Hierarchy of calibrated global models reveals improved distributions and fluxes of biogeochemical tracers in models with explicit representation of iron

Environmental Research Letters

LETTER

Hierarchy of calibrated global models reveals improved distributions and fluxes of biogeochemical tracers in models with explicit representation of iron

Wanxuan Yao, Karin F Kvale, Eric Achterberg, Wolfgang Koeve and Andreas Oschlies

GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, D-24105, Kiel, Germany

E-mail: wyao@geomar.de

Keywords: iron model, nutrient pathways, scavenging, calibration

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
23 July 2019

REVISED
23 September 2019

ACCEPTED FOR PUBLICATION
9 October 2019

PUBLISHED
29 October 2019

Abstract

Iron is represented in biogeochemical ocean models by a variety of structurally different approaches employing generally poorly constrained empirical parameterizations. Increasing the structural complexity of iron modules also increases computational costs and introduces additional uncertainties, with as yet unclear benefits. In order to demonstrate the benefits of explicitly representing iron, we calibrate a hierarchy of iron modules and evaluate the remaining model-data misfit. The first module includes a complex iron cycle with major processes resolved explicitly, the second module applies iron limitation in primary production using prescribed monthly iron concentration fields, and the third module does not explicitly include iron effects at all. All three modules are embedded into the same circulation model. Models are calibrated against global data sets of NO_3, PO_4 and O_2 applying a state-of-the-art multi-variable constraint parameter optimization. The model with fully resolved iron cycle is marginally (up to 4.8%) better at representing global distributions of NO_3, PO_4 and O_2 compared to models with implicit or absent parameterizations of iron. We also found a slow down of global surface nutrient cycling by about 30% and a shift of productivity from the tropics to temperate regions for the explicit iron module. The explicit iron model also reduces the otherwise overestimated volume of suboxic waters, yielding results closer to observations.

1. Introduction

Historically, global biogeochemical models have been becoming more complex. There is some justification for this; for example by a reduced misfit in global models using multiple phytoplankton functional types versus single-phytoplankton models (Friedrichs *et al* 2007). However, increasing structural complexity requires careful assessment, if anything is to be learned from the modeling study (Anderson 2005). Models are also becoming more complex with respect to biogeochemical processes. Similar model-data misfits remaining after parameter optimization are possible for both structurally simple and complex biogeochemical ocean models (Kriest 2017), but resulting nutrient pathways, i.e. routes taken by nutrients between particulate and dissolved pools, can differ substantially (Friedrichs *et al* 2007, Löptien and Dietze 2017). Biogeochemical differences arising due to differences in nutrient pathways can increase with climate change forcing (e.g. Kvale *et al* 2015, Laufkötter *et al* 2016), further underscoring the need for careful assessment as part of model development. Differences in nutrient pathways may also be relevant when modeling higher trophic levels, including fish (Pauly and Christensen 1995, Stock *et al* 2017).

Iron modeling is an excellent example of the challenges mentioned above. An insufficient supply of the trace metal iron limits primary production in over one-third of the surface ocean (e.g. Boyd *et al* 2007, Boyd and Ellwood 2010, Moore *et al* 2013) and also controls di-nitrogen fixation in nitrate-limited ocean

regions (Schlosser *et al* 2014). Iron is therefore considered an important regulator of the strength of the soft tissue pump (Volk and Hoffert 1985) and, via atmospheric CO_2, of the climate (Joos *et al* 1991). Consequently, earth system models (including all CMIP5 models; Laufkötter *et al* 2015) include some parameterization of the marine iron cycle. However, a comparison of 13 global ocean biogeochemistry models against a compilation of dissolved iron (dFe) measurements showed that all of the models had clear deficits in reproducing many aspects of the observed patterns (Tagliabue *et al* 2016). The authors emphasized that iron scavenging parameters are particularly poorly constrained. Iron scavenging (the adsorption of dFe onto particle surfaces) is commonly hand-tuned to achieve a good model fit to observed dFe concentrations (resulting in global mean model water column concentrations of 0.58 ± 0.14 μmol m^{-3} across the range of models, Tagliabue *et al* 2016). Tuning a model to dFe using the iron scavenging parameterization has unclear consequences for model behavior. Pasquier and Holzer (2017) calibrated a simple steady-state model against multiple data fields (dFe, PO_4, $Si(OH)_4$ and chlorophyll) and found that they could achieve a similar model misfit (differing by less than 1%) when applying a wide range of assumed external iron source strengths, however, resulting in differing parameter values of the iron module. The Pasquier and Holzer (2017) study advanced our understanding of the sensitivity of simulated nutrient distributions to specific parameterizations of the iron cycle. The impact of including iron models of different structural complexity on the cycling of nutrients in a seasonally varying model environment is as yet un-quantified.

Nickelsen *et al* (2015) introduced a dynamic iron module into an earth system model of intermediate complexity and hand-tuned model parameters against surface macro-nutrient observations and sparse observations of iron concentrations. The module resolved major components of the marine iron cycle, e.g. iron sources including aerosol deposition, detrital remineralization and sedimentary release, and iron sinks including biological uptake, iron scavenging and colloid formation. The focus of this earlier effort was on exploring the impact of an explicit iron cycle on model sensitivities to environmental change, while the hand-tuning ensured that overall model performance for present-day ocean state was not affected negatively by the addition of the iron module.

In this study we present calibrations of three variants of a global model of ocean biogeochemical cycles (dynamic iron cycle, Nickelsen *et al* 2015; prescribed iron mask, Keller *et al* 2012; without iron, disabled iron limitation in Keller *et al* 2012). Model variants share the same physical circulation but differ in their representation of the micro-nutrient iron. We calibrate against oceanic observations of NO_3, PO_4 and O_2, using a recently developed framework (Kriest *et al* 2017). Our aim was to assess whether, and to what extent, the incorporation of Fe-related processes in the model improves the model skill of simulating NO_3, PO_4 and O_2 as well as global indicators of biogeochemical cycles (described below). While it is generally assumed that the inclusion of an explicit iron cycle improves the capability of marine biogeochemical models to simulate distributions and fluxes of biogeochemical tracers also other than iron, this has, to our knowledge, not yet been demonstrated in a quantitative manner.

2. Materials and methods

2.1. Model description

Our ocean model is a coupling of the University of Victoria Earth System Climate Model (UVic ESCM, version 2.9; Weaver *et al* 2001, Eby *et al* 2013) with the Transport Matrix Method (TMM; Khatiwala 2007) architecture (an 'offline' physical circulation, Kvale *et al* 2017, see appendix). For the biogeochemical ocean model component, we deploy different versions of the Kiel Marine Biogeochemistry Module (KMBGC), which differ in their representations of iron. The first one utilizes a full dynamic iron cycle (FeDyn; Nickelsen *et al* 2015), while the second one utilizes prescribed monthly dFe surface concentration fields (FeMask; Keller *et al* 2012), which is output from the Biology Light Iron Nutrient and Gas model (BLING), coupled with MOM4p1 circulation (Galbraith *et al* 2010), regridded to UVic2.9 ESCM grid by Keller *et al* (2012), and the third one includes no representation of the iron cycle (NoFe). NoFe is created by turning off the iron control on primary production in FeMask. Detailed descriptions of FeDyn can be found in Nickelsen *et al* (2015); FeMask and NoFe in Keller *et al* (2012). In FeDyn, we also include hydrothermal vents as an additional source of dFe to the deep ocean. The hydrothermal iron flux data was compiled by Tagliabue *et al* (2010). The parameters chosen for calibration are shown in table 1 (see the appendix for details of parameter selection).

2.2. Calibration framework

The calibration framework used in this study is adapted from Kriest *et al* (2017). It contains three components; TMM, a biogeochemical model (either FeDyn, FeMask, or NoFe) and the Covariance Matrix Adaption Evolution Strategy (CMA-ES) (see appendix, Hansen 2006, Kriest *et al* 2017). CMA-ES is an evolutionary algorithm mimicking natural selection to search for an 'individual' (a set of parameter values) that minimize the data-model misfit. It starts with a random set of individuals (the first generation) and evaluates the model-data misfit of each individual. Individuals with lowest misfit are more likely to be present in the next generation, but random mutations are also permitted, that explore the search space with a mutation rate and intensity computed from the convergence behavior of previous generations. In our

Table 1. The parameters chosen for each calibration, their observational value and calibration boundary settings. The parameters chosen for each calibration are marked with √.

Parameter	Description	Observation	Cali. bound	Uncalibrated value			Unit
				NoFe	FeMask	FeDyn	
a	Maximum growth rate at 0 °C	0.04–0.60[a]	0.30–0.75	0.600√	0.600	0.600	d^{-1}
μ_P^*	Microbial loop recycling rate		0.001–0.035	0.015√	0.015√	0.015	d^{-1}
K_{Fe}^P	Ordinary phyto. half-satu. of Fe-limitation	0.035–1.14[b,c]	0.04–1.14	—	0.1√	—	μmol Fe m^{-3}
K_{Fe}^{Pmax}	Maximum K_{Fe}^P	0.035–1.14[b,c]	0.04–1.14	—	—	0.4√	μmol Fe m^{-3}
K_{Fe}^D	Diazotrophs half-satu. of Fe-limitation	6.29–17.2[d,e]	0.05–1.50	—	0.1√	0.1√	μmol Fe m^{-3}
g_Z^0	Maximum grazing rate at 0 °C	0.06–1.9[f]	0.2–2.0	0.4√	0.4√	0.4√	d^{-1}
ω_{Det}^i	Sinking speed slope	0.024–0.107[g,h]	0.001–0.160	0.06√	0.06√	0.06√	d^{-1}
K_{Fe}^{org}	POM dependent scavenging	0.079–6.62[i,j]	0.225–6.750	—	—	0.45√	
$R_{O:N}$	Molar O: N ratio	8.1–11.2[k]	8.0–12.0	10√	10√	10√	

[a] Le Quere *et al* (2005).
[b] Timmermans *et al* (2004).
[c] Price *et al* (1994).
[d] Bucciarelli *et al* (2013) unpublished data.
[e] Jacq *et al* (2014).
[f] Calbet and Landry (2004).
[g] Berelson (2001).
[h] Jackson *et al* (2015).
[i] Honeyman *et al* (1988).
[j] Lerner *et al* (2016).
[k] Körtzinger *et al* (2001).

study, each individual is represented by a set of parameters used for a 3000 model-year model spin-up and subsequent computation of the model-data misfit. The CMA-ES stops when a pre-defined upper limit of the number of generations is reached or when successive generations do not yield a further reduction in the model-data misfit. The individual with the lowest model-data misfit is then the calibrated parameter set. All model spin-ups are performed with the same transport matrices as used by Kvale *et al* (2017, derived from online simulations of the UVic ESCM; see appendix).

2.3. Misfit function

Following Kriest *et al* (2017), we define the misfit (J_T) of the model as the sum of the volume weighted Root Mean Square Error (RMSE) between simulated and observed annual mean concentrations of the tracers NO_3, PO_4 and O_2 (World Ocean Atlas 2013; Garcia *et al* 2013b, 2013a). Volume weighting of the RMSE puts relatively more weight on the ocean interior than the surface ocean model-data misfit. The misfit function J_T is calculated as:

$$J_T = \sum_{j=1}^{M} J_j = \sum_{j=1}^{M} \frac{1}{o_j} \sqrt{\sum_{i=1}^{N} (m_{i,j} - o_{i,j})^2 \frac{V_i}{V_T}}, \; M = 3, \tag{1}$$

where $j = 1,..,M = 3$ denotes tracers NO_3, PO_4 and O_2; and $i = 1, \ldots, N$ denotes the location of each model grid cell, $N = 87307$ for the UVic ESCM; V_T is the total volume of the model ocean and V_i is the volume of the respective grid cell; o_j is the global average observed concentration of the respective tracer; $m_{i,j}$ and $o_{i,j}$ are the modeled and observed concentrations of tracer j at location i, respectively. Note that, in order to allow for a fair comparison between models with and without an iron cycle, dFe is not included in the misfit function.

Table 2. A comparison of the misfit from optimized model spin-ups. Values for uncalibrated FeDyn (FeDyn0) are calculated diagnostically (by us) and have not been used in the tuning process of Nickelsen *et al* (2015). The minimum value of each column is marked in bold.

Experiment	J_{NO_3}	J_{PO_4}	J_{O_2}	J_T
NoFe	0.1575	0.1305	0.1786	0.4667
FeMask	0.1537	0.1293	0.1803	0.4632
FeDyn	**0.1457**	**0.1254**	**0.1730**	**0.4442**
Hand-tuned				
FeDyn0	(0.1838)	(0.1389)	(0.1635)	(0.4862)

3. Results and discussion

3.1. Improvement of misfit and biogeochemical indicators after calibration

For the dynamic iron model (FeDyn) we find a marked improvement by 8.6% for the total misfit after calibration, compared to the hand-tuned version of FeDyn (FeDyn0). Individual misfits J_{NO_3} and J_{PO_4} are reduced by 20.7% and 9.7%, respectively (table 2), compared to FeDyn0. This significant gain in the model's ability to reproduce the observational pattern of NO_3 and PO_4, comes at the expense of an increasing misfit J_{O_2} (5.8%), i.e. the ability to reproduce the observed patterns of O_2, after calibration. In our

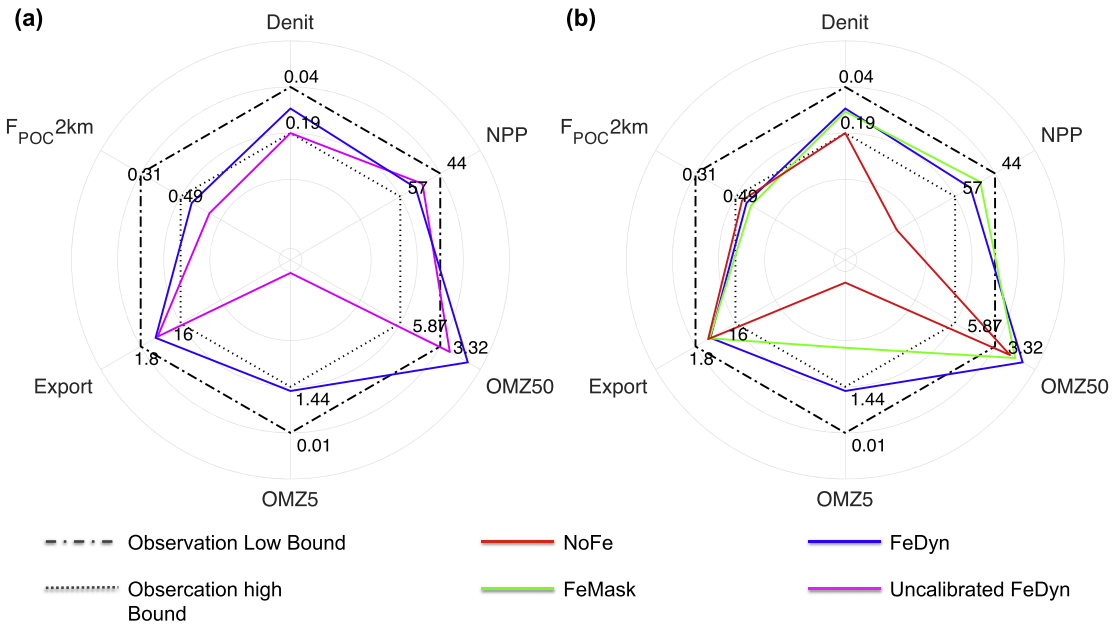

Figure 1. Model global annual fluxes and oxygen minimum zone size against their respective observational data. We plot six global indicators on lineal axis pointing to the pole and the units are GtC yr^{-1} (NPP), PgN yr^{-1} (Denit), GtC yr^{-1} (F_{POC} 2 km), GtC yr^{-1} (Export), ‰(OMZ5), % (OMZ50). The dashed–dotted (dotted) line denotes the lower (upper) boundary of the observational value. The solid lines present the model annual mean values. (a) Shows the comparison between un-calibrated FeDyn (magenta) with calibrated (blue); (b) Shows the comparison between three calibrated models, FeDyn, FeMask (green) and NoFe (red); the observational data are from Carr *et al* (2006, NPP), Kriest and Oschlies (2015, Denit, compilation of rates from other studies referred to in their table 2), Honjo *et al* (2008, F_{POC} 2 km and Export), Lutz *et al* (2007, Export), Dunne *et al* (2007, Export) and Garcia *et al* (2013a, WOA). Here we use 2.5 (7.5) mmol m^{-3} as the lower (upper) boundary value for OMZ5 and 40 (60) mmol m^{-3} as the lower (upper) boundary value for OMZ50 Conversion between different elements followed C: N: P = 106: 16: 1.

calibration, the objective is to minimize the overall misfit, which produced a trade-off between individual components of the misfit function, leading to an increase in J_{O_2}. In order to assess whether a better overall misfit against observations of biogeochemical tracer concentrations improves the representation of the marine biogeochemical fluxes, we diagnosed several global indicators, i.e. net primary production (NPP, GtC yr^{-1}), export production (Export, GtC yr^{-1}), flux of organic carbon at a depth of 2 km (F_{POC} 2 km, GtC yr^{-1}), denitrification (Denit, PgN yr^{-1}), the volume fraction of the ocean with oxygen concentrations less than 5 mmol m^{-3} (OMZ5, ‰) and the volume fraction of ocean with oxygen concentrations less than 50 mmol m^{-3} (OMZ50, %), and compared them to observational or synthetic data. Three out of six indicators (Denit, F_{POC} 2 km, OMZ5) show improvements for FeDyn against the hand-tuned FeDyn0 by getting closer to the observational data range (figure 1(a)). The estimations of Export are similar. Although NPP is increased by 4.5% after calibration, both the calibrated and hand-tuned models fall within the range of observational data. While the calibration improves the simulation of the OMZ5 by decreasing its volume by 73.6%, a reduction by 42.4% of OMZ50 size worsens the description of the volume fraction of hypoxic waters where O_2 is less than 50 mmol m^{-3}.

3.2. The calibrated dynamic iron model has the best skill

When comparing the misfits from different calibrations, FeDyn with a full dynamic iron cycle has a misfit J_T 4.8% smaller than NoFe, and 4.1% smaller than FeMask (table 2). Also for the individual components (J_{NO_3}, J_{PO_4}, and J_{O_2}) FeDyn provides the smallest misfit against observations. The most significant difference is in J_{NO_3}, with FeDyn having a 7.5 (5.2)% smaller J_{NO_3} than NoFe (FeMask). This improvement may be related to the improvement in OMZ5, and associated Denit, in FeDyn, described below. However, even the smallest J_T (FeDyn) still amounts to about 15% of global mean tracer concentrations of each tracer, which is on par with the calibration result from Kriest *et al* (2017, conducted using a somewhat simpler biogeochemistry coupled to a different circulation) and considerably larger than the 5% obtained for PO_4 by Pasquier and Holzer (2017). This is not unexpected because their study utilized an inverse model coupled with a data-assimilated, steady circulation (Primeau *et al* 2013), where PO_4 had been used already as part of the constraint for obtaining a circulation that can optimally represent the tracer field. The misfit for FeMask and FeDyn are around 15.4% and 15.6% of global mean tracer concentrations, respectively. This indicates that the differences between our biogeochemical models are smaller than the differences between either model and observations. We believe that this relates to biases in our physical circulation model

Table 3. Calibrated parameter value and their resulting fluxes ranges with 1% of misfit. Here we show the range of parameter values and simulated NPP and Export diagnosed from all simulations that yield not more than a 1% increase of the misfit function with respect to its global minimum. The parameter values that are kept fixed during the calibration are put in parentheses. The unit of NPP and Export is GtC yr^{-1}. See table 1 for the units of the individual parameters.

	NoFe		FeMask		FeDyn	
Parameter	Value	Range[a]	Value	Range[a]	Value	Range[a]
a	0.698	0.650–0.780	(0.060)		(0.060)	
μ_P^*	0.0157	0.0142–0.0212	0.0012	0.0000–0.0020	(0.015)	
K_{Fe}^{P}	—		0.050	0.045–0.071	—	
K_{Fe}^{Pmax}	—		—		0.585	0.520–0. 598
K_{Fe}^{D}	—		0.406	0.387–0.494	0.377	0.326–0.479
g_Z^0	1.282	1.223–1.474	0.668	0.551–0.691	0.567	0.526–0.685
ω_{Det}^{i}	0.065	0.063–0.072	0.062	0.058–0.064	0.060	0.059–0.065
K_{Fe}^{org}	—		—		0.427	0.208–0.510
$R_{O:N}$	10.44	10.18–10.91	9.54	9.34–9.62	10.502	10.27–10.58
Fluxes						
NPP	75.9	70.1–81.44	49.1	44.0–49.9	52.0	50.4–54.5
Export	6.3	5.8–6.4	6.8	6.4–7.0	7.1	6.7–7.2

[a] Here is obtained by sampling model runs within 1% range of smallest misfit for the respected calibration.

(see figure A1, available online at stacks.iop.org/ERL/14/114009/mmedia), which would also be consistent with the considerably smaller PO_4 misfit obtained by Pasquier and Holzer (2017) for their PO_4-constrained circulation field. In our model, annually-averaged basin vertical profiles of macro-nutrients and O_2 (figure A1) display the circulation bias reported in Kvale *et al* (2017); a deep ocean (>2000 m) that is generally too old (according to the radiocarbon), which produces an overestimation of NO_3 and PO_4, and an under-estimation of O_2 in the deep ocean basins. The Pacific O_2 profile is affected by an over-estimate of deep vertical mixing in the Southern Ocean (not shown). Even with calibration, these biases cannot be completely overcome.

For the six biogeochemical indicators, half of them (Denit, OMZ5 and NPP) improve with respect to observations after the inclusion of iron modules, two (F_{POC} 2 km and OMZ50) become worse and Export was essentially unchanged (figure 1(b)). The indicators most sensitive to iron are NPP and OMZ5: experiment FeDyn resulted in a reduction of NPP by 31.4% and OMZ5 by 71.9% compared to NoFe. Between the two models with iron, FeDyn shows lower F_{POC} 2 km (3.6%), OMZ5 (50.6%) and OMZ50 (23.5%) than FeMask. The former two values reflect improvements compared to observations, while the latter does not. When considering all indicators equally, the simulation of biogeochemical indicators is improved when iron modules are included, and a fully dynamic iron cycle outperforms a model with an iron mask. Both misfits (table 2) and biogeochemical indicators are improved when a full dynamic iron cycle is used.

An interesting feature of the calibrations is the different behavior of OMZ50 and OMZ5 upon the inclusion of an explicit representation of the iron cycle, namely an improved agreement of the OMZ5 volume with observations while the agreement deteriorates for the OMZ50 volume. Both volumes decreased after the calibration of biogeochemical parameters (improving the OMZ5 bias but exacerbating the underestimation of the OMZ50 volume with respect to NoFe). The OMZ volume is determined by a combination of physical and biogeochemical processes (Kriest *et al* 2012, Kriest and Oschlies 2015). The circulation is held constant across all model experiments, with only the biogeochemical parameters allowed to vary. The calibrations utilize global misfit weighted by ocean volume, which gives little consideration to the OMZ volume that comprises less than 4% of the global ocean. However, OMZ5 waters have a special role as featuring the only permanent sink of NO_3 (via denitrification) in our model. The OMZ5 volume therefore strongly impacts the global nitrogen inventory and associated nitrate distributions (Landolfi *et al* 2013, Kvale *et al* 2019). The calibration against observed O_2 and NO_3 distributions emphasized the improvement of J_{NO_3}, via improved OMZ5 volume. In contrast, there is no such control via simulated NO_3 distributions on the OMZ50 volume.

3.3. Calibrated parameters and uncertainties

The calibrated parameter values and their range within a 1% increase in misfit around the minimum of the misfit function J_T for the respective calibrations define a measure of uncertainty of the calibrated parameters and are provided in table 3. Also shown are the resulting uncertainties in NPP and Export. All parameters and fluxes are constrained within a relatively small range. The more tightly constrained parameters are the increase of sinking speed with depth and the molar O: N ratio. Their uncertainty ranges amount to 9.7%–13.8% and 2.9%–7.0% in the different

calibrations, with calibration results for NoFe having the largest uncertainty. This is due to the higher misfit of NoFe, which, via our 1% misfit criterion increases the absolute misfit difference and therefore may lead to a larger range of parameter values consistent with a 1% misfit increase. The less constrained parameters are the microbial loop recycling rate, the diazotrophic half-saturation of iron and the iron scavenging rate. However, those parameters also have a relatively large prior parameter range used as input in the calibration. Interestingly, estimates of the tracer fluxes, such as NPP and Export are relatively well constrained, despite large uncertainties in individual parameters (bottom rows of table 3). For each model, the uncertainty range in NPP is diagnosed from the ensemble of simulations and parameter combinations with a misfit function value less than 1% larger than the minimum of the cost function. The thus estimated uncertainties amount to 7.9%–14.9% for NPP and 7.0%–9.5% for Export (bottom rows of table 3).

An encouraging result of the calibration is that NPP and Export, which are not explicitly included in the misfit function, turn out to be well constrained and in good agreement with independent observational evidence. Some optimal parameters appear to be portable across model configurations (e.g. see the close agreement between ω^{i}_{Det} values). However, other parameters are model dependent (e.g. calibrated g^{0}_{Z} values vary up to 126% between models).

3.4. The impact of including an iron module on surface ocean nutrient pathways

The different models reveal different surface ocean biogeochemical pathways (figure 2). Although the use of volume-weighted model-data misfits in the model calibration emphasizes fitness in the deep ocean, the biogeochemical parameters selected for the calibration reflect mostly upper-ocean processes. The differences in NPP are the most prominent, with NPP being much lower in FeDyn and FeMask (52.0 and 49.1 GtC yr^{-1}) compared to NoFe (75.9 GtC yr^{-1}). However, the corresponding total nutrient recycling (sum of the microbial loop recycling, zooplankton excretion and remineralization) also slows, which leads to relatively similar export production rates (6.3–7.1 GtC yr^{-1}) out of the euphotic zone (130 m) for all three calibrated models. The similarity of Export estimations between calibrations is likely linked to the calibration setup, which uses the same circulation to achieve the same observational distribution of tracer concentrations and, therefore, is likely to produce similar values of export production. Kriest (2017) calibrated two no-iron prognostic models with identical physical circulation against observational data, and found similar values of export production of between 6.0 and 7.4 GtC yr^{-1}. In our calibrations, the inclusion of iron seems to slightly increase Export, nevertheless our global number (7.1 GtC yr^{-1} in FeDyn) is still lower than the steady state model calibration from Pasquier and Holzer (2017, 9.5–11.0 GtC yr^{-1}) using a different circulation model. Our 1% misfit range of NPP and Export of respective calibrations can be found in table 3.

The three calibrated model configurations result in similar Export through different surface nutrient pathways. Zooplankton in our model grazes not only on phytoplankton but also on detritus and itself. The net nutrient flux from zooplankton to detritus represents sloppy feeding plus zooplankton mortality minus grazing on detritus. Increasing the zooplankton grazing rate increases the flux in both directions, with the net effect reducing the percentage of nutrient that goes directly from zooplankton to detritus, e.g. NoFe has the highest grazing rate and only 20.1% of NPP goes to detritus through zooplankton, comparing to 25.3% in FeMask and 27.9% in FeDyn. A higher grazing rate also reduces the standing stock of phytoplankton and reduces phytoplankton loss via mortality. Hence, despite the differences in NPP, detritus production rates are similar in the three calibrations (3.3 PgN yr^{-1} in NoFe, 3.2 PgN yr^{-1} in FeMask and 3.2 PgN yr^{-1} in FeDyn). In all our calibrations, the remineralization rate at 0 °C and its dependence on temperature is identical. The higher remineralization flux in NoFe is due to the high productivity in the tropical (warmer) region in that configuration (figures 3(a) and (d)). The higher remineralization of detritus keeps nutrients in the surface ocean. In combination with the higher phytoplankton growth rate and higher grazing rate this accelerates the surface ocean nutrient cycling (47.3% faster when comparing NoFe to FeDyn).

Differences in surface ocean nutrient cycling and pathways such as identified here may have substantial implications when projecting future changes to ecosystem services, e.g. sustainability and resilience of fisheries, since catches are closely related to NPP, trophic energy pathways and transfer efficiencies (Stock *et al* 2017).

3.5. Geographical shifts in NPP, Export and surface NO_3

The geographical distributions of simulated marine biogeochemical fluxes also differs across calibrations. The NPP reductions per area are 150 gC m^{-2} yr^{-1} for the Pacific, 30 gC m^{-2} yr^{-1} for the Atlantic, and 60 gC m^{-2} yr^{-1} for the Indian Ocean, when comparing FeDyn to NoFe. The Pacific Ocean is most impacted by iron limitation, since it hosts three major HNLC regions, the subpolar North Pacific, tropical Pacific, and part of the Southern Ocean (Cullen 1995, Pitchford 1999). However, the reductions of NPP in the Southern Ocean and North Pacific are small compared to the tropical Pacific (see figure 3(a)). The large reduction of NPP in the tropical Pacific in FeDyn is due to the reduction of nutrient recycling as compared to the NoFe model (described in section 3.4), where a higher phytoplankton growth

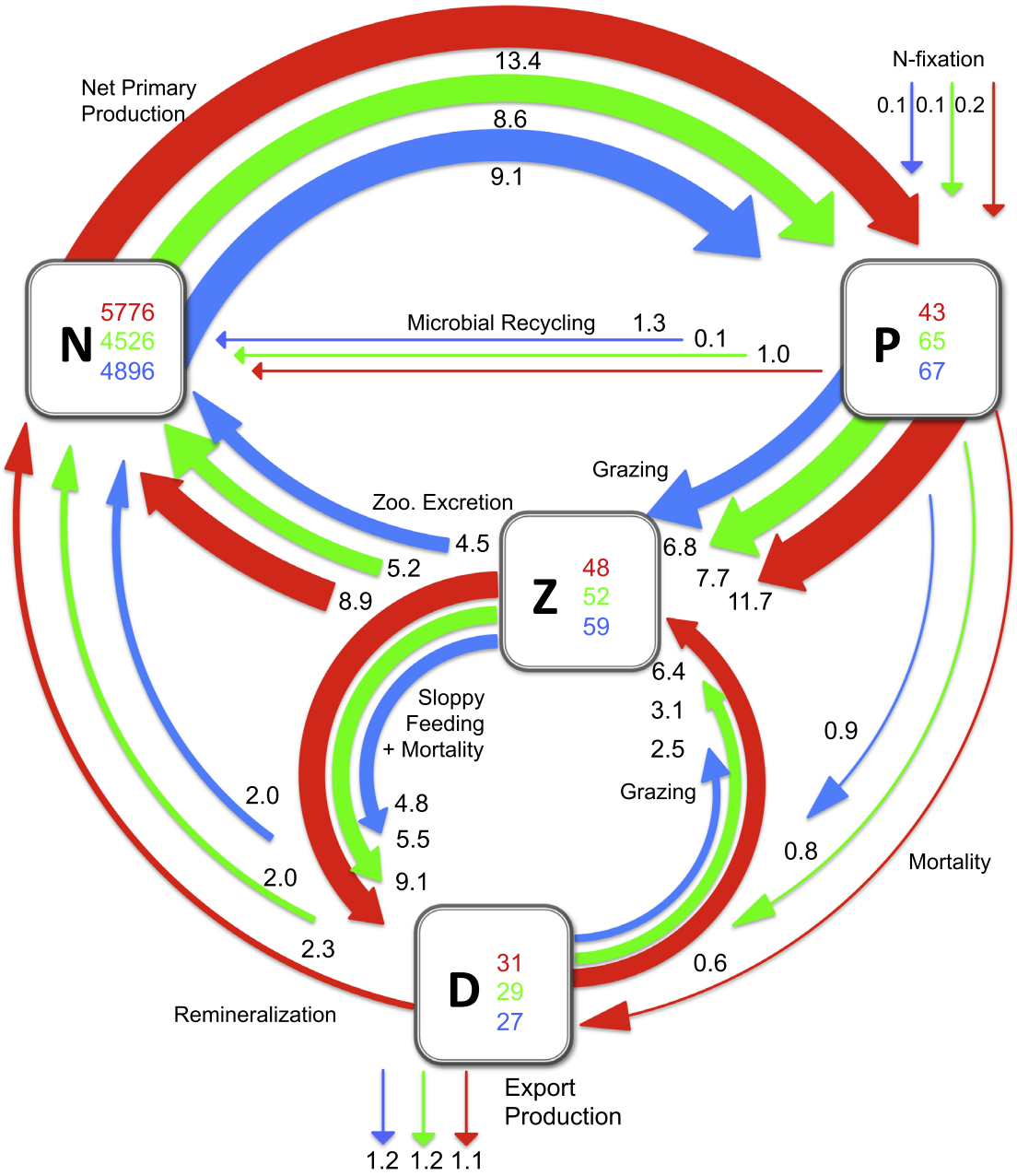

Figure 2. A NPZD plot comparing the surface nutrient pathway in three calibrated models. NoFe is red, FeMask is green and FeDyn is blue. The boxes represent nutrient (N), phytoplankton (P), zooplankton (Z) and detritus (D) and their inventory in the euphotic zone for each calibration are marked with numbers (unit PgN) in the box with respective color. The thicknesses of the arrows is approximately proportional to the associated flux and all fluxes are in unit of PgN yr^{-1}).

rate and higher grazing rate accelerate nutrient recycling in the warm tropical waters in NoFe. This recycling is reduced in FeDyn via iron limitation on primary production, as well as this model having generally lower biological rates. Reduction of NPP in the tropical Pacific, caused by the inclusion of a dynamic iron module and regional iron limitation, results in higher annual mean euphotic zone NO_3 concentrations (by up to 8 mmol m^{-3}) in the east Pacific upwelling region (see figure 3(c)). This residual NO_3, which has not been exported into the Pacific OMZ, can be utilized further downstream of the equatorial upwelling system, where iron is less limiting (e.g. the subtropical regions). Meanwhile, higher global Export in FeDyn (see figure 3(b)) produces lower annual-mean euphotic zone NO_3 concentrations (by up to 14 mmol m^{-3}) in the rest of the surface ocean (e.g. equatorial Atlantic, see figure 3(c)).

Bopp *et al* (2013) report that the different CMIP5 models show little agreement in simulated trends in primary production in eastern equatorial regions despite all models accounting for iron limitation. In our calibrated iron models (FeDyn and FeMask), we also see different changes in NPP and NO_3 concentrations in this region (figure 3), which suggests that the available observations of NO_3, PO_4 and O_2 together with a volume-weighted

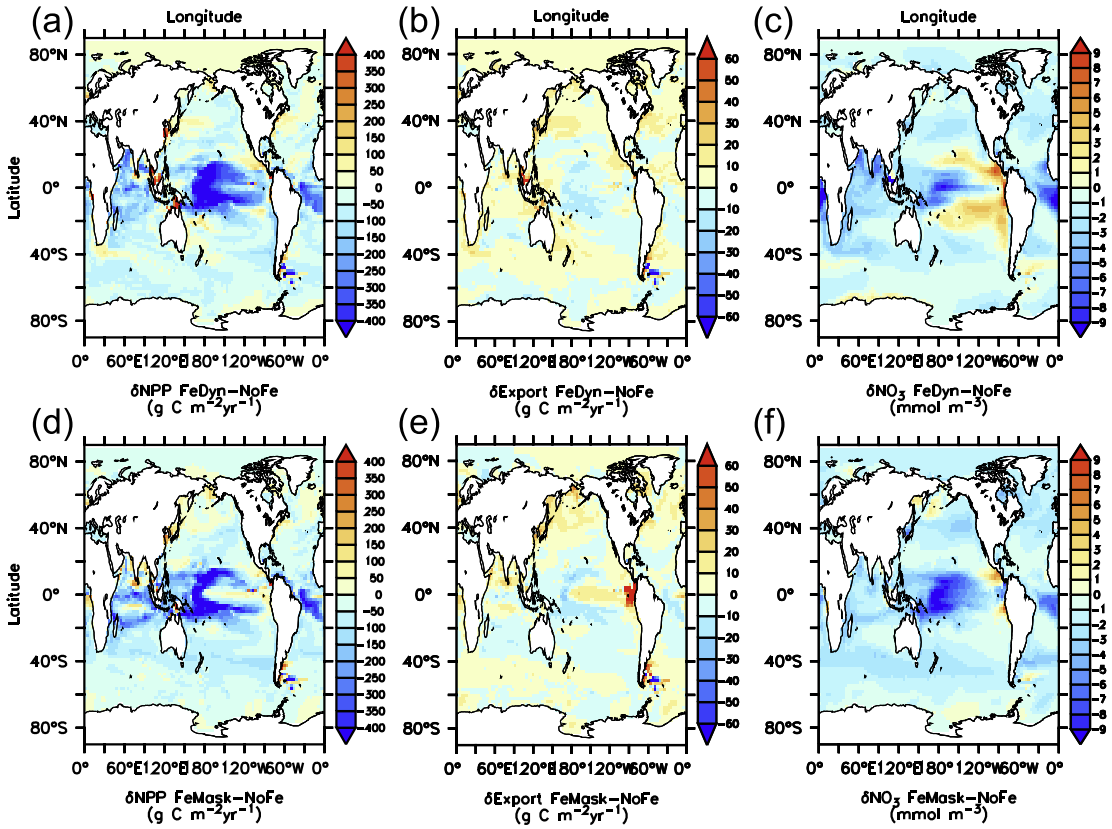

Figure 3. Difference of annual mean NPP, Export and euphotic zone NO_3 concentration between calibrations. Note that after iron limitation is introduced, not all regions reveal a reduced NPP (e.g. temperate region in the Southern Ocean, North Pacific and East Equatorial Pacific). This is due to the strong grazing control on phytoplankton in those regions in NoFe.

cost function might not be sufficient to constrain the model dynamics in this region.

4. Conclusions and future work

This study describes a range of advantages of explicitly including iron in a global biogeochemical ocean model by comparing calibrated models that employ a hierarchy of different iron implementations. The models with explicit consideration of iron perform better compared to the model without iron in two respects: (a) improved representation of macro-nutrient and oxygen distributions with around a 4.5% reduction of total model misfit, and (b) generally improved estimations of global indicators of biogeochemical cycles. The inclusion of a calibrated iron module produces more realistic global NPP, which is around 33% smaller than in the model without iron. It also produces a more realistic suboxic volume (O_2 is less than 5 mmol m^{-3}), which is reduced by 72.0 (43.3)% in the dynamic iron model (iron mask model) compared to the no-iron model. A more realistic suboxic volume benefits the global nitrate distribution by constraining denitrification rates and is achieved in the iron models via a combination of changes in both tropical export production and oxygen demand for organic particle remineralization.

We therefore conclude that it is worthwhile to include a dynamic iron cycle in biogeochemical models for research questions relevant to distributions of macro-nutrients, oxygen, NPP and very low oxygen thresholds. However, modeling of water masses with somewhat higher oxygen concentrations (regions with O_2 less than 50 mmol m^{-3}), remains problematic. Furthermore, for each individual model, calibration improved the model performance against chosen metrics compared to the hand-tuned model, regardless of the structural model complexity. In our case, this improvement was generally larger than the differences between the calibrated models differing in the representation of the iron cycle. This is an important point, because not only does inclusion of iron improve the model performance, but also does the calibration itself offer substantial model improvements that cannot generally be achieved with hand-tuning. Systematic model calibration also reveals, which parameters are portable across biogeochemical model structures (in our instante, the rate of increase of the sinking speed with depth) and which are not (e.g. the grazing rate).

We also found that the iron scavenging parameter in our calibration is less well constrained then other parameters. This may be improved when we use direct iron measurements as an extra observational constraint. However, in order to better reproduce observed patterns of dissolved iron, we may need a

more sophisticated ligands parameterization, as indicated by Völker and Tagliabue (2015). Ligands can keep dissolved iron in solution instead of being scavenged (Gledhill 2012), which can directly impact the iron scavenging parameter estimation. As Buck *et al* (2015, 2017) point out, ligands have different iron-binding strengths and life spans; Pham and Ito (2018, 2019) suggest ligands may be responsible for subsurface maxima of dissolved iron. Our model assumes a constant ligand concentration, which probably underestimates the control of ligands on our model biogeochemistry. It would be interesting to calibrate a more complex representation of ligands in the model in the future and to investigate if this can further improve the model performance.

The surface nutrient pathways differ across calibrated models, particularly with respect to total NPP. These differences arise from both the differences in model structure and the calibrated parameter values. Whether different nutrient pathways result in significantly different model responses to transient forcing (i.e. a changing climate) remains an open question. The fact that different optimal nutrient pathways arise in models of different complexity and model structure despite an identical calibration objective, demonstrates that surface processes such as microbial loop recycling cannot straightforwardly be constrained by seasonally cycling patterns of biogeochemical tracer concentrations. Lastly, the high sensitivity of simulated tropical Pacific NPP and Export to model structure emphasizes the importance of iron limitation in this region.

After the benefit of including an explicit representation of the iron cycle has been demonstrated in the current study, future work will investigate the effect of including direct iron measurements in the calibration process of such models. The impact of calibration will also be studied with respect to the simulated biogeochemical responses under a changing climate.

Acknowledgments

This work is supported by the Helmholtz Research School for Ocean System Science and Technology (HOSST, Germany) at GEOMAR Helmholtz Centre for Ocean Research Kiel (VH-KO-601). We acknowledge discussions within the Biogeochemical Modeling and Marine Chemistry research units at GEOMAR, Dalhousie University, and University of Liverpool. We in particular thank Iris Kriest (GEOMAR), Alessandro Tagliabue (Liverpool) and Allison Chuan (Dalhousie) for continuous support. We also acknowledge the computing resources provided by the North German Supercomputing Alliance (HLRN) and Christian-Albrechts-University of Kiel (CAU).

Data availability statement

The model data that support the findings of this study will be openly available on OceanRep—GEOMAR Repository (https://data.geomar.de/thredds/catalog/open_access/ yao_et_al_2019_erl/catalog.html). The source code of TMM is available at https://doi.org/10.5281/zenodo.1246300. The code of our calibration framework are available from the corresponding author upon reasonable request.

ORCID iDs

Karin F Kvale https://orcid.org/0000-0001-8043-5431

References

Anderson T R 2005 Plankton functional type modelling: running before we can walk? *J. Plankton Res.* **27** 1073–81

Berelson W M 2001 Particle settling rates increase with depth in the ocean *Deep Sea Res.* II **49** 237–51

Bopp L *et al* 2013 Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models *Biogeosciences* **10** 6225–45

Boyd P W and Ellwood M J 2010 The biogeochemical cycle of iron in the ocean *Nat. Geosci.* **3** 675–82

Boyd P W *et al* 2007 Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions *Science* **315** 612–7

Bucciarelli E, Ridame C, Sunda W G, Dimier-Hugueney C, Cheize M and Belviso S 2013 Increased intracellular concentrations of DMSP and DMSO in iron-limited oceanic phytoplankton Thalassiosira oceanica and Trichodesmium erythraeum *Limnol. Oceanogr.* **58** 1667–79

Buck K N, Lohan M C, Sander S G, Hassler C and Pižeta I 2017 Editorial: organic Ligands—a key control on trace metal biogeochemistry in the ocean *Frontiers Marine Sci.* **4** 1–2

Buck K N, Sohst B and Sedwick P N 2015 The organic complexation of dissolved iron along the U.S. GEOTRACES (GA03) North Atlantic section *Deep Sea Res.* II **116** 152–65

Calbet A and Landry M R 2004 Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems *Limnol. Oceanogr.* **49** 51–7

Carr M-E *et al* 2006 A comparison of global estimates of marine primary production from ocean color *Deep Sea Res.* II **53** 741–70

Cullen J J 1995 Status of the iron hypothesis after the open-ocean enrichment experiment1 *Limnol. Oceanogr.* **40** 1336–43

Dunne J P, Sarmiento J L and Gnanadesikan A 2007 A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor *Glob. Biogeochem. Cycles* **21** n/a–n/a

Eby M *et al* 2013 Historical and idealized climate model experiments: an intercomparison of Earth system models of intermediate complexity *Clim. Past* **9** 1111–40

Friedrichs M A M *et al* 2007 Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups *J. Geophys. Res.* **112** C08001

Galbraith E D, Gnanadesikan A, Dunne J P and Hiscock M R 2010 Regional impacts of iron-light colimitation in a global biogeochemical model *Biogeosciences* **7** 1043–64

Garcia H E, Locarnini R A, Boyer T P, Antonov J I, Baranova O K, Zweng M M, Reagan J R and Johnson D R 2013b *World Ocean Atlas 2013. Vol 4: Dissolved Inorganic Nutrients (Phosphate, Nitrate, Silicate)* ed S Levitus and A Mishonov vol 4 (Washington, DC : US Government Printing Office) p 25 (www.nodc.noaa.gov/OC5/indprod.html)

Garcia H E, Locarnini R A, Boyer T P, Antonov J I, Mishonov A V, Baranova O K, Zweng M M, Reagan J R and Johnson D R 2013a *World Ocean Atlas 2013. Vol 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation* NOAA Atlas NESDIS 75 vol 3 (Washington, DC: US Government Printing Office) p 27 (www.nodc.noaa.gov/OC5/indprod.html)

Gledhill M 2012 The organic complexation of iron in the marine environment: a review *Frontiers Microbiol.* **3** 1–17

Griffies S M, Harrison M J, Pacanowski R C and Rosati A 2004 A technical guide to MOM4 *GFDL Ocean Group Tech. Rep.* **5** 371 (www.gfdl.noaa.gov/~fms)

Hansen N 2006 The CMA evolution strategy: a comparing review *Towards a New Evolutionary Computation* vol 192 ed L Jose A *et al* (Berlin: Springer) pp 75–102

Holland W R, Chow J C and Bryan F O 1998 Application of a third-order upwind scheme in the NCAR ocean model* *J. Clim.* **11** 1487–93

Honeyman B D, Balistrieri L S and Murray J W 1988 Oceanic trace metal scavenging: the importance of particle concentration *Deep Sea Res.* A **35** 227–46

Honjo S, Manganini S J, Krishfield R A and Francois R 2008 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** 217–85

Jackson G A, Checkley D M and Dagg M 2015 Settling of particles in the upper 100m of the ocean detected with autonomous profiling floats off California *Deep Sea Res.* I **99** 75–86

Jacq V, Ridame C, L'Helguen S, Kaczmar F and Saliot A 2014 Response of the unicellular diazotrophic cyanobacterium crocosphaera watsonii to iron limitation *PLoS One* **9** e86749

Joos F, Sarmiento J L and Siegenthaler U 1991 Estimates of the effect of Southern Ocean iron fertilization on atmospheric CO_2 concentrations *Nature* **349** 772–5

Keller D P, Oschlies A and Eby M 2012 A new marine ecosystem model for the university of victoria earth system climate model *Geosci. Model Dev.* **5** 1195–220

Khatiwala S 2007 A computational framework for simulation of biogeochemical tracers in the ocean *Glob. Biogeochem. Cycles* **21** n/a–n/a

Körtzinger A, Hedges J I and Quay P D 2001 Redfield ratios revisited: removing the biasing effect of anthropogenic CO_2 *Limnol. Oceanogr.* **46** 964–70

Kriest I 2017 Calibration of a simple and a complex model of global marine biogeochemistry *Biogeosci. Discuss.* 1–28

Kriest I and Oschlies A 2015 MOPS-1.0: towards a model for the regulation of the global oceanic nitrogen budget by marine biogeochemical processes *Geosci. Model Dev.* **8** 2929–57

Kriest I, Oschlies A and Khatiwala S 2012 Sensitivity analysis of simple global marine biogeochemical models *Glob. Biogeochem. Cycles* **26** n/a–n/a

Kriest I, Sauerland V, Khatiwala S, Srivastav A and Oschlies A 2017 Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0) *Geosci. Model Dev.* **10** 127–54

Kvale K F, Khatiwala S, Dietze H, Kriest I and Oschlies A 2017 Evaluation of the transport matrix method for simulation of ocean biogeochemical tracers *Geosci. Model Dev.* **10** 2425–45

Kvale K F, Meissner K J and Keller D P 2015 Potential increasing dominance of heterotrophy in the global ocean *Environ. Res. Lett.* **10** 074009

Kvale K F, Turner K E, Landolfi A and Meissner K J 2019 Phytoplankton calcifiers control nitrate cycling and the pace of transition in warming icehouse and cooling greenhouse climates *Biogeosciences* **16** 1019–34

Landolfi A, Dietze H, Koeve W and Oschlies A 2013 Overlooked runaway feedback in the marine nitrogen cycle: the vicious cycle *Biogeosciences* **10** 1351–63

Laufkötter C *et al* 2015 Drivers and uncertainties of future global marine primary production in marine ecosystem models *Biogeosciences* **12** 6955–84

Laufkötter C *et al* 2016 Projected decreases in future marine export production: the role of the carbon flux through the upper ocean ecosystem *Biogeosciences* **13** 4023–47

Lerner P *et al* 2016 Corrigendum to 'Testing models of thorium and particle cycling in the ocean using data from station GT11-22 of the U.S. GEOTRACES North Atlantic section' [Deep-Sea Res. I 113 (2016) 57-79] *Deep Sea Res.* I **118** 101

Löptien U and Dietze H 2017 Effects of parameter indeterminacy in pelagic biogeochemical modules of Earth System Models on projections into a warming future: the scale of the problem *Glob. Biogeochem. Cycles* **31** 1155–72

Lutz M J, Caldeira K, Dunbar R B and Behrenfeld M J 2007 Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean *J. Geophys. Res.* **112** C10011

Moore C M *et al* 2013 Processes and patterns of oceanic nutrient limitation *Nat. Geosci.* **6** 701–10

Nickelsen L, Keller D P and Oschlies A 2015 A dynamic marine iron cycle module coupled to the University of Victoria Earth System Model: the Kiel Marine Biogeochemical Model 2 for UVic 2.9 *Geosci. Model Dev.* **8** 1357–81

Pasquier B and Holzer M 2017 Inverse-model estimates of the oceanas coupled phosphorus, silicon, and iron cycles *Biogeosciences* **14** 4125–59

Pauly D and Christensen V 1995 Primary production required to sustain global fisheries *Nature* **374** 255–7

Pham A L D and Ito T 2018 Formation and maintenance of the GEOTRACES subsurface-dissolved iron maxima in an ocean biogeochemistry model *Glob. Biogeochem. Cycles* **32** 932–53

Pham A L D and Ito T 2019 Ligand binding strength explains the distribution of iron in the north atlantic ocean *Geophys. Res. Lett.* **46** 7500–8

Pitchford J 1999 Iron limitation, grazing pressure and oceanic high nutrient-low chlorophyll (HNLC) regions *J. Plankton Res.* **21** 525–47

Price N M, Ahner B A and Morel F M M 1994 The equatorial Pacific Ocean: Grazer-controlled phytoplankton populations in an iron-limited ecosystem1 *Limnol. Oceanogr.* **39** 520–34

Primeau F W, Holzer M and DeVries T 2013 Southern Ocean nutrient trapping and the efficiency of the biological pump *J. Geophys. Res.: Oceans* **118** 2547–64

Schlosser C, Klar J K, Wake B D, Snow J T, Honey D J, Woodward E M S, Lohan M C, Achterberg E P and Moore C M 2014 Seasonal ITCZ migration dynamically controls the location of the (sub)tropical Atlantic biogeochemical divide *Proc. Natl Acad. Sci.* **111** 1438–42

Stock C A, John J G, Rykaczewski R R, Asch R G, Cheung W W L, Dunne J P, Friedland K D, Lam V W Y, Sarmiento J L and Watson R A 2017 Reconciling fisheries catch and ocean productivity *Proc. Natl Acad. Sci.* **114** E1441–9

Tagliabue A *et al* 2016 How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Global Biogeochem. Cycles* **30** 149–74

Tagliabue A *et al* 2010 Hydrothermal contribution to the oceanic dissolved iron inventory *Nat. Geosci.* **3** 252–6

Timmermans K R, van der Wagt B and de Baar H J W 2004 Growth rates, half-saturation constants, and silicate, nitrate, and phosphate depletion in relation to iron availability of four large, open-ocean diatoms from the Southern Ocean *Limnol. Oceanogr.* **49** 2141–51

Volk T and Hoffert M I 1985 Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes *The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present* Chapman Conference Papers, 1984 (*Geophysical Monograph* vol 32) ed E T Sundquist and W S Broecker (Washington, DC: American Geophysical Union) pp 99–110

Völker C and Tagliabue A 2015 Modeling organic iron-binding ligands in a three-dimensional biogeochemical ocean model *Mar. Chem.* **173** 67–77

Weaver A J *et al* 2001 The UVic earth system climate model: model description, climatology, and applications to past, present and future climates *Atmos. Ocean* **39** 361–428