https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Tuning ocean biogeochemistry for an Earth system model: approaches, challenges and impact on model performance

Iris Kriest[1], Tronje Kemena[1], and Haichao Guo[1,2]

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstrasse 1-3, 24148 Kiel, Germany
[2]Department of Oceanography, School of Ocean and Earth Science and Technology (SOEST), University of Hawaii at Mānoa, 1000 Pope Road, Honolulu, HI, 96822 USA

Correspondence: Iris Kriest (ikriest@geomar.de)

Abstract. The parameterisation of biogeochemical models, when simulated within global ocean models, poses many challenges, among them those related to the calibration of rate constants (parameters) to the different functional groups. Global coupled general circulation models are computationally expensive to run, which limits the number of experiments and increases the risk of issues such as unrealistic tracer distributions. Efficient "surrogate" circulations and methods to accelerate the spin up of global models have now become available, and provide the possibility to simulate a biogeochemical model globally with only moderate computational resources. When coupled to an optimisation algorithm these surrogate models even allow the calibration of model parameters against observations in a coherent and systematic way. We here investigate whether biogeochemical model parameters, that were objectively calibrated (optimised) in an efficient surrogate circulation against a wide range of observations, can be transferred to the same biogeochemical model embedded in an Earth system model (ESM), without loosing too much of the model's improvement through optimisation. Comparison of biogeochemical results obtained from two circulation model environments shows that insolation as well as circulation can play a role especially for simulated primary production, and, regionally, surface concentrations such as chlorophyll; however, comparison to an earlier biogeochemical setup of the ESM shows that biogeochemical model parameters play an even larger role on certain biogeochemical processes. Export production is mainly affected by circulation, in line with earlier model studies. Deep particle flux is, however, mostly affected by the biogeochemical model parameters, especially the particle flux length scale. Given the relatively large role of biogeochemical model parameters, their optimisation in a surrogate circulation and subsequent transfer to the ESM leads to an improved performance of the ESM. This suggests that prior calibration of parameters in "light-weight" circulations can support model development, before performing computationally expensive simulations with ESMs. Model improvement with respect to organic matter, for which only sparse observations exist, becomes most visible when applying non-parametric metrics, that avoid interferences from data patchiness and episodic events. These metrics, which also allow for a clearer distinction between the performance of different biogeochemical models, merit further exploitation with regard to model assessment and calibration.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

1 Introduction

Global biogeochemical ocean models rely on many assumptions, among them those related to the model parameters (constants). Simulated organic model components (variables), such as phytoplankton, zooplankton or detritus, usually integrate over many organism groups, characterised by different growth rates, decay constants, etc.. These parameters therefore represent a value representative of an "average global organism". Hence, they are associated with a large parametric uncertainty, in contrast to physical or chemical constants such as the gravitational constant or the molar gas constant. During model development the biogeochemical parameters are often modified to obtain a better match of simulated biogeochemical variables to observed quantities, such as global nutrient distributions from climatologies or surface chlorophyll obtained from remote sensing.

The simulated global distribution of biogeochemical variables results from the combined effects of the biogeochemical model as well as the circulation model, with the latter introducing very long, millennial time scales to the coupled system (Wunsch and Heimbach, 2008). Because of this, a full adjustment (an equilibrium solution) of the model to its biogeochemical and physical setup requires simulations of thousands of years; only then will the model be independent of its initial conditions and allow an unbiased assessment against observations.

Because of their many components Earth system models (ESMs) are characterised by a large computational demand, which hinders a systematic tuning and sensitivity analysis of ocean biogeochemistry (and related parameters), especially after a complete spin-up towards equilibrium. Despite increasing computational resources, biogeochemical models implemented in ESMs or in highly resolved ocean models therefore mostly lack a systematic sensitivity analysis with respect to their model parameters. Calibration, i.e., the objective optimisation of model parameters to obtain a better match to observations, is the exception rather than the rule. Instead, model parameters are often set by "expert judgement" (Seferian et al., 2020). Model tuning - the manual adjustment of model parameters or boundary conditions to obtain a better match to observations - is carried out either in separate experiments, or "underway", i.e., during model spin-up (e.g., Lindsay et al., 2014; Stock et al., 2020). As summarised by Seferian et al. (2020), who examined the setup and performance of a large suite of ESMs "There is no consensus between modelling groups on how model calibration or tuning takes place in the model preparation. Depending on modelling group, the calibration or tuning is either included in the model development or during the spin-up procedure [...]". However, even tuning and/or structural biogeochemical model development do not guarantee improved model performance for all model tracers and diagnostics, as shown in the comprehensive study by Seferian et al. (2020).

Unfortunately, biogeochemical ocean models embedded in ESMs differ in many aspects with regard to the circulation and the biogeochemical model, which results in a large range of simulated model outcomes; a situation that has not much improved in the past decade. For example, recent studies often show a large range of global primary and export production (e.g., Seferian et al., 2020; Doney et al., 2024). Also, Walker and Palevsky (2025) have shown that there is a large variability in simulated present day and future deep particle flux, which is a good predictor of carbon storage in the ocean (Wilson et al., 2022). The large variety of combinations of physical and biogeochemical model configurations complicates inter-model comparison, as it is hardly possible to trace differences in model outcome back to either circulation or biogeochemical model properties. This

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

was also discussed by Seferian et al. (2020), who noted that "[...] it remains difficult to identify which model updates are responsible for a given improvement."

Simpler circulation models, that are more coarsely resolved or omit details of circulation and/or aspects of the Earth system can provide a more efficient way to tune or even calibrate (through objective optimisation) parameters of a biogeochemical model in the presence of large scale circulation.

Tools are now available that allow the relatively fast and efficient spin-up of global biogeochemical models (e.g., Khatiwala, 2007; Khatiwala et al., 2012; Khatiwala, 2024). These can be combined with objective optimisation, to adjust the parameters of a biogeochemical model such that it improves its fit to a variety of observations (e.g., Kriest et al., 2023). Spinning up and tuning a biogeochemical ocean model in a computationally more efficient environment is, to some extent, similar to the offline (ocean-only) tuning approaches applied to biogeochemical models of ESMs (Seferian et al., 2020; Yool et al., 2021). The adjusted biogeochemical model (and, potentially, its equilibrated three-dimensional state) can then be transferred back into the computationally more demanding ESM. However, optimisation of biogeochemical model parameters can partially adapt to and compensate for circulation biases (Pasquier et al. , 2023), with potential consequences for its transfer to a different circulation. Hence, there is no guarantee that a biogeochemical model, that was calibrated in one circulation, will perform well in a different circulation (Kriest et al., 2020; Pasquier et al. , 2023), and adjustments may have to be made again, when simulating the biogeochemical model in the ESM (e.g., Sellar et al., 2019; Yool et al., 2021).

To investigate the problems associated with biogeochemical model calibration for its implementation in an ESM, we here investigate the following two key questions:

- What are the effects of transferring a biogeochemical model that was calibrated in an efficient (surrogate) circulation to a state-of-the art circulation model embedded in an ESM?
- Does the calibrated model, when simulated within the ESM, finally perform better than an earlier version of the same model?

Our approach is based on a biogeochemical model whose parameters were calibrated against observations in a "light-weight" global climatological ocean model that represents ocean circulation at a relatively coarse spatial scale, but in a computationally efficient way (Khatiwala et al., 2012). Calibration was carried out in a systematic way, using a quasi-evolutionary optimisation algorithm and observations relevant for seven model state variables, after a model spin-up of 3000 years (Kriest et al., 2023). We then applied a version of this calibrated biogeochemical model to the ocean climatology of an Earth system model (ESM), that resolves the spatial scale and physical processes with much more detail, but whose spin-up is also computationally more demanding. The results of the modified ESM are compared to those of an earlier model version, that was simulated with different biogeochemical model parameters (Chien et al., 2022a). Both configurations of the ESM were simulated under pre-industrial and historical conditions, following a spin-up of 500 years. To avoid any effects arising from model adjustment during spin-up, and to allow for a straightforward comparison of different model versions, simulations with the light-weight model were carried out over the same length as with the ESM. Using this set of experiments, where the same biogeochemical

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

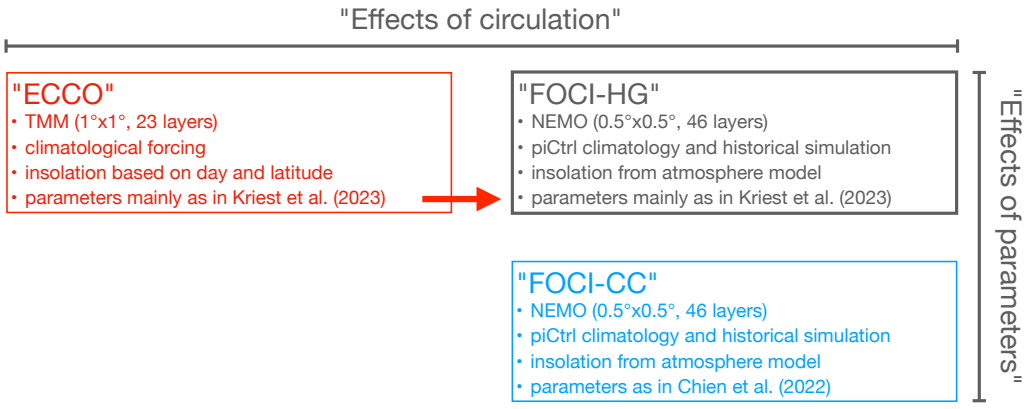

Figure 1. Schematic showing the experimental setup of this study,

model configuration is applied in different model environments (circulations) or in the same circulation, but with two different sets of model parameters, we thereby examine the effects of

- Transitioning the calibrated model from the coarser, light-weight circulation to the ESM ("Effects of circulation")
- Exchanging biogeochemical model parameters in the ESM under pre-industrial and historical conditions ("Effects of parameters")

The inter-model comparison is complemented by an extensive comparison of model results to observations, which also aims at identifying the most useful metrics for model assessment and calibration.

2 Methods

To address the above questions we apply two different circulation models dubbed in short "ECCO" and "FOCI", and two different sets of biogeochemical model parameters. One set of parameters was derived from an optimisation against observed biogeochemical tracers in circulation "ECCO", with a few minor modifications. This set of parameters was applied in both circulations ECCO and FOCI. The second set of model parameters was applied in an earlier configuration of FOCI by Chien et al. (2022a). By doing so, we can examine the effects of model circulation (or environment) via comparison of ECCO and FOCI using the same set of model parameters, and the effect of circulation by using the same circulation FOCI, but with the two different parameter sets. The approach is sketched in Fig. 1, and below we present the details of circulations, the biogeochemical parameter set, simulation setup and approaches to the assessment of model performance.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

2.1 Two circulation models

As an "offline" method the Transport Matrix Method (TMM) represents advection and mixing in the form of transport matrices (TMs), that were calculated from an ocean circulation model simulated to steady state. The method has been described in detail by Khatiwala (2007, 2018), and we here only describe the main features of this climatological "light-weight" circulation, which is hereafter dubbed "ECCO". The ECCO project provides circulation fields that yield a best fit to hydrographic and remote sensing observations over the period 1992 through 2001 (Stammer et al., 2004). The model is set up with a horizontal resolution of 1° and 23 levels in the vertical, varying between 10 m at the surface and 500 m in the deep ocean. In the biogeochemical model setup within ECCO insolation at the sea surface is calculated from day and latitude using standard astronomical formulas and assuming a constant planetary albedo (which includes the cloud cover fraction). Because of its high computational efficiency, TMM can be used to quickly spin up biogeochemical models into steady state; it therefore is also applicable to more costly optimisation of biogeochemical parameters at a global scale.

The physical ocean model component of the more complex ESM "FOCI" (Flexible Ocean and Climate Infrastructure) is built on NEMO version 3.6 (Madec and the NEMO team, 2016), with a nominal global ocean resolution of 1/2° on a tripolar grid (ORCA05) and 46 vertical levels, their thickness varying between 6 m at the surface to 250 m in the deep ocean. The ocean (and land) model is coupled to the atmospheric model ECHAM6.3, which features a detailed representation of atmospheric processes, among them the dynamic representation of cloud cover. Details of NEMO, its coupling to ice, the atmosphere, and land are described in Matthes et al. (2020). In contrast to the TMM, the ocean model explicitly calculates advective and diffusive transport, as well as exchange of momentum, heat, freshwater fluxes at the sea surface, thereby enabling it to simulate the ocean's response to changes in various types of external forcing (such as wind speed or ice coverage), and hence climate change. On the other hand, its finer resolution (compared to ECCO), and the explicit computation of ocean physics, together with other Earth system components, cause a much larger computational demand compared to ECCO. Hence, sensitivity experiments with biogeochemical model parameters, or even their optimisation are not feasible with this type of model. In the following we refer to this circulation as "FOCI".

A brief comparison of annual maximum mixed layer depths (MLD), defined as the depth where the density σ_θ differs 0.01 compared to that at 10 m depth, reveals some differences between ECCO and FOCI. Density and hence MLD in ECCO was calculated from monthly mean temperature and salinity. Areas of deep mixing occur in ECCO mainly in the Greenland and Norwegian Sea, as well as in the Southern Ocean between about 60°S to 40°S (Fig. 2, left panel). The tropics and subtropics are in general characterised by MLDs of maximal 50 m.

FOCI, because of its more dynamic physical processes and higher spatial and vertical resolution, shows a larger variability. Here, we calculated the maximum MLD in two different ways, with one of them considering temporal changes well below monthly means. Defining MLD in the same way as in ECCO (i.e., from monthly mean temperature and salinity) shows areas of deep mixing in the Weddell Sea, as well as in the Southern Ocean south-west of Australia (middle panel in Fig. 2). Compared to ECCO, MLDs deeper than 100 m occur in a larger domain in the southern and northern hemisphere. In total, deep mixing in FOCI appears somewhat more pronounced and extended than in ECCO. Large differences occur, however, when diagnosing

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

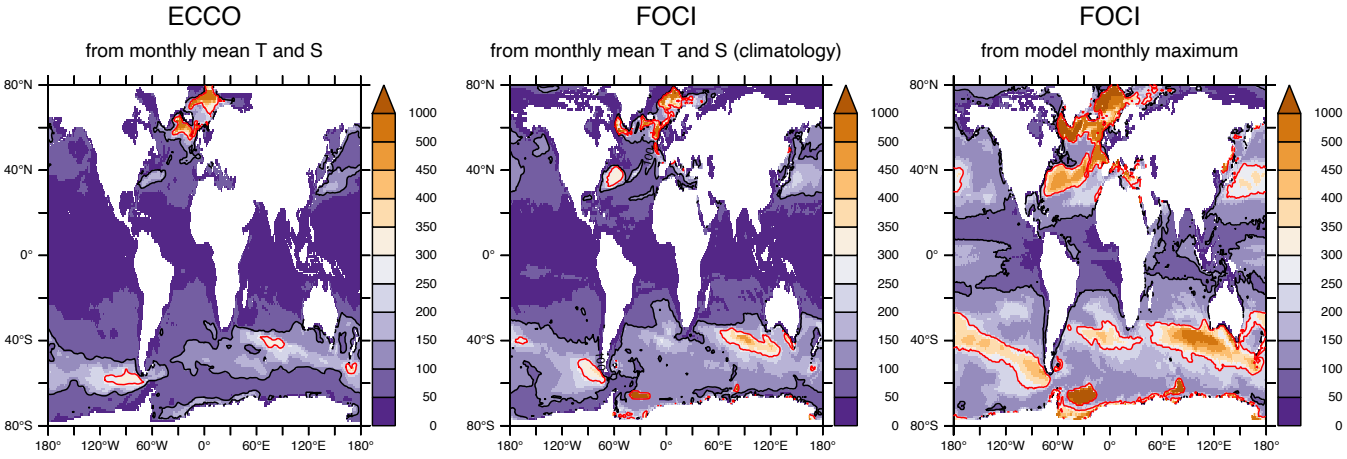

Figure 2. Annual maximum mixed layer depths calculated from σ_θ for ECCO and FOCI-HG. For ECCO, σ_θ has been calculated from monthly mean T and S, and the maximum depth of these twelve monthly values has been chosen. For FOCI we evaluated maximum mixed layer depth in two different ways: (i) as for ECCO from monthly mean T and S, where T and S were averaged for each month over the period 2005 and 2014 and (ii) from maximum monthly MLD of the period from 2005 to 2014 calculated online during model simulation. Hence, while (i) represents a treatment (or diagnostic) similar to that of ECCO, (ii) captures the full temporal variability of mixed layer depths in FOCI. Contour lines indicate levels of 100 m (black) and a depth of $100\,e^{1/b}$, where $b = 1.022589$ is the prescribed particle flux exponent of the model runs (red). Hence, the red line described the depth at which - at least theoretically - particle flux has declined to $1/e = \approx 37\%$ of that at 100 m (or above which $\approx 63\%$ of the export flux has been remineralised).

the maximum mixed layer depth from the monthly maxima determined online over a 10-year period (right panel in Fig. 2). This type of diagnostic also captures episodic events of deep mixing. The pronounced temporal variability of FOCI becomes evident, and indicates that over 10 years various deep mixing events may occur; also, maximum MLD is generally deeper than that calculated from monthly mean temperature and salinity. The cause for a potential overestimation of MLDs in FOCI can be found in a pronounced warm bias in the Southern Ocean, particularly in the upper to intermediate layers, as documented by Zeller and Martin (2024). This bias is linked to an inaccurate representation of Antarctic Circumpolar Current. The reduced vertical density gradient associated with this warm bias can promote overly deep mixed layers. In our simulations, this is consistent with the systematically deeper MLDs in the Southern Ocean in FOCI compared to the ECCO reanalysis.

A further difference that may affect the performance of biogeochemistry in ECCO and FOCI can be found in irradiance that fuels primary production at the sea surface. As noted above, for the simulation of biogeochemistry in ECCO we calculate photosynthetically available radiation (PAR) as a function of latitude and day, assuming cloud cover that is constant in space and time. In contrast, FOCI features a dynamic atmosphere that simulates variable cloud cover (see, e.g., Mauritsen et al., 2019). The effects of clouds are clearly visible in Fig. 3, which compares PAR in the two different model environments. Due to its simple parameterisation, PAR simulated in the ECCO configuration increases monotonically with latitude at summer (austral summer) solstice (June and December, respectively), whereas in FOCI PAR shows a quite strong spatial variance, with

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

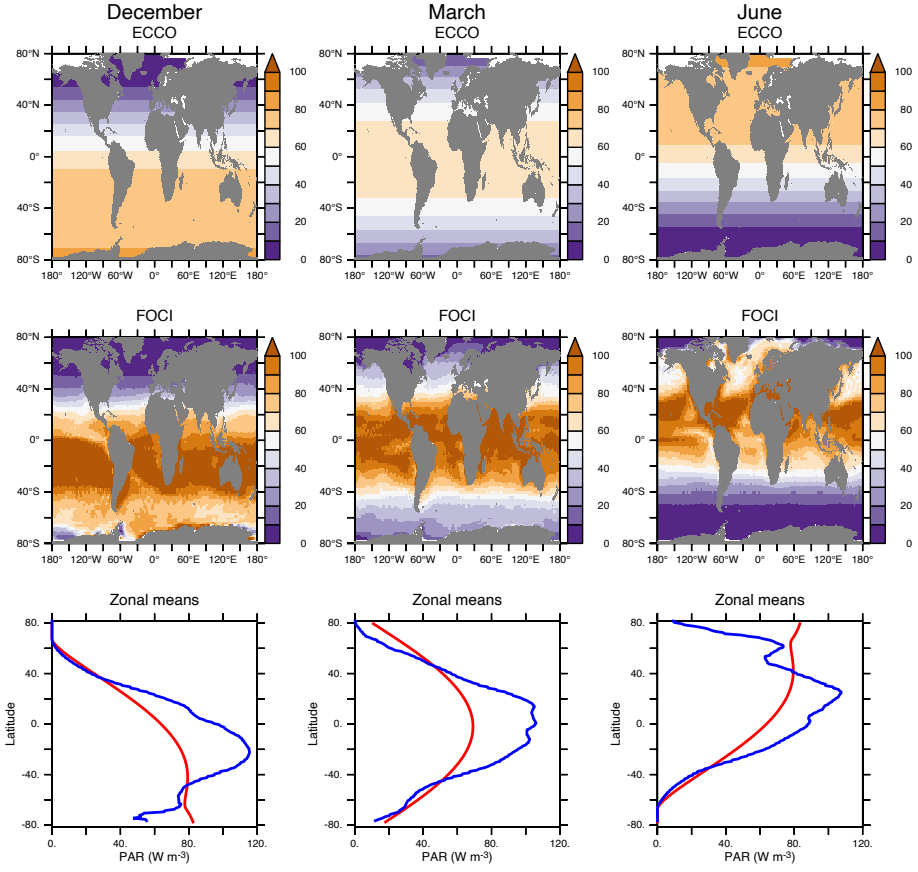

Figure 3. Monthly mean PAR (W m^{-2}) at the sea surface in FOCI-MOPS and ECCO-MOPS, for December, March and June. Values for FOCI-MOPS are climatological means from 2005-2014, those of ECCO-MOPS only for a single year. Lower panels show zonal mean values for FOCI-MOPS (blue) and ECCO-MOPS (red).

light decreasing even around summer (austral summer) solstice beyond $\approx 40°$N ($40°$S). Also, FOCI shows higher PAR in the tropics and subtropics. One explanation for the spatial variation in FOCI-MOPS can be found in cloud coverage, which is typically much higher in high latitudes than in tropics, and does not vary much with season, thereby reducing insolation in regions beyond $40°$N ($40°$S).

In summary, the two circulation models vary in many aspects, among them horizontal and vertical resolution, surface forcing, deep mixing and insolation, all of which may affect the biogeochemical model performance.

2.2 The biogeochemical model

In all setups we apply model "MOPS" (Kriest and Oschlies, 2015), which features the biogeochemical cycling of phosphorus, nitrogen and oxygen through phytoplankton, zooplankton, dissolved and particulate organic matter. Carbon is coupled to the cycling of these elements by assuming a constant C:P ratio of organic matter of 117 mol C:mol P, without any feedbacks of

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

carbon on the biogeochemical turnover of the core elements phosphorus, nitrogen and oxygen. MOPS and its predecessors have been originally designed for, and coupled to, the TMM, and various sensitivity experiments with respect to the effect of circulation (Kriest and Oschlies, 2013; Kriest et al., 2020) and biogeochemical model parameters (e.g., Kriest et al., 2012) have been carried out. The computational efficiency of the TMM also allowed the objective optimisation of MOPS' model parameters against observations of nutrients and oxygen (e.g., Kriest et al., 2017, 2020). A calibration of MOPS against observations of organic components, phosphate, nitrate and oxygen, within the circulation of ECCO has been described in detail by Kriest et al. (2023), and resulted in reduced biases of several tracers, especially of those of dissolved organic matter. In particular, calibration led to an adjustment of the particle flux length scale b to a deeper flux penetration, a higher half-saturation irradiance I_c of phytoplankton, a lower grazing rate μ_{ZOO} of zooplankton and a much lower release rate of dissolved organic matter σ_{DOP} (Kriest et al., 2023). These calibrated parameters form the basis for the new model simulations with FOCI presented here, with some slight modifications (see also Table 1).

MOPS by default applies an upstream scheme to calculate the flux of detritus through the water column. However, especially with a coarse vertical resolution, this scheme can become very diffusive, resulting in an additional numerical downward transport of detritus, and hence a longer realised remineralisation length scale than prescribed (Kriest and Oschlies, 2011). The numerical effect of the upstream scheme arises from the assumption that detritus is distributed homogeneously within a vertical grid box. It increases for very coarse grids. In consequence, when diagnosing the particle flux exponent b from simulated particle flux, numerical diffusion by an upstream scheme will lead to a reduction in (accomplished) b, compared to the (desired) b defined a priori by the parameter. The effect can be avoided by the assumption of an "implicit" profile of detritus within a vertical box that is consistent with the model's parameterisation of detritus' decay and sinking speed (Kriest and Oschlies, 2011). Because our model comparison addresses two model environments with very different vertical resolutions (23 layers in ECCO vs. 46 layers in FOCI; i.e. a twice as good a resolution in FOCI), we here additionally examine the effect of vertical discretisation with two different numerical configurations of MOPS coupled to ECCO. One configuration applies the (default) upstream scheme (hereafter dubbed "ECCO-U"), and one applies the parameterisation of implicit profiles by Kriest and Oschlies (2011, "ECCO-I"). Both configurations apply the same forcing and biogeochemical parameters. In the following, the general coupled system of ECCO and MOPS (regardless of the numerical representation of detritus) is here referred to as "ECCO-MOPS".

The first setup and simulation of MOPS in FOCI analysed in this study was presented by Chien et al. (2022a), and we here only briefly comment on its configuration, and refer to that study for a complete description, values of model parameters and assessment of model performance. In the following we refer to it as "FOCI-CC". We note that the biogeochemical model parameters of the configuration by Chien et al. (2022a) were also partly based on a previous calibration of MOPS against nutrients and oxygen (see Kriest et al., 2020), but adjusted for differences in circulation (Chien et al., 2022a, see also Table 1). The second simulation of MOPS in FOCI presented here is based on the same set of model parameter as ECCO-I and ECCO-U and is referred to as "FOCI-HG" (see Table 1). Both FOCI-CC and FOCI-HG apply an upstream scheme for the sinking of detritus. The general coupled system of FOCI and MOPS (regardless of the parameters applied) is, in the following, referred to as "FOCI-MOPS".

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Table 1. Biogeochemical model parameters of MOPS simulated with ECCO (TMM) and with FOCI. We here only list those that differ among the configurations, and refer the reader to Chien et al. (2022a) or Kriest et al. (2023) for the other model parameters of MOPS.

Parameter	FOCI-CC (Chien et al., 2022a)	FOCI-HG, ECCO this study	Meaning	Unit
Phytoplankton and nitrogen fixation				
μ_{NFix}	1.889	2.272	max. N-fixation rate	μmol N m^{-3} d^{-1}
I_{c}	9.653	22.139	half-sat. constant (PAR)	W m^{-2}
K_{PHY}	0.031	0.500	half-sat. constant (nutrients)	mmol P m-3
Zooplankton				
μ_{ZOO}	1.893	2.385	max. grazing rate	d^{-1}
K_{ZOO}	0.086	0.097[a]	half-sat. constant (phytoplankton)	mmol P m-3
Dissolved organic matter				
σ_{DOP}	0.150	0.000	fraction produced by sloppy grazing etc.	
λ_{DOP}	0.170	0.185	decay rate	y^{-1}
Detritus				
b	1.413	1.023	particle flux exponent	
Remineralisation and denitrification				
$R_{-\mathrm{O2:P}}$	165.08	169.25	stoichiometry	mol O_2:mol P
DIN_{min}	15.978	16.000	min. DIN for denitrification	mmol N m-3
K_{DIN}	23.104	23.084	half-sat. constant (DIN)	mmol N m-3
K_{O2}	1.066	1.146	half-sat. constant (O_2)	mmol O_2 m-3

[a] Note that this value was fixed to 0.086 mmol P m^{-3} in Kriest et al. (2023).

In MOPS, once sinking detrital phosphorus (and associated nitrogen and carbon) reaches the sea floor (the lower model boundary) it is partly buried, i.e., lost to the (pelagic) model system (see Kriest and Oschlies, 2013, for details). To prevent the model ocean from loosing phosphorus during its simulation, it has to be supplied back. Earlier versions of MOPS coupled to TMM scaled the re-supply of buried phosphorus and associated elements to the freshwater river runoff (Kriest and Oschlies, 2013), while in FOCI this supply is delivered via globally uniform supply at the sea surface (Chien et al., 2022a). We note that the supply by river runoff (as in MOPS coupled to TMM) introduces an implicit transfer of matter to the Atlantic Ocean, which, in proportion to its volume, receives a relatively large amount of runoff; this feature of coupling between burial and runoff can also introduce quite long time-scales to a coupled ocean-atmosphere system (Roth et al., 2014). To align all model versions with respect to the resupply of phosphorus, nitrogen and carbon we configured ECCO-MOPS presented here in the same way as applied in FOCI-MOPS, i.e., by supplying buried matter at the sea surface.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

2.3 Experimental setup and model analysis

Altogether we analyse six different experiments of MOPS coupled to a global circulation, two with ECCO-MOPS (namely, ECCO-U and ECCO-I) and two with FOCI-MOPS (FOCI-CC and FOCI-HG), each of the latter with two different forcings.

In the following, we describe the forcings used in these configurations. For the simulation of FOCI-MOPS we adopted the approach by Chien et al. (2022a): Following a 1500-year physical spin-up, MOPS coupled to FOCI was simulated for 665 years with repeating pre-industrial forcing (see also Chien et al., 2022a). Branching off from year 500 of this simulation, historical simulations following the CMIP6 protocol with prescribed pCO_2 were carried out for 165 years, representative of year 1850 to year 2014. These experiments are referred to as "pictrl" and "hist".

To consider the combined effects of circulation and biogeochemistry, earlier simulations and calibrations with ECCO-MOPS typically applied a spin up of 3000 years (or even longer) before model analysis and the evaluation of the cost function (e.g., Kriest and Oschlies, 2015; Kriest et al., 2020, 2023). For the present study, we aligned with the shorter spin up of FOCI-MOPS, and simulated ECCO-MOPS only over 665 years, prescribing an atmospheric pCO_2 of 280 ppm.

Because of the relatively short spin up length, deep ocean biogeochemistry may not have adjusted to the combined effects of the physical and biogeochemical model setup. We therefore mostly restrict our model analysis to the top 2000 m of each model domain, unless stated otherwise. Our analysis of biogeochemical properties is based on annual mean tracer concentrations and fluxes. Of these, we calculated the average over last 10 years (year 656 to year 665) of each model simulation. For the historical simulations with FOCI-MOPS, these correspond to the period from 2005 to 2014.

As noted above, FOCI-MOPS is based on a tripolar, irregular grid. Analysis of global averages, integrals and fluxes (Figures 6, 7, S3 and S4) was carried out on the original grid of FOCI-MOPS, using information on grid box volumes and areas. For model-data comparison (all other plots, as well as performance statistics presented in Table S1 and Table S2) we mapped output from FOCI-MOPS onto a regular grid of 1°, while maintaining its vertical resolution. Hence, for these comparisons the horizontal resolutions of ECCO-MOPS and FOCI-MOPS are the same, while the vertical grid differs.

2.4 Assessment of model performance

For the assessment of model performance we applied several metrics, all based on the post-processed model results mentioned above. Observational data used to assess model performance are the same as those used for optimisation of ECCO-MOPS in Kriest et al. (2023). The data sets include observations of nutrients and oxygen from climatologies (Garcia et al., 2006a, b), surface chlorophyll data derived from remote sensing (Malin, 2013), data for zooplankton biomass (Moriarty and O'Brien, 2013), for particulate organic phosphorous (POP; Martiny et al., 2014) and a compilation of dissolved organic phosphorus (DOP). Data sources and mapping of DOP observations, as well as conversion and further treatment of data is described in Kriest et al. (2023). In particular, all data have been mapped onto the ECCO grid. We note that observations of zooplankton, POP and especially DOP are very sparse in space and time. Therefore, they contain a high level of episodicity and patchiness, i.e., they are likely influenced by episodic and/or local phenomena such as storms, eddies and other (sub)mesoscale or short-term events.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

In addition to these tracers we also compare model results of alkalinity and DIC (for assessment of historical simulations) and preanthropogenic DIC (for preindustrial simulations) to the corresponding data by GLODAPv2.2016b (Lauvset et al., 2016). These data were mapped onto the ECCO grid in the same way as for phosphate or oxygen described above. Conversion of concentrations from μmol kg^{-1} (in GLODAP and FOCI-MOPS) to μmol L^{-1} was done by assuming a constant density of 1.024 kg L^{-1}.

We complement the assessment of simulated tracer concentrations by comparison of simulated detritus sinking flux to particle flux determined by sediment traps (Mouw et al., 2016). We only consider observed particle flux from sediment traps located at a minimum depth of 200 m which were deployed at least one year. Again we note that the observations may be subject to a high level of spatial variability, with particle flux being affected by various physical processes such as eddies, filaments, lateral advection etc.

2.4.1 Parametric performance indicators

Parametric performance indicators for each of the nine model tracers X include the RMSE(X), the non-biased RMSE'(X), the models' standard deviation divided by standard deviation of observations, and the correlation coefficient R(X), as described in Taylor (2001). We further calculate the global bias B(X). RMSE and B have units of tracers, and may vary over several orders of magnitude. Striving for a single, scalar metric across all tracers requires some normalisation. We therefore also calculate the normalised global bias absolute B(X)* (|B(X)| divided by the observed mean) and normalised RMSE(X)* (RMSE(X) divided by the observed mean). Being dimensionless, we can then add B(X)* and RMSE(X)* across multiple tracers X and obtain two scalar metrics $\sum_{X=1}^{N}$ B*(X) and $\sum_{X=1}^{N}$ RMSE*(X) that indicate overall performance of each model simulation. Because especially the organic tracers suffer from the sparseness, patchiness and episodicity we also evaluated non-parametric performance indicators, which are less likely to be influenced by these factors in two ways:

2.4.2 Non-parametric performance indicators based on the distribution of discrete concentration classes

As in Chien et al. (2022a), for all common model-data pairs at $i = 1, ..., n$ locations tracer concentrations x_i were been binned into $j = 1, ..., 50$ classes over a typical range. Their frequency (PDF) was calculated for the model (PDF$(x_j)_{\mathrm{mod}}$) and observations ((PDF$(x_j)_{\mathrm{obs}}$; see also Figures S8 to S11). From these discrete distributions we calculated the Hellinger distance HD(X) for a tracer type X

$$\mathrm{HD}(X)_{\mathrm{PDF}} = \sqrt{0.5 \sum_{j=1}^{j=50} \left(\mathrm{PDF}(x_j)_{\mathrm{mod}}^{0.5} - \mathrm{PDF}(x_j)_{\mathrm{obs}}^{0.5}\right)^2} \quad , \tag{1}$$

and the Bhattacharyya distance BD(X)

$$\mathrm{BD}(X)_{\mathrm{PDF}} = -\ln\left(1 - \mathrm{HD}(X)_{\mathrm{PDF}}^2\right) \quad , \tag{2}$$

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Like $\mathrm{B}(X)^*$ and $\mathrm{RMSE}(X)^*$ these metrics are are independent of the magnitude and unit of a tracer type X, and can therefore be summed over all N tracers to obtain an overall performance indicator of each model simulation, namely $\sum_{X=1}^{N} \mathrm{HD}(X)_{\mathrm{PDF}}$ and $\sum_{X=1}^{N} \mathrm{BD}(X)_{\mathrm{PDF}}$.

We note that the evaluation of non-parametric metrics based on PDFs can depend on the width of concentration classes (or the number of classes applied over the tracer range). To address this issue, we additionally calculated HD and BD based on kernel density estimates, as explained in the next section.

2.4.3 Non-parametric performance indicators based on kernel density estimates (KDEs)

The distributions of observed and simulated tracer concentrations x were calculated with kernel density estimates (KDEs) for every point i where both model and data are available. We first define a (dimensionless) kernel $k(x_i)$ around each point where both model and data exist:

$$\mathrm{k}(x, x_i) = \frac{1}{4\pi^2} \exp\left(-0.5 \left(\frac{x - x_i}{\beta}\right)^2\right) \quad , \tag{3}$$

where x_i is the concentration of model or observation at point i, and β is the bandwidth. The bandwidth was estimated using the rule-of-thumb by Silverman (1986), based on the full range of observations:

$$\beta = 0.9 \frac{\min\left(\sigma(x), IQR(x)/1.35\right)}{n^{1/5}} \quad , \tag{4}$$

where $\sigma(x)$ is the standard deviation, $IQR(x)$ the interquartile range of observations, and n is the total number of common model-data points. The bandwidth β has the same units as the tracer.

The KDE for a specific (arbitrary) concentration x of tracer type X was then calculated via

$$\mathrm{KDE}(X) = \frac{1}{n\,\beta} \sum_{i=1}^{i=n} k(x, x_i) = \frac{1}{n\,\beta} \sum_{i=1}^{i=n} \frac{1}{4\pi^2} \exp\left(-0.5 \left(\frac{x - x_i}{\beta}\right)^2\right) \quad , \tag{5}$$

We note that, because of the division by β, the units of $\mathrm{KDE}(X)$ are the inverse of tracer units. Figures S14 to S19 depict the KDEs calculated in this way. Following the evaluation of the KDEs, we evaluated the non-parametric metrics HD and BD similar to their evaluation from the discrete concentration classes by integration. We first determine the upper ($x_{\max}$) and lower ($x_{\min}$) boundary for integration from the maximum and minimum concentration of the entire ensemble of model-data pairs, providing a total range of $\Delta_x = x_{\max} - x_{\min}$. Integration is then carried out from over $x_{\min} - 0.05\Delta_x$ to $x_{\max} + 0.05\Delta_x$:

$$\mathrm{HD}(X)_{\mathrm{KDE}} = \sqrt{0.5 \int_{x_{\min} - 0.05\Delta_x}^{x_{\max} + 0.05\Delta_x} \left(\mathrm{KDE}(X)_{\mathrm{mod}}^{0.5} - \mathrm{KDE}(X)_{\mathrm{obs}}^{0.5}\right)^2 dx} \quad . \tag{6}$$

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

As for the discrete case we evaluate the Bhattacharyya distance from the Hellinger distance:

$$\mathrm{BD}(X)_{\mathrm{KDE}} = -\ln\left(1 - \mathrm{HD}(X)^2_{\mathrm{KDE}}\right) \quad , \tag{7}$$

Again, both BD and HD are dimensionless metrics, and we can sum $\mathrm{BD}(x)$ and $\mathrm{HD}(x)$ over all N tracers, thereby obtaining $\sum_{X=1}^{N} \mathrm{HD}(x)_{\mathrm{KDE}}$ and $\sum_{X=1}^{N} \mathrm{BD}(x)_{\mathrm{KDE}}$.

The integrated quadratic distance IQD is a potential third and robust metric to compare different distributions. It is defined by the squared difference of the cumulative distributions of model and observations, and can be applied to both the discrete PDFs described in section 2.4.2 as well as to KDE's. However, it is not dimensionless - to devise a single metric across all tracers requires normalisation by a quantity such as observed global mean or range of tracer concentrations. We therefore only include it in our analysis in a cursory manner, but would like to note that this metric would benefit from further exploration.

3 Results

In section 3.1 we first analyse the differences that occur due to the transfer of MOPS from ECCO to FOCI by comparing results from FOCI-HG/pictrl to those of ECCO-U and ECCO-I. All three simulations applied exactly the same set of biogeochemical model parameters, and differences between FOCI-HG and one of the two ECCO simulations are expected to occur mainly from physical properties such as circulation, mixing, insolation, etc.. Differences caused by the vertical grid are addressed by including ECCO-I in the comparison, which moderates the effects of numerical diffusion of detritus. We focus on the preindustrial setup of FOCI-HG to align with the surface pCO_2 of ECCO-U and ECCO-I, but note that differences between FOCI-HG/pictrl and FOCI-HG/hist are generally small.

In section 3.2 we analyse differences between FOCI-CC/pictrl and FOCI-HG/pictrl, which are expected to occur only from differences in the biogeochemical model parameters. Again, we note that differences between the preindustrial and historical scenarios are small.

In section 3.3, we analyse through the comparison of FOCI-CC/hist and FOCI-HG/hist with observations whether the new set of model parameters in FOCI-HG has led to any improvement of model performance.

3.1 "Effects of circulation": FOCI-HG/pictrl vs. ECCO-U and ECCO-I

Transferring MOPS from ECCO to the FOCI model environment (FOCI-HG/pictrl) results in less primary production at high latitudes, but more in the subtropics (Fig. 4; compare red lines with medium blue lines). One explanation for this could be differences between the insolation simulated by both model environments. As noted above, due to its explicit representation of cloud cover, FOCI-HG simulates much less PAR north (south) of 40°N (40°S) than ECCO (see Fig. 3), which, of course, reduces primary production in that region. Deeper mixing in FOCI (Fig. 2) may additionally contribute to the reduction in primary production in the northern North Atlantic and Southern Ocean.

On the other hand, FOCI-HG/pictrl simulates higher insolation in the low latitudes, which slightly increases production. However, the effect of insolation is less visible in the low latitudes, because these can be considered as largely nutrient limited

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

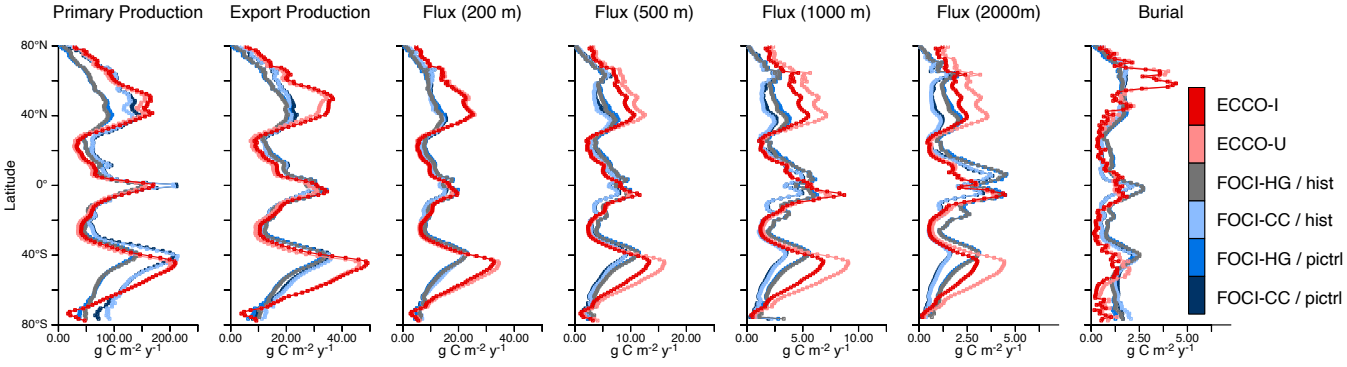

Figure 4. Zonal mean annual fluxes of different model experiments. See also Fig. S1 for zonal integrals.

(see also Fig. 5), whereas the high latitudes are mostly light limited. ECCO-U shows a somewhat lower primary production compared to ECCO-I (Fig. 4), which can be explained by the numerical diffusion of detritus, causing a lower realised particle flux exponent b, and hence less effective nutrient recycling near the surface layers.

Compared to ECCO-I and ECCO-U export production (here defined as particle flux at a depth of 100 m) at high latitudes is far less both versions of FOCI (FOCI-CC/pictrl and FOCI-HG/pictrl), indicating that this flux is mostly controlled by physics, and less by the biogeochemical model parameters. This finding is in agreement with earlier studies that applied different circulations and model parameters for the same model used in our study (Kriest et al., 2020, 2023).

The differences in export production propagate down to the deep ocean as differences in sinking flux. Both ECCO-U and ECCO-I show similar high mesopelagic fluxes at 200 m, but below that depth these signals start to diverge. Here, the less diffusive transport of ECCO-I reduces the numerical (downward) transport of detritus within a vertical box. Eventually, this leads to a flux profile that is often closer to that realised by FOCI-HG, which applies a better vertical resolution, and is thus less prone to numerical diffusion (see also Fig. S2). The effects are less visible in the tropics, where high upwelling velocities counteract the effect of detritus sinking speed.

The vast majority of the ocean area lies in regions between about 40°S and 40°N. So while the large differences of zonal average fluxes at high latitudes in Fig. 4 may look impressive, for the global integral small differences in subtropical and tropical areas will become influential, especially for deeper fluxes. This is exemplified in Fig. S1, which shows smaller differences in zonally integrated fluxes; it is also evident in Fig. 6 showing the global deviations in absolute fluxes. Although both ECCO configurations simulate much higher vertical flux at and below 500 m in high latitudes, and only marginally less in the tropics and subtropics, at a global scale the differences between the two ECCO simulations and FOCI-HG/pictrl becomes less or even change sign, when the subtropical regions with only small differences, but vast extent play a role.

The more vigorous, high-latitude production in both ECCO-I and ECCO-U is reflected in slightly higher plankton concentrations (Fig. 5), while the low latitudes show the opposite, namely higher plankton in FOCI-HG/pictrl than ECCO-U or ECCO-I. Overall, FOCI-HG/pictrl therefore shows a less variable spatial distribution of plankton. Concentrations of POP and

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

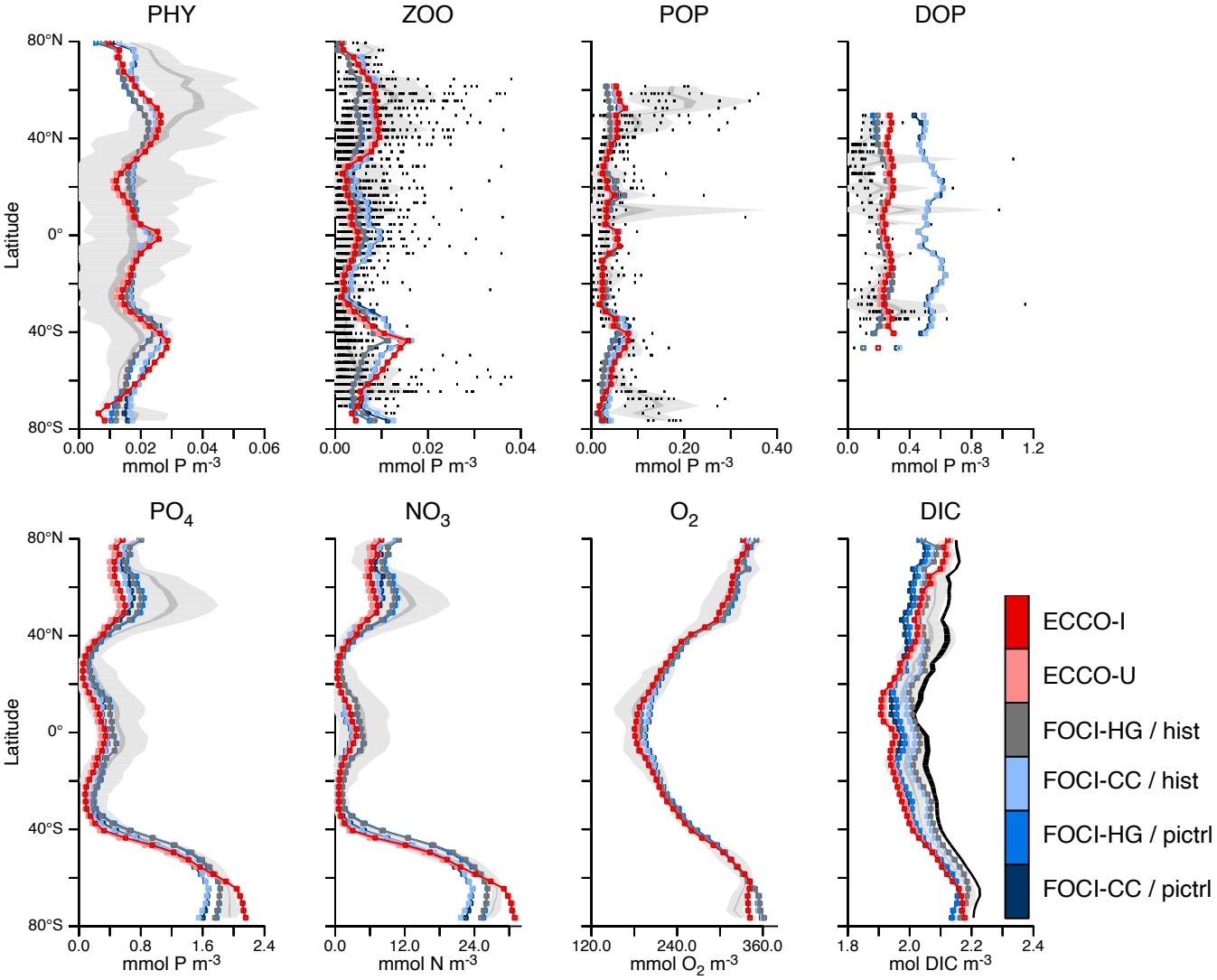

Figure 5. Zonal averages of biogeochemical tracers in the upper 100 m (upper 10 m for phytoplankton). Dark shaded areas: mean of observations ± 0.1 standard deviations. Light shaded areas: mean ± one standard deviation. The light and dark grey area in the panel for DIC denotes observed preanthropogenic DIC, the black area contemporary DIC. Small black symbols: individual observations (zooplankton, POC and DOC only). Coloured lines with symbols: model means, with the models sampled at locations of observations.

DOP are very similar in all model configurations. In all cases, simulated organic tracers fit the observations quite well, with FOCI-HG/pictrl apparently showing a slightly better match to observed phytoplankton in the subtropics.

The muted response of production and phytoplankton in FOCI-HG/pictrl likely also prevents nutrient depletion in many areas and leads to a better match to observed phosphate and nitrate. Oxygen, whose concentration at the sea surface is mainly determined by wind and temperature, is very similar in all model simulations, and shows an almost perfect match to observations.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

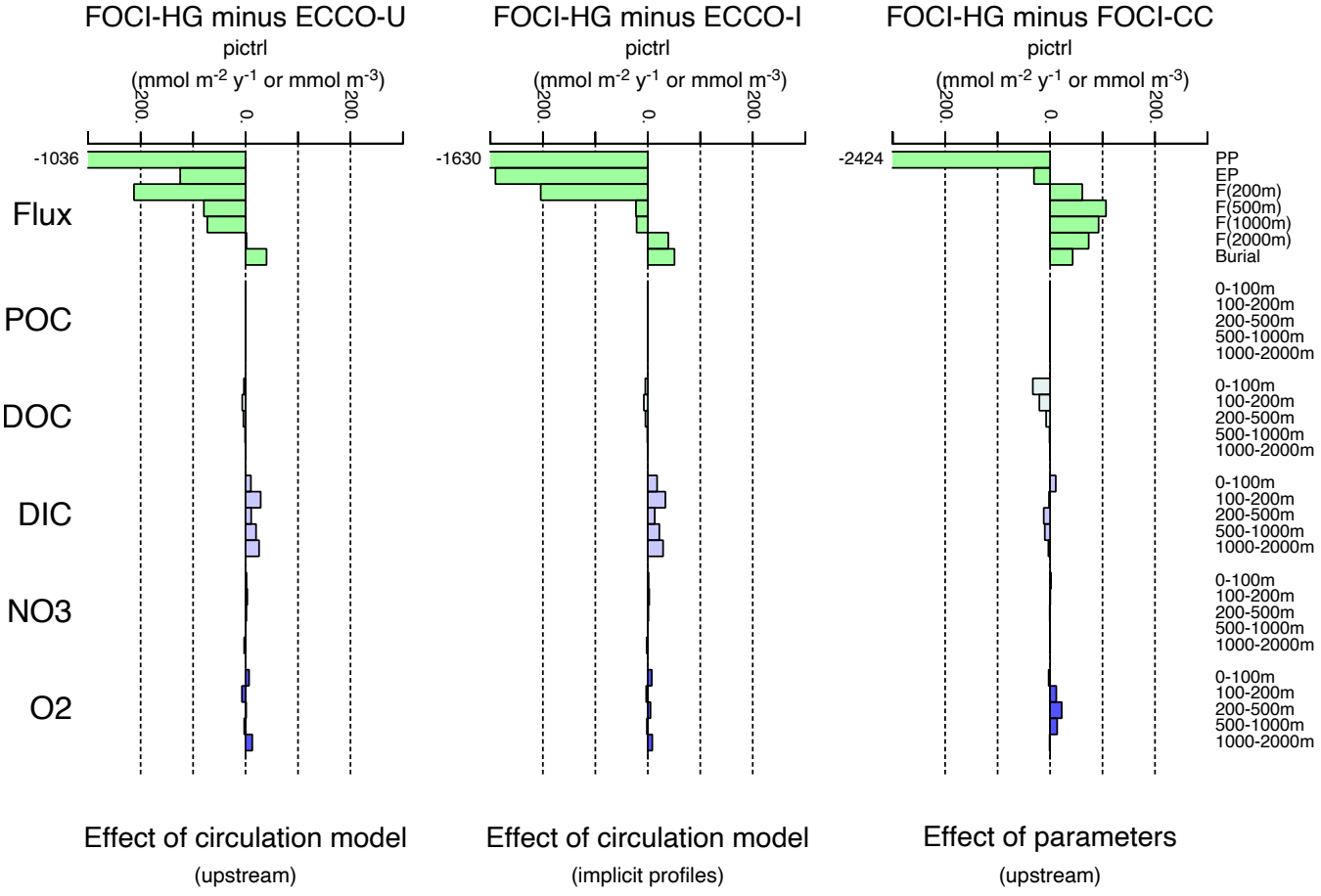

Figure 6. Absolute effects of changes in model setup and forcing for different properties. The two panels indicate the difference between FOCI-HG/pictrl minus ECCO-U/pictrl (left), FOCI-HG/pictrl minus ECCO-I/pictrl (middle), and FOCI-HG/pictrl minus FOCI-CC/pictrl (right). All differences have been evaluated over the last 10 years of a simulation period. Diagnostics from top to bottom: global average annual flux (primary production, sinking fluxes and burial), POC, DOC, DIC, nitrate and oxygen averaged over different vertical domains. The large effect of model changes on primary production exceeds the scale and is indicated by numbers. See also Figures S3 and S4 for fluxes and concentrations of individual model simulations.

Upper-ocean DIC, in contrast, disagrees to some extent between FOCI-HG/pictrl and ECCO-U/ECCO-I in the high latitudes. The ECCO simulations show a relatively good match to observations (which could be attributed to the data-assimilation applied in the configuration; see section 2.1) All model simulations underestimate preanthropogenic DIC in the tropics and subtropics. The negative bias of surface DIC by FOCI has already been noted by Chien et al. (2022a), who discussed a potentially insufficient e-folding export and dissolution length scale of $CaCO_3$ as potential source. This length scale is the same in FOCI-CC, FOCI-HG as well as in the ECCO simulations, and the mismatch between observed and simulated DIC could possibly be avoided by an adjustment of the dissolution length scale.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Extending the comparison between FOCI-HG and ECCO-I (or ECCO-U) further downwards shows only small differences in zonal means of the different models: there are slightly higher nutrient concentrations in FOCI-HG/pictrl and, in some areas, lower oxygen concentrations (Fig. S5). At a global scale, the absolute differences are quite small (Fig. 6).

Because DIC is characterised by large absolute concentrations which are typically above 2 mol m^{-3}, the largest absolute differences in tracer concentrations between FOFI-HG/pictrl and ECCO-I occur for DIC (Fig. 6. However, the relative difference in DIC between FOCI-HG/pictrl and ECCO-I is small, especially when compared to the relative differences of organic tracer concentrations or nutrients (Fig. 7). For example, FOCI-HG/pictrl exhibits surface nitrate concentrations that are about 30% higher than those of ECCO-U or ECCO-I.

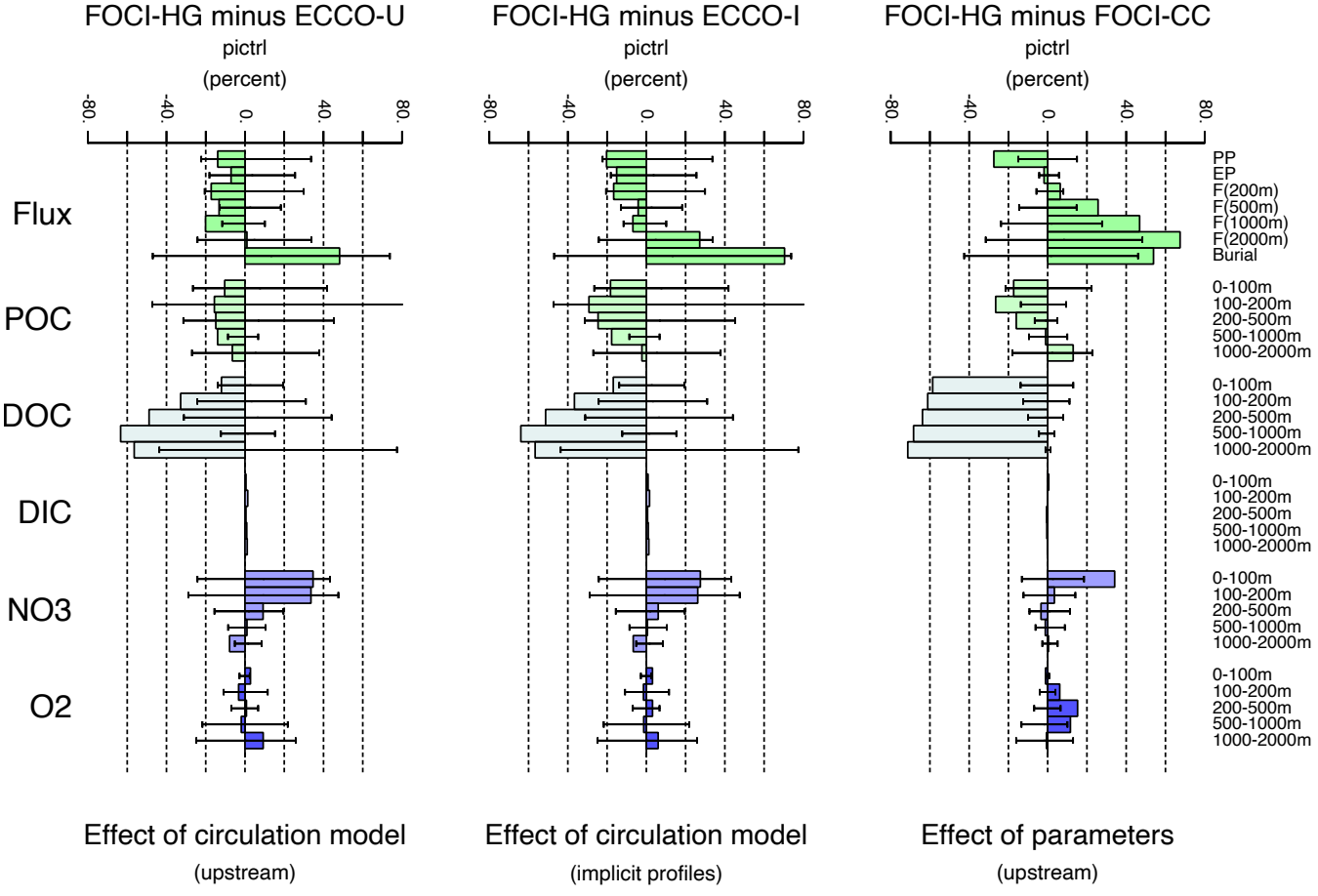

Figure 7. As Fig. 6, but showing the relative effects of changes in model setup and forcing for different properties (all differences are expressed as percent of the subtrahend). The whisker bars in the left and middle panel indicate the the largest positive and negative deviations from the optimal configuration in the cross-validation experiments by Kriest et al. (2020, see text for further details).

Summarising, some differences occur after the transition of MOPS from one model environment (ECCO) to the other (FOCI). Those in primary production can be attributed to disparaties in insolation and/or deep mixing. The parameterisation of

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

physics and circulation in the two model environments has an impact on export production, which also projects onto shallow mesopelagic fluxes. Deeper fluxes are additionally affected by the numerical vertical grid, where the coarse resolution of ECCO can cause an additional vertical transport of detritus. This numerical effect can be moderated or even reversed to some extent by choosing a more advanced numerical scheme for detritus sinking. Absolute tracer concentrations differ less between the two global model configurations, the largest effects occurring in shallow waters. We note that the quite short spinup time of the models of only 665 years impedes an influence of physical model characteristics on tracers in deep waters, and also only the large scale distribution of tracers.

3.2 "Effects of parameters": FOCI-HG/pictrl vs. FOCI-CC/pictrl

FOCI-HG/pictrl and FOCI-CC/pictrl apply the same physical setup and environment, and differ only in biogeochemical model parameters (Table 1), so a comparison of their outcomes sheds light onto the impact of parameters on biogeochemical cycling. Because of its lower light affinity (higher I_c; see Table 1) FOCI-HG/pictrl shows a much lower primary production in high latitudes than FOCI-CC/pictrl (Fig. 4). In addition, a higher maximum grazing rate μ_{ZOO} in FOCI-HG exercises additional top-down control on phytoplankton, leading to lower concentrations (Fig. 5) and hence lower primary production.

As already noted above, there are indications that export production depends mainly on physics; it is thus almost the same in both biogeochemical model configurations, with values in FOCI-HG being only slightly lower. In contrast, particle flux in the deep ocean is much higher in FOCI-HG, which can be clearly traced back to its longer remineralisation length scale ($b = 1.023$ in FOCI-HG compared to $b = 1.413$ in FOCI-CC; see Table 1).

At a global scale, the effect of different model parameters in FOCI-CC and FOCI-HG on primary production clearly exceeds the effect of different insolation in ECCO and FOCI (Fig. 6 and 7). While the transfer of MOPS from ECCO to FOCI leads to a reduction of global average primary production by 1036 mmol C m^{-2} y^{-1} for ECCO-U (Fig. 6, left panel) or even 1630 mmol C m^{-2} y^{-1} for ECCO-I (Fig. 6, middle panel), the difference between FOCI-HG and FOCI-CC is now about twice as large (2424 mmol C m^{-2} y^{-1}; Fig. 6, right panel). On the other hand, export production in both FOCI-HG and FOCI-CC is nearly the same (despite very different particle flux exponents), which again confirms the importance of circulation for this diagnostic. The increase in remineralisation length scale in FOCI-HG ($b = 1.023$) compared to FOCI-CC ($b = 1.413$) leads, however, to a strong increase in deep particle flux, that is more than 40% higher at the base of the mesopelagic (1000 m; see Fig. 7). Hence, except for benthic burial, deep particle flux is therefore more affected by a change in b than by a change in the physical environment.

A consequence of the more muted growth and cycling in FOCI-HG are lower plankton concentrations especially in the high latitudes, which are associated with higher surface nutrient concentrations (Fig. 5). The quite low nutrient affinity of phytoplankton in FOCI-HG is almost not reflected in the phytoplankton concentrations of the largy oligotrophic subtropics. Beside its light-limitation, phytoplankton in MOPS appears to be top-down controlled, as indicated by lower zooplankton:phytoplankton ratios of FOCI-HG/pictrl whenever light is the the limiting factor (Fig. S6).

Deeper remineralisation in FOCI-HG, leads to slightly lower subsurface and mesopelagic (100-500 m) phosphate concentrations, and slightly higher mesopelagic oxygen concentrations (Figures 6, 7, S4 and S5). Mesopelagic nitrate concentrations

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

are about the same in both configurations, which can be explained by the deeper remineralisation (which releases less nitrate in the mesopelagic) in FOCI-HG, but also less denitrification (hence less nitrate consumption) in oxygen minimum zones. The most striking difference between FOCI-HG/pictrl and FOCI-CC/pictrl can obviously be found in surface DOP. Due to the very low production fraction of DOP (Table 1) FOCI-HG/pictrl simulates far lower DOP concentrations (Figures 5 and 6), which agrees much better with the observations. The overall effect on global carbon storage caused by the differences between FOCI-CC and FOCI-HG is small: combined DIC, detrital carbon and DOC in FOCI-CC/pictrl amounts to 36,970 Pg C and in FOCI-HG/pictrl 36,895 Pg C, i.e. FOCI-HG/hist has stored about 75 Pg C (2‰) carbon less. The effect is even smaller for carbon stored below 1000 m, i.e. outside areas of deep mixing, resulting in a difference of less than 9 Pg C (about 0.3‰) less - despite the fast sinking and greater deep carbon flux in FOCI-HG. We note that these numbers may be influenced by the relatively short spin-up time of only 665 years; a longer spin-up time may increase both the amount of carbon stored, as well as the differences between the two model configurations.

Summarising, at a global scale, the influence of parameter values is approximately as important for tracer concentrations and fluxes as the effect of the physical environment (see also Figures 6 and 7), with the exception of export production and DIC, which seem to be more affected by the physical environment. In general, and not surprisingly, large changes in parameters, that emerged from ECCO-MOPS' optimisation against inorganic and organic quantities (Kriest et al., 2023), also cause large changes in the biogeochemical cycling when these are implemented in FOCI-MOPS; the direction of changes in simulated quantities can broadly be inferred from the direction of parameter changes (e.g., an increase in I_c causing less production, or a decrease in b causing larger deep particle flux). As optimisation aims at a better fit to observed tracers, we can expect MOPS to perform better with the new parameters in FOCI - indeed, the visual comparison (Figures 5, S3, and S4) suggests a better match of FOCI-HG to observed quantities. We provide a more quantitative assessment of these improvements in next section.

3.3 Performance of the revised model

A more stringent criterion for model improvement (or deterioration) than visual comparison can be found in scalar parametric or non-parametric metrics. A common way to present the former is via Taylor plots, which depict correlation, RMSE' (i.e., the RMSE with the bias subtracted), and the variance of model quantities normalised by that of observations. However, especially RMSE and correlation coefficient may remain inconclusive when applied to observations of a high sparsity and episodicity. This is also the case in the present study, which shows that for organic components the correlation of all model results to observations is almost equally bad (between about 0 and 0.4; see Fig. S7 and Table S1), while it is very good for the inorganic tracers (the correlation coefficient is always well above 0.9).

The normalised RMSE* (RMSE divided by observed mean) of organic tracers suggests a much larger error than that of inorganic tracers (Fig. 8, left two panels). It is largest for DOP of FOCI-CC, which, as shown above, simulates a large positive bias for this tracer. Being a dimensionless quantity, RMSE* allows to calculate the sum of this metric $\sum \mathrm{RMSE}^*(x)$ over all tracers, thereby arriving at a single scalar metric, similar to that applied during optimisation of MOPS (Kriest et al., 2023). When excluding the error for DOP (which is obviously very large in FOCI-CC), FOCI-HG still performs slightly better than FOCI-CC with respect to this combined metric (Fig. 8). However, the range of $\sum \mathrm{RMSE}^*(x)$ over all model simulations is only

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

less than 1% of the mean, i.e., this metric is not very sensitive with respect to changes in circulation model or biogeochemical model setup. One explanation for this low sensitivity is that this metric was already used in the original calibration of MOPS by Kriest et al. (2023). Further, as noted above, a large fraction of the mismatch in organic tracers can be assigned to mismatches in the pattern (the spatial correlation; see also Fig. S7 and Table S1) in all models. Kriest et al. (2023) showed that these remain even after optimisation, and suggested that they arise from the fact that possibly no mechanistic model will exactly match natural episodic events and fine scale physics.

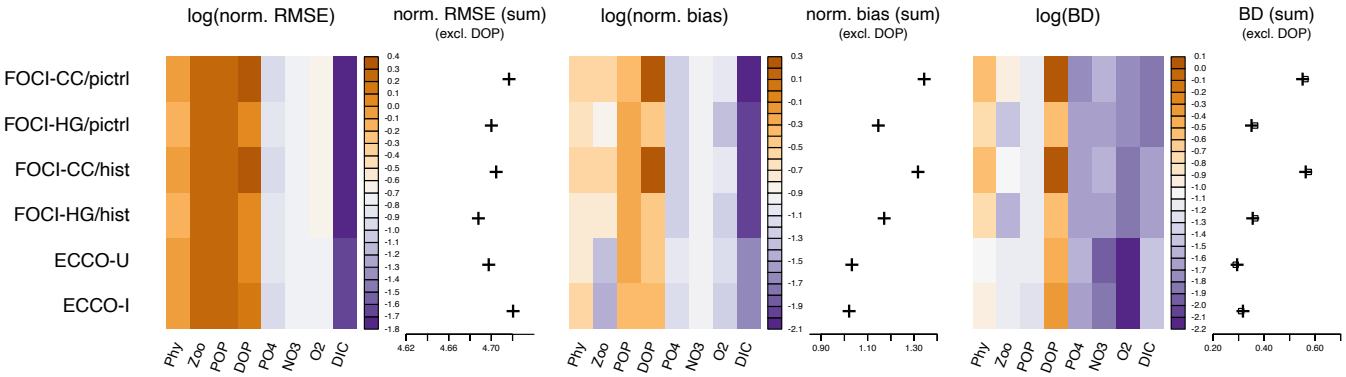

Figure 8. Statistics of model performance. Left two panels: RMSE* (RMSE divided by observed mean; see section 2.4.1). Middle two panels: normalised absolute bias B* (absolute bias divided by observed mean; ; see section 2.4.1). Right two panels: Bhattacharyya distance BD, evaluated from KDEs (; see section 2.4.3). For each metric the colour code shows the log-transformed metric for each model and tracer x. The symbols "+" in the right sub-panels indicate the sum over all tracers except DOP for each individual model, i.e. $\sum_{X=1}^{8} \mathrm{RMSE}^*(X)$, $\sum_{X=1}^{8} \mathrm{B}^*(X)$ and $\sum_{X=1}^{8} \mathrm{BD}^*(X)_{\mathrm{KDE}}$ (evaluated from KDEs). For BD we also indicate $\sum_{X=1}^{8} \mathrm{BD}^*(X)_{\mathrm{PDF}}$ evaluated from PDFs ; see section 2.4.2. See Tables S1 and S2 for numerical values of various metrics, Figures S8 to S11 for PDFs and Figures S14 to S17 for KDEs.

The normalised bias B^* (bias divided by the observed mean) neglects any error in local patterns, and is thus less sensitive to the sparsity and episodicity of observations. For plankton and POP it varies between $\approx 3\%$ (zooplankton in ECCO-I) to almost 58% (FOCI-HG/hist), and is close to 200% for DOP simulated by FOCI-CC (Fig. 8, middle panels). Overall, and excluding DOP the combined normalised bias $\sum \mathrm{B}^*(x)$ varies by almost 28% of the mean across all models. This metric also clearly indicates a superior performance of FOCI-HG for the preindustrial as well as for the historical configuration. However, FOCI-HG is outcompeted by ECCO-U and ECCO-I, which is not surprising, as MOPS' model parameters were calibrated in this circulation.

The Bhattacharya distance BD (see section 2.4.3), which also neglects mismatches in spatial patterns shows more similar performance metrics for different tracer types (Fig. 8, two panels on the right). For this metric, FOCI-HG/hist shows a considerable improvement compared to FOCI-CC of phytoplankton (from $\mathrm{BD}_{\mathrm{KDE}} = 0.314$ in FOCI-CC/hist to $\mathrm{BD}_{\mathrm{KDE}} = 0.164$ in FOCI-HG/hist), zooplankton (from $\mathrm{BD}_{\mathrm{KDE}} = 0.095$ to $\mathrm{BD}_{\mathrm{KDE}} = 0.030$) and, of course, DOM ($\mathrm{BD}_{\mathrm{KDE}} = 1.195$ and $\mathrm{BD}_{\mathrm{KDE}} = 0.277$). Simulated particulate organic matter (POM) – which, in our model comparison, is the sum of detritus, phytoplankton and half of the zooplankton – deteriorates somewhat in FOCI-HG from 0.067 to 0.075 (see also Table S2).

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

When adding BD of all tracers except DOP ($\sum_{X=1}^{8} \mathrm{BD}^*(X)_{\mathrm{KDE}}$) we find a variation of 66% of the mean, rendering BD the most sensitive of all metrics applied so far. $\sum_{X=1}^{8} \mathrm{BD}^*(X)_{\mathrm{PDF}}$, when evaluated from histograms, is equally sensitive, with the spread of combined metrics being 70% of the model mean (see small squares in Fig. 8). The sum of Hellinger distances $\sum_{X=1}^{8} \mathrm{HD}^*(X)$ is less sensitive (Table S2). Nevertheless, according to this combined metric, FOCI-HG again performs much better that FOCI-CC, although it is still outperformed by ECCO-I and ECCO-U. We note that the differences between FOCI-CC and FOCI-HG would be even larger if DOP was included in the overall metric (see also Table S2).

So far, we examined the model fit to observations that were mostly part of the original calibration of MOPS by Kriest et al. (2023). As shown above and in earlier studies the particle flux exponent b can be very influential for the large-scale distribution of nutrients and oxygen (e.g., Kwon and Primeau, 2006; Kriest et al., 2012; Kriest and Oschlies, 2015), of DIC and alkalinity (Kwon and Primeau, 2008), and strongly impact the atmospheric pCO_2 (Kwon et al., 2009). At the same time it determines simulated deep particle flux (Kriest and Oschlies, 2013), indicating that the large-scale distribution of especially nutrients and oxygen and the particle flux are related. We here finally investigate whether the improvement of model performance for FOCI-HG is also reflected in the comparison between simulated and observed particle flux, by applying the same performance metrics as we did for dissolved and particulate tracers.

Applying parametric metrics such as RMSE and the correlation coefficient shows that both FOCI-CC/hist and FOCI-HG/hist perform similarly good (or bad), with only a weak correlation to observations and a large RMSE (Fig. 9). Again, as for observations of organic tracers, this can likely be explained with the sparsity of data (the data set consist only of 295 data points for direct model-data comparison), which are also heavily influenced by (sub)mesoscale features such as mixing events, filaments, etc., thereby creating a large pattern error. However, the bias of FOCI-HG (-14.7 mmol C m^{-2} y^{-1}) is less than one fifth of the bias simulated by FOCI-CC (-80.5 mmol C m^{-2} y^{-1}), which is mostly due to a better match in deep particle flux (Fig. 9).

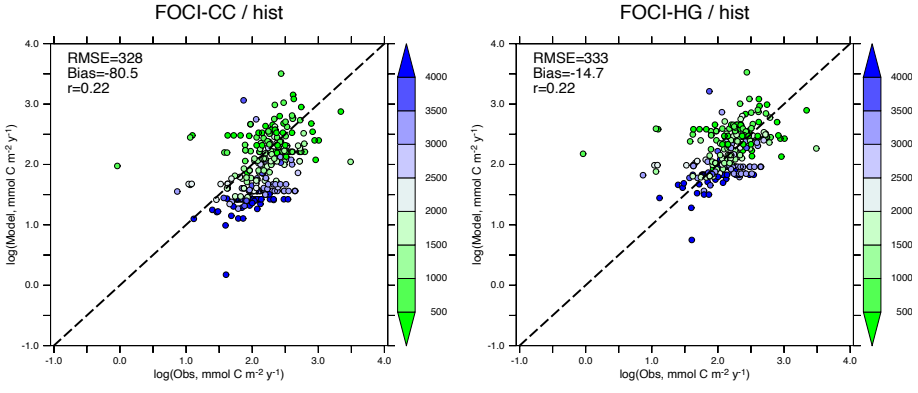

Figure 9. Simulated vs. observed particle flux (log-transformed) for FOCI-CC/hist and FOCI-HG/hist. Colour indicates trap depth. Numbers in the upper right of each panel provide parametric metrics of untransformed data.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Non-parametric metrics derived via KDEs also show a much better performance of FOCI-HG. Regardless of the bandwidth applied and the individual metric (HD, BD or IQD), FOCI-HG/hist always results in a metric that is at least 26% (HD with a bandwidth of 30) less than that of FOCI-CC; in case of the Bhattachryya distance, FOCI-HG improves even down to only 40% to 55% that of FOCI-CC (Fig. 10). The improvement is less pronounced when evaluating the three non-parametric metrics from PDFs (Fig. S20), but even here FOCI-HG always performs better than FOCI-CC.

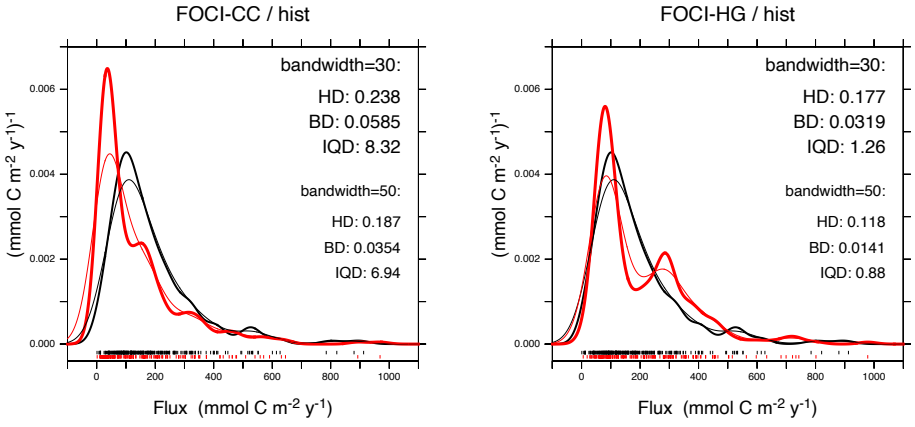

Figure 10. Kernel density estimates (KDE) of particle flux from observations (black) and simulations FOCI-CC/hist and FOCI-HG/hist. KDEs of simulated and observed tracer concentrations have been calculated as explained in section 2.4.3. Numbers in the upper right of each panel show the the Hellinger distance (HD), the Bhattacharyya distance (BD), and the integrated quadratic distance (IQD), evaluated as explained in section 2.4.2. Thick lines and large letters show the KDEs evaluated with a bandwith of 30 mmol C m^{-2} y^{-1} (which is close to Silverman's rule-of-thumb value of 30.88 mmol C m^{-2} y^{-1}) and thin lines and small letters those with a bandwidth of 50 mmol C m^{-2} y^{-1} (close to the bandwidth of 50.48 mmol C m^{-2} y^{-1} evaluated with Scott's rule-of-thumb). Small bars along the x-axis depict the individual data points.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

4 Discussion

4.1 Effects of circulation and numerics - can we transfer an optimised biogeochemical model to a different circulation?

When transferring MOPS' parameters, optimised in a light-weight circulation, to the ESM we found dissimilarities in primary production. These can partly be traced back to differences in incoming solar radiation and deep mixing, which are effective mainly in the high latitudes. The different environments for the biogeochemical model cause a range of 18% of the mean of FOCI-HG/pictrl and ECCO-I (Δ circ in Table 2). The difference between FOCI-HG/pictrl and ECCO-U is smaller (11%), illustrating the effect of numerical diffusion and subsurface remineralisation in a relatively coarsely resolved model geometry such as ECCO.

This range between ECCO-U/ECCO-I and FOCI-HG is smaller than the range found by Kriest et al. (2020), who carried out cross-validation experiments with model MOPS in five different circulations (in form of TMs), but applying the same calculation of insolation. Swapping the circulation while keeping the parameter set, Kriest et al. (2020) found a range of global primary production of 21-25% of the mean, which is larger than the range of 11% or 18% obtained in the present study (see also Table 2). The larger effect of circulation in the study by Kriest et al. (2020) can possibly be explained by the fact that the TMs applied by Kriest et al. (2020) represented circulations with very different physical properties (e.g., in the area of deep mixing, outcrop of water masses such as SAMW, or ideal age of NADW); also, Kriest et al. (2020) carried out their analysis after a spin-up time of 3000 years, which allows the combined effects of circulation and biogeochemistry to become more evident. Hence, while we regard differences in insolation to play a large role for the difference between simulated production of FOCI-HG and ECCO-U/ECCO-I, the results by Kriest et al. (2020) suggest that circulation can also exert a large influence on this diagnostic.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Table 2. Global annual primary production (PP), export production (EP) and particle flux in 1000 and 2000 m of different model simulations. Where available, the range (as percent of the mean) of flux due to circulation (Δ circ), due to biogeochemical parameterisation (Δ bgc) and numerical implementation of particle flux (Δ num) is given, as well as the range across all model experiments (Δ all). For model simulations analysed in this study Δ all refers to the range across FOCI-CC/pictrl, FOCI-HG/pictrl, ECCO-U and ECCO-I, Δ circ to the range between FOCI-HG/pictrl and ECCO-U / ECCO-I, Δ par to the range between FOCI-CC/pictrl and FOCI-HG/pictrl and Δ num to the range between ECCO-U and ECCO-I.

	PP		EP		Flux 1000m		Flux 2000m	
	mean (Pg C y^{-1})	range (%)	mean (Pg C y^{-1})	range (%)	mean (Pg C y^{-1})	range (%)	mean (Pg C y^{-1})	range (%)
This study:								
FOCI-CC/pictrl	38.3		6.9		0.76		0.41	
FOCI-HG/pictrl	27.8		6.8		1.12		0.68	
ECCO-U	31.0		7.0		1.37		0.64	
ECCO-I	33.5		7.7		1.18		0.51	
Δ all		32		14		55		49
Δ circ		11 / 18		4 / 13		20 / 5		7 / 50
Δ bgc		32		2		38		50
Δ num		8		9		15		23
FOCI-CC/hist	37.7		6.8		0.75		0.40	
FOCI-HG/hist	27.5		6.7		1.09		0.66	
Kriest et al. (2020)	45.5		6.6		1.00		0.41	
Δ all		40		17		43		70
Δ circ		21-25		13-16		13-17		28-31
Δ bgc		14-16		2-7		24-30		37-43
Najjar et al. (2007), Δ circ			13	77				
Pasquier et al. (2023), Δ circ			15	100				
Buchanan et al. (2018)			7.5					
Δ all				32				
Δ circ				13-19				
Seferian et al. (2020, CMIP5), Δ all	42.5	157	7.6	124				
Seferian et al. (2020, CMIP6), Δ all	38.8	92	7.5	75				
Doney et al. (2024) Δ all[(1)]	41.5	22	6.1	19	0.65	37		
Walker and Palevsky (2025) Δ all			7.6	55	1.1	117		

[(1)] Doney et al. (2024) report one standard deviation

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

The physical properties of FOCI and ECCO have only a moderate effect of 4% on global integrated export production when comparing FOCI with ECCO-U, that applies an upstream scheme, but we find a larger impact of 13% of when considering ECCO-I, that features an implicit numerical representation of particle sinking (Table 2). The disparity between ECCO-U and ECCO-I highlights the importance of vertical discretisation, causing a difference of 9% in export production (Δnum, Table 2). Effects of model parameters (Δbgc) on export production are almost negligible (2%). However, compared to other biogeochemical diagnostics such as global primary production or deep particle flux, export production is affected least by any change in model setup (circulation, numerics, parameters), as evident from the total variation across all model setups (Δall) of 14%.

The cross-validation experiments by Kriest et al. (2020) also showed a dominant effect of circulation on export production, namely a variation between 13 and 16% of the mean - which is considerable, compared to an overall variation (of both biogeochemical and physical changes) of 17% in their study (Table 2). As for primary production, this large variation can possibly be attributed to the very different circulations tested by Kriest et al. (2020); in the present study, despite differences in deep mixing (Fig. 2), the circulations of ECCO and FOCI seem to be more similar.

One potential explanation for the similarity of ECCO and FOCI circulation is that both were tuned or shown to match observed features. Surface fluxes of the ECCO circulation model, that formed the basis of TMs and forcing applied in this study, were adjusted so that the model simulation became consistent with the observations of temperature and salinity (Stammer et al., 2004). We can therefore expect the circulation of ECCO to provide a quite realistic representation of ocean circulation, despite some potentially missing features such as very deep mixing in the Southern Ocean (see also section 2.1 and Table S1 of Kriest et al., 2020). Likewise, the circulation of FOCI also represents many features of the ocean well, among them the AMOC, mean climate in terms of ocean temperatures, as well as SST variability in the tropical Pacific, but shows a cold bias in the North Atlantic, and a warm bias in the Southern Ocean (Matthes et al., 2020). As noted above, the latter might be the cause for too deep mixing in that area (Zeller and Martin, 2024). We thus hypothesize that the adjustment and tuning of circulations in both FOCI and ECCO led to a somehow similar physical framework, that causes only little disruptions in export production when MOPS was transferred between the two environments.

The effects of circulation on export production between 4 and 16% found in this study and by Kriest et al. (2020) are, however, small compared to those found in earlier model comparisons. In a pioneering study, Najjar et al. (2007) applied a very simple biogeochemical model to 12 different circulation models, and found a range in global particle export from $\approx 7-17$ Pg C y^{-1}, with a mean of 13 Pg C y^{-1}; i.e. a range that is about 77% of the mean. Najjar et al. (2007) attributed this to "differences in the parameterisations of lateral diffusive mixing and mixed layer process", and further noted a strong correlation of particle export with ^{14}C in circumpolar deep water (CDW), as well as with the concentration of DOC in the upper 50 m. Even when restricting their analysis to a subset of only five models, that showed the best match to observed ^{14}C in CDW, they found a large range of $\approx 7-14$ Pg C y^{-1}, that is, about 70% of the mean. The cause for the strong impact of circulation on export production in the study by Najjar et al. (2007) may, to some extent, be found in the simplicity of their biogeochemical model, that relied on an adjustment of simulated to observed nutrients for the calculation of export production. This may have dampened the feedback effects of shallow (or fast) remineralisation on surface production.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

More recently, Pasquier et al. (2023) optimised 12 parameters of a simple biogeochemical model coupled to two different circulations against observed nutrients, oxygen, alkalinity and DIC. For the two physical configurations they found very different sets of optimal model parameters, with the largest deviations occurring for optimal particle decay rates, which varied up to six-fold between the different experiments. Differences in model parameters and/or circulation then led to a steady-state global export (particle flux at 100 m) that varied between 7.4 and 22 Pg C $^{-1}$, a range that is even larger than the variation of 100% observed by Najjar et al. (2007). However, it cannot be fully ruled out that a part of the large variation in optimal parameters and hence export production in that study is due to the optimisation scheme, which does not guarantee that the algorithm converges to a global minimum; a feature that could also explain the very high export production and large remaining mismatches to observed tracer distributions for one of the coupled models.

While Najjar et al. (2007) did not believe that nutrient restoring and model simplicity had a large effect on the sensitivity of simulated export production to circulation, a recent study by Buchanan et al. (2018) provides evidence that more complex models tend to be more robust with regard to changes in the physical model setup. Buchanan et al. (2018) tested a quite simple prognostic model that does not rely on nutrient restoring, in six different biogeochemical and six different physical configurations. Variations in biogeochemistry affected details of elemental stoichiometry and the representation of particle flux; physical variations were based on the forcing derived from different Earth system models. They found a total variation due to the physical setup between 1 and 1.4 Pg C y^{-1} (13% and 19% of the mean) for the different biological configurations, close to the present study and the one by Kriest et al. (2023). Furthermore, the most "complex" biogeochemical model parameterisation applied by Buchanan et al. (2018) showed the lowest sensitivity to physics.

Summarising, we hypothesize that the sensitivity of export production especially to changes in circulation and forcing may depend on biogeochemical model complexity, and be moderated by the feedbacks embedded in fully prognostic biogeochemical models.

Our study suggests that the further fate of export production, namely its transfer to the ocean interior, is strongly determined by the biological model configuration and by the numerical representation of particle sinking. At 1000 m, circulation only plays a negligible role of 20% (ECCO-U) or 5% (ECCO-I) for particle flux (Table 2). Beyond that depth the numerical parameterisation becomes even more important, with ECCO-I showing a far lower flux than either ECCO-U or FOCI-HG (Table 2).

To summarise, although the transfer of optimised MOPS from ECCO to FOCI gives rise to some considerable regional changes at the surface, which are to some extent due to the calculation of incoming solar radiation, many of these differences level out at about 100 m, resulting in more or less the same values for export production; this may depend, however, on the similarity of circulations. A potential solution to avoid the former source of disparity would be to force the light-weight ECCO model by incoming solar radiation from the more complex FOCI model. Especially when optimising parameters such as the light-saturation constant I_c, this could yield quite different optimal parameters. A large influence on model results is exerted by the numerical scheme that parameterises particle flux through the water column. By comparing FOCI-HG to ECCO-U we can not rule out an impact of the vertical grid and the associated numerical diffusion on particle flux to the ocean interior. To account for this, it may be worthwhile to consider the numerical scheme applied in ECCO-I (which is very light-weight, and

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

requires only minor code adjustments) for MOPS both models environments, and also include this parameterisation in future model optimisations. By doing so, the effects of numerical diffusion through vertical grid spacing could be largely avoided in both models, possibly allowing for a fairer inter-model comparison.

4.2 Effects of parameters on model performance - does the revised ESM benefit from model optimisation and subsequent transfer?

As shown above, biogeochemical model parameters can play a considerable role for the concentrations and turnover of organic matter at the surface. The effects can often be traced back to the respective parameter changes, such as the production fraction of DOM or the light affinity of phytoplankton.

Because of the adjusted model parameters in FOCI-HG/hist, global primary production declines strongly with respect to FOCI-CC/hist, down to a value of 27.5 Pg C y^{-1}. This value is below observed estimates (between 43.5 and 68 Pg C y^{-1}; Doney et al., 2024), but still within the range of models applied in that study (see their Fig. 3) or that of CMIP6 models (between 22 and 57 Pg C y^{-1}; Seferian et al., 2020). We note that we rather consider MOPS as a "net model", that does not account for the very fast recycling of nutrients via the microbial loop. Implementing this (as, for example, discussed and parameterised in Oschlies and Schartau, 2005; Schmittner et al., 2005, 2008) could likely enhance primary production without affecting export production too strongly.

Surprisingly, both configurations of FOCI-MOPS (FOCI-CC and FOCI-HG) show quite similar export production, despite applying very different sets of biogeochemical model parameters, and thereby simulating very different global primary production. The relative robustness of simulated export production may be explained by feedback processes between the production of organic matter within the euphotic zone and its subsequent export to subsurface layers, where it is remineralised. The recycled nutrients then again provide a source for production the euphotic zone.

This feedback can be exemplified by some simple calculations for the two model configurations of FOCI-MOPS. Particle export out of the euphotic zone is determined by sinking speed at a given depth, $w(z)$. For a depth of 100 m it is set in the model through the particle flux exponent b via $w(100) = r/b \times 100$ m, where r is the remineralisation rate of detritus (see also Kriest and Oschlies, 2008). Multiplication with detritus concentration at that depth then provides the particle flux. Both FOCI-HG and FOCI-CC apply a constant remineralisation rate r of 0.05 d^{-1}. With $b = 1.023$ FOCI-HG therefore applies a sinking speed at 100 m of 4.9 m d^{-1}, that is, detritus in FOCI-HG sinks about 40% faster than that of FOCI-CC ($w(100) = 3.5$ m d^{-1}). The faster sinking in FOCI-HG is, however, accompanied by a detritus concentration that is about 30% less than that of FOCI-CC (see also Fig. 7 for average POC between 0 and 100 m), thereby almost offsetting the faster sinking, and resulting in almost the same export production. In summary, a slow sinking speed will result in large production and standing stocks of organic matter at the surface, which enhances the export flux simply by its concentration; at the same time, the flux will be reduced through the slow sinking speed. On the contrary, a fast sinking speed will enhance the flux, but at the same time reduce it through low concentrations of sinking organic matter.

Although simulating too low primary production, global export of 6.7 to 6.9 Pg C y^{-1} of all FOCI simulations (see Table 2) is still within the range of observed estimates of 8.2 ± 2.78 Pg C y^{-1} (Doney et al., 2024), and within those of other model

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

studies (e.g., Seferian et al., 2020; Doney et al., 2024). Because of its relatively low primary production, the e-ratio (export divided by production) of FOCI-CC and FOCI-HG is rather high (0.18 and 0.24, respectively) compared to values reported by Doney et al. (2024).

Despite the similar export production, FOCI-CC and FOCI-HG simulate very different deep particle flux, which can be traced back to the particle flux parameter b. The dependence of deep particle flux on b (and potentially other parameter changes) is in agreement with the results by Kriest et al. (2020). A global flux at 1000 m of 0.75 Pg C y^{-1} in FOCI-CC/hist is rather low, but within the range of other model studies, which vary between 0.4 to 1.6 Pg C y^{-1} (Walker and Palevsky, 2025). It agrees with the model range of 0.65 ± 0.24 Pg C y^{-1} reported by Doney et al. (2024), but is too low compared to their observed mean of 0.95 ± 0.25 Pg C y^{-1}. In contrast, particle flux at 1000 m simulated by FOCI-HG/hist agrees very well with this observed estimate (see also Table 2). The better agreement of deep particle flux simulated by FOCI-HG is also confirmed in direct comparison with sediment traps. Although a local comparison of simulated and observed flux via the Pearson correlation coefficient or RMSE does not reveal any superior performance of FOCI-HG (Fig. 9), the negative bias in FOCI-HG/hist is substantially lower than that of FOCI-CC/hist, and all non-parametric metrics clearly show a better performance of FOCI-HG.

Direct comparisons of simulated particle flux with observations via parametric metrics have been inconclusive in other model-data comparisons as well: for example, Gehlen et al. (2006) and Schwinger et al. (2016) could not detect large differences when applying these metrics to assess model with various particle sinking schemes. Likewise, when systematically varying b between 1.287 and 0.858 (about the range applied in this study), Kriest and Oschlies (2013) did not find large changes in correlation between simulated and observed particle flux. Obviously, non-parametric metrics are much better suited for comparison against sediment trap data, that can be affected by episodic or local phenomena such as mixing events, eddies, etc.

The changes in biogeochemical cycling caused by the adjusted model parameters in FOCI-HG affect the representation of model tracers, as depicted by a slight improvement in the sum over normalised errors (RMSE*; Fig. 8). However, the improvement is relatively small or absent for nutrients and DIC (which already showed a very good correlation in FOCI-CC), whereas oxygen improves with respect to all parametric metrics (see Fig. S7 and Table S1). With correlation coefficients for inorganic tracers well above 0.9, all model versions presented in this paper are well within those reported in other studies (e.g., Seferian et al., 2013; Kwiatkowski et al., 2014; Buchanan et al., 2018; Seferian et al., 2020). A correlation between observed and simulated inorganic tracers that is already very good at the global scale has been observed earlier (Seferian et al., 2020), and leaves little room for further improvement, which may explain the low sensitivity of this metric to changes in model parameters. We note that, on long time scales, inventories of tracers that exchange through model boundaries, such as nitrate and oxygen, may provide good constraints on model performance (e.g., Kriest and Oschlies, 2015; Buchanan et al., 2018), as also indicated by the quite high sensitivity of the normalised bias to model changes in Fig. 8.

With correlation coefficients for phytoplankton between 0.32 and 0.35 FOCI-CC and FOCI-HG are in the range of other model studies (e.g., Seferian et al., 2013; Kwiatkowski et al., 2014; Seferian et al., 2020, but note that some of these studies referred to specific vertical domains at the surface for model assessment). Simulated higher trophic levels have been assessed less frequently. Petrik et al. (2022) compared mesozooplankton simulated by 6 ESMs vs. a global data set adjusted for temporal,

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

spatial and sampling biases. For log-transformed biomass they found a correlation between -0.19 to 0.44. Carrying out the same log-transform, and comparing against the data base also used in this study, we find correlation coefficients between 0.32 to 0.33 for all models, i.e., well within the range reported by Petrik et al. (2022), but overall with little difference among the model configurations.

The above comparison shows that, especially for organic tracers, parametric metrics such as the correlation coefficient and RMSE provide quite little information on model performance, and non-parametric metrics, such as the Bhattacharyya distance or the IQD may be much more suited to assess and compare different models. This is also depicted in Fig. 11, which shows the normalised range of the respective metrics value over all model simulations for different metrics applied in this study Obviously, RMSE, RMSE* and R provide the least information on model performance, whereas the bias is more sensitive to changes in model environment and configuration. We note that the sensitivity of the bias of nitrate and oxygen can be even larger if the model was spun up over a longer time, and had more time to accumulate (or loose) tracers based on the combined effect of physics and biogeochemistry. Of the non-parametric metrics either the Bhattacharyya distance or IQD show the largest sensitivity across a wide range of tracers, while the Hellinger distance is less sensitive; we note that the higher sensitivity of the Bhattacharyya distance may arise from the fact that is is based on a log-transform of the squared Hellinger distance, and not restricted to values between 0 and 1. Therefore, the Bhattacharyya distance or the IQD seem to be more suited for model intercomparison; however, for model assessment across a range of tracers and diagnostics the IQD would require a normalisation.

5 Conclusions

We have examined whether the improved performance of a biogeochemical ocean model, which was optimised against observations in an efficient "surrogate" circulation, is maintained when this model is simulated within a more complex and computationally demanding Earth system model (ESM). Our results suggest that the revised parameter set leads to improved biogeochemical model performance in the ESM. This demonstrates that parameter optimisation conducted within a simpler and computationally efficient circulation framework, such as the TMM, can be beneficial when transferred to an ESM, where the underlying dynamics are more complex and computationally expensive to simulate.

Incoming solar radiation, mixing and circulation, together with model parameters, all affect primary production and biogeochemical cycling at the surface. Their influence on organic and inorganic model tracers is less pronounced, unless very large changes in model parameters are imposed (as in the case for DOM). Overall, the improvement due to optimised model parameters outweighs the deterioration that comes from a change in model environment.

Export production is only slightly affected by circulation, which could be explained by the similarity of the circulations applied, in addition to feedback processes between production, export, and subsurface remineralisation. It is is nearly identical in both versions of the ESM where circulation environments are the same. On the other hand, deep particle flux is heavily influenced by the detritus sinking speed. Remarkably, the model fit to sediment trap data, that were not part of the metric

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

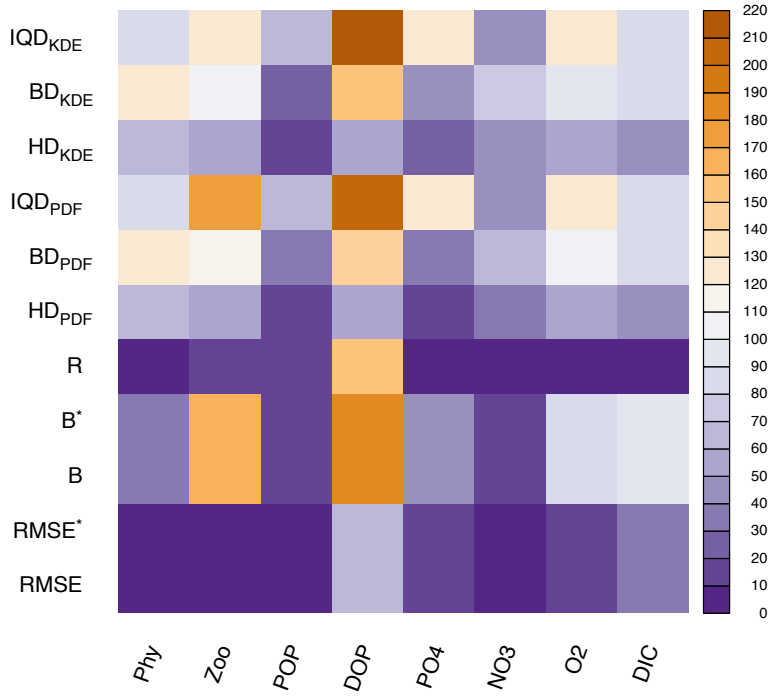

Figure 11. Sensitivity to model changes of several performance indicators for different tracers. Colour indicates the range of the metric across all model experiments divided by the mean (as percent). Non-parametric metrics (upper 6 rows) are described in sections 2.4.2 and 2.4.3. Parametric metrics include correlation coefficient R, normalised bias B*, bias B, normalised RMSE* and RMSE as described in section 2.4.1.

during optimisation, improves after optimisation; this improvement persists even when the optimised model is transferred to the ESM, shows a much better performance with regard to global estimates and local observations of particle flux.

Beside standard metrics such as the RMSE, that were also applied during optimisation, a posteriori analysis of several non-parametric metrics, such as the Bhattacharyya distance and the IQD showed that these may be much better suited to evaluate biogeochemical models with respect to scattered data sets, that are likely to be influenced from episodic events and mesoscale features.

To guide future biogeochemical model development and its implementation in global, computationally demanding models, the following strategies may be worthwhile pursuing:

- The large role of insolation for primary production, and its potential impact on parameter optimisation, can be accounted for by forcing the light-weight optimisation model with the same insolation (model) as the ESM
- Given circulations that are not too different, we hypothesise that export production will be similar in the two environments. With the given particle flux curve that in our study is uniform within a model configuration, biogeochemical parameter changes have almost no effect on export production. Furthermore, as both primary production and deep par-

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

ticle flux seem to be independent of export production, we consider this flux as a relatively insignificant diagnostic, in line with other studies (Wilson et al., 2022).

- Deep particle flux is highly sensitive to the applied particle flux length scale. Because the particle flux profile is also influential for the long-term distribution of inorganic tracers, a good match to observed particle flux may be indicative for the overall model performance, and we recommend the use of this diagnostic (and potential further, dense and high-quality data sets such as particle abundances; Kiko et al., 2022) in further model calibration and assessment.

- The low sensitivity of standard metrics such as the RMSE and correlation may result in inconclusive results when comparing simulated organic tracers and fluxes to observations. Here we recommend to further exploit, and potentially apply, non-parametric metrics for model assessment and calibration.

- Finally, we note that the results of the present study have been obtained after a relatively short spin up time of the models. The resulting output and assessment can therefore not fully address the combined effects of biogeochemistry and circulation, especially with regard to the model bias, which has been shown to be very informative. However, recent advances facilitate a spin-up of biogeochemical ocean models in a fraction of the time required by a "normal" (forward) spin-up (Khatiwala, 2024), and can strongly support model calibration even on long time scales.

Code and data availability. The FOCI model code by Chien et al. (2022a) is available at https://doi.org/10.5281/zenodo.6772175 (Chien et al., 2022b). Code used for simulations FOCI-HG is structurally the same as in Chien et al. (2022a), with some modifications to account for diagnostic output, compute environment and model parameters. This code and further model output can be found under https://doi.org/10.5281/zenodo.19048349 (Kriest et al., 2026a). For most up-to-date version of the TMM and MOPS see code repository by Samar Khatiwala: https://github.com/samarkhatiwala/tmm. All model output relevant to this paper, as well as scripts for postprocessing, model analysis and plots are provided with additional information at https://hdl.handle.net/20.500.12085/16d3928c-0b30-11f1-a2ca-005056a30ade (Kriest et al., 2026b). This link also includes the three subroutines (two for FOCI, one for TMM) that define the biogeochemical model parameters of FOCI-CC, FOCI-HG and ECCO-I/ECCO-U.

Author contributions. HG configured the revised ESM and carried out the simulations with support from TK. IK carried out the simulations with TMM, analysed and plotted all model results. All authors contributed to the writing of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This work was funded by the European Union under grant agreement no. 101083922 (OceanICU) and UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10054454, 10063673, 10064020,

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

10059241, 10079684, 10059012, 10048179]. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The authors wish to acknowledge use of the PyFerret program for analysis and graphics in this paper. PyFerret is a product of NOAA's Pacific Marine Environmental Laboratory. (Information is available at http://ferret.pmel.noaa.gov/Ferret/.) The authors gratefully acknowledge the computing time provided on the supercomputer "Emmy"/"Grete" at NHR-Nord@Göttingen as part of the NHR infrastructure , as well as on the high-performance computer "Lise" at the NHR centre NHR@ZIB. This center is jointly supported by the Federal Ministry of Education and Research and the state governments participating in the NHR (www.nhr-verein.de). We thank Markus Schartau for helpful discussions about the value and applicability of KDE's and related metrics for model assessment.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

References

Buchanan, P. J., Matear, R. J., Chase, Z., Phipps, S. J., and Bindoff, N. L.: Dynamic Biological Functioning Important for Simulating and Stabilizing Ocean Biogeochemistry, Global Biogeochemical Cycles, 32, 565–593, https://doi.org/10.1002/2017GB005753, 2018.

Chien, C.-T., Durgadoo, J. V., Ehlert, D., Frenger, I., Keller, D. P., Koeve, W., Kriest, I., Landolfi, A., Patara, L., Wahl, S., and Oschlies, A.: FOCI-MOPS v1 – integration of marine biogeochemistry within the Flexible Ocean and Climate Infrastructure version 1 (FOCI 1) Earth system model, Geoscientific Model Development, 15, 5987–6024, https://doi.org/10.5194/gmd-15-5987-2022, 2022a.

Chien et al.: FOCI-MOPS v1 – integration of marine biogeochemistry within the Flexible Ocean and Climate Infrastructure version 1 (FOCI 1) Earth system model, Zenodo [code], https://doi.org/10.5281/zenodo.6772175, 2022b

Doney, S. C., Mitchell, K. A., Henson, S. A., Cavan, E., DeVries, T., Gruber, N., Hauck, J., Mouw, C. B., Müller, J. D., and Primeau, F. W.: Observational and Numerical Modeling Constraints on the Global Ocean Biological Carbon Pump, Global Biogeochemical Cycles, 38, e2024GB008 156, https://doi.org/10.1029/2024GB008156, 2024.

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.

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.: Transport Matrix Method software for ocean biogeochemical simulations, https://doi.org/10.5281/zenodo.1246300, 2018.

Khatiwala, S.: Efficient spin-up of Earth System Models using sequence acceleration, Science Advances, 10, eadn2839 (, https://doi.org/10.1126/sciadv.adn2839, 2024.

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.

Kiko, R., Picheral, M., Antoine, D., Babin, M., Berline, L., Biard, T., Boss, E., Brandt, P., Carlotti, F., Christiansen, S., Coppola, L., de la Cruz, L., Diamond-Riquier, E., Durrieu de Madron, X., Elineau, A., Gorsky, G., Guidi, L., Hauss, H., Irisson, J.-O., Karp-Boss, L., Karstensen, J., Kim, D., Lekanoff, R. M., Lombard, F., Lopes, R. M., Marec, C., McDonnell, A. M. P., Niemeyer, D., Noyon, M., O'Daly, S. H., Ohman, M. D., Pretty, J. L., Rogge, A., Searson, S., Shibata, M., Tanaka, Y., Tanhua, T., Taucher, J., Trudnowska, E., Turner, J. S., Waite, A., and Stemmann, L.: A global marine particle size distribution dataset obtained with the Underwater Vision Profiler 5, Earth System Science Data, 14, 4315–4337, https://doi.org/10.5194/essd-14-4315-2022, 2022.

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.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Kriest, I. and Oschlies, A.: Swept under the carpet: organic matter burial decreases global ocean biogeochemical model sensitivity to remineralization length scale, Biogeosciences, 10, 8401–8422, https://doi.org/10.5194/bg-10-8401-2013, 2013.

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, 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, 2017.

Kriest, I., Kähler, P., Koeve, W., Kvale, K., Sauerland, V., and Oschlies, A.: One size fits all? Calibrating an ocean biogeochemistry model for different circulations, Biogeosciences, 17, 3057–3082, https://doi.org/10.5194/bg-17-3057-2020, 2020.

Kriest, I., Getzlaff, J., Landolfi, A., Sauerland, V., Schartau, M., and Oschlies, A.: Exploring the role of different data types and timescales in the quality of marine biogeochemical model calibration, Biogeosciences, 20, 2645–2669, https://doi.org/10.5194/bg-20-2645-2023, 2023.

Kriest, I., Kemena, T., and Guo, H.: Tuning ocean biogeochemistry for an Earth system model: approaches, challenges and impact on model performance, Zenodo [code], https://doi.org/10.5281/zenodo.19048349, 2026a.

Kriest, I., Kemena, T., and Guo, H.: Supplementary data for: Tuning ocean biogeochemistry for an Earth system model: approaches, challenges and impact on model performance [dataset]. GEOMAR Helmholtz Centre for Ocean Research Kiel [distributor]. https://hdl.handle.net/20.500.12085/16d3928c-0b30-11f1-a2ca-005056a30ade, 2026b

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. and Primeau, F.: Optimization and sensitivity of a global biogeochemistry ocean model using combined in situ DIC, alkalinity, and phosphate data, J. Geophys. Res., 113, 2008.

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.

Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the $1^\circ \times 1^\circ$ GLODAP version 2, Earth System Science Data, 8, 325–340, https://doi.org/10.5194/essd-8-325-2016, 2016.

Lindsay, K., Bonan, G. B., Doney, S. C., Hoffman, F. M., Lawrence, D. M., Long, M. C., Mahowald, N. M., Moore, J. K., Randerson, J. T., and Thornton, P. E.: Preindustrial-Control and Twentieth-Century Carbon Cycle Experiments with the Earth System Model CESM1(BGC), Journal Of Climate, 27, 8981–9005, https://doi.org/10.1175/JCLI-D-12-00565.1, 2014.

Madec, G. and the NEMO team: NEMO ocean engine, Tech. rep., Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No 27 Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No 27, https://www.nemo-ocean.eu/doc/, 2016.

Malin, F.: GMIS - MODIS-AQUA Monthly climatology sea surface Chlorophyll-a concentration (9km) in mg m^{-3}., Dataset, European Commission, Joint Research Centre (JRC), http://data.europa.eu/89h/51b9459f-aa6c-4160-9754-3e203c9c99b8, 2013.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Martiny, A., Vrugt, J., and Lomas, M.: Concentrations and ratios of particulate organic carbon, nitrogen, and phosphorus in the global ocean, Sci. Data, p. 1:140048, https://doi.org/10.1038/sdata.2014.48, 2014.

Matthes, K., Biastoch, A., Wahl, S., Harlass, J., Martin, T., Bruecher, T., Drews, A., Ehlert, D., Getzlaff, K., Krueger, F., Rath, W., Scheinert, M., Schwarzkopf, F. U., Bayr, T., Schmidt, H., and Park, W.: The Flexible Ocean and Climate Infrastructure version 1 (FOCI1): mean state and variability, Geosci. Model Dev., 13, 2533–2568, https://doi.org/10.5194/gmd-13-2533-2020, 2020.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenéz-de-la Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2, Journal of Advances in Modeling Earth Systems, 11, 998–1038, https://doi.org/10.1029/2018MS001400, 2019.

Moriarty, R. and O'Brien, T.: Global distributions of mesozooplankton abundance and biomass - Gridded data product (NetCDF) - Contribution to the MAREDAT World Ocean Atlas of Plankton Functional Types, Earth System Science Data, 5, 45–55, https://doi.org/10.5194/essd-5-45-2013, 2013.

Mouw, C. B., Barnett, A., McKinley, G. A., Gloege, L., and Pilcher, D.: Global ocean particulate organic carbon flux merged with satellite parameters, Earth Sys. Sci. Data, 8, 531–541, https://doi.org/10.5194/essd-8-531-2016, 2016.

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., 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.

Oschlies, A. and Schartau, M.: Basin-scale performance of a locally optimized marine ecosystem model, J. Mar. Res., 63, 335–358, 2005.

Pasquier, B., Holzer, M., Chamberlain, M. A., Matear, R. J., Bindoff, N. L., and Primeau, F. W.: Optimal parameters for the ocean's nutrient, carbon, and oxygen cycles compensate for circulation biases but replumb the biological pump, Biogeosciences, 20, 2985–3009, https://doi.org/10.5194/bg-20-2985-2023., 2023

Petrik, C. M., Luo, J. Y., Heneghan, R. F., Everett, J. D., Harrison, C. S., and Richardson, A. J.: Assessment and Constraint of Mesozooplankton in CMIP6 Earth System Models, Global Biogeochemical Cycles, 36, e2022GB007 367, https://doi.org/10.1029/2022GB007367, 2022.

Roth, R., Ritz, S. P., and Joos, F.: Burial-nutrient feedbacks amplify the sensitivity of atmospheric carbon dioxide to changes in organic matter remineralisation, Earth Syst. Dynam., 5, 321–343, https://doi.org/10.5194/esd-5-321-2014, 2014.

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.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Schmittner, A., Oschlies, A., Matthews, H. D., and Galbraith, E. D.: Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business-as-usual CO_2 emission scenario until year 4000 AD, Global Biogeochem. Cy., 22, https://doi.org/10.1029/2007GB002953, 2008.

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., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J. G., Li, H., Long, M. C., Luo, J. Y., Nakano, H., Romanou, A., Schwinger, J., Stock, C., Santana-Falcón, Y., Takano, Y., Tjiputra, J., Tsujino, H., Watanabe, M., Wu, T., Wu, F., and Yamamoto, A.: Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6, Current Climate Change Reports, 6, 95–119, https://doi.org/10.1007/s40641-020-00160-0, 2020.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, https://doi.org/10.1029/2019MS001739, 2019.

Silverman, B.: Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC, London, 1986.

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.

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., Laufkotter, C., Paulot, F., and Zadeh, N.: Ocean Biogeochemistry in GFDL's Earth System Model 4.1 and Its Response to Increasing Atmospheric CO2, Jour. Adv. Model. Earth Systems, 12, https://doi.org/10.1029/2019MS002043, 2020.

Taylor, K.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, https://doi.org/10.1029/2000JD900719, f1a, 2001.

Walker, S. L. and Palevsky, H. I.: Ocean Carbon Export Flux Projections in CMIP6 Earth System Models Across Multiple Export Depth Horizons, Global Biogeochemical Cycles, 39, e2024GB008 329, https://doi.org/10.1029/2024GB008329, e2024GB008329 2024GB008329, 2025.

Wilson, J. D., Andrews, O., Katavouta, A., Virissimo, F. d. M., Death, R. M., Adloff, M., Baker, C. A., Blackledge, B., Goldsworth, F. W., Kennedy-Asser, A. T., Liu, Q., Sieradzan, K. R., Vosper, E., and Ying, R.: The biological carbon pump in CMIP6 models: 21st century trends and uncertainties, Proc. Natl. Acad. Sci., 119, https://doi.org/10.1073/pnas.2204369119, 2022.

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.

https://doi.org/10.5194/egusphere-2026-3423
Preprint. Discussion started: 15 July 2026

Yool, A., Palmiéri, J., Jones, C. G., de Mora, L., Kuhlbrodt, T., Popova, E. E., Nurser, A. J. G., Hirschi, J., Blaker, A. T., Coward, A. C., Blockley, E. W., and Sellar, A. A.: Evaluating the physical and biogeochemical state of the global ocean component of UKESM1 in
870 CMIP6 historical simulations, Geoscientific Model Development, 14, 3437–3472, https://doi.org/10.5194/gmd-14-3437-2021, 2021.

Zeller, M. and Martin, T.: On warm bias and mesoscale dynamics setting the Southern Ocean large-scale circulation mean state, Ocean Modelling, 191, 102 426, https://doi.org/10.1016/j.ocemod.2024.102426, 2024.