified by our objective parameter calibration (see Table 1).

Average global primary production of the models lies well within the range of observed estimates (Carr et al., 2006), and depends slightly more on circulation than on biogeochemical parameters, similar to the results obtained in the sensitivity study

19

by Schmittner et al. (2005). That study included a wide range of sinking and mixing parameterisations, which might explain the larger variation compared to our results. Applying three different circulations to one biogeochemical model, Seferian et al. (2013) found a spread of primary production which comparable to our experiments.

Quite many global model studies analysed the impact of circulation on global export production. Najjar et al. (2007) found an effect of circulation more than ten times larger than in the present study, which can likely be explained by the nutrient-restoring approach applied in their simple model. Using a more complex biogeochemical model, export production in the study by Seferian et al. (2013) varied only by ≈ 3 Pg C y-1, which is closer to the effects of circulation found in our study. The large effects observed by Schmittner et al. (2005) can again likely be ascribed to the wide range of parameterisations tested.

Using a biogeochemical model similar to the one applied by Najjar et al. (2007), Kwon et al. (2009) found an increase in export production of ≈ 5 Pg C y^{-1} when increasing b from 1.1 to 1.4 (about the range tested in our study). Again, this can possibly be explained by the nutrient restoring approach, which does not account for the interplay between particle export and remineralisation in surface and subsurface layers.

Therefore, our model experiments show a lower sensitivity of global export production on biogeochemical parameters or circulation than the previous studies. Some part of this difference could be explained by the large variation in physical model setup (in the study by Schmittner et al., 2005), or by the very different structure of the biogeochemical model applied (Najjar et al., 2007; Kwon et al., 2009). The large sensitivity of export production in the study by Najjar et al. (2007) also reflects on the range of deep particle flux (as diagnosed from export production times 0.052), which is almost ten times higher than in the present study.

Our experiments suggest that even though circulation does play a large role for export production, deep particle flux is mainly determined by parameter b. As a consequence, the sensitivity of export production to circulation noted by Najjar et al. (2007) and in the present study does not necessarily imply an equally large sensitivity of deep particle flux (and resulting remineralisation and deep oxygen consumption) on physical model features, which is somehow in contrast to the conclusions drawn by Najjar et al. (2007).

4.3 The effect of model complexity

To summarise, in all cases studied here biogeochemical parameter optimisation can help to narrow down the model uncertainty induced by circulation and biogeochemical parameters. Even if parameters optimised in one circulation are later transferred to a different circulation the resulting spread is mostly around 50% that of the model intercomparison presented by Bopp et al. (2013). However, that study included models that diverged not only in physics, but also in the biogeochemical structure, which might introduce another source of variability. To have a closer look at this we finally contrast the results of model MOPS, which we consider as a model of intermediate complexity, with an equivalent optimisation of a much simpler model "RetroMOPS" presented by Kriest (2017).

RetroMOPS is a four-component model that simulates only phosphate, nitrate, oxygen and dissolved organic phosphorus, but includes the same structural form for particle flux and remineralisation as model MOPS. When coupled to MIT28, and optimised against the same data set and misfit function presented above, the performance and global fluxes of RetroMOPS

are very similar to MOPS (in the same circulation; Table 3 and Kriest, 2017)). For primary production the difference between RetroMOPS and MOPS is about as large as when MOPS is simulated with different parameter sets within a given circulation. One reason for this is the fact that the optimisation of RetroMOPS by Kriest (2017) aimed only at the parameters related to particle sinking and remineralisation, but not at parameters related to phytoplankton growth and loss terms; an additional optimisation of these parameters may likely have produced smaller differences.

Therefore, after optimisation a simple model can perform quite well with respect to large-scale biogeochemical quantities, in agreement with earlier findings (Kriest et al., 2012; Kwiatkowski et al., 2014; Galbraith et al., 2015), illustrating the benefit of parameter optimisation: optimisation allows for a "fair" comparison of models of different complexity (after each model has been tuned to match some desired quantity best); it can therefore also help to decide about the necessary level of model complexity.

5 Conclusions

Optimisation of a global biogeochemical ocean model coupled to five different circulations achieved a good fit to observed nutrients and oxygen with partly different biogeochemical parameters. We identified three parameters that depend significantly on characteristic features of circulation, as summarised in Figure 13. Areas of deep ventilation in the North Atlantic and in the Southern Ocean determine how much oxygen is supplied to the ocean via air-sea gas exchange and subsequent mixing. As a consequence, optimisation of the model in circulations with vigorous ventilation triggers a high oxygen demand of remineralisation during optimisation. Fast turnover and mixing of NADW, as expressed through low ideal age, affects the parameter responsible for the timescale and vertical extent of remineralisation of sinking particles. Here, models characterised by relatively young waters in the NADW adjust to deeper sinking and remineralisation, with consequences for the large scale distribution of phosphate. Finally, the combined outcrop area of SAMW and AAIW determines the optimal maximum rate of nitrogen fixation. This may be explained with the role of the Southern Ocean as "mixer" for waters of different origin and nitrogen deficit (Tuerena et al., 2015). The extent and properties of waters originating from this region (and their fixed nitrogen deficit), in conjunction with denitrification within the OMZs then set the stage for the competitive advantage of cyanobacteria in tropical and subtropical waters. Models with a small outcrop area of these waters benefit from a low maximum rate of nitrogen fixation, and vice versa. In conclusion, when aiming for a good fit to large-scale biogeochemical quantities, key properties of the underlying circulation model should be considered; depending on region and tracer of interest, these key properties may, however, be different from those of our study.

Cross-validation experiments showed that the optimal parameters could be swapped between the different circulations to a limited extent. Parameters affect biogeochemical model performance in different ways: while the stoichiometric demand of oxygen during remineralisation affects, for example, the ocean oxygen inventory, the particle flux parameter b determines the large-scale distribution of nutrients, in line with earlier studies (Bacastow and Maier-Reimer, 1991; Kwon and Primeau, 2006; Kriest et al., 2012).

21

Compared to other intercomparisons of global coupled models tuned more subjectively, our overall variation of biogeochemical key properties is at least 50% smaller, with different contributions from circulation and biogeochemical parameters. For example, export production seems to be mainly determined by circulation, while deep particle flux is determined almost entirely by the particle flux parameter b. Other biogeochemical diagnostics are affected more or less equally by circulation and biogeochemical parameters. Finally, OMZ volume is very sensitive to changes in circulation and biogeochemical parameters, and varies most strongly across all model experiments.

However, models considered in global intercomparisons usually differ also in their biogeochemical structure and complexity. Our experiments suggest that after optimisation the differences due to model structure are much smaller than those due to model parameters or circulation. This indicates that a simpler model can perform as well as a more complex model (with respect to the metrics and diagnostics applied here), similar to the results obtained by Kriest et al. (2012), Kwiatkowski et al. (2014) and Galbraith et al. (2015). It also illustrates how biogeochemical parameter optimisation can aid model development: whenever new components or parameterisations are introduced to a global model, this new model has to be tuned in order to match observations on global or regional scales. Often, the choice of appropriate parameters is not easy, and requires extensive testing and sensitivity analysis. Automatic (algorithmic) optimisation can make calibration more efficient, simplifying the search for a good model match to observed quantities. For a given misfit function it can also support decisions about the necessary level of model complexity.

Therefore, our experiments suggest that global biogeochemical ocean models benefit from optimisation, even if this was carried out in a circulation differing from that of the "target" circulation. However, to date there is no guarantee that a model showing a good fit to observed quantities in steady state (i.e., when simulated in a preindustrial or present day, climatological physical forcing) will exhibit the correct response when applied to transient scenarios reflecting future climate change. As shown here and in earlier studies (Cocco et al., 2013; Cabre et al., 2015; Löptien and Dietze, 2019), OMZs seem to be particularly sensitive to both biogeochemical and physical parameters. Accounting for the match between simulated and observed OMZs during optimisation can reduce the model spread in steady state (Sauerland et al., 2019). Extending the simulation of optimal models to future states could then inform us about their sensitivity to changes in circulation and forcing, and may provide a better constraint on their uncertainties. Thus, the study presented here serves as a first step to unravel the uncertainties associated with the divergence of global biogeochemical model performance and uncertainty.

23

References

Anderson, L. and Sarmiento, J.: Redfield ratios of remineralization determined by nutrient data analysis, Global Biogeochem. Cy., 8, 65–80, https://doi.org/10.1029/93GB03318, 1994.

Bacastow, R. and Maier-Reimer, E.: Dissolved organic carbon in modeling oceanic new production, Global Biogeochem. Cy., 5, 71–85, https://doi.org/10.1029/91GB00015, 1991.

Bopp, L., Resplandy, L., Orr, J., Doney, S., Dunne, J., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Seferian, R., Tjiputra, J., and Vichi, M.: 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, 2013.

Cabre, A., Marinov, I., Bernadello, R., and Bianchi, D.: Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends, Biogeosciences, 12, 5429–5454, https://doi.org/10.5194/bg-12-5429-2015, 2015.

Carr, M.-E., Friedrichs, M., Schmeltz, M., Aitac, M., Antoine, D., Arrigo, K., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., , Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Quere, C. L., Lohrenz, S., Marra, J., lino, F. M., Moore, K., Morel, A., Reddy, T., J.Ryan, Scardi, M., T.Smyth, Turpie, K., Tilstone, G., Waters, K., and Yamanaka, Y.: A comparison of global estimates of marine primary production from ocean color, Deep-Sea Res. Pt. II, 53, 741–770, https://doi.org/10.1016/j.dsr2.2006.01.028, 2006.

Cocco, V., Joos, F., Steinacher, M., Frölicher, T., Bopp, L., Dunne, J., Gehlen, M., Heinze, C., Orr, J., Oschlies, A., Schneider, B., Segschneider, J., and Tjiputra, J.: Oxygen and indicators of stress for marine life in multi-model global warming projections, Biogeosciences, 10, 1849–1868, https://doi.org/10.5194/bg-10-1849-2013, 2013.

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

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

Dietze, H. and Loeptien, U.: Revisiting "nutrient trapping" in global coupled biogeochemical ocean circulation models, Global Biogeochem. Cy., 27, 265–284, https://doi.org/10.1002/gbc.20029, 2013.

Dong, S., Sprintall, J., Gille, S., and Talley, L.: Southern Ocean mixed-layer depth from Argo float profiles, Journal Of Geophysical Research, 113, https://doi.org/10.1029/2006JC004051, 2008.

Dunne, J. P., Sarmiento, J. L., and Gnanadesikan, A.: A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor, Global Biogeochem. Cy., 21, https://doi.org/10.1029/2006GB002907, 2007.

Dutay, J., Bullister, J. L., Doney, S. C., Orr, J. C., Najjar, R., Caldeira, K., Campin, J., Drange, H., Follows, M., Gao, Y., Gruber, N., Hecht, M. W., Ishida, A., Joos, F., Lindsay, K., Madec, G., Maier-Reimer, E., Marshall, J. C., Matear, R. J., Monfray, P., Mouchet, A., Plattner, G., Sarmiento, J., Schlitzer, R., Slater, R., Totterdell, I. J., Weirig, M., Yamanaka, Y., and Yool, A.: Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models, Ocean Modell., 4, 89–120, https://doi.org/10.1016/S1463-5003(01)00013-0, 2002.

Duteil, O. and Oschlies, A.: Sensitivity of simulated extent and future evolution of marine suboxia to mixing intensity, Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL046877, 2011.

Duteil, O., Schwarzkopf, F. U., Böning, C. W., and Oschlies, A.: Major role of equatorial current system in setting oxygen levels in the eastern tropical Atlantic Ocean: A high resolution model study, Geophys. Res. Lett., 41, 2033–2040, https://doi.org/10.1002/2013GL058888, 2014.

Eugster, O. and Gruber, N.: A probabilistic estimate of global marine N-fixation and denitrification, Global Biogeochem. Cy., 26, https://doi.org/10.1029/2012GB004300, 2012.

Galbraith, E. D., Dunne, J. P., Gnanadesikan, A., Slater, R. D., Sarmiento, J. L., Dufour, C. O., de Souza, G. F., Bianchi, D., Claret, M., Rodgers, K. B., and Marvasti, S. S.: Complex functionality with minimal computation: Promise and pitfalls of reduced-tracer ocean biogeochemistry models, Jour. Adv. Model. Earth Systems, 7, 2012–2028, https://doi.org/10.1002/2015MS000463, 2015.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., and Antonov, J. I.: World Ocean Atlas 2005, Vol. 4: Nutrients (phosphate, nitrate, silicate), in: NOAA Atlas NESDIS 64, edited by Levitus, S., U.S. Government Printing Office, Wash.,D.C., 2006a.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., and Antonov, J. I.: World Ocean Atlas 2005, Vol. 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, in: NOAA Atlas NESDIS 63, edited by Levitus, S., U.S. Government Printing Office, Wash.,D.C., 2006b.

Gehlen, M., Bopp, L., Emprin, N., Aumont, O., Heinze, C., and Ragueneau, O.: Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model, Biogeosciences, 3, 521–537, https://doi.org/10.5194/bg-3-521-2006, 2006.

Getzlaff, J. and Dietze, H.: Effects of increased isopycnal diffusivity mimicking the unresolved equatorial intermediate current system in an earth system climate model, Geophys. Res. Lett., 40, 2166–2170, https://doi.org/10.1002/grl.50419, 2013.

Griffies, S., Harrison, M., Pacanowski, R., and Rosati, A.: A Technical Guide to MOM4, GFDL Ocean Group Technical Report 5, NOAA/Geophysical Fluid Dynamics Laboratory, 2008.

Guidi, L., Legendre, L., Reygondeau, G., Uitz, J., Stemmann, L., and Henson, S. A.: A new look at ocean carbon remineralization for estimating deepwater sequestration, Global Biogeochemical Cycles, 29, 1044–1059, https://doi.org/10.1002/2014GB005063, 2015.

Hansen, N.: The CMA evolution strategy: a comparing review, in: Towards a new evolutionary computation. Advances on estimation of distribution algorithms, edited by Lozano, J. A., Larranaga, P., Inza, I., and Bengoetxea, E., pp. 75–102, Springer, 2006.

Hansen, N. and Ostermeier, A.: Completely Derandomized Self-Adaptation in Evolution Strategies, Evolutionary Computation, 9, 159–195, https://doi.org/10.1162/106365601750190398, 2001.

Holland, W., Chow, J., and Bryan, F.: Application of a third-order upwind scheme in the NCAR Ocean Model, Journal Of Climate, 11, 1487–1493, https://doi.org/10.1175/1520-0442, 1998.

Holzer, M., Primeau, F., DeVries, T., and Matear, R.: The Southern Ocean silicon trap: Data-constrained estimates of regenerated silicic acid, trapping efficiencies, and global transport paths, Journal Of Geophysical Research, 119, 313–331, https://doi.org/10.1002/2013JC009356, 2014.

Honjo, S., Manganini, S. J., Krishfield, R. A., and Francois, R.: Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983, Prog. Oceanogr., 76, 217–285, https://doi.org/10.1016/j.pocean.2007.11.003, 2008.

Ilyina, T., Six, K., Segschneider, J., Maier-Reimer, E., Li, H., and nez Riboni, I. N.: Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, Jour. Adv. Model. Earth Systems, 5, 1–29, https://doi.org/10.1029/2012MS000178, 2013.

Iudicone, D., Rodgers, K., Stendardo, I., Aumont, O., Madec, G., Bopp, L., Mangoni, O., and Ribera d'Alcala, M.: Water masses as a unifying framework for understanding the Southern Ocean Carbon Cycle, Biogeosciences, 8, 1031–1052, https://doi.org/10.5194/bg-8-1031-2011, 2011.

Khatiwala, S.: A computational framework for simulation of biogeochemical tracers in the ocean, Global Biogeochem. Cy., 21, GB3001, https://doi.org/10.1029/2007GB002923, 2007.

Khatiwala, S.: Fast spin up of ocean biogeochemical models using matrix-free Newton-Krylov, Ocean Modell., 23, 121–129, https://doi.org/10.1016/j.ocemod.2008.05.002, 2008.

Khatiwala, S.: Transport Matrix Method software for ocean biogeochemical simulations, available at: https://doi.org/10.5281/zenodo.1246300 (last access: 10 February 2020), 2018.

Khatiwala, S., Primeau, F., and Holzer, M.: Ventilation of the deep ocean constrained with tracer observations, Earth Planet. Sci. Lett., 325–326, 116–125, https://doi.org/10.1016/j.epsl.2012.01.038, 2012.

Kriest, I.: Calibration of a simple and a complex model of global marine biogeochemistry, Biogeosciences, 14, 4965–4984, https://doi.org/10.5194/bg-14-4965-2017, 2017.

Kriest, I. and Oschlies, A.: On the treatment of particulate organic matter sinking in large-scale models of marine biogeochemical cycles, Biogeosciences, 5, 55–72, https://doi.org/10.5194/bg-5-55-2008, 2008.

Kriest, I. and Oschlies, A.: Numerical effects on organic-matter sedimentation and remineralization in biogeochemical ocean models, Ocean Modell., 39, 275–283, https://doi.org/10.1016/j.ocemod.2011.05.001, 2011.

Kriest, I. and Oschlies, A.: MOPS-1.0: towards a model for the regulation of the global oceanic nitrogen budget by marine biogeochemical processes, Geosci. Model Dev., 8, 2929–2957, https://doi.org/10.5194/gmd-8-2929-2015, www.geosci-model-dev.net/8/2929/2015/, 2015.

Kriest, I., Oschlies, A., and Khatiwala, S.: Sensitivity analysis of simple global marine biogeochemical models, Global Biogeochem. Cy., 26, https://doi.org/10.1029/2011GB004072, 2012.

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

Kvale, K., Khatiwala, S., Dietze, H., Kriest, I., and Oschlies, A.: Evaluation of the Transport Matrix Method for simulation of ocean biogeochemical tracers, Geosci. Model Dev., 10, 2425–2445, https://doi.org/10.5194/gmd-10-2425-2017, 2017.

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschoen, M., Enright, C., Halloran, P. R., Le Quere, C., de Mora, L., Racault, M. F., Sinha, B., Totterdell, I. J., and Cox, P. M.: iMarNet: an ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework, Biogeosciences, 11, 7291–7304, https://doi.org/10.5194/bg-11-7291-2014, 2014.

Kwon, E. Y. and Primeau, F.: Optimization and sensitivity study of a biogeochemistry ocean model using an implicit solver and in situ phosphate data, Global Biogeochem. Cy., 20, https://doi.org/10.1029/2005GB002631, 2006.

Kwon, E. Y., Primeau, F., and Sarmiento, J.: The impact of remineralization depth on the air-sea carbon balance, Nature Geoscience, 2, https://doi.org/10.1038/NGEO612, 2009.

Letscher, R., Moore, J. K., Teng, Y.-C., and Primeau, F.: Variable C : N : P stoichiometry of dissolved organic matter cycling in the Community Earth System Model, Biogeosciences, 12, 209–221, https://doi.org/10.5194/bg-12-209-2015, 2015.

835 Li, X. and Primeau, F.: A fast Newton–Krylov solver for seasonally varying global ocean biogeochemistry models, Ocean Modell., 23, 13–20, https://doi.org/10.1016/j.ocemod.2008.03.001, 2008.

Löptien, U. and Dietze, H.: Reciprocal bias compensation and ensuing uncertainties in model-based climate projections: pelagic biogeochemistry versus ocean mixing, Biogeosciences, 16, 1865–1881, https://doi.org/10.5194/bg-16-1865-2019, 2019.

Lutz, M., Caldeira, K., Dunbar, R., and Behrenfeld, M. J.: Seasonal rhythms of net primary production and particulate organic carbon flux
840 to depth describe biological pump efficiency in the global ocean, J. Geophys. Res., 113, C10011, https://doi.org/10.1029/2006JC003706, 2007.

Marconi, D., Sigman, D. M., Casciotti, K., Campbell, E., Weigand, M., Fawcett, S., Knapp, A., Rafter, P., Ward, B., and Haug, G.: Tropical Dominance of N_2 Fixation in the North Atlantic Ocean, Glob. Biogeochem. Cyc., 31, 1608–1623, https://doi.org/10.1002/2016GB005613, 2017.

845 Marsay, C. M., Sanders, R., Henson, S., Pabortsava, K., Achterberg, E., and Lampitt, R.: Attenuation of sinking particulate organic carbon flux through the mesopelagic ocean, Proc. Nat. Acad. Sci., 112, 1089–1094, https://doi.org/10.1073/pnas.1415311112, 2015.

Marshall, J., Adcroft, A., Hill, C., Perelman, L., and Heisey, C.: A finite-volume, incompressible Navier-Stokes model for studies of the ocean on parallel computers, J. Geophys. Res., 102, 5733–5752, https://doi.org/10.1029/96JC02775, 1997.

Matsumoto, K., Sarmiento, J. L., Key, R. M., Aumont, O., Bullister, J. L., Caldeira, K., Campin, J., Doney, S. C., Drange, H., Dutay, J., Fol-
850 lows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Lindsay, K., Maier-Reimer, E., Marshall, J. C., Matear, R. J., Monfray, P., Mouchet, A., Najjar, R., Plattner, G., Schlitzer, R., Slater, R., Swathi, P. S., Totterdell, I. J., Weirig, M., Yamanaka, Y., Yool, A., and Orr, J. C.: Evaluation of ocean carbon cycle models with data-based metrics, Geophys. Res. Lett., 31, https://doi.org/10.1029/2003GL018970, 2004.

Najjar, R. G., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S. C., Dutay, J.-C., Follows, M., Gruber, N., Joos, F., Lindsay, K.,
855 Maier-Reimer, E., Matear, R., Matsumoto, K., Monfray, P., Mouchet, A., Orr, J. C., Plattner, G.-K., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Weirig, M.-F., Yamanaka, Y., and Yool, A.: 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 Biogeochem. Cy., 21, https://doi.org/10.1029/2006GB002857, 2007.

Niemeyer, D., Kriest, I., and Oschlies, A.: The effect of marine aggregate parameterisations on global biogeochemical model performance,
860 Biogeosciences Disc., p. in review, https://doi.org/10.5194/bg-2019-122, 2019.

Palter, J., Sarmiento, J., Gnanadesikan, A., Simeon, J., and Slater, R.: Fueling export production: nutrient return pathways from the deep ocean and their dependence on the Meridional Overturning Circulation, Biogeosciences, 7, 3549–3568, https://doi.org/10.5194/bg-7-3549-2010, 2010.

Paulmier, A., Kriest, I., and Oschlies, A.: Stoichiometries of remineralisation and denitrification in global biogeochemical ocean model,
865 Biogeosciences, 6, 923–935, https://doi.org/10.5194/bg-6-923-2009, 2009.

Primeau, F. and Deleersnijder, E.: On the time to tracer equilibrium in the global ocean, Ocean Sci., 5, 13–28, https://doi.org/10.5194/os-5-13-2009, 2009.

Rafter, P., DiFiore, P., and Sigman, D.: Coupled nitrate nitrogen and oxygen isotopes and organic matter remineralization in the Southern and Pacific Oceans, J. Geophys. Res., 118, 4781–4794, https://doi.org/10.1002/jgrc.20316, 2013.

870 Rafter, P. A., Sigman, D. M., Charles, C. D., Kaiser, J., and Haug, G. H.: Subsurface tropical Pacific nitrogen isotopic composition of nitrate: Biogeochemical signals and their transport, Global Biogeochem. Cy., 26, https://doi.org/10.1029/2010GB003979, 2012.

Sallee, J.-B., Speer, K., Rintoul, S., and Wijfels, S.: Southern Ocean thermocline ventilation, Journal Of Physical Oceanography, 40, 509–529, https://doi.org/doi:10.1175/2009JPO4291.1, 2010.

Sallee, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A.J.S., Bracegirdle, T.J., and Wang, Z.: Assessment of Southern Ocean mixed-layer depths in CMIP5 models: Historical bias and forcing response, Journal Of Geophysical Research, 118, 1845–1862, https://doi.org/doi:10.1002/jgrc.2015, 2013.

Sauerland, V., Kriest, I., Oschlies, A., and Srivastav, A.: Multiobjective Calibration of a Global Biogeochemical Ocean Model Against Nutrients, Oxygen, and Oxygen Minimum Zones, Jour. Adv. Model. Earth Systems, https://doi.org/10.1029/2018MS001510, 2019.

Schmittner, A., Oschlies, A., Giraud, X., Eby, M., and Simmons, H. L.: A global model of the marine ecosystem for long-term simulations: Sensitivity to ocean mixing, buoyancy forcing, particle sinking, and dissolved organic matter cycling, Global Biogeochem. Cy., 19, GB3004, https://doi.org/10.1029/2004GB002283, 2005.

Schwinger, J., Goris, N., Tjiputra, J., Kriest, I., Bentsen, M., Bethke, I., Ilicak, M., Assmann, K., and Heinze, C.: Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1), Geosci. Model Dev., 9, 2589–2622, https://doi.org/10.5194/gmd-9-2589-2016, 2016.

Seferian, R., Bopp, L., Gehlen, M., Orr, J., Ethe, C., Cadule, P., Aumont, O., Salas y Melia, D., Voldoire, A., and Madec, G.: Skill assessment of three earth system models with common marine biogeochemistry, Clim. Dyn., 40, 2549–2573, https://doi.org/10.1007/s00382-012-1362-8, 2013.

Seferian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J., Marti, O., Dunne, J. P., Christian, J., Doney, S., Ilyina, T., Lindsay, K., Halloran, P., Heinze, C., Segschneider, J., Tjiputra, J., Aumont, O., and Romanou, A.: Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment, Geosci. Model Dev., 9, 1827—1851, https://doi.org/10.5194/gmd-9-1827-2016, 2016.

Somes, C., Oschlies, A., and Schmittner, A.: Isotopic constraints in the pre-industrial oceanic nitrogen budget, Biogeosciences, 10, 5889–5910, https://doi.org/10.5194/bg-10-5889-2013, 2013.

Stammer, D., Ueyoshi, K., Köhl, A., Large, W. G., Josey, S. A., and Wunsch, C.: Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation, J. Geophys. Res., 109, https://doi.org/10.1029/2003JC002082, 2004.

Taylor, K.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Tuerena, R. E., Ganeshram, R. S., Geibert, W., Fallick, A. E., Dougans, J., Tait, A., Henley, S. F., and Woodward, E. M. S.: Nutrient cycling in the Atlantic basin: The evolution of nitrate isotope signatures in water masses, Glob. Biogeochem. Cyc., 29, 1830–1844, https://doi.org/10.1002/2015GB005164, 2015.

Weaver, A. and Eby, M.: On the numerical implementation of advection schemes for use in conjunction with various mixing parameterizations in the GFDL ocean model, Jour. Phys. Oceanogr., 27, 369–377, https://doi.org/10.1175/1520-0485, 1997.

Weaver, A., Eby, M., Wiebe, E., Bitz, C., Duffy, P., Ewen, T., Fanning, A., Holland, M., MacFadyen, A., Matthews, H., Meissner, K., Saenko, O., Schmittner, A., Wang, H., and Yoshimori, M.: The UVic Earth System Climate Model: Model description, climatology, and applications to past, present and future climates, Atmos.-Ocean, 39, 361–428, https://doi.org/10.1080/07055900.2001.9649686, 2001.

Weber, T., Cram, J., Leung, S., DeVries, T., and Deutsch, C.: Deep ocean nutrients imply large latitudinal variation in particle transfer efficiency, Proc. Natl. Acad. Sci., 113, 8606–8611, https://doi.org/10.1073/pnas.1604414113, 2016.

Wunsch, C. and Heimbach, P.: Estimated decadal changes in the North Atlantic meridional overturning circulation and heat flux 1993–2004, Jour. Phys. Oceanogr., 36, 2012–2024, https://doi.org/10.1175/JPO2957.1, 2006.

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

Table 1. General setup, performance and results of optimisations expressed as N, the number of generations (each with a population size of 10) until convergence of the optimisation algorithm, the minimum misfit J^* (the target of optimisation), bias of global average nitrate and oxygen, and OMZ volume (for OMZs defined by $O_2 < 50$ and $O_2 < 80\,\mathrm{mmol\,m^{-3}}$), optimal parameters and their uncertainties. To determine parameter uncertainty, we defined the group Ω of model simulations whose misfit J_k deviates less than 1% from the optimal misfit J^* ($J_k/J^* - 1 \leq 0.01$). For each parameter the first value gives the optimal parameter, averaged over the last generation, and the range of all individuals of Ω is given in squared brackets. See section 2.1 for more information on the setup of the offline circulations. [†] optimisation ECCO* was stopped at generation 100, when the minimum misfit varied less than 10^{-4} around the optimal value.

	MIT28*		UHigh*		U20*		U17.5*		ECCO*	
Ocean model:	MIT28		UVic		UVic		UVic		ECCO	
Horizontal resolution:	$2.8° \times 2.8°$		$3.6° \times 1.8°$		$3.6° \times 1.8°$		$3.6° \times 1.8°$		$1.0° \times 1.0°$	
Vertical resolution:	15		19		19		19		23	
Global metrics and optimisation performance:										
N	119		144		125		130		100[†]	
J^*	0.439		0.401		0.431		0.489		0.366	
Bias NO$_3$ (mmol m^{-3})	-0.27		0.68		0.30		0.52		0.28	
Bias O$_2$ (mmol m^{-3})	3.03		1.63		3.56		5.29		0.08	
Bias OMZ Vol. 50 (10^{15} m^3)	-19.0		-5.7		3.1		12.6		-23.2	
Bias OMZ Vol. 80 (10^{15} m^3)	-44.0		16.8		11.7		22.7		-40.7	
Parameters [Range]										
b	1.39	[1.3-1.5]	1.27	[1.2-1.3]	1.40	[1.3-1.5]	1.44	[1.3-1.6]	1.46	[1.4-1.5]
$R_{-\mathrm{O2:P}}$	173.7	[166-178]	161.5	[158-166]	161.5	[158-167]	167.0	[159-175]	151.1	[150-154]
μ_{NFix}	1.19	[1.1-1.8]	2.98	[2.1-3.0]	1.00	[1.0-1.4]	1.98	[1.5-3.0]	2.29	[1.5-3.0]
$DIN_{\min}$	15.8	[12.2-16.0]	16.0	[6.2-16.0]	15.9	[7.0-16.0]	16.0	[12.0-16.0]	16.0	[10.2-16.0]
K_{DIN}	32.0	[17.2-32.0]	26.5	[17.4-31.4]	31.9	[25.8-32.0]	32.0	[21.3-32.0]	23.1	[21.7-32.0]
K_{O2}	1.01	[1.0-8.5]	13.58	[2.1-14.3]	7.15	[1.0-8.0]	5.62	[1.5-16.0]	1.07	[1.0-16.0]

Table 2. Correlation coefficient r for the regression of three optimal model parameters against physical diagnostics. The outcrop area of dense waters in the northern ($> 40°$N) and southern ($> 40°$S) hemisphere is given for two different density intervals. Outcrop area for deep mixing in the North Atlantic ($> 40°$N) and Southern Ocean ($> 40°$S) is given for two different criteria of maximum deep mixing (200m and 400m). Mean water mass age is given for four different water masses. See text for further details. Significant correlations ($p < 0.05$) are denoted by bold face.

Parameter	Physical diagnostic			
	Area MLD North Atlantic		Area MLD Southern Ocean	
	200m	400m	200m	400m
$R_{-\mathrm{O2:P}}$	**0.916**	**0.950**	0.875	0.465
μ_{NFix}	-0.492	-0.405	-0.332	-0.263
b	-0.280	-0.425	-0.499	-0.210
	Area Outcrop North		Area Outcrop South	
	26.5-27.5	>27.5	26.5-27.5	>27.5
$R_{-\mathrm{O2:P}}$	-0.329	0.460	-0.641	0.792
μ_{NFix}	0.449	-0.361	**0.890**	-0.746
b	-0.287	0.050	-0.538	0.202
	Ideal Ages			
	NADW	CDW	NPDW	ETP
$R_{-\mathrm{O2:P}}$	-0.358	0.273	0.787	-0.266
μ_{NFix}	-0.083	-0.380	-0.857	-0.070
b	**0.914**	0.136	0.069	0.334

Table 3. Mean (across all experiments and optimal models) and range of variation of global biogeochemical model properties and fluxes across different parameter sets (circulation constant; Δ Par), and different circulations (parameters constant; Δ Circ), as well as across the five different optimal models MIT28*, ECCO*, uHigh*, U20* and U17.5* (Δ Opt). Δ Mod shows the difference between MIT28* of this study and model RetroMOPS of Kriest (2017). Observed oxygen and OMZ volume from Garcia et al. (2006b, mapped onto ECCO grid); observed global flux ranges derived from estimates by Carr et al. (2006, primary production), Dunne et al. (2007, export production), Lutz et al. (2007, export production; radiogen. calib.), Honjo et al. (2008, mean particle flux), Guidi et al. (2015, particle flux), Eugster and Gruber (2012, median fixed nitrogen loss) and Somes et al. (2013, fixed nitrogen loss of best data-constrained model).

	Mean (All)	Mean (Opt)	Δ Mod	Δ Par	Δ Circ	Δ Opt	Δ All
Global mean O_2 (observed: 174.17 mmol m^{-3})							
This study	177.2	177.6	0.1	24.1	24.3	6.8	43.0
Bopp et al. (2013)							95.0
OMZ Volume (50 mmol m^{-3}; observed: 57.0×10^{15} m^3)							
This study	54.9	52.1	0.7	39.4	55.4	39.0	67.2
Bopp et al. (2013)							212.5
OMZ Volume (80 mmol m^{-3}; observed: 119.1×10^{15} m^3)							
This study	122.3	112.8	0.6	112.0	119.8	73.3	145.9
Bopp et al. (2013)							328.9
Fixed N loss (observed estimates: 52-76 Tg N y^{-1})							
This study	67.7	67.4	1.9	35.1	33.8	17.4	44.9
Somes et al. (2013)[$]			53.6				
Primary Production (observed estimates: 40-60 Pg C y^{-1})							
This study	47.4	47.2	9.1	8.17	12.27	5.65	18.95
Schmittner et al. (2005)[§]				20.1	24.5		48.5
Seferian et al. (2013)					8.64		
Bopp et al. (2013)							47.8
Export Production (observed estimates: 4.6-9.6 Pg C y^{-1})							
This study	6.86	6.86	0.23	0.49	1.08	1.05	1.20
Schmittner et al. (2005)[§]				2.2	8.4		11.2
Najjar et al. (2007)					10		
Kwon et al. (2009)[†]				$\approx$5			
Seferian et al. (2013)					3.0		
Bopp et al. (2013)							3.2
Sedimentation (2000m) (observed estimates: 0.33-0.43 Pg C y^{-1})							
This study	0.36	0.36	0.02	0.16	0.06	0.16	0.19
Najjar et al. (2007)[‡]					0.52		

[$] experiments 1-3; [§] For ΔCirc we have omitted the experiment with $K_b = 0$ of the study by Schmittner et al. (2005), and refer only to experiments 2, 8 and 12. For ΔPar we report the difference between experiments 12 and 13, and for ΔAll the maximum spread across experiments 2 to 13. [†] We refer to Figure 2b of Kwon et al. (2009), but consider only a range of $1.1 \leq b \leq 1.4$ for ΔPar, to be comparable to our range of b. [‡] Calculated from export production $\times (75/2000)^{0.9} = 0.052$.
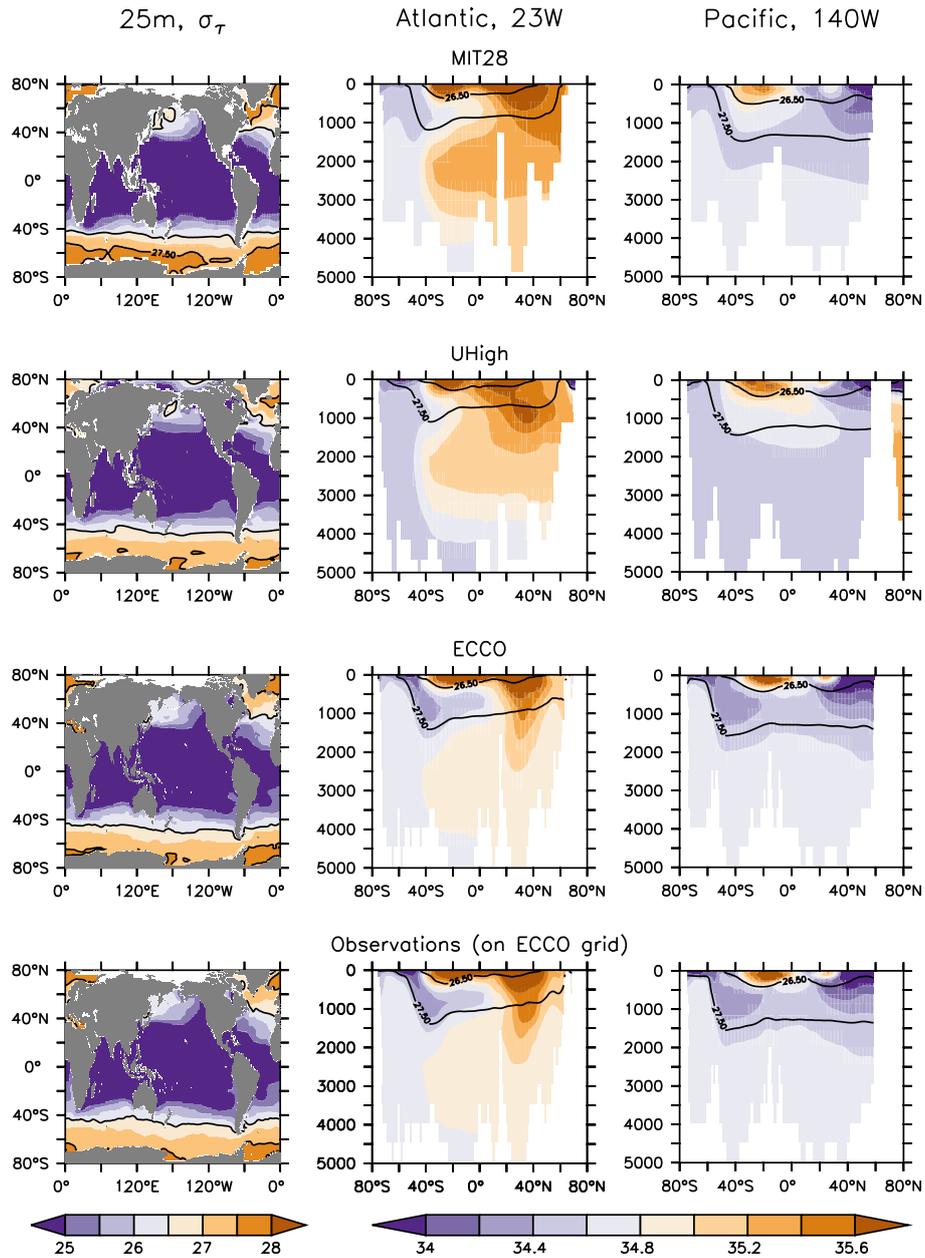
Figure 1. Density (σ_Θ) at 25 m (left panels), and salinity along sections at 23°W and 140°W (middle and right panels), of forcing from (top to bottom) MIT28, UHigh and ECCO circulation. The lower panel shows observations mapped onto ECCO geometry. Density has been derived from annual mean potential temperature and salinity. Contour lines highlight isopycnal of $\sigma_\theta = 26.5$ and $\sigma_\theta = 27.5$.
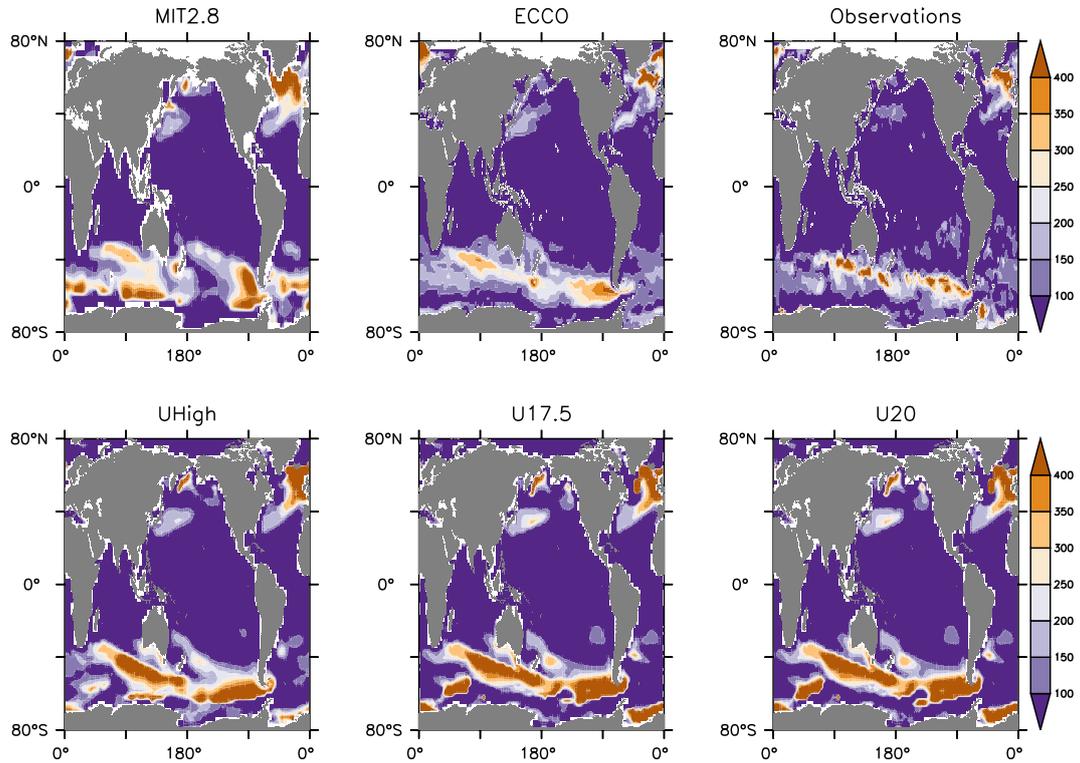
Figure 2. Annual maximum mixed-layer depths of models and observations (gridded onto ECCO model geometry). Mixed-layer depths have been determined from a density difference of $\sigma_\theta = 0.03$ (density calculated from monthly mean temperature and salinity). Square pattern in high latitudes denotes regions with an ice coverage of at least 50%.
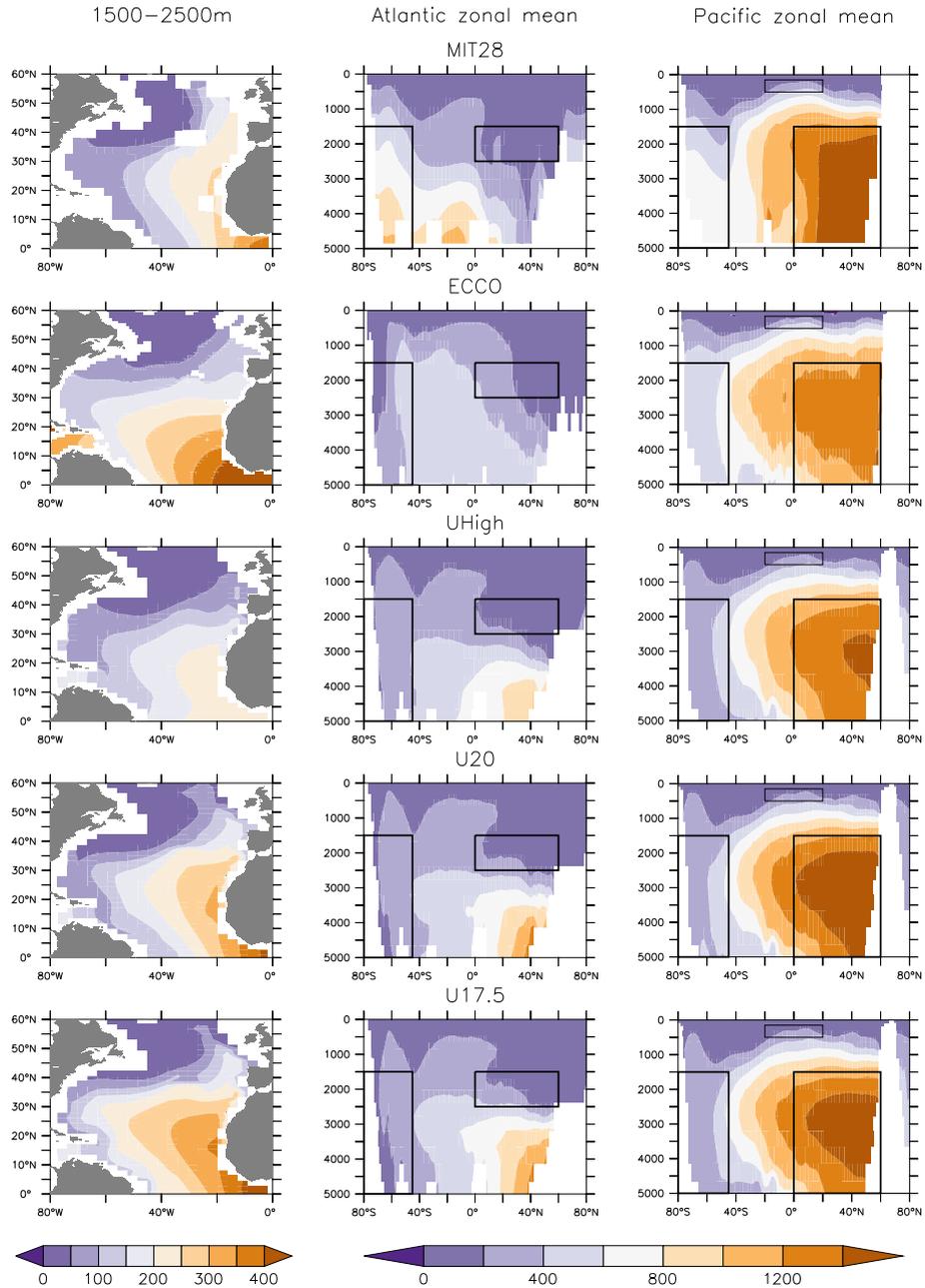
Figure 3. Simulated ideal age (years) averaged over 1500-2500 m in the North Atlantic (NADW; left panels), and as zonal mean for the Atlantic (mid panels) and the Pacific (right panels). Boxes indicate regions of NADW, NPDW, CDW and ETP (see Fig. S2 for detailed plot).
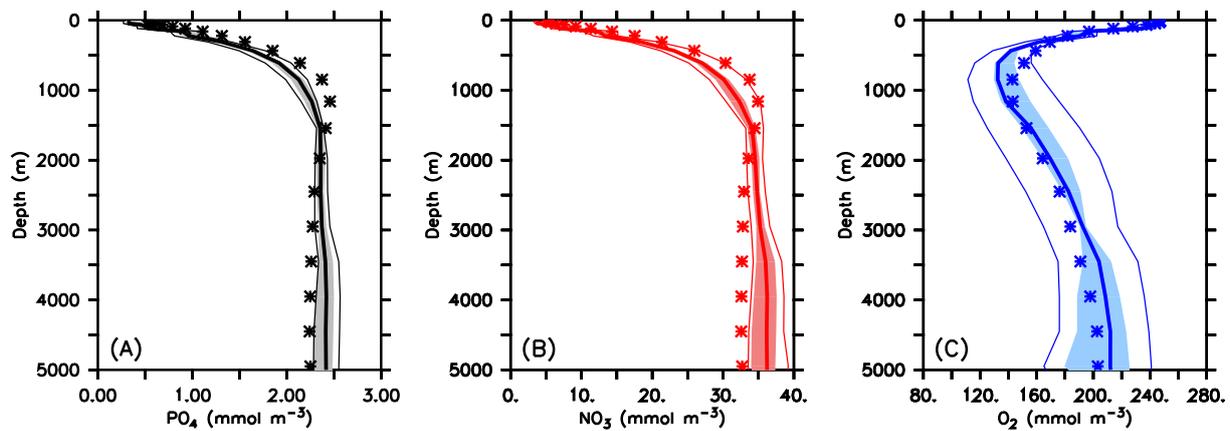
Figure 4. Global mean vertical profiles of phosphate (A), nitrate (B) and oxygen (C). Thin lines denote range across all 25 model experiments, shaded areas the range across the five optimal model experiments and the thick lines the average of optimal model experiments. Stars denote observed profiles. Prior to averaging, all model have been regridded onto the ECCO grid, using nearest neighbour filling in the vertical, and linear interpolation horizontally.
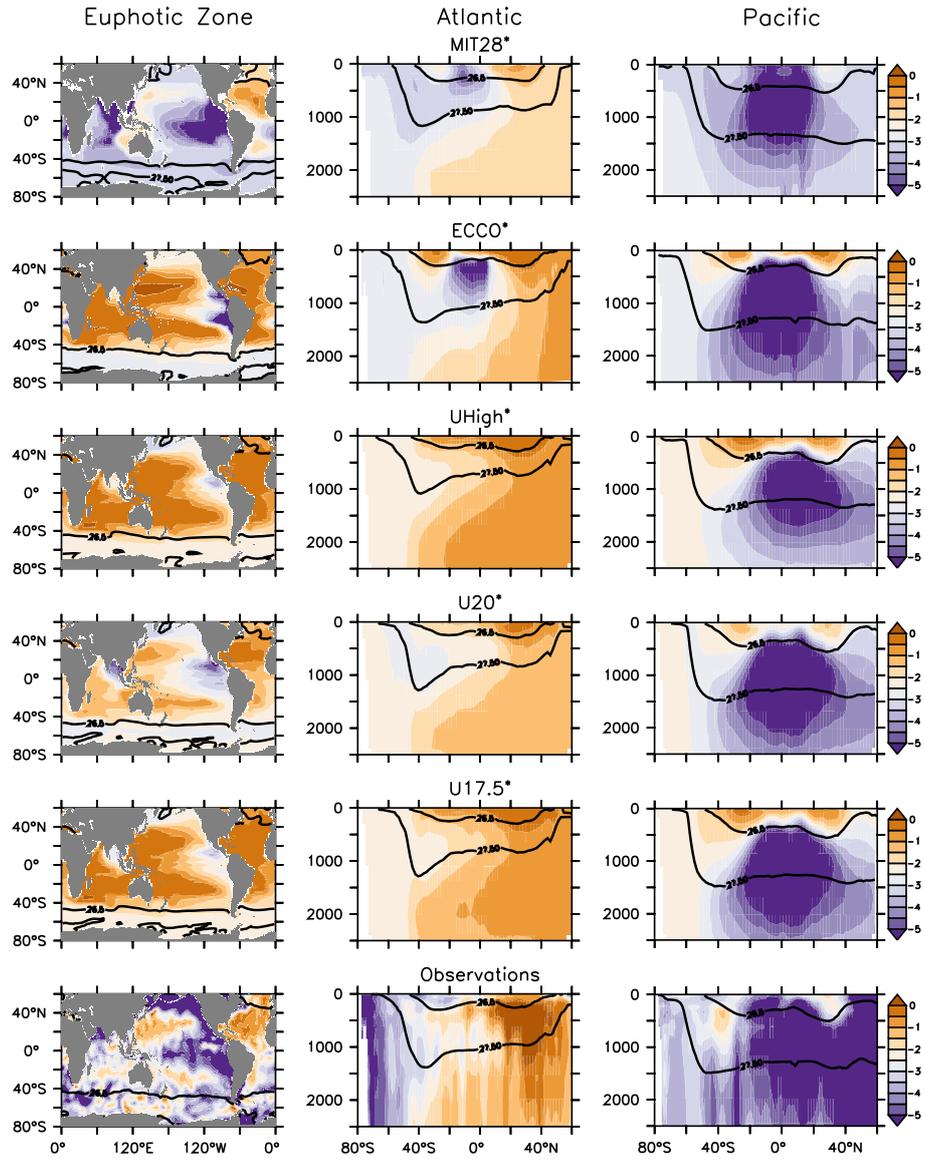
Figure 5. Excess nitrate N* (N* = NO$_3$ − 16 × PO$_4$) of optimal model configurations, averaged over the euphotic zone (left panels), and of zonal mean nitrate and phosphate in the Atlantic and Pacific (middle panels and right panels, respectively). Lines denote density of $\sigma_\theta = 27$ and $\sigma_\theta = 27.5$.
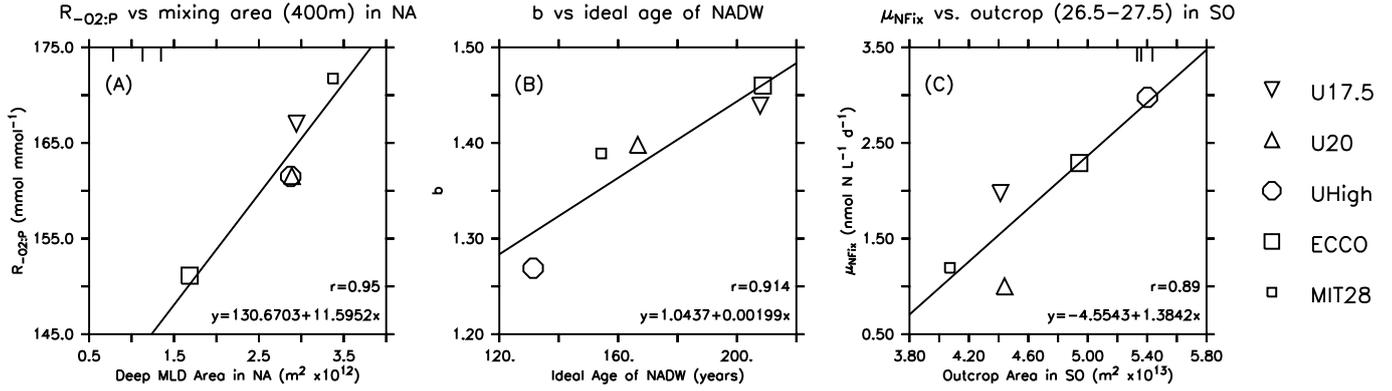
Figure 6. Optimal parameters for which the correlation of Table 2 is significant at $p < 0.05$, plotted against physical diagnostics. Symbols indicate the different model optimisations. Small squares: MIT28*. Large squares: ECCO*. Circles: UHigh*. Triangles: U20*. Inverted triangles: U17.5*. Lines denote the regression of optimal parameters against the respective circulation diagnostic. Vertical bars at the upper plot boundary indicate observed diagnostic.
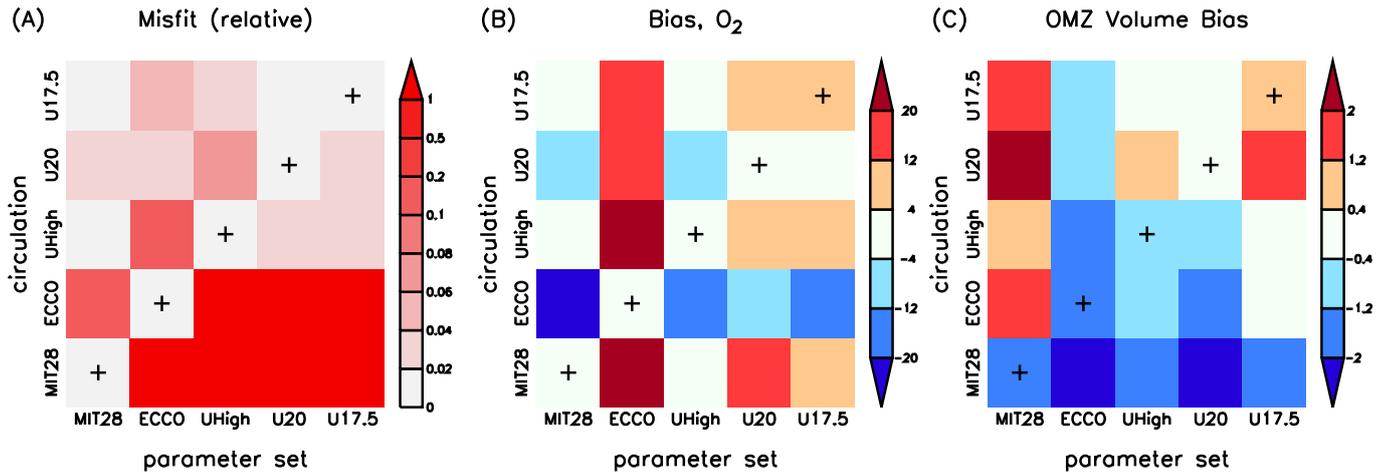

Figure 7. Three different model diagnostics for all cross-validation experiments with parameter set i and circulation j: (A) normalised misfit $J_{i,j}/J_j^* - 1$, where J_j^* is the lowest misfit for each circulation j. (B) oxygen bias (model minus observation, [mmol m^{-3}]). (C) OMZ volume (as percent of total ocean volume) bias. OMZs are defined by 50 mmol m^{-3}. The x-axis denotes the optimal parameter sets of MIT28*, ECCO*, UHigh*, U20*, U17.5* and the y-axis the circulation. Pluses along the main diagonal indicate the optimal model simulation for each circulation ($i = j$), i.e., MIT28*, ECCO*, UHigh*, U20*, U17.5*.
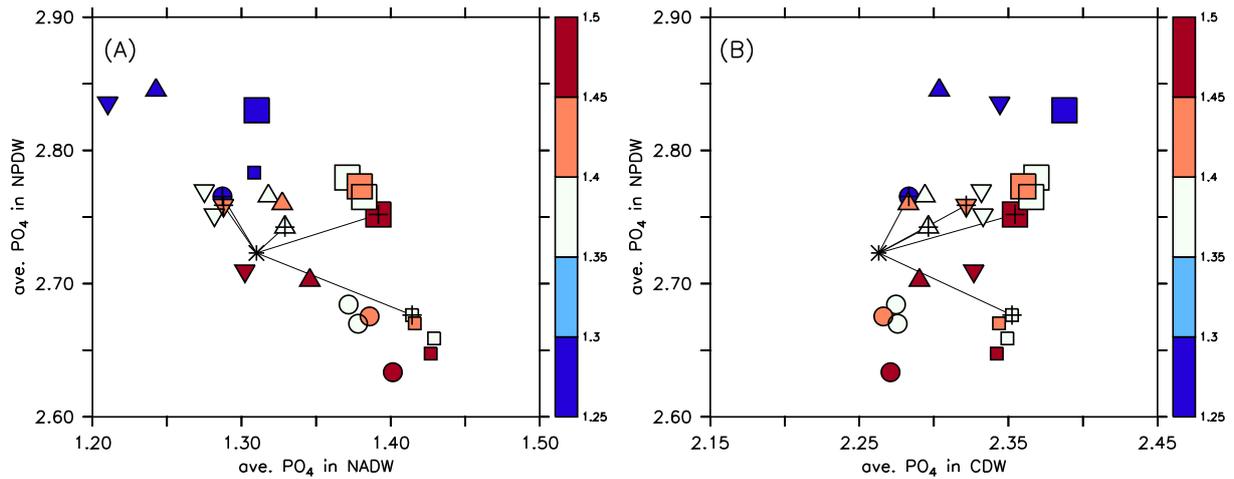
Figure 8. Average phosphate in the northern North Pacific Deep Water (NPDW; between 0°N and 60°N, and 1500-5000 m; see Text) plotted against average phosphate in the northern North Atlantic Deep Water (NADW; between 0°N and 60°N, and 1500-2500m; panel (A)) or Circumpolar Deep Water (CDW; south of 45°S, 1500-5000 m; panel (B)), of all 25 model experiments. Small squares: MIT28 circulation. Large squares: ECCO circulation. Circles: UHigh circulation. Triangles: U20 circulation. Inverted triangles: U17.5 circulation. (See also symbol legend in Fig. 6.) The colour indicates the value of parameter b. Pluses indicate the optimal parameter set. Stars indicate observed values. Thin black lines extending from the stars denote the distance between observation and each optimal model configuration.
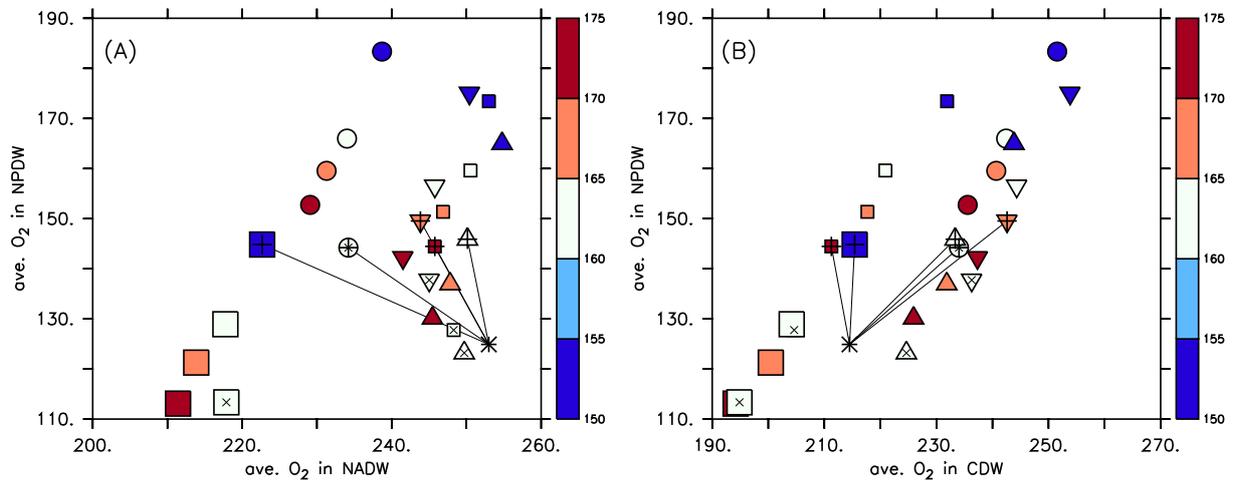
**Figure 9.** As Figure 8, but for average oxygen and parameter $R_{-O2:P}$ (colour scale).
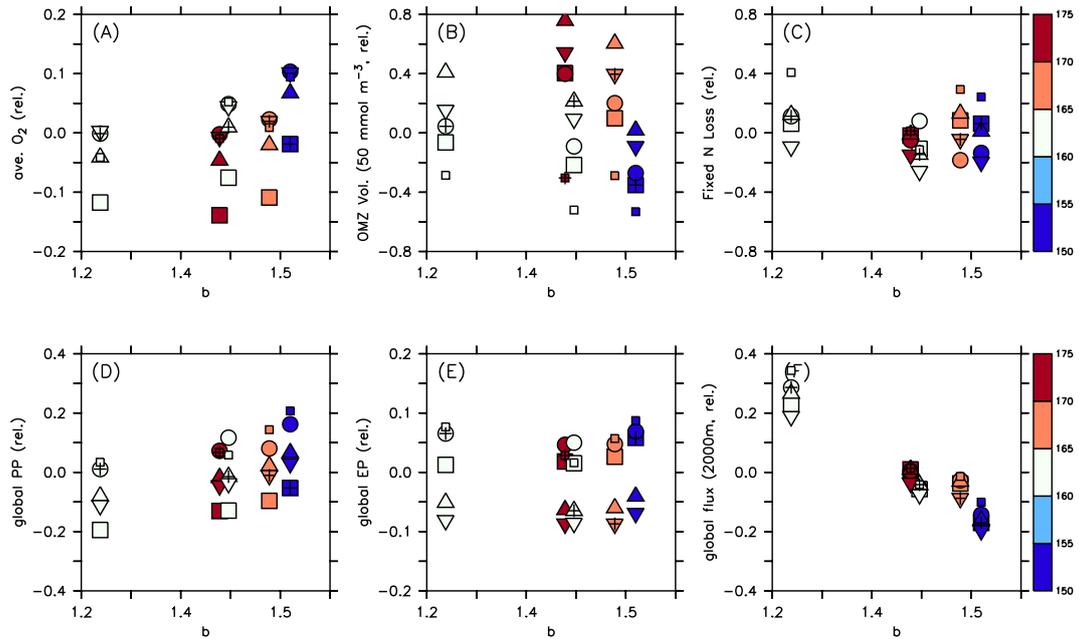
Figure 10. Normalised biogeochemical diagnostics plotted against parameter b: (A) Global average oxygen, (B) OMZ volume defined by concentrations < 50 mmol m^{-3}, (C) global fixed N loss, (D) global primary production, (E) export production and (F) organic particle flux at 2000 m. All diagnostics X expressed as relative deviation to the mean of the five optimal model simulations ($X_{i,j}/\bar{X}_i^* - 1$), where j and i denote different combinations of circulation j and parameter set i, and $\bar{X}_j^*$ the average of all optimal model configurations (see Table 3 for values). Colour denotes value of parameter $R_{-\mathrm{O2:P}}$. Symbols denote circulation: Small squares: parameter set of MIT28*. Large squares: ECCO*. Circles: UHigh*. Triangles: U20*. Inverted triangles: U17.5*. (See also symbol legend in Fig. 6.) Optimal model configurations are indicated by pluses.
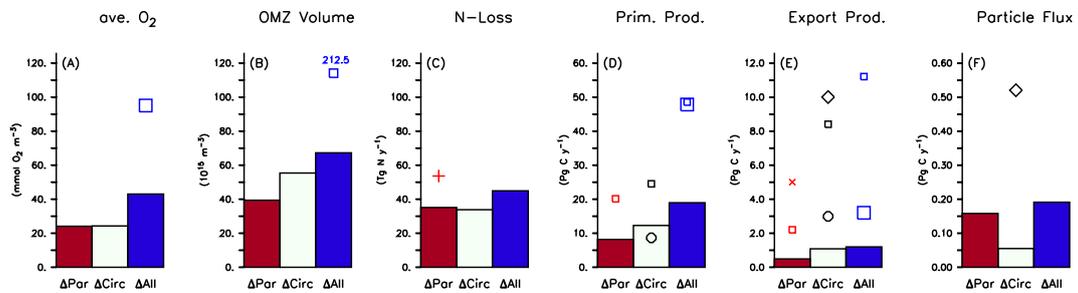

Figure 11. Effect of variation in biogeochemical parameters (ΔPar), circulation (ΔCirc) and across all model experiments (ΔCirc) on (A) global average oxygen, (B) OMZ volume defined by a concentration of < 50 mmol m^{-3}, (C) global fixed nitrogen loss, (D) global primary production, (E) export production and (F) organic particle flux at 2000 m, as listed in Table 3. Symbols denote values from Bopp et al. (2013, large squares), Schmittner et al. (2005, small squares), Seferian et al. (2013, circles), Najjar et al. (2007, diamonds), Somes et al. (2013, plus) and Kwon et al. (2009, cross).
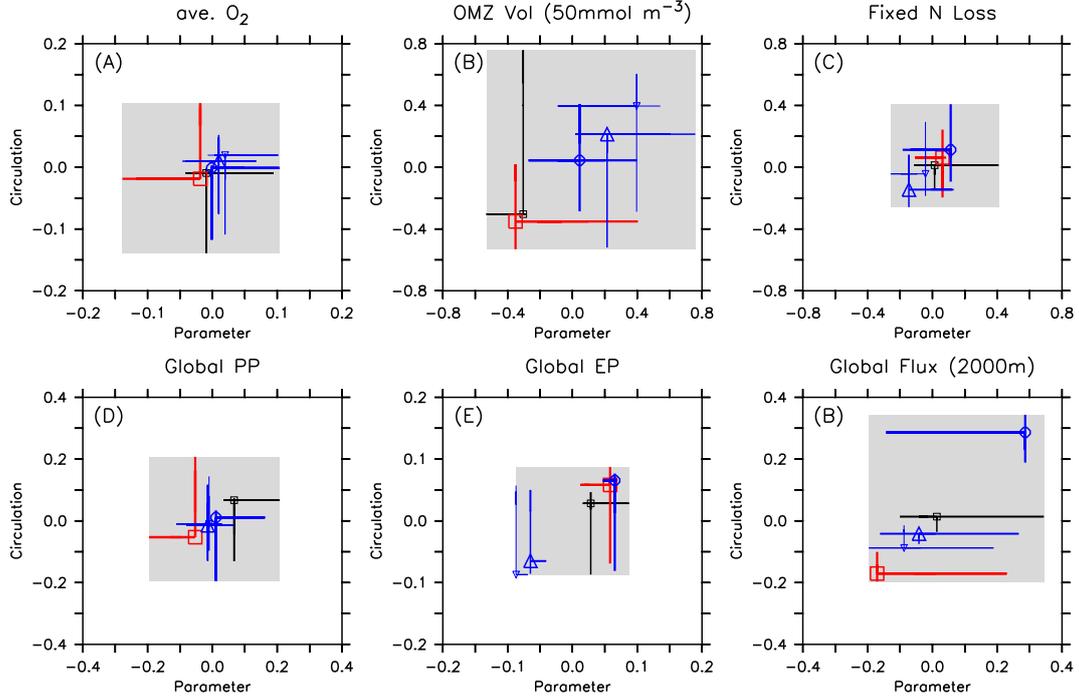
Figure 12. Effect of variation in circulation and biogeochemical parameters on normalised diagnostics. (A) global average oxygen, (B) OMZ volume defined by a concentration of < 50 mmol m^{-3}, (C) global fixed nitrogen loss, (D) global primary production, (E) export production and (F) organic particle flux at 2000 m. All diagnostics X are expressed as relative deviation to mean of the five optimal model simulations $(X_{i,j}/\bar{X}_j^* - 1)$, where j and i denote different combinations of circulation j and parameter set i, and $\bar{X}_j^*$ is the average of all optimal model configurations (see Table 3 for values). Each panel shows the range of the normalised diagnostic when the parameter set is kept constant and circulation varied (vertical lines), and when circulation is kept constant and the parameter set is varied (horizontal lines). Symbols denote the optimal model configuration. Colour denotes circulation and optimal parameter set. Black: circulation MIT28 or parameter set of MIT28*. Thick blue: UHigh/UHigh*. Medium blue: U20/U20*. Thin blue: U17.5/U17.5*. Red: ECCO/ECCO*. Symbols denote optimal model configurations (see also symbol legend in Fig. 6). The grey-shaded area shows total variation across all 25 model experiments, corresponding to Δ All in Table 3. Note that the maximum variation due to parameter within each circulation (ΔPar of Table 3) is given by the longest horizontal line in each panel. Likewise, the maximum variation due to circulation within each parameter set (ΔCirc of Table 3) is given by the longest vertical line.
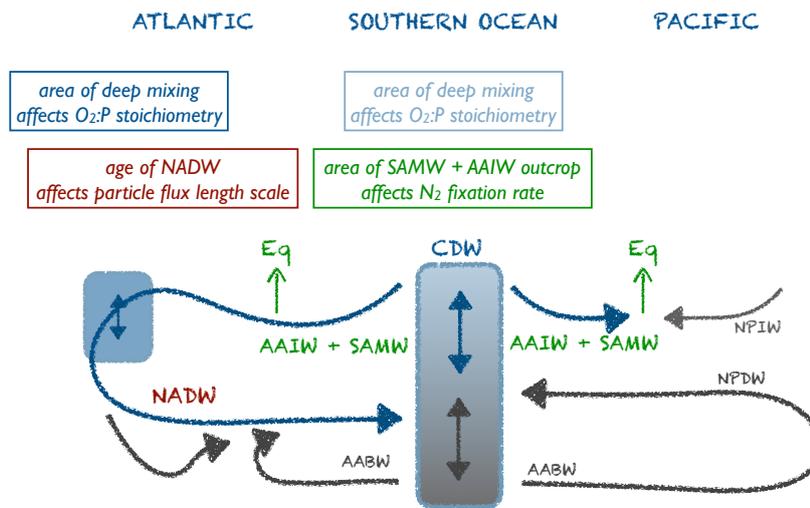
Figure 13. Cartoon depicting the simplified large scale circulation pattern of the Atlantic and Pacific Ocean, and the dependence of three biogeochemical parameters $R_{-O2:P}$, b and μ_{NFix} of physical properties.