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

To cite this article: Wanxuan Yao et al 2019 Environ. Res. Lett. 14 114009

View the article online for updates and enhancements.
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

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 NO₃, PO₄ and O₂ 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 NO₃, PO₄ and O₂ 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 CO₂, 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, PO₄, 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 NO₃, PO₄ and O₂, 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 NO₃, PO₄ and O₂ 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 √.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Observation</th>
<th>Cali. bound</th>
<th>Uncalibrated value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>Maximum growth rate at 0 °C</td>
<td>0.04–0.60$^a$</td>
<td>0.30–0.75</td>
<td>0.600√</td>
<td>0.600</td>
</tr>
<tr>
<td>$\mu_p$</td>
<td>Microbial loop recycling rate</td>
<td>0.001–0.035</td>
<td>0.015√</td>
<td>0.015√</td>
<td>μmol Fe m$^{-3}$</td>
</tr>
<tr>
<td>$k_{py}$</td>
<td>Ordinary phyto. half-saturation of Fe-limitation</td>
<td>0.035–1.14$^{b,c}$</td>
<td>0.04–1.14</td>
<td>—</td>
<td>0.1$^c$</td>
</tr>
<tr>
<td>$k_{py}^{Fe}$</td>
<td>Maximum $k_{py}$ $^b$</td>
<td>0.035–1.14$^{b,c}$</td>
<td>0.04–1.14</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$k_{dz}$</td>
<td>Diazotrophs half-saturation of Fe-limitation</td>
<td>6.29–17.2$^d$</td>
<td>0.05–1.50</td>
<td>—</td>
<td>0.1$^e$</td>
</tr>
<tr>
<td>$g_0$</td>
<td>Maximum grazing rate at 0 °C</td>
<td>0.06–1.9$^d$</td>
<td>0.2–2.0</td>
<td>0.4√</td>
<td>0.4√</td>
</tr>
<tr>
<td>$\omega_s$</td>
<td>Sinking speed slope</td>
<td>0.024–0.107$^{b,h}$</td>
<td>0.001–0.160</td>
<td>0.06√</td>
<td>0.06√</td>
</tr>
<tr>
<td>$K_{pm}$</td>
<td>POM dependent scavenging</td>
<td>0.079–6.62$^{d,i}$</td>
<td>0.225–6.750</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$R_{ON}$</td>
<td>Molar O: N ratio</td>
<td>8.1–11.2$^k$</td>
<td>8.0–12.0</td>
<td>10√</td>
<td>10√</td>
</tr>
</tbody>
</table>

$^g$ Berelson (2001).

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.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$J_{SO_3}$</th>
<th>$J_{PO_4}$</th>
<th>$J_{O_2}$</th>
<th>$J_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoFe</td>
<td>0.1575</td>
<td>0.1305</td>
<td>0.1786</td>
<td>0.4667</td>
</tr>
<tr>
<td>FeMask</td>
<td>0.1537</td>
<td>0.1293</td>
<td>0.1803</td>
<td>0.4632</td>
</tr>
<tr>
<td>FeDyn</td>
<td>0.1457</td>
<td>0.1254</td>
<td>0.1730</td>
<td>0.4442</td>
</tr>
<tr>
<td>Hand-tuned</td>
<td>0.1838</td>
<td>0.1389</td>
<td>0.1635</td>
<td>0.4862</td>
</tr>
</tbody>
</table>

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:

$$J_T = \sum_{j=1}^{M} J_j = \sum_{j=1}^{M} \frac{1}{\Omega_j} \sqrt{\sum_{i=1}^{N} (m_{ij} - o_{ij})^2 V_i / V_T}, \quad M = 3,$$

(1)

where $j = 1,.., M = 3$ denotes tracers NO$_3$, PO$_4$ and O$_2$; and $i = 1, ..., N$ 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; $o_{ij}$ is the global average observed concentration of the respective tracer; $m_{ij}$ and $o_{ij}$ 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.

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 $J_{SO_3}$ 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 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).
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 $J_D$. 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), export production (Export), and denitrification (Denit) using a somewhat simpler biogeochemistry coupled to a circulation that can optimally represent the tracer concentrations of each tracer, which is on par with the calibration used already as part of the constraint for obtaining a 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 figure 2, Honjo et al (2008, $F_{\text{POC}}$ 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 OMZ30 Conversion between different elements followed C: N: P = 106: 16: 1. 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 $J_D$ 4.8% smaller than NoFe, and 4.1% smaller than FeMask (table 2). Also for the individual components ($J_{\text{NO}_3}$, $J_{\text{PO}_4}$, and $J_D$), FeDyn provides the smallest misfit against observations. The most significant difference is in $J_{\text{NO}_3}$, with FeDyn having a 7.5 (5.2)% smaller $J_{\text{NO}_3}$ than NoFe (FeMask). This improvement may be related to the improvement in OMZ5, and associated Denit, in FeDyn, described below. However, even the smallest $J_D$ (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.
(see figure A1, available online at stacks.iop.org/ERL/14/114009/mmedia), which would also be consistent with the considerably smaller PO4 misfit obtained by Pasquier and Holzer (2017) for their PO4-constrained circulation field. In our model, annually-averaged basin vertical profiles of macro-nutrients and O2 (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 NO3 and PO4 and an under-estimation of O2 in the deep ocean basins. The Pacific O2 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 (FPOC: 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 FPOC 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 NO3 (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 O2 and NO3 distributions emphasized the improvement of fNO3, via improved OMZ5 volume. In contrast, there is no such control via simulated NO3 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 JT 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

### Table 3. Calibrated parameter value and their resulting fluxes ranges with 1% of misfit.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>NoFe</th>
<th>FeMask</th>
<th>FeDyn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluxes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPP</td>
<td>75.9</td>
<td>70.1–81.4</td>
<td>49.1</td>
</tr>
<tr>
<td>Export</td>
<td>6.3</td>
<td>5.8–6.4</td>
<td>6.8</td>
</tr>
</tbody>
</table>

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

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 prior parameter range used as input in the calibration.

Half-saturation of iron and the iron scavenging rate.

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).

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

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
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
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 instance, 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 than 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 TR 2005 plankton functional type modelling: running before we can walk? J. Plankton Res. 27 1073–81
Berelson WM 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 PW and Ellwood MJ 2010 The biogeochemical cycle of iron in the ocean Nat. Geosci. 3 675–82
Buccionelli E, Ridame C, Sunda WG, D¨um¨urner-Huguey C, Cheze M and Belviso S 2013 Increased intracellular concentrations of DMSP and DMSO in iron-limited oceanic phytoplankton Thalassiosira oceanica and Trichodendsium erythraeum Limnol. Oceanogr. 58 1667–79
Buck KN, Lohan MC, Sander SG, Hassler C and Pihlaja I 2017 Editorial: organic Ligands—a key control on trace metal biogeochemistry in the ocean Frontiers Marine Sci. 4 1–2
Carr ME et al 2006 A comparison of global estimates of marine primary production from ocean color Deep Sea Res. II 53 741–70
Cullen JJ 1995 Status of the iron hypothesis after the open-ocean enrichment experiment1 Limnol. Oceanogr. 40 1336–43
Dunne JP, Sarmiento JL 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 MAM et al 2007 Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups J. Geophys. Res. 112 C08001
Galbraith ED, Gnanadesikan A, Dunne JP and Hiscock MR 2010 Regional impacts of iron-light colimitation in a global biogeochemical model Biogeosciences 7 1043–64