I thank the four referees for their detailed comments. Below is my point-by-point response to them (as already published on BGs discussion website). This is followed by a marked-up version of the revised manuscript, that includes all the according changes I made, plus some minor changes with respect to wording. In the revised version I also discuss the recent paper by Galbraith et al. (2015) who compared the outcome of three different biogeochemical models, simulated in a common earth system model. This paper just recently came my attention.

Reply to referee 1:

I thank referee 1 for his/her constructive and helpful comments. Below are my answers to the comments, indicated by "IK:"

"I found the comparison interesting and informative, although the main conclusion from the comparison does not differ much from an earlier work by Kriest et al., 2010."

IK: Indeed the results obtained in this study point into the same direction as those of Kriest et al. (2010), the exception being that here via optimisation I use a more thorough and objective scan of the parameter space, with many different combinations of parameters. I will try to point this out in a revised version of the paper.

"The choice of the parameter sets to be optimised vs to be retained from earlier studiesseems arbitrary. Is there any particular reason or criteria by which the parameter sets are chosen for the optimization?"

IK Yes, there is - I am sorry that this did not come across properly: The first optimisation (MOPSˆoS; Kriest et al., 2017) was basically meant as a test for the framework, and a proof-of-concept. I therefore chose a very wide range of parameter types, across all trophic levels, and acting on different time and space scales (see also Kriest et al., 2012, for the large scale effect of "b"). As it turned out, mainly the surface parameters couldn't be constrained by the present misfit function; so I exchanged those by parameters related to deeper processe (MOPSˆoD). The selection of those was motivated by the large uncertainty regarding extent and expansion of oxygen minimum zones in models (Cocco et al., 2013; Cabre et al, 2013), and by the fact, that very little knowledge exists about these parameters (or even parameterisations). As it turned out, some of these were quite difficult to constrain, probably owing to the small volume of OMZs, relative to total ocean volume. Given that in RetroMOPS especially the parameters for DOP production and decay have to act as a shortcut for the entire complex surface ecosystem, I chose to optimise those, in addition to "b" and the -O2:P stoichiometry (as control). I will try to make this approach clearer in a revised version of the paper.

"How would the results differ if the author optimized all of the parameters presented in Table 1 simultaneously? Remineralized nutrients are eventually transported to the euphotic layer and become available to support primary production. Therefore all parameters are interrelated with each other. In other words, optimal values for Ic, KPHY, μzoo, and κzoo would be different with the newly optimized values for b*, KO2, KDIN, etc. This could affect the model evaluations as well."

IK: I agree that it would be most desirable to optimise all parameters at once. However, given what was found by Ward et al. (2010), and the possible interdependence of some of these parameters, this did not seem feasible to me. I will explain the choice of parameters in a revised version of the manuscript.

"Evaluation of DOP simulated from RetroMOPS remains qualitative. Although it is not sufficient, global datasets of DOM were presented in some previous studies including Letscher and Moore, 2015. How does the simulated DOP from the two models compare with the observations in terms of its distributions and concentrations? Why can't DOP be used as an extra constraint for the optimization in this study?"

IK: It can, and should, to my opinion, be part of the misfit function. However, as the global DOP data distribution is very sparse (compared to nutrients and oxygen), this may require a special treatment or weighting

in the misfit function. We are currently working on different ways and methods to account for DOP in model calibration. A thorough treatment and discussion of this would, I fear, exceed the scope of this paper, which is already quite diverse.

"In many parts of the manuscript, statements are quite qualitative. For example, in conclusions, the author wrote “results in a slightly better fit to observed tracers, and in a much better fit to observed estimates of ...” The author should provide some quantitative assessment. How good is it compared to other experiments within this study and also compared with other studies?"

IK: I will try to be more specific in the conclusion section of the revised version of this paper.

"Table 3: I don’t understand what rθ(Ω) represents. How is a priori range determined? How should I interpret the values? Is it discussed in the text?"

IK: This is the range of of parameter values of "very good" individuals, i.e. those model solutions, whose misfit is not worse than 1 permil of the best individual, divided by the "allowed" parameter range. For those parameters with a very wide range (or a priori uncertainty), the chance of having a huge spread in final solutions might just be larger - this is why I showed this normalised value of Rθ(Ω). I will try to explain this better in a revised version of the manuscript.

"Fig. 6: It is interesting that the six panels are all different in terms of the pairs of the parameter values corresponding to the lowest misfit and bias. Does it mean that the optimal values for the two parameters (as presented in Table 3) are not optimal anymore if only subsets of the tracers are used as a constraint or if the bias is used as a cost function?"

IK: Yes: e.g., if we chose only oxygen RMSE for the misfit function, we would probably end up with a different best set of parameters than when only using nitrate.

Technical corrections

"Equation (2): This formulation does not look like the Martin curve. Is the term z(j+1) right?"

IK: No, it wasn’t. It should indeed be z(k+1). Thank you for spotting this! "Equation (11): Please check the last term"

IK: The 1/bar(o_j) was in the wrong place: it should have been after the first sum of the second term. Thank you for drawing my attention to this.

"Page #8, line #24: typo" IK: Will be corrected.

"Page #12, line #1: There is no such a term like preformed waters. Perhaps change it to “reduces preformed DOP in subducted waters”?"

IK: Yes, thank you!

Reply to referee 2:

"Specific comment: Page 5, line 10-11: "The discretised flux divergence, that can actually be remineralised to phosphate and nitrate with available oxidants (oxygen and/or nitrate) Deff, (j), is then determined by...." - organic matter cannot be remineralized to nitrate by nitrate. In the absence of oxygen and the presence of nitrate, organic matter will be remineralized by denitrification (reducing nitrate to N2 while oxidizing the organic carbon), but the organic nitrogen in the organic matter will be remineralised and released as ammonium. This can in turn be oxidized to nitrate in the presence of oxygen (nitrification). Under low oxygen concentration, this process can occur parallel to

denitrification (coupled nitrification-denitrification). But in the absence of oxygen, ammonium will not be oxidised to nitrate. I am not sure if this is important in this context, but the sentence as it stands now is misleading."

IK: I thank Friederike Hoffmann for drawing my attention to this quite erroneous and misleading expression in the paper. Indeed, in MOPS organic matter is denitrified to phosphate and N2 (without any intermediate reduced N compounds). To avoid any confusion, I will skip to "phosphate and nitrate"

Reply to referee 3:

I thank referee 3 for his/her thorough reading and instructive comments. Below are my answers to the comments, indicated by "IK:"

"Regarding the model intercomparison, in the first instance, this is hampered by limitations in the traceability between the models. For instance, both models have DOP (results of which are later shown to be interesting), but (a) MOPS remineralises this at a single rate while RetroMOPS has two rates for different depths, and (b) this rate is around double that in RetroMOPS compared to MOPS. Similarly, the models both differ in value for other shared parameters, and even whether parameters are optimised between the models. Given that there is a degree of overlap in the components of both models, I would expect, firstly, that these would be as aligned as much as possible, and, secondly, if unaligned, that it would be fully explained why not. A passing remark on parameters being “hand-tuned” is not enough. The result of all of this is that the intercomparison seems less systematic and more arbitrary than it otherwise could be. Additionally, the manuscript’s discussion does not contextualise itself in a way that would help readers understand how what has been learned from this work could be applied to other models (the complex model here is still relatively simple, for instance).

IK: I will try to present the approach I followed in a more concise way, and will discuss the finding (Given this particular misfit function, a model that shortcuts biology at the surface van perform almost as good as a model with more detail) before the background of Kwiatkowski et al. (2014; see also below). I will further slightly restructure the paper to better emphasize the main points.

"Since the comparison between the two models is a major focus of the paper, at a minimum, I would expect a revised manuscript to better explain the seeming discrepancies in the choices made here, as well as more effort put into the intercomparison. Ideally, additional optimisations might be done to narrow the gaps between the models and permit a more complete intercomparison that would be of greater value to the marine biogeochemistry community."

IK: I am sorry that the degree of alignment between MOPS and RetroMOPS did not show up clearly enough: e.g., both models are exactly the same with respect to nutrient and light sensitivity of primary production. Also, the decay rate constants and oxidant sensitivities are both the same. The different parameterisations of DOP production and decay arises because this component in RetroMOPS acts as a kind of "shortcut" of the grazing-remineralization cycle in MOPS. In fact, the two different degradation rate constants were introduced to allow an even greater similarity between MOPS and RetroMOPS, allowing RetroMOPS a more flexible mimicking of a potential fast nutrient turnover. I will try to explain this better, and discuss potential insights and consequences in more detail, in a revised version of the manuscript. Given that circulation, and also the formulation of the misfit function are most probably two other, highly important factors in biogeochemical model calibration, and that optimisation is a computing-time consuming issue, I would nevertheless prefer to keep this set of experiments presented here, and in future runs concentrate on physical and observational errors and choices.

==Abstract==

"Lines 6-7: The paper introduces two models of differing complexity, but beyond noting that the simpler model does almost as well as the complex

one (which is not a given; e.g. Kwiatkowski et al., Biogeosciences, 2014), nothing else is said; more generally, this is more widely true of the manuscript — a better case / explanation should be made for presenting both the simple and complex models; especially as the two model structures are not traceably related to one another (though I'm sure that they share subroutines for specific processes)"

IK: This comment by the reviewer addresses two points: (a) the (lack of) introduction of RetroMOPS, as well as (b) its performance with respect to the skill metrics to biogeochemical tracers.

(a) When developing RetroMOPS I tried to include as many features of previous simple models global biogeochemical models (e.g., Bacastow and Maier-Reimer, 1991; Kriest et al., 2010, 2012), while at the same time maintaining MOPS's remineralisation scheme. The resulting "compromise" RetroMOPS is therefore not directly traceable to MOPS; however, the more gradual transition between model complexity (N, N+DOP, NP+DOP, NPZ+DOP, NPZD+DOP) presented in Kriest et al. (2010, 2012) might, to some extent, provide some more insight into effects of adding complexity on model skill. The revised version will include some more discussion on the current results before the background of this paper.

(b) I have tried to point out in the paper that it depends on the research question to be addressed with a model, which metric to choose; this will in turn determine which model is best suited to address that question. So far, the optimizations presented here indicate that a more complex model does not necessarily outperfom a simpler one. I will try to emphasize this even more in a revised version of the manuscript. However, I do not think that the results obtained by Kwiatkoski et al. (2014) contradict this finding, and suggest to put my results into a wider context by referring to their work: Kwiatkowski et al. (2014) compared six different global biogeochemical models, coupled to NEMO (1x1 spatial resolution and 75 vertical levels), and simulated over 118 years, against data sets of surface pCO2, DIC, alkalinity, DIN, Chl a and primary production. The models of vary in complexity from seven to 57 compartments, and thus also in their computational demand by almost a factor of five.

To assess model skill they ranked the models with respect to spatial correlation between and variance of model and observations. In general, the more complex models perform better with respect to simulated variance, but the simpler models better with respect to spatial correlation. Although no model outperforms all models across all metrics, they conclude that "Results suggest that little evidence that higher biological complexity implies better model performance in reproducing observed global-scale bulk properties."

This conclusion may be even more obvious when taking into account the ability of the different diagnostics to distinguish among the models: For example, spatial correlation of DIN ranges only betwee 0.79 to 0.94. Even more, for DIN, alkalinity and DIC normalized standard deviations vary less than 10% around the average standard deviation. Excluding these diagnostics from the model assessement would result in an advantage for the simpler models (MEDUSA or HADOCC) with respect to spatial correlation and a quite good performance of these model with respect to standard deviation (sum of ranks 10 and 9 for MEDUSA and HADOCC, respectively, compared to seven and 10 for PlankTOM6 and PlankTOM10). Finally, some of the models differ only very slightly in their performance (e.g., a difference between r=0.93 and r=0.92 for spatial correlation of DIC), in my opinion hampering the applicability of ranking.

Although it is clear that intermediate complexity models such as HADOCC cannot represent the level of detail embedded in more complex models, and that it cannot be ruled out that "more complex models can in future provide additional insight based on ongoing measurements and data syntheses", so far the model evaluation with respect to the bulk, biogeochemical observations such as dissolved inorganic tracers or chlorophyll does not seem to indicate any superiority of more complex models on a global scale. Although Kwiatkowski et al. (2014) apply very different temporal and spatial scales (given by the much shorter model spinup and focus on surface diagnostics), the results obtained with RetroMOPS and MOPS corroborate their findings. As noted by them, future availability of more complex data sets, such as different plankton groups, or particle distributions, will provide further insight about

the level of model complexity required, given the research question to be addressed with a model.

==Introduction==

"Lines 7-16: While I understand the need to keep manuscripts to a reasonable length, this paragraph is extremely dense and confusing; ideally, the concepts it describes should be clearly spelled out"

IK: I agree, and will add some sentences on the different methods applied.

"Lines 7-10: In particular, this list of tools is mentioned in passing without any contextualising information; a sentence on each would be helpful to readers unfamiliar with them"

IK: I agree, and will add some sentences on the different methods applied.

"Lines 17-18: This point about simulation time is slightly confusing here, since the preceding text is talking about accelerated models."

IK: Indeed, it is confusing and misleading. I will skip the reference to simulation time.

"Lines 27-31: As you latterly introduce oxidant dependent decay, it may be helpful to expand briefly on what parameters and processes Kriest et al. (2017) looked at"

IK: I will add some explanation on decay parameters and processes.

"Line 31: This sentence could do with an example, for instance “(e.g. parameter X was found to play a role in vertical distributions of process Y)”"

IK: I agree, and will add an example of the effect of b on large spatial phosphate distribution, with reference to Kriest et al. (2012)

"Pg. 3, line 2: “have been popular” – presumably this refers to use in CMIP, etc.?"

IK: Not necessarily to CMIP (complexity of many current CMIP models is more similar to MOPS), but to models used by e.g., Bacastow and Maier-Reimer, 1991; Matear and Hirst, 2003; Kwon et al., 2006; Dutkiewicz et al., 2006-

==Methods==

"Pg. 3, line 28: You say “A fraction” but don’t give a value; other parameter values are listed; what’s this one?"

IK: The fraction buried depends on the deposition rate onto the sediment (Kriest and Oschlies, 2013); I will add a brief description on it in the revised version of the paper.

"Pg. 3, line 30: “Non-buried detritus is resuspended into the water column” – this sounds intriguing, care to expand? For instance, resuspension over how much of the water column? What about the benthic boundary layer?"

IK: In fact, it is resuspended evenly in the last bottom box (i.e., there is no BBL). Effects of this have been investigated extensively in Kriest and Oschlies, 2013.

"Pg. 3-4: Simple flow schematics of the two models would probably be helpful"

IK: I will add two flow charts of MOPS and RetroMOPS.

"Pg. 4, lines 19-20: It is not immediately clear to me that the absence of the explicit plankton reservoirs in MOPS means that remineralisation would be too slow in RetroMOPS; arguably, the storage of material within particulate reservoirs for a period oftime might instead act to slow down remineralisation back to useable nutrient; in fact, is that not what Table 3 reports for its optimised value for this parameter?"

IK: Yes, indeed - thank you for the comment. I will add some discussion on this in the revision. See below, your comments and my reply re. Conclusions section: Given the rather fast turnover rates of DOP observed by Hopkinson et al. (2002), I did not find it appropriate to have a specific slowdown of remineralization.

"Pg. 5, equation 4: Why not just expand on what s_O2(j) and s_DIN(j) are here instead of sending the reader to another manuscript?"

IK: I will add some more description on this in the revised version (see also my response to Lines 27-31).

"Pg. 5, equations 4-5: These seem to imply that you calculation potential remineralisation, then calculate the possible remineralisation given O2 and DIN, then apply the latter at level k but apply the remainder at level k+1; how does this deal with the situation where level k+1 has insufficient oxidant?; that doesn't seem all that unlikely in OMZs (though with low vertical resolution as here, this may be less of a concern); more obviously, why mess around with where the remineralisation flux occurs and not just stop the sinking flux from remineralising?; for instance, couldn't the calculation of D(j) not use O2 and DIN to affect how the export flux is attenuated?; i.e. when there's no oxidant, remineralisation cannot occur vs. remineralisation occurs, where the required oxidant is taken from is dealt with afterwards"

IK: In the case layer k+1 also has insufficient oxidants, the organic matter will propagate further downwards, until it reaches sufficient oxidants, or the sea floor to be buried. Using this scheme I tried to be as close as possible to MOPS, where the explicit detritus sinks with its prescribed sinking speed, but only remineralises when it encounters enough oxygen or nitrate. So I think I have parameterised the model as suggested by reviewer 3; however, Eqn. 3 first presents the more general case (without oxidant dependency), which has been widely used in former simple global models.

"Pg. 6, line 2: In saying "nitrogen fixation balances the simulated loss", this implies a direct connection which does not appear to be the case in equation 6; instead, the model losses and gains inevitably come into a balance, but they are not directly linked (some other models do make this connection)"

IK: In the model nitrogen fixation balances denitrification on large time and space scales. It depends on biogeochemical parameters and circulation, how fast the two processes are connected (see also Kriest and Oschlies, 2015). I will add "in the long term" in a revised version of the manuscript, and add some words on the potential spatial distinction.

"Pg. 6, equations 7-10: With unwieldy equations like these, underbraces can be helpful in providing a quick reference for the reader as to the identity of the terms"

IK: This is a very good suggestion, thank you!

"Pg. 6, section 2.3: This overlooks any statement as to the performance of the physical model, even one that simply cites a source on this; given that the whole ocean misfit is used as an optimisation target, letting readers know that there's not a strong ventilation bias in the model ocean might be useful; this has a relationship with the next point ... Pg. 6, line 26: Is the Marshall et al. (1997) reference is the source of the circulation state used here?; or is it based on a more recent simulation?"

IK: This is the source of the circulation - see also Khatiwala et al. (2004) and Khatiwala (2008). The TM was derived from a 2.8×2.8 global configuration of the MIT model with 15 vertical layers, forced with monthly mean climatological fluxes of heat and freshwater, and subject to a weak restoring of surface temperature and salinity to observations. The circulation is detailed in Dutkiewicz et al (2005) and its configuration is similar to that applied in the Ocean Carbon-cycle Model Intercomparison Studies (OCMIP) (Orr et al, 2002). Circulation has been assessed within the OCMIP-2 project against a series of diagnostics and observations, such as T, S, and MLD (Doney et al. 2004), CFCs (Dutay et

al. 2002; Matsumoto et al. 2004) and radiocarbon (Matsumoto et al.2004; also Graven et al. 2012). These studies suggest a good overal perfomance comparable to other models, with some weaknesses (too much North Pacific intermediate waters, AABW water formation only in Drake passage; unrealistic spreading of the CFC-11 signal into the interior of the deep ocean in the deep western boundary current of the Atlantic), and strengths (e.g., mode water formation in the Antarctic). Depending on diagnostic applied, waters may appear too young in that model, although this is influenced by the upper boundary condition of the respective age tracer (Koeve et al., 2012). I will add some sentences on this in the discussion, but would prefer to direct the interested reader directly to these papers.

"Pg. 7, line 1: “After 3000 years most tracers have approached steady state” – This is an oddly loose definition of equilibrium; you could instead refer to the stabilisation of misfit J (e.g. that it fixes to N decimal places)"

IK: I have not checked for the stabilisation of the misfit for all model simulations, but I agree, this is a very useful information to have. In fact, J for the two optimal runs is stable (at least up to e-4) after 3000 years. I will add a plot of the transient misfit function of these runs to the supplement, and note, that this might - to some extent - depend on the parameter settings (cf Kriest and Oschlies, 2015).

"Pg. 7, lines 24-25: While the normalisation to global concentration should help with N and P (since they are related linearly through the ocean), does this overplay or underplay O2?; this doesn’t show the same sort of relationships (for obvious reasons); also, you don’t mention AOU at all – would this be a good alternative misfit target?"

IK: N and P in this model are not linearly related throughout the ocean, because P is conserved (and only affected by "Redfieldian" processes), while N (either nitrate or fixed nitrogen) may change due to nitrogen fixation and denitrification. Both contribute to 20%-30% to the misfit function, while oxygen contributes about 40-50% (see also Kriest et al., 2017, Figures 4, 10, 13). With regard to "overplay" or "underplay" of oxygen, I think it depends on what we are interested in: if OMZs, we might even wantto stress this tracer in the misfit function. Yes, AUO (or EOU) could be a very useful diagnostic for the misfit function; likewise preformed nutrients.

"Pg. 8, line 15: I might not have waited until the last sentence of this section to explain about the default parameter set"

IK: I will move this to the beginning of the paragraph.

"Pg. 8, section 2.7: Why is R_-O2:P optimised in MOPS but not in RetroMOPS?; I can’t immediately see why this isn’t an option"

IK: In MOPS’ optimisations, R_-O2:P did not show any significant deviations from its default value of 170, so I decided to keep this fixed, and only vary those parameters, that relate to the parts of RetroMOPS that are very different to MOPS (see above). I will try to explain this better in the revision.

"Pg. 8, section 2.7: MOPS is optimised with a reduced data set but RetroMOPS is not; this seems like a strange omission considering the same underlying issue affects both models; it is again symptomatic of the disparities between the models being intercompared"

IK: Because the reduced data set did not show any large changes in estimated parameters or misfit of MOPS, and because I assume that other issues might be more important to investigate in an optimisation context (circulation; components and form of the misfit function), I decided not to spend computational resources on RetroMOPS with reduced data set. This "reduced data" experiment was mainly aimed at investigating the potential effect of circulation in the equatorial Pacific on the parameter estimate. Possibly due to its small area, the effect is almost negligible.

==Results and Discussion==

"Pg. 9, line 10: Missing “of”, i.e. “Because of optimisation, MOPS¿oD results . . .”" IK: Will be corrected.

"Pg. 9, line 17: Typo on “threshold”" IK: Will be corrected.

"Pg. 9, line 20: The statement “impose a threshold” is unclear; do you mean that denitrification could not occur below this concentration?"

IK: Yes. I will rephrase this.

"Pg. 9, lines 25-34: What does this omitting of the Equatorial Pacific to the misfit in this region?; is it better or worse than when it’s included in the global misfit function?"

IK: The misfit for the equatorial Pacific becomes slightly worse (by about 3%) when omitting this region from the misfit function. I will add a few sentences on this in the revision.

"Pg. 9, lines 25-34: Also, what about the reverse situation where only the Equatorial Pacific (and / or OMZs more generally) is used for tuning?; if you tried that, perhaps a passing remark on it would be interesting"

IK: No, I did not try this as I suspect it is to a large extent the physics (e.g., missing equatorial jets) that causes the BGC misfit here.

"Pg. 10, lines 8-17: You note in the manuscript that “nitrogen fixation counteracts denitrification” but, as mentioned above, there’s no direct connection in the model (e.g. unlike some models that represent the former implicitly as a function of the explicit latter); in the context of (dis-)equilibrium, I don’t have a feel for the relative rates of the two processes in the work here; I guess I’m wondering if certain combinations of parameter values promote or diminish equilibration time; I suspect this is unlikely, but optimisation can take models to strange places"

IK: As we have shown (Kriest et al., 2015) sinking speed ("b") is one parameter that connects regions of N-loss with regions of N-gain (of course, before the background of circulation); DOM and POM and its lability will most likely be another candidate.

However, by year 3000 many of the models should have approached (more or less) equilibrium (Kriest and Oschlies, 2015, Figures 2 and 3).

"Pg. 10, lines 20-22: Per previous remarks on circulation, how good is the ocean’s ventilation?; reporting CFC or (especially) C-14 performance earlier would help (even if this reiterates previous work)"

IK: See my answer above; I will add some sentences on this in the revision, but would like to keep my focus on the biogeochemistry in the current paper.

"Pg. 10, line 23: One of the lambdas in the bracketed comment should be the surface DOP remineralisation rate; also, only one of them is given units"

IK: Yes, the second one. I will correct this and add the unit.

"Pg. 11, line 15: Does “direct evaluation of steady state” mean that they calculated the steady state solution analytically?"

IK: No, not analytically, but using the Newton scheme involving the model’s Jacobian. I will rephrase this in the revision of the manuscript.

"Pg. 11, lines 15-16: “. . . may still exhibit some drift . . .”?; it would probably be helpful to make this clearer, or possibly quantify it (e.g. in terms of misfit J fit; see my earlier remark); by all means use a fixed simulation duration, but knowing what this means for the misfit measure would be useful (e.g. its drift rate at this point)"

IK: See above: I will add figures on the transient of the misfit function to the supplement.

"Pg. 12, line 4: separate this last part of Section 3.3 into Section 3.4?; as it’s on the comparison of MOPS and RetroMOPS, it would make a clear section; it might also afford an location to delve a little deeper

into the complexity issue that's currently rather glossed over in the draft manuscript"

IK: Yes, thank you very much for this suggestion - I will do that.

"Pg. 12, line 27: remove spare comma to get ". . . (Table 4), and indicates that these tracers . . ."

IK: Will be corrected.

==Conclusions==

"Pg. 13, line 12: Regarding the use of observational DOP, can you clarify somewhere in the text how homogeneous DOP is?; i.e. is a single remineralisation timescale likely to be representative?"

IK: Hopkinson et al. (2002) applied a multi-G model to incubations of DOP sampled in surface waters of the middle Atlantic Bight, and measured decay constants for the very labile fraction of 0.22 per day (79 and 29-252 per year), with a range of 0.08-0.70 per day (29-252 per year). The labile fraction was characterized by a decay constant of ~0.02 per day (~7 per year), with a range of 0.002-0.053 per day (7.2 and 0.72-19 per year). The very labile and labile fraction constituted 32% and 50% of total DOP, respectively. RetroMOPS presented here focuses on the dominant labile fraction; its maximum possible rate for DOP decay for optimization is 7.2 per year, the observed average decay rate of the labile DOP in Hopkinson et al. (2002). Note that, however, most of the simulated ocean is far off the productive shelf areas; further, DOP in the model mimicks a variety of biogeochemical components (possibly even bacteria, or other non-sinking dead organic particles), and thus the observations may not be directly transferable to simulated DOP. In a three-step optimization process Letscher et al. (2015a) optimized a global model of semi-labile and refractory DOM against observations, and found rates of 0.016 per year for semilabile DOP at the surface, and 0.22 per year for semilabile DOP in the mesopelagical. Production and turnover rates for refractory DOP were very small, except for an additional photo-oxidation rate of 0.07 per year. The optimum decay rate of 0.47 per year found in this study is within the range estimated by Letscher2015; also, the nonrequirement of fast surface turnover agrees with their results, which point towards lower remineralization of DOP at the sea surface. I will add more details on the range of potential decay rates, and my particular choice for boundaries in a revised version of the paper.

==Acknowledgements==

"Pg. 13: Is this paper part of a special issue or wider celebration of the life of Ernst Maier-Reimer?; if so, an earlier note in the introduction would seem to be in order; if not, it may be worth making the rationale for this tribute a little clearer (e.g. note Maier-Reimer's recent passing)"

IK: It is part of a special issue in memory of Ernst Maier-Reimer, and there is an introduction by Christoph Heinze. I here just wanted to add my personal acknowledgment.

==Figures and Tables==

Table 1: Why is the ostensibly fixed parameter DINmin very slightly different in the two RetroMOPS runs?

IK: This was a typo, I correct this to 15.80 for RetroMOPS^r

Reply to referee 4:

I thank referee 4 for his/her critical yet helpful comments. Below is my reply, indicated by "IK"

"My overall feeling is that the paper is badly lacking in focus. Reading through I was always struggling to understand what major point the author was hoping to make. Is it that the simple model is nearly as good as the complex model, or is it that different parts of the model are better constrained by different kinds of observations? At the moment the

article reads as if two separate (and somewhat poorly developed) stories have been combined into one, with very little thought as to what connects them. I think that the author either needs to pick one theme, and develop it better, or needs to do a much better job of finding a narrative thread linking the two themes together. It is up to the author to identify how that might be achieved."

IK: I am sorry that the paper appears to be so unfocused, and will try to explain my reasoning better here, and in a revised version of the paper. In short, given the sometimes high structural complexity of global biogeochemical models there are only sparse observations to constrain them, the two main findings of the paper:

- complex, biological dynamics are not well constrained by a rather biogeochemical misfit to nutrients and oxygen

- the simple model performs almost as good as the more complex one, with respect to the given misfit function are somehow connected. Calibrating a complex model would possibly require either a much more complex misfit function (with respect to observations; e.g. using Chl a, observations of zooplankton abundance, and DOP). Given that

- models of higher complexity, such as MOPS, are usually applied to research questions that relate to more biogeochemical issues (such as ocean carbon inventory, or deoxygenation)

- these models are expensive in terms of computing time, thereby hampering exploitation of model (parameter) sensitivity and skill in spun up state, and

- more "sophisticated" data sets are sparse, and many of the observations may contain a high uncertainty, or noise

I find it important to raise some awareness about the necessary level of model complexity, and the uncertainty associated with model structure and parameters. In some cases it may be more approriate to spend more time on carefully exploiting the parameter space instead of adding more complexity. Of course, this tightly relates to the research question addressed with the model. I will do my best to render the paper more focused and clear in a revised version.

Specific comments

Abstract

"Line 7: "a complex seven-component model (MOPS), and a very simple two-component model (RetroMOPS)" and "The simple model, which contains only nutrients and dissolved organic phosphorus (DOP)". RetroMOPS clearly has four components: PO4, NO3, O2 and POM."

IK: Yes, thank you. This will be corrected.

"Line 13: Please do a better job of explaining what is "the global bias"."

IK: I will add "(global inventory of oxygen and fixed nitrogen)"

1 Introduction

"Line 29: "[Kriest et al. (2017)] showed that annual mean tracer concentrations do not provide much information on parameters related to the dynamic biological processes taking place in the euphotic zone". Should be "annual mean tracer concentrations did not provide much information", as I am not convinced this is a general result for all models."

IK: It would be interesting to see other models when facing optimisation against the same misfit (volume weighted RMSE of annual mean nutrients and oxygen); Until then, I agree, it should be "did".

2.2.1 Primary production

"Equation 1: Why use the mean phytoplankton concentration at all? It would be more consistent with the rest of the model (i.e. Equation 5) to convolve the specific growth rate and the phytoplankton concentration

into a single growth rate of the phytoplankton population (mmol P m-3 d-1)."

IK: I used this particular decomposition to clearly illustrate the specific assumptionsthat may be inherent in simple models such as RetroMOPS (similar to Kriest et al., 2010). In addition, deriving the specific growth rate from MOPS, and transferring this to RetroMOPS, would involve accounting for nutrient concentration and limitation - which in turn depend on the remineralisation rate and sinking speed. I therefore chose this way of aligning both models, and would prefer to keep it that way. Note that the resulting specific growth rates (including limitation by nutrients, temperature and light) between both models are not too different: 0.1021 d-1 (RetroMOPS) and 0.1267 d-1 (MOPS). I will add a sentence on this in the revision.

2.2.2 The fate of primary production: Export, DOP production and remineralisation

"Line 19: "DOP then decays to phosphate and nitrate". To me it would make sense to call it POM."

IK: POM would be something that sinks, which clearly distinguishes it from DOM.

"Line 19: "To allow for a potential, fast recycling loop at the surface, RetroMOPS parameterises an additional decay rate". Presumably this is inspired by (Oschlies 2001), but why would this be necessary in the absence of assimilated primary production observations?"

IK: There are three reasons why I have embedded this fast recycling loop: first, DOP production and decay in RetroMOPS has to mimick all dynamic surface processes of MOPS, so I initially expected it to require a specific degradation rate constant for the surface. As it turned out, this is not necessary (this parameter during optimisation was reduced to nearly zero). Second, at a later stage it might indeed be interesting (and helpful) to include primary production into the misfit function, with possibly different resulting best parameters surface DOP decay. Finally, data by Hopkinson et al (2002) indicate that DOP recycling rates may be much higher than commonly applied in global models. I will add some discussion on this in a revised version of the paper.

"Equation 4: I think a bit more could be said about the interdependence of sO2(j)and sDIN(j). For example, their sum forms a coefficient for remineralisation, so it is important to note that their sum is constrained between 0 and 1."

IK: I will comment more on this function in a revised version of the paper.

2.5 Misfit function

"Equation 11: I am a bit confused by how the misfit function and its components are defined. In particular, I cannot see how o ÌDˇ (the global average observed concentration of the respective tracer) is included in the RHS. "

IK: This was a mistake by me; The RHS was missing 1/bar(o_j) after the first sum, but it should have been after J(j). Thank you for drawing my attention to this.

"Also, it seems that the model is being compared to gridded observations, instead of observational equivalents being extracted at the spatiotemporal locations of the observations. As the gridding process will introduce its own set of errors, this choice needs some justification."

IK: Although regridding the observations onto the coarser model grid removes much of the variability in the observations, this procedure is much more efficient (in terms of computing time) during the optimisation process. Further, by following this approach the model is not penalised for its apparent lack of resolution. It could be worthwhile adding the variance, that arises from the regridding process, and the variance in the data themselves, as weight to the misfit function. However, in an earlier study (Kriest et al., 2010) we could not find any large effects of this on model assessment. Testing different misfit function with

respect to observational data sets, weighting schemes, etc., will be subject of follow-up work, but possibly exceed the scope of this paper.

2.6 Optimisation of MOPS

"Line 15: I don’t think including results from the hand-tuned model brings anything of value to the paper."

IK: I disagree: as "hand-tuning" still seems to be common practice in global biogeochemical modelling, I think it merits some presentation and discussion.

3.1 Optimisation of MOPS

"Line 4: Fig 1 (rather than Fig S1)?"

IK: Both - this will be changed to "Figures 1 and S1".

"Line4:10%and1%(ratherthan10%and1âU˚e¸/âU˚e¸âU˚e¸)?"

IK: Yes, thank you for noting this!

"Line 12 (and throughout the text): "reduced denitrification". It is probably safer to avoid the word reduced except with regard to the chemical process."

IK: I agree, and will exchange "reduced" by "lower".

3.3 Optimisation of RetroMOPS

"Line 21: "The misfit to phosphate (Fig. 8, lower left panel) shows an elongated valley in the two-dimensional projection". I do not see a valley in this Figure. The misfit slopes down from the top-right towards the bottom left, but there is nothing to suggest it slopes back up again after reaching a minimum."

IK: I was referring to the - admittedly few - green and red points in the lower right corner, that indicate an increase in the misfit function. I will replace "shows" by "indicates".

Calibration of a simple and a complex model of global marine biogeochemistry

Iris Kriest[1]

[1]GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany

Correspondence to: Iris Kriest (ikriest@geomar.de)

Abstract.

The assessment of the ocean biota's role in climate ~~climate~~ change is often carried out with global biogeochemical ocean models that contain many components, and involve a high level of parametric uncertainty. Because many data that relate to tracers included in a model are only sparsely observed, assessment of model skill is often restricted to tracers that can be easily measured and assembled. Examination the models' fit to climatologies of inorganic tracers, after the models have been spun up to steady state, is a common, but computationally expensive procedure to assess model performance and reliability. Using new tools that have become available for global model assessment and calibration in steady state, this paper examines two different model types - a complex seven-component model (MOPS), and a very simple ~~two-component~~ four-component model (Retro-MOPS) - for their fit to dissolved quantities. Before comparing the models, a subset of their biogeochemical parameters has been optimised against annual mean nutrients and oxygen. Both model types fit the observations almost equally well. The simple model, which contains only nutrients, oxygen and dissolved organic phosphorus (DOP), is sensitive to the parameterisation of DOP production and decay. The spatio-temporal decoupling of nitrogen and oxygen, and processes involved in their uptake and release, renders oxygen and nitrate valuable tracers for model calibration. In addition, the non-conservative nature of these tracers (with respect to their upper boundary condition) introduces the global bias (fixed nitrogen and oxygen inventory) as a useful additional constraint on model parameters. Dissolved organic phosphorous at the surface behaves antagonistically to phosphate, and suggests that observations of this tracer - although difficult to measure - may be an important asset for model calibration.

1 Introduction

Global biogeochemical ocean models are now routinely used to assess the ocean biota's role in climate change. Although these models have become ever more complex with respect to the number of biogeochemical tracers they contain, they are often calibrated only against a subset of their components, mostly nutrients, oxygen, and components of the carbon cycle (e.g. Bacastow and Maier-Reimer, 1991; Ilyina et al., 2013; Cocco et al., 2013; Cabre et al., 2015).

There has been an intensive discussion about the necessary level of marine ecosystem model complexity, mostly on a theoretical basis, or in a local or regional context (e.g., Anderson, 2005; Le Quere, 2006; Flynn, 2006; Anderson, 2006; Leles et al., 2016; Shimoda and Arhonditsis, 2016). It remains an open question whether additional complexity is of advantage for repre-

senting biogeochemical processes and tracers on a global scale ~~(i.e., for processes acting on rather long time and large space scales; e.g., Kr~~ (i.e., for processes acting on rather long time and large space scales). For example, Kriest et al. (2010, 2012) found no large differences when comparing model skill with respect to oxygen and phosphate across a range of models of different complexity, but quite large effects of parameter settings when applying a coarse examination of the parameter space.

However, a thorough and dense scan of the parameter space would be required for a fair assessment of the virtues of models of different complexity. Such a scan usually requires many model evaluations, which, given long equilibration time scales of coupled global models (Khatiwala, 2008; Wunsch and Heimbach, 2008; Primeau and Deleersnijder, 2009; Siberlin and Wunsch, 2011), is difficult to carry out. For assessment of only surface properties and processes a short model ~~spinup~~ spin-up may be sufficient; however, on a global scale, many centuries to millennia of coupled model simulations are necessary, in order to remove the drift in biogeochemical tracer fields and fit to observed properties (Kriest and Oschlies, 2015; Seferian et al., 2016).

Only recently tools have become available that allow for a reduction in simulation times, such as the Transport Matrix Method (TMM, Khatiwala et al., 2005), which replaces the ocean circulation model with an efficient "offline" circulation, or methods that solve for steady state tracer fields using Newton's method~~, which requires the inverse~~. The latter require either the inversion of the Jacobian ~~(Kwon and Primeau, 2006, 2008),~~ (e.g., Kwon and Primeau, 2006), or apply matrix-free Newton-Krylov (MFNK, Khatiwala, 2008; Li and Primeau, 2008) ~~, or surrogate-based optimisation (Priess et al., 2013). The~~ to compute the steady-state solution. Surrogate-based optimisation replaces the original and computationally expensive model by a so-called surrogate, which is created from a less accurate but computationally cheaper model. The latter is corrected to reduce the misalignment between the two solutions. Priess et al. (2013) applied this method, together with the TMM to recover parameters of a simple global biogeochemical model; the surrogate in their case consisted of shorter (decades) spin-ups.

The gain in computational efficiency resulting from these methods can then be used for a systematic calibration of global biogeochemical models. For example, Kwon and Primeau (2006, 2008) used global climatological data sets of phosphate, inorganic carbon, and alkalinity to calibrate a simple global biogeochemical model. The misfit between observed and simulated phosphate was used by DeVries et al. (2014) to calibrate parameters related to particle properties in a simple two-component, nutrient-restoring model. In a similar approach Holzer et al. (2014) optimised parameters for opal production and dissolution against observed silicate. Letscher et al. (2015) switched between a complex and a simple model of ocean biogeochemistry to estimate production and decay rates of dissolved organic phosphorus on a global scale.

All these biogeochemical models employed in global parameter estimates were of a low level of biogeochemical complexity. One reason for this restriction might be ~~the large number of tracers in more complex models, which increases simulation time. Another problem is~~ associated with the variety of time scales associated with more complex models. Piwonski and Slawig (2016) used MFNK to evaluate the steady state of simple and complex biogeochemical models. They noted that "[...] for more complex models the Newton method requires more attention to solver parameter settings [...]" (Piwonski and Slawig, 2016), which may be related to the highly nonlinear structure of these models. The nonlinearity, and the large number of parameters, also complicates their simultaneous optimisation (Ward et al., 2010). On a global scale, these problems are amplified by the sparsity of observations of organism groups, particularly of higher trophic levels. Observations of dissolved inorganic con-

stituents, on the other hand, are much more frequent, and therefore provide more information on the spatio-temporal variability of these tracers.

Recently, Kriest et al. (2017) combined the TMM with an estimation of distribution algorithm (Covariance Matrix Adaption Evolution Strategy, CMA-ES), to ~~optimize~~ optimise six biogeochemical parameters of a seven component model against global climatologies of annual mean phosphate, nitrate and oxygen. They showed that annual mean tracer concentrations ~~do~~ did not provide much information on parameters related to the dynamic biological processes taking place in the euphotic zone, but that parameters related to ~~larger time- and space scales~~ long time and large space scales (e.g., the remineralisation length scale or so-called "Mar be estimated from these observations. ~~To follow up on this, I here~~ The large uncertainty associated with surface parameter estimates can be attributed to the relatively small volume of the surface ocean, leading to a misfit that is dominated by deep ocean observations.

Replacing the misfit function by a metric that targets at the surface ocean, and/or contains additional observations that provide information on plankton could be one way to resolve this indeterminacy. Alternatively, one could skip these parameters from optimisation and focus on parameters more tightly connected to the meso- and bathypelagic ocean. A more drastic measure is lies in downscaling the biogeochemical model to a simpler system, that only contains components with a counterpart in global, quasi-synoptic data sets. The latter procedure may help to elucidate, which level of complexity is required to represent and investigate global distribution and patterns of biogeochemical tracers.

This paper examines the the latter two potential solutions: Firstly, I investigate if parameters related to oxidant-dependent decay in the mesopelagial are better constrained by this type of misfit function. This is done by replacing four parameters of the optimisation carried out by Kriest et al. (2017) by parameters related oxidant-affinity of remineralisation, and - to account for the possible alterations in fixed nitrogen turnover - by the maximum nitrogen fixation rate.

~~Given~~ Secondly, given the successful parameter optimisation of simpler models noted above, and also to acknowledge the fact, that these models have been popular and quite successful in global simulations of ocean biogeochemistry (e.g. Bacastow and Maier-Rei this paper ~~further~~ presents an optimised model, which has been derived from downscaling the seven-component model MOPS (Kriest and Oschlies, 2015; Kriest et al., 2017) to a model that retains only three abiotic dissolved tracers (phosphate, nitrate, and oxygen) and one biotic tracer (dissolved organic phosphorus; DOP). This new model, which I refer to as "RetroMOPS", includes the oxidant-dependency of MOPS, but is otherwise very similar to models applied earlier in global models. In contrast to some of these models (Marchal et al., 1998; Najjar et al., 2007) it assumes no relaxation to observed tracer fields, but simulates changes in surface production fully prognostic, as in Bacastow and Maier-Reimer (1991); Maier-Reimer (1993); Matear and Hirst (2003); Parekh et al. (2005).

After a brief presentation of model MOPS (Kriest and Oschlies, 2015), the downscaled model RetroMOPS is introduced, followed by an outline on circulation, optimisation and experimental design (section 2). In section 3 results from optimisation of both MOPS and RetroMOPS are presented and discussed. The paper closes with conclusions drawn from these experiments.

2 Models, experiments, and optimisations

2.1 Model MOPS

Model MOPS (Kriest and Oschlies, 2015) is based on phosphorus, and simulates seven compartments. Phosphate, phytoplankton, zooplankton, dissolved organic phosphorus (DOP) and detritus are calculated in units of mmol P m^{-3}. Oxygen is coupled to the P-cycle with a constant stoichiometry given by $R_{-\mathrm{O2:P}}$. Aerobic remineralisation of organic matter follows a saturation curve, with half-saturation constant K_{O2}. Aerobic remineralisation ceases when oxygen declines; at the same time, denitrification takes over, as long as nitrate is available above a defined threshold, DIN_{min}. Like the oxic process, suboxic remineralisation follows a saturation curve for oxidant nitrate, with half-saturation constant K_{DIN}. MOPS does not explicitly resolve the different oxidation states of inorganic nitrogen (nitrite, N_2O, ammonium), but assumes immediate coupling of the ~~the~~ different processes involved in nitrate reduction, the end-product being dinitrogen (see also Paulmier et al., 2009; Kriest and Oschlies, 2015). All organic components are characterised by a constant N:P stoichiometry of $d = 16$. Loss of fixed nitrogen is balanced by a simple parameterisation of nitrogen fixation by cyanobacteria, which relaxes the nitrate-to-phosphate ratio to d with a time constant, μ^*_{NFix}. Detritus sinks with a vertically increasing sinking speed: $w = a\,z$. Assuming a constant degradation rate r, in equilibrium this would result in a particle flux curve given by $F(z) \propto z^{-b}$, with $b = r/a$. For better comparison with values of b derived from observations (e.g. Martin et al., 1987; Van Mooy et al., 2002; Buesseler et al., 2007), and with the simpler model RetroMOPS (see below), a is expressed in terms of b (assuming constant, nominal $r = 0.05$ d^{-1}). A fraction of detritus deposited at the sea floor (at the bottom of the deepest vertical box) is buried instantaneously in some hypothetical sediment. The fraction buried depends on the deposition rate onto the sediment. Non-buried detritus is resuspended into the last box of the water column, where it is treated as regular detritus. The phosphorus budget is closed on an annual time scale through resupply via river runoff. More details about the biogeochemical model and parameters, and their effects on model behaviour can be found in Kriest and Oschlies (2013) and Kriest and Oschlies (2015).

2.2 Model RetroMOPS

MOPS' structure has been simplified by skipping the explicit simulation of phytoplankton, zooplankton, and detritus ~~.~~ (see Fig. S1). The remaining equations of, and functional relationships between, phosphate, nitrate, oxygen and DOP have been parameterised similar to MOPS. Because the downscaled model resembles so many features of earlier biogeochemical models simulated in a global context (e.g., Bacastow and Maier-Reimer, 1991; Maier-Reimer, 1993; Matear and Hirst, 2003), but keeps the oxidant dependency of MOPS, the model is named "RetroMOPS".

2.2.1 Primary production

Like MOPS, RetroMOPS calculates primary production only in the euphotic zone, which, in the current configuration, is confined to the upper two numerical layers ($k_{\mathrm{EZ}} = 2$, $z = 0{-}120$ m). ~~Phytoplankton~~ As in Kriest et al. (2010) and Kriest et al. (2012) phytoplan is parameterised with a constant concentration of $\overline{PHY} = 0.02$ mmol P m^{-3}, which is the mean phytoplankton concentration

in the upper 120 m of two optimised model setups MOPS^{oS} and MOPS^{oD} (see below). Using this constant phytoplankton concentration RetroMOPS calculates light- and nutrient dependent primary production P in each layer k as:

$$P(k) = \begin{cases} \mu_{\mathrm{PHY}}\,\overline{PHY}\,\min\left(f(I(k)),\,\frac{L(k)}{K_{\mathrm{PHY}}+L(k)}\right) & : \quad k \le k_{\mathrm{EZ}} \\ 0 & : \quad k > k_{\mathrm{EZ}} \end{cases} \tag{1}$$

where $f(I(k))$ defines light-limitation, μ_{PHY} is the temperature-dependent maximum growth rate of phytoplankton, and L determines the limiting nutrient: $L(k) = \min(\mathrm{PO_4}(k), \mathrm{DIN}(k)/d)$ (see Kriest and Oschlies, 2015, for more details). Note that, with the given parameters for nutrient and light affinity, the resulting specific growth rates ($P(k)/\mathrm{PHY}$) of optimised MOPS and RetroMOPS are quite similar (0.127 d^{-1} for MOPS and 0.102 d^{-1} for RetroMOPS).

2.2.2 The fate of primary production: Export, DOP production and remineralisation

Instead of resolving heterotrophic processes (zooplankton grazing, excretion and egestion) at the sea surface explicitly, in RetroMOPS a fraction $\cancel{\sigma_{\mathrm{DOP}}}$ σ of organic matter fixed photosynthetically is immediately released as dissolved organic phosphorous, DOP. DOP then decays to phosphate and nitrate with a constant rate $\cancel{\lambda_{\mathrm{DOP}}}\lambda$. To allow for a potential, fast recycling loop at the surface, RetroMOPS parameterises an additional decay rate, $\cancel{\lambda_{\mathrm{sDOP}}}\lambda_{\mathrm{s}}$, that affects DOP only in the first two layers. By doing so, the model mimics multiple DOP fractions with different decay rates, as observed by Hopkinson et al. (2002). The remaining fraction of production, $\cancel{1-\sigma_{\mathrm{DOP}}}1-\sigma$, of each layer in the euphotic zone is exported to the layers below, where it immediately remineralises to nutrients, following a power-law of depth. The discretised form for flux F into box j from all (surface) source layers k, with $1 \le k \le k_{\mathrm{EZ}}$ is then given by:

$$F(j) = \sum_{k=1}^{k=k_{\mathrm{EZ}}} P(k)\,(1-\sigma_{\mathrm{DOP}})\,\Delta z(k) \left(\frac{z(j)}{z(j+1)}\,\frac{z(j)}{z(k+1)}\right)^{-b} \qquad \text{for} \quad j > k\,, \tag{2}$$

where $\Delta z(k)$ denotes the thickness of a numerical (source) layer, and $z(j)$ is the depth of the upper boundary of layer j. The flux divergence, $D = \mathrm{d}F/\mathrm{d}z$, for any box j in discretised form is defined by

$$D(j) = \frac{F(j-1) - F(j)}{\Delta z(j)} \tag{3}$$

Neglecting oxidant dependency of decay, the entire flux divergence $D(j)$ would be released as phosphate and nitrate, with equivalent oxidant consumption. It is, however, possible that oxidants become depleted at some location. Earlier models in this case continued the degradation of organic matter, thereby implicitly assuming unspecified oxidants (e.g., Marchal et al., 1998; Matear and Hirst, 2003; Najjar et al., 2007; Kriest et al., 2010, 2012). In contrast, RetroMOPS, like MOPS, accounts for suppression of remineralisation (oxic/suboxic) in the absence of sufficient oxidants, by assuming saturation curves for the limitation by either oxygen or nitrate. The amount of organic matter available for oxidation is given by the decay of dissolved

organic matter, ~~λ_{DOP} DOP~~ λDOP, and by the flux divergence, $D(j)$ (Eqn. 3). The discretised flux divergence, that can actually be remineralised ~~to phosphate and nitrate~~ with available oxidants (oxygen and/or nitrate), $D^{\mathrm{eff}}(j)$, is then determined by

$$D^{\mathrm{eff}}(j) = D(j)\,(s_{\mathrm{O2}}(j) + s_{\mathrm{DIN}}(j)) \qquad , \tag{4}$$

where $s_{\mathrm{O2}}(j)$ and $s_{\mathrm{DIN}}(j)$ represent the oxidant limitation terms, ~~as detailed in Kriest and Oschlies (2015, Equations 15-27).~~ expressed as saturation curves l_{O2} and l_{NO3} for either oxygen (oxic remineralisation) or nitrate (denitrification), with half-saturation constants K_{O2} and K_{DIN}. Denitrification is further inhibited by oxygen via $(1 - l_{\mathrm{O2}})$, resulting in $s_{\mathrm{O2}}(j) + s_{\mathrm{DIN}}(j) = 1$; (see also Equations 15-27 of Kriest and Oschlies, 2015). In all models oxic remineralisation only takes place down to a a lower threshold of $\mathrm{O_2} = 4\mathrm{mmol\ m^{-3}}$. The lower threshold for denitrification is determined by parameter DIN_{min}, and subject to optimisation.

The ~~remaining~~ flux divergence that cannot be remineralised under the given concentrations of oxidants is added as additional flux divergence to the layer below:

$$D(j+1) \quad = \quad D(j+1) + (D(j) - D^{\mathrm{eff}}(j))\,\frac{\Delta z(j)}{\Delta z(j+1)} \tag{5}$$

where again $D^{\mathrm{eff}}(j+1)$ is evaluated. In the bottom layer the remaining flux that has not been remineralised in the water column eventually enters the sediment.

2.2.3 Benthic exchanges

Models that implicitly assume unspecified oxidants often prescribe a zero boundary flux, i.e. all organic matter in the last bottom box is degraded instantaneously (e.g., Marchal et al., 1998; Matear and Hirst, 2003; Najjar et al., 2007; Yool et al., 2011). Both MOPS and RetroMOPS have to take "leftover" organic matter flux into account, that arrives undegraded at the sea floor because of incomplete remineralisation in the water column. The explicit detritus compartment in MOPS allows for only partial burial at the sea floor, which may result in detritus accumulation in the deepest model box (see Kriest and Oschlies, 2013). Because there is no such detritus compartment in RetroMOPS, all flux arriving at the sea floor is buried immediately. Therefore, MOPS and RetroMOPS differ with respect to their lower boundary condition.

2.2.4 Nitrogen fixation

~~In~~ Both RetroMOPS and MOPS ~~nitrogen fixation balances the simulated loss of fixed nitrogen via denitrification. Both models~~ do not explicitly simulate cyanobacteria, but assume zero net growth of these organisms, parameterised as an immediate release of fixed nitrogen as nitrate:

$$S^{NFix}_{DIN}(k) = \begin{cases} \mu^*_{\mathrm{NFix}} f_1(\mathrm{T}(k)) f_2(\mathrm{DIN}(k), \mathrm{PO_4}(k)) & : \quad k \le k_{\mathrm{EZ}} \\ 0 & : \quad k > k_{\mathrm{EZ}} \end{cases} \tag{6}$$

f_1 parameterises the temperature dependence of nitrogen fixation with a second order polynomial approximation of the function by Breitbarth et al. (2007). f_2 regulates the relaxation of the nitrate:phosphate ratio towards the global observed stoichiometric ratio of $d = 16$. ~~μ^*_{NFix} is the~~ The maximum nitrogen fixation ~~of the parameterized cyanobacteria population(mmol N m^{-3}d^{-1}; see Ki~~ μ^*_{NFix} (mmol N m^{-3}d^{-1}) parameterises an implicit cyanobacteria population. In the long term nitrogen fixation balances the simulated loss of fixed nitrogen via denitrification, although they may occur in distant areas (see Kriest and Oschlies, 2015, for more details

2.2.5 Source-minus-sinks

Combining the above mentioned processes and interactions, the time rate of change in each layer k for phosphate, nitrate, oxygen, and DOP due to biogeochemical processes are

$$\begin{aligned}
S^{\mathrm{PO4}}(k) &= -P(k) \underbrace{-P}_{\text{production}} + \lambda_{\mathrm{sDOP}}\,\mathrm{DOP}\,\underbrace{\lambda_{\mathrm{s}}\,\mathrm{DOP}}_{\text{surface decay}} + D(k) + \lambda_{\mathrm{DOP}}\,\mathrm{DOP}(k)\,s_{\mathrm{O2}}(k) + s_{\mathrm{DIN}}(k)\,\underbrace{[D + \lambda\,\mathrm{DOP}]}\,[s_{\mathrm{O2}} \\
S^{\mathrm{DOP}}(k) &= \sigma_{DOP}\,P(k)\,\underbrace{\sigma_{DOP}\,P}_{\text{release}} - \lambda_{\mathrm{sDOP}}\,\mathrm{DOP}\,\underbrace{\lambda_{\mathrm{s}}\,\mathrm{DOP}}_{\text{surface decay}} - \lambda_{\mathrm{DOP}}\,\mathrm{DOP}(k)\,s_{\mathrm{O2}}(k) + s_{\mathrm{DIN}}(k)\,\underbrace{\lambda\,\mathrm{DOP}\,[s_{\mathrm{O2}} + s_{\mathrm{DIN}}]}_{\mathrm{d} \ldots} \\
S^{\mathrm{O2}}(k) &= R_{-O2:P}\,P(k)\,\underbrace{R_{-O2:P}\,P}_{\text{production}} - R_{-\mathrm{O2:P}}\,\lambda_{\mathrm{sDOP}}\,\mathrm{DOP}\,\underbrace{R_{-\mathrm{O2:P}}\,\lambda_{\mathrm{s}}\,\mathrm{DOP}}_{\text{surface decay}} - R_{-\mathrm{O2:P}}\,D(k) + \lambda_{\mathrm{DOP}}\,\mathrm{DOP}^*\,s_{\mathrm{O} \ldots} \\
S^{\mathrm{DIN}}(k) &= -d\,P(k)\,\underbrace{-d\,P}_{\text{production}} + S^{NFix}_{DIN}(k)\,\underbrace{S}_{\text{N-fixation}} + d\,\lambda_{\mathrm{sDOP}}\,\mathrm{DOP}\,\underbrace{d\,\lambda_{\mathrm{s}}\,\mathrm{DOP}}_{\text{surface decay}} + D(k) + \lambda_{\mathrm{DOP}}\,\mathrm{DOP}\,s_{\mathrm{O2}}(k)
\end{aligned}$$

Summarising, RetroMOPS is similar to model "N-DOP" of Kriest et al. (2010, 2012), to the phosphorus component of the model presented by Parekh et al. (2005), or to the models presented by Bacastow and Maier-Reimer (1991) and Maier-Reimer (1993), the exception being details of primary production at the sea surface, and the explicit parameterisation of oxidant-dependent remineralisation. By assuming constant cyanobacteria biomass, and a relaxation of the nitrate:phosphate ratio via immediate release of fixed nitrogen, its parameterisation of nitrogen fixation is similar to the one described by Maier-Reimer et al. (2005) and Ilyina et al. (2013).

2.3 Circulation and physical transport

All model simulations apply the Transport Matrix Method (TMM; Khatiwala, 2007, `github.com/samarkhatiwala/tmm`) for tracer transport, with monthly mean transport matrices (TMs), wind, temperature and salinity (for air-sea gas exchange) derived from a 2.8° global configuration of the MIT ocean model, with 15 levels in the vertical~~(Marshall et al., 1997).~~ ,

as described in Marshall et al. (1997) and Dutkiewicz et al. (2005). The circulation model was forced with climatological annual cycles of wind, heat and freshwater fluxes, and subject to a weak restoring of surface temperature and salinity to observations. Its configuration is similar to that applied in the Ocean Carbon-Cycle Model Intercomparison Studies (OCMIP) (Orr et al., 2000), which has been assessed against observations of temperature, salinity and mixed layer depth (Doney et al., 2004), CFCs (Dutay et al., 2002; Matsumoto et al., 2004) and radiocarbon (Matsumoto et al., 2004; Graven et al., 2012). Overall, its performance is comparable to other global models.

Using this efficient offline approach, a time step length of 1/2 day for tracer transport and 1/16 day for biogeochemical interactions, simulation of 3000 years requires about 0.5-1.5 hrs on 4 nodes (24 core Intel Xeon Ivybridge) at a High Performance Computing Centre (`www.hlrn.de`). After 3000 years most tracers have approached steady state ~~.~~ (see also Kriest and Oschlies, 2015, for and the transient of the misfit function becomes very small (see Fig. S2). The last year is used for model analysis and evaluation of the misfit function.

2.4 Optimisation algorithm

Optimisation of parameters is carried out using an Estimation of Distribution Algorithm, namely the Covariance Matrix Adaption Evolution Strategy (CMA-ES; Hansen and Ostermeier, 2001; Hansen, 2006). The application of this algorithm to the coupled biogeochemistry-TMM framework has shown good performance with respect to quality and efficiency (in terms of function evaluations), and is described only briefly below. More details about the algorithm, its setup and coupling to the global biogeochemical model can be found in Kriest et al. (2017).

Let n be the number of biogeochemical parameters to be estimated. In each iteration ("generation") the algorithm defines a population of λ individuals (biogeochemical parameter vectors of length n), with $\lambda = 10$ (derived from the default parameter $\lambda = 4 + 3\ln(n)$, Hansen and Ostermeier, 2001). The candidate vectors are sampled from a multi-variate normal-distribution, which ~~generalizes~~ generalises the usual normal distribution, also known as Gaussian distribution, from $\mathbb{R}$ to the vector space $\mathbb{R}^n$.

Following the simulation of these λ individual model setups to steady state (3000 years), the misfit function is evaluated, and information of the current, as well as previous generations is used to update the probability distribution in $\mathbb{R}^n$ such that the likelihood to sample good solutions increases. Usually, the realisation of the probability distribution update ensures that information of former solutions fades out slowly, resisting for several iterations. Therefore, the population (the number of model simulations per generation) in CMA-ES is smaller, and of less computational demand, than in classical evolutionary algorithms. Nevertheless, CMA-ES can still, to a certain degree, perform well with misfit functions characterised by a rough topography (Kriest et al., 2017).

2.5 Misfit function

As in Kriest et al. (2017) the misfit to observations J is defined as the root-mean-square error RMSE between simulated and observed annual mean phosphate, nitrate, and oxygen concentrations (Garcia et al., 2006a, b), mapped onto the three-dimensional model geometry. Although regridding the observations onto the coarser model geometry removes some of the

variability, this method is computationally more efficient in an optimisation framework. Also, a sensitivity study with a similar coupled model showed that accounting for the variance inherent in the observational data, and arising from regridding did not have a large influence on the misfit (Kriest et al., 2010).

Deviations between model and observations are weighted by the volume of each individual grid box, V_{i}, expressed as fraction of total ocean volume, V_{T}. The resulting sum of weighted deviations is then normalised by the global mean concentration of the respective observed tracer:

$$J = \sum_{j=1}^{3} J(j) \frac{1}{\overline{o_j}} = \sum_{j=1}^{3} \frac{1}{\overline{o_j}} \sqrt{\sum_{i=1}^{N} (m_{i,j} - o_{i,j})^2 \frac{V_i}{V_{\mathrm{T}}}} \tag{11}$$

$j = 1, 2, 3$ indicates the tracer type and $i = 1, ..., N$ are the model locations for $N = 52749$ model grid boxes. $\overline{o_j}$ is the global average observed concentration of the respective tracer. $m_{i,j}$ and $o_{i,j}$ are model and observations, respectively. By weighting each individual misfit with volume, J serves more as a long time-scale geochemical estimator, in contrast to a misfit function that e.g. focuses on (rather fast) turnover in the surface layer, or resolves the seasonal cycle.

2.6 Optimisation of MOPS

~~Building upon the~~ Based on a "hand-tuned", a priori setup of MOPS (Kriest and Oschlies, 2015), which hereafter is referred to as MOPS^{r}, Kriest et al. (2017) presented an optimisation of mostly surface-related parameters ~~presented in Kriest et al. (2017, "OBS-NARR optimisation~~ (hereafter referred to as MOPS^{oS}). They chose a very wide range of parameter types, across all trophic levels, and acting on different time and space scales. In that optimisation many of the surface parameters were difficult to constrain, because of a misfit function that consists mostly of observations in the deep ocean. Optimisation MOPS^{oD} ~~presented here aims at calibrating parameters related to processes that directly affect the oxidants nitrate and oxygen in subsurface layers. In MOPS^{oD} the optimal parameters of MOPS^{oS} for light and nutrient affinity of phytoplankton, zooplankton grazing and its mortality are retained, and parameters relevant for deep aerobic and anaerobic remineralisation are subject to change during optimisation (Table 1)~~presented here applies the same metric, but focuses on parameters in subsurface waters. The selection of parameters to be optimised is motivated by the large uncertainty regarding extent and expansion of oxygen minimum zones in models (Cocco et al., 2013; Cabre et al., 2015), and because little knowledge exists about their values, or even parameterisations.

Parameter K_{O2} determines the affinity of the aerobic remineralisation to oxygen, and the gradual transition from this process to denitrification (see Eqns. 15 and 20 of Kriest and Oschlies, 2015). K_{DIN} determines the affinity of denitrification to nitrate. Parameter ~~DINmin~~ DIN_{min} defines the lower threshold for the onset of denitrification. MOPS^{oD} also optimises the maximum rate of nitrogen fixation, μ^*_{NFix}, which balances fixed nitrogen loss through denitrification. The fifth and sixth parameter to be estimated are the oxygen requirement per mole phosphorus remineralised, $R_{-O2:P}$, and the flux (or remineralisation) length scale, b. Upper and lower boundaries of parameters to be optimised have been set to a rather wide range (Table 1), to allow optimisation to explore a wide range of potential parameters. The optimal parameters of MOPS^{oS} for light and nutrient affinity

of phytoplankton, zooplankton grazing and its mortality are retained in MOPS^{oD} (Table 1). Therefore optimisation MOPS^{oD} builds upon a previous tuning of surface processes.

Most of the processes affected by the parameters to be optimised take place in suboxic waters, e.g. of the eastern equatorial Pacific (EEP). Given the coarse model geometry, it is possible that circulation dynamics are not represented well in the model. To investigate the influence of observations ~~entering the misfit function~~ within this region on misfit function and parameter estimates, MOPS^{oD} is repeated with a reduced data set, that excludes the ~~eastern equatorial Pacific (~~EEP (here: east of 140°W, between 10°S and 10°N) from the misfit function. This optimisation is named MOPS$^{oD}_{*}$. ~~In the following, results from the optimised models MOPS^{oS}, MOPS^{oD} and MOPS$^{oD}_{*}$ are compared to a reference experiment, MOPS^{r}, which represents a "hand-tuned", a priori setup of this model.~~

2.7 Optimisation of RetroMOPS

In model RetroMOPS processes such as grazing of phytoplankton, and its subsequent release of organic or inorganic phosphorus are parameterised via a single component, DOP. Because DOP production and decay regulate the partitioning between sinking and dissolved organic matter, optimisation RetroMOPS^{o} targets at these parameters, namely ~~σ_{DOP}, λ_{sDOP} and λ_{DOP}.~~

~~While σ_{DOP}~~σ, λ_s and λ. While σ, as parameter that regulates the export ratio, may be more or less well constrained, ~~λ_{sDOP} and λ_{DOP}~~ λ_s and λ both include a variety of processes, which may act on time scales of days to years. ~~In a set of nine a priori sensitivity experiments the effect of these parameters on the misfit function is explored by varying λ_{DOP}~~ Hopkinson et al. (2002) applied a multi-G model to incubations of DOP sampled in surface waters of the middle Atlantic Bight, and measured decay constants for the very labile fraction (32% of total DOP) of ≈ 80 y^{-1}, with a range of 3-254 y^{-1}. Half of total DOP was in the labile fraction and characterised by a decay constant of ≈ 7 y^{-1}, ranging from 0.8-43 y^{-1}. However, these observations may not be directly transferable to globally simulated DOP, because most of the simulated ocean is far off the productive shelf areas; further, DOP in RetroMOPS is assumed to mimic a variety of biogeochemical components and processes. In a three-step optimisation study Letscher et al. (2015), who optimised a global model of semi-labile and refractory DOM against observations estimated rates of 0.016 y^{-1} for semilabile DOP at the surface, and 0.22 y^{-1} for semilabile DOP in the mesopelagial, i.e. much lower than suggested by Hopkinson et al. (2002). Summarising, the potential decay rate of the very labile to semi-labile fraction varies over several orders of magnitude, from $O(0.01) - O(100)$ y^{-1}.

Optimisation of RetroMOPS focuses on the dominant labile to semi-labile fraction, but allows for some potential fast turnover rates of DOP at the sea surface (towards the values observed by Hopkinson et al., 2002). To obtain a first impression on model sensitivity towards these parameters, a set of nine a priori experiments, that vary λ between 0.18 y^{-1} and 0.72 y^{-1} ~~, and λ_{sDOP}~~ and λ_s between 0 y^{-1} and 0.36 y^{-1} ~~(see table~~has been carried out (Table 2)~~. The results of these sensitivity experiments~~ , which provide a guidance for upper and lower boundaries ~~of optimisation (table 1) . The~~ for optimisation of RetroMOPS. To nevertheless explore the full range of potential decay rates, the maximum possible rate ($\lambda + \lambda_s$) for optimisation is set to 7.2 y^{-1}, towards the average decay rate of the labile DOP observed by Hopkinson et al. (2002). Optimised RetroMOPS^{o} will be compared to the sensitivity experiment with the lowest misfit (~~$\lambda_{sDOP} = 0$, $\lambda_{DOP} = 0.36$)is used for comparison with the optimised RetroMOPS, and referred to~~ $\lambda_s = 0$, $\lambda = 0.36$), which is denoted as RetroMOPS^{r}.

The explicit representation of detritus in MOPS may result in considerable numerical diffusion (particularly on coarse vertical grids as used here; see also Kriest and Oschlies, 2011) and thus in a different estimate of optimal b then when applying a direct flux curve, such as in RetroMOPS. Therefore, b is included as fourth parameter to be optimised. The effect of explicit vs. implicit flux description on parameter estimate will be discussed in more detail below.

All other parameters (primary production, oxidant-dependent remineralisation, stoichiometry) have been fixed to those obtained in optimisations MOPS^{oS} and MOPS^{oD} (Table 1). By doing so, optimisation RetroMOPS^{o} builds upon previous optimisations of the more complex MOPS. While it might be desirable to optimise all parameters of RetroMOPS at once, this study rather aims at investigating to what extent a simpler model can serve as a shortcut to the more complex one, given the applied misfit function and observations.

3 Results and discussion

3.1 ~~Optimisation~~ Optimal remineralisation parameters of MOPS

Both ~~$R_{\mathrm{O2:p}}$~~ $R_{\mathrm{O2:P}}$ and b are constrained very well by the observations, as indicated by a well-defined minimum of the misfit function (Fig. S3), and ~~the~~ a narrow, almost gaussian distribution of the best 10% to 1~~‰~~ % of parameters (Fig. 1~~,~~). On the other hand, parameters related to remineralisation or nitrogen fixation are determined with lower accuracy. This is also reflected in the rather wide range of candidate solutions within 1‰ of the best misfit, which vary between 10% to 20% of their assigned a priori range range (Table 3). Thus, in the presence of noise inherent in the observations, some parameters could only be estimated within a quite wide range of uncertainty, a feature that has already been addressed in a one-dimensional model by Löptien and Dietze (2015). So far, the potential consequences of this parametric uncertainty for other metrics (such as extent of oxygen minimum zones, OMZ) and possibly transient scenarios (e.g., their impact on simulated future evolution of OMZ volume) are not known.

The good determination of b by dissolved inorganic tracers is agreement with earlier studies ~~(Kwon and Primeau, 2006; Kriest et al., 2017 Parameters related to oxidant-dependent remineralisation approach the lower (K_{O2}) or upper (K_{DIN}, DIN_{min}) boundary, with a rather wide, skewed distribution.The rate for~~ that applied the same model (Kriest et al., 2017; Schartau et al., 2017). Its optimal value is very close to that obtained in MOPS^{oS}, i.e. higher than the value estimated by Kwon and Primeau (2006). Optimisation of maximum nitrogen fixation rate shows a slightly skewed distribution, but suggests an overall good estimate of this parameter.

~~Fixed nitrogen loss and gain depend on parameters for oxidant-dependency of remineralisation : In MOPS^{oS}, both fluxes are very high (Fig. 2), and outside the observed range (Table 4). Because optimisation MOPS^{oD} results in a strongly reduced affinity to, and higher threshold of, nitrate, its pelagic fixed nitrogen loss is almost halved, and now agrees with observed global estimates (Table 4). Further, as a result of reduced denitrification, the nitrate deficit in the eastern equatorial Pacific is smaller; however, at the cost of a small underestimate of observed oxygen in this region (Fig. 3) . The latter is a consequence of the now very low half-saturation constant for oxygen uptake (Table 3). Overall, optimisation of parameters related to the oxidant affinity of oxic and suboxic remineralisation leads to a slightly improved fit to tracer concentrations, to $J^* = 98\%$ of that of MOPS^{oS}~~

~~(Table 3) , and to a better agreement with observed estimates of global biogeochemical fluxes (Table 4)~~Optimal parameters for oxidant-dependent remineralisation also show wide, skewed distributions, with their mode near the lower (K_{O2}) or upper (K_{DIN}, DIN_{min}) boundary.

~~Optimisation MOPS^{oD} results in a high treshold~~ The high thresholds for the limitation of denitrification ~~, with K_{DIN} and DIN_{min} close to their upper boundaries. The increase protects~~ protect nitrate from becoming depleted in the upwelling regions, particularly the eastern equatorial Pacific, and ~~resembles~~ resemble results obtained by Moore and Doney (2007): To prevent their model from reproducing unrealistically low nitrate values in this region, they had to impose a threshold of 32 mmol NO_3 m^{-3} for the occurrence of denitrification. An explanation for this requirement of a high nitrate threshold might be found in the representation of the equatorial intermediate current system in coarse resolution models, which can result in spurious zonal oxygen gradients (Dietze and Loeptien, 2013; Getzlaff and Dietze, 2013). It is possible that the optimisation of biogeochemical parameters attempts to ameliorate these effects, which are in fact caused by the parameterisation of physics.

To further investigate the impact of this region on the parameter estimate, an additional optimisation was carried out, that ~~applies~~ targets at the same set of parameters~~to be optimised~~, but omits the eastern equatorial Pacific from the calculation of the misfit function. This optimisation MOPS$^{\mathrm{oD}}_{*}$ generates a lower threshold of nitrate for the onset of denitrification, and a higher maximum nitrogen fixation rate (Table 3), resulting in slightly enhanced fixed nitrogen turnover, particularly in the eastern equatorial Pacific (Fig. 2). ~~Global fixed nitrogen loss increases by about 20%, towards the upper limit of observed estimates (Table 4).~~ Compared to MOPS^{oD} the estimates of K_{DIN} and DIN_{min} become more uncertain with respect to the best 10% to 1‰ individuals, and ~~show an even~~ even show a bimodal distribution (Fig. S4, Table 3). The uncertainty in parameter estimates can be related to the missing data in regions of simulated denitrification. ~~Summarising, although the~~ Because the misfit function excludes the EEP it is lower then when considering the entire ocean (Table 3). A posteriori evaluation of misfit to the entire data set results in a misfit of 0.439, the same as for MOPS^{oD}.

Global fixed nitrogen turnover depends on parameters for oxidant-dependency of remineralisation: In MOPS^{oS}, both denitrification and nitrogen fixation are very high (Fig. 2), and outside the observed range (Table 4). Because of the reduced affinity to nitrate, in MOPS^{oD} pelagic fixed nitrogen loss is almost halved, and now agrees with observed global estimates (Table 4). Further, as a result of lower denitrification, the nitrate deficit in the eastern equatorial Pacific ~~, and~~ is smaller, but at the cost of a small underestimate of observed oxygen in this region (Fig. 3). The latter is a consequence of the now very low half-saturation constant for oxygen uptake (Table 3). In MOPS$^{\mathrm{oD}}_{*}$ the constraint on nitrate affinity is again relaxed, resulting in an enhancement of fixed nitrogen turnover by about 20%, towards the upper limit of observed estimates (Table 4).

Overall, optimising parameters related to the oxidant affinity of oxic and suboxic remineralisation leads to a slightly improved fit to tracer concentrations, to $J^{*} = 98\%$ of that of MOPS^{oS} (Table 3), and to a better agreement with observed estimates of global biogeochemical fluxes (Table 4). Although the eastern equatorial Pacific, and potential unresolved processes in simulated circulation, ~~evoke only relatively small effects on~~ has no effect on global misfit, its effect on some parameter estimates, ~~these nevertheless result~~ however, results in an increase in global fixed nitrogen loss of about 20%.

3.2 ~~Sensitivity of~~ A shortcut for surface biology: RetroMOPS~~to DOP production and decay~~

Given that parameters related to surface biology were difficult to constrain in MOPS^{oS}, and, within a certain range, exert only a small influence on the fit to global tracer distributions (Kriest et al., 2017), this section examines if RetroMOPS, as a model that parameterises surface biology in a much simpler way, suffices to represent biogeochemical tracer fields. Starting from growth and decay parameters optimised in MOPS, sensitivity experiments and optimisation search for optimal parameters for DOP production and decay, that mimic the surface nutrient turnover of MOPS.

3.2.1 Sensitivity to DOP production and decay

In RetroMOPS fast DOP recycling ~~leads to the~~ results in higher primary production, export production, and deep organic particle flux, especially in the equatorial upwelling regions (Fig. 4). While this has only a small effect on vertically or globally averaged phosphate concentrations (Figures 5 and 6), it ~~leads to a a~~ causes a large underestimate of nitrate in the ocean (Figures S6 and 6). The underestimate can be explained by the tight coupling between production, export and denitrification, which leads to higher denitrification and global fixed N-loss (Fig. 4), and thus a larger nitrate deficit (Fig S6) in the eastern equatorial Pacific~~. This is~~ , in agreement with effects hypothesised and investigated by Landolfi et al. (2013).

In ~~the model nitrogen fixation counteracts fixed nitrogen loss through denitrification. In contrastto nitrogen fixation , which~~ contrast, nitrogen fixation is not much affected by DOP turnover rates~~, global fixed nitrogen loss increases with increasing DOP decay rate (Fig. 4).~~ . The imbalance between nitrogen losses and gains suggests that the models even after 3000 years of simulation are not yet in equilibrium~~. It~~ , which might be explained by the large spatial scales between regions of fixed nitrogen loss and gain~~, in conjunction with the slow turnover rates of the biogeochemical processes~~. The divergence increases with ~~increasing recycling rates~~ higher DOP recycling rates (and thus larger denitrification), indicating that there is no unique equilibration time scale for one and the same model, but that it depends on biogeochemical parameters associated with sinking and remineralisation of organic matter, as observed earlier (Kriest and Oschlies, 2015). The resulting ~~long spinup times~~ requirement for long spin-up times for a complete model adjustment, their dependence on biogeochemical parameters, and the model's nonlinearity during ~~spinup~~ spin-up (Kriest and Oschlies, 2015), complicate model calibration and assessment, in addition to those factors already investigated by Seferian et al. (2016). It emphasises the need for a thorough assessment of trade-offs between model complexity and computational demand, and the possibility to examine the parameter space in sufficient detail.

The effect of DOP recycling on oxygen concentrations differs from its effect on nitrate. With fast recycling DOP is remineralised mostly at its place of production, and does not contribute much to oxygen consumption in deep waters (see also Fig S5). As a consequence, deep oxygen concentrations are high, particularly in the northern North Pacific (Fig. 5), and global average oxygen is overestimated by more than 10% (Fig. 6). Slow DOP recycling, in contrast, leads to less organic matter remineralisation in ~~preformed,~~ well-ventilated waters, but more remineralisation in deep waters. This in turn results in an underestimate of global mean oxygen of almost 10% (for ~~$\lambda_{DOP} = 0.18$~~$\lambda = 0.18$ y^{-1} and ~~$\lambda_{DOP} = 0$~~$\lambda_s = 0$ y^{-1}), which is somewhat surprising, given that production and export in this scenario are the lowest of all simulations (Fig. 4).

Overall, the best fit to observed inorganic tracer concentrations is achieved with moderate DOP recycling (Table 2, Fig. 5).

Most likely because of its fixed inventory, phosphate contributes to less than 1/3 of the misfit function, and is quite insensitive to changes in DOP recycling rate (Fig. 6). Nitrate and oxygen play a larger role for model fit, because their inventory can adapt to changing biogeochemistry. The misfit to nitrate and oxygen increases more or less ~~increases~~ in concert with their bias (Fig. 6). Therefore, these tracers with their flexible inventory provide some very useful constraints on DOP recycling rates.

Slow DOP recycling increases DOP concentrations at the surface, particularly in the ACC and in the northern North Atlantic (Fig. 5) ~~, and simulated concentrations largely~~ towards concentrations that exceed the observations (Yoshimura et al., 2007; Raimbault et al., 2008; Torres-Valdes et al., 2009; Letscher and Moore, 2015). Only the simulation with quite fast DOP recycling of ~~$\lambda_{\mathrm{DOP}} = 0.72$~~$\lambda = 0.72$ y^{-1} and ~~$\lambda_{\mathrm{sDOP}} = 0.36$~~$\lambda_{\mathrm{s}} = 0.36$ y^{-1} results in reasonable concentrations of DOP - but at the cost of too high phosphate concentrations along these sections, and a too high global misfit (Table 2), a too low nitrate and too high oxygen inventory (Figures 5 and 6). Therefore, it should be noted that despite the relatively good fit of RetroMOPS^{r}, it nevertheless suffers from a potential mismatch to DOP, which so far is not included in misfit evaluation.

3.3 ~~Optimisation of RetroMOPS~~

3.2.1 Optimal parameters for DOP cycling in RetroMOPS

All four parameters of RetroMOPS^{o} are ~~quite~~ well constrained by the observations, as indicated by the narrow, almost gaussian distribution around the optimal parameter (Figures 7, S7, and Table 3). Optimisation reduces the decay rate for surface DOP, ~~λ_{sDOP}~~λ_{s}, to almost zero, i.e., in RetroMOPS there seems to be no requirement for fast DOP turnover at the surface, similar to the results obtained by Letscher et al. (2015). The optimal ~~DOP~~ total remineralisation rate of DOP ($\lambda + \lambda_{\mathrm{s}}$) is about 0.5 y^{-1}, ~~and the~~ more than twice as high as the recycling rate estimated by Letscher et al. (2015), but lower than the rates observed by Hopkinson et al. (2002). The optimal fraction of primary production released as DOP, σ, is 73% ~~, resulting in a slightly higher turnover as compared to the reference scenario RetroMOPS^{r}. Optimal σ_{DOP}~~ and agrees very well with $\sigma = 0.74$ obtained by Kwon and Primeau (2006); however, their optimal DOP decay rate was twice as high (1 y^{-1}).

When optimising a simple biogeochemical model similar to RetroMOPS against observed phosphate, Kwon and Primeau (2006) noted a correlation between DOP production fraction and decay rate, impeding the simultaneous estimation of these parameters. On the contrary, in optimisation RetroMOPS^{o} both ~~σ_{DOP}~~ σ and the DOP decay rates seem to be rather well constrained. An analysis of the different components of the misfit function, similar to Fig. 4 of Kwon and Primeau (2006), helps to resolve this apparent contradiction. For this, in Fig. 8 ~~misfit (~~the total misfit J and its components $J(j)$ of Eqn. 11~~) and~~, as well as the bias of the best 5% of all ~~individual is mapped against σ_{DOP}~~ individuals are mapped against σ and DOP decay timescale ~~$\tau = 1/(\lambda_{\mathrm{DOP}} + \lambda_{\mathrm{sDOP}})$~~$\tau = 1/(\lambda + \lambda_{\mathrm{s}})$.

Note that the analysis depicted in Fig. 8 differs from that of Kwon and Primeau (2006) in several aspects: Firstly, their global biogeochemical model was fully equilibrated (due to their direct evaluation of steady state via Newton's method), whereas simulations of RetroMOPS may still exhibit some drift in nitrogen inventory (see subsection 3.2.1 and supplement). Second, Kwon and Primeau (2006) evaluated model sensitivity at $b = 1$, while Fig. 8 displays a region $\pm 5\%$ around optimal ~~b~~$b = 0.98$. Thirdly, Fig. 8 maps only the misfit of solutions realised by the optimisation routine, while Kwon and Primeau (2006) analysed

the entire parameter space at $b = 1$. ~~Finally~~Most important, the misfit function applied here is based on three components, with very different properties and associated time scales (see above), which can be of advantage for parameter estimation.

The misfit to phosphate (Fig. 8, lower left panel) ~~shows~~ indicates an elongated valley in the two-dimensional projection on DOP decay timescale τ (years) and DOP production fraction ~~σ_{DOP}~~σ, and resembles Fig. 4 of Kwon and Primeau (2006). Indeed, one of the lowest misfits to phosphate is achieved with about the same set of parameters as in Kwon and Primeau (2006), namely $\tau \approx 1$, ~~$\sigma_{\mathrm{DOP}} \approx 0.73$~~$\sigma \approx 0.73$. However, nitrate and oxygen show a different, and, partly, antagonistic, pattern: the best fit to observed nitrate is achieved with rather high values of ~~$\sigma_{\mathrm{DOP}} \approx 0.8$~~ $\sigma \approx 0.8$ and τ between about 1-2 years, while the best fit to oxygen is obtained with ~~$\sigma_{\mathrm{DOP}} \approx 0.7$~~ $\sigma \approx 0.7$ and $\tau \approx 1.5$ years. The superposition of the different components of the misfit function leads to a unique optimum ~~of~~ at $\tau = 2$ (~~$\lambda_{\mathrm{DOP}} = 0.47$ and $\lambda_{\mathrm{DOP}} = 0.02$) and $\sigma_{\mathrm{DOP}} = 0.73$~~ $\lambda = 0.47$ and $\lambda_s = 0.02$) and $\sigma = 0.73$ (Table 3). Thus, oxygen and nitrate can provide some useful, independent information on these parameters.

~~As noted above, the advantage of including nitrate and oxygen in the misfit function is that, in contrast to phosphate, the~~ This can partly be explained by their non-conservative nature. As noted in section 3.2.1 the inventory of these tracers may change freely according to model parameterisation. The resulting bias to observations thus adds two important components to the misfit function, both of which are independent: while high DOP turnover (as simulated by low τ) biases nitrate low (Fig. 8, upper mid panel), the same value leads to an overestimate of oxygen (Fig. 8, upper right panel; see also Fig. 6). This behaviour can be explained with the different processes and boundary conditions for the two tracers already noted ~~above~~in section 3.2.1: a high DOP turnover leads to higher fluxes and a tighter coupling of production and denitrification in upwelling waters, causing a nitrate deficit in the model (see above, and Fig S6). On the ~~contrary~~other hand, it reduces ~~DOP in preformed~~ preformed DOP in subducted waters e.g., ~~in~~ the Southern Ocean, thereby decreasing aerobic remineralisation and oxygen consumption in these waters on their passage towards, e.g., the northern North Pacific. The latter process increases oxygen particularly in deep waters (Fig S5).

To summarise, including nitrate and oxygen as non-conservative tracers in the misfit function helps to resolve parameters related to DOP production and decay on long time scales. This can be explained by the different pathways of DOP originating from upwelling regions or subducted water masses in the high latitudes, and is confirmed by the analysis of sensitivity experiments presented in section 3.2.1. However, a better fit to observed phosphate seems to come at the expense of a mismatch to observed DOP concentration. It remains to be investigated, if a simultaneous fit to observed inorganic and organic phosphorus is possible.

3.2.2 Comparison of MOPS and RetroMOPS

The optimal $b = 0.98$ of RetroMOPS^{o} is lower than that of MOPS^{oS} and MOPS^{oD}. This may be partially explained with the absence of numerical diffusion of detritus in RetroMOPS. As shown by Kriest and Oschlies (2011), in models that explicitly simulate detritus sinking with an upstream scheme ~~,~~ the assumption of homogenous ~~distribution of detritus~~ detritus distribution in each vertical grid box causes an additional, usually downward transport of detritus. This results in an effective b which is about 10-20% smaller (corresponding to faster sinking) than the nominally prescribed b. Optimisation of MOPS accounts for

this additional numerical transport by increasing b (= reducing sinking velocity) by some amount. Therefore, optimal b of MOPS without any influence numerical of diffusion would likely be around 1.1-1.2, i.e. closer to $b = 0.98$ of RetroMOPS^{o}. Considering this effect, optimal b of MOPS^{oD} and, in particular, RetroMOPS^{o} agree with the optimal value of $b = 1$ found by Kwon and Primeau (2006).

Despite its generally lower fluxes, fixed nitrogen loss in the eastern equatorial Pacific is higher in RetroMOPS^{o} than in MOPS^{oD} (Fig. 2), resulting in a nitrate deficit in this region. ~~The deficit is comparable to that of MOPS^{oS}, i.e. of a model simulation with default parameters for oxidant dependent processes (Fig. 3).~~ Likely, the instantaneous remineralisation of sinking material inherent in the direct flux parameterisation of RetroMOPS ~~, which~~ causes a tighter spatial coupling between production, sinking, remineralisation and upwelling (see also section 3.2.1). It has been suggested earlier that the production of slowly degradable organic matter above upwelling regions and/or oxygen minimum zones may help to decouple these processes, and avoid a runaway effect of nitrate loss (Landolfi et al., 2013; Dietze and Loeptien, 2013). The very low optimal value for surface DOP turnover ~~λ_{sDOP}~~ λ_s found in this study, and also in the study by Letscher et al. (2015) supports this finding.

~~The total misfit to observed dissolved tracer concentrations of RetroMOPS^{o} is only about 4% higher than that of MOPS^{oD}, i.e. RetroMOPS can perform almost as well as MOPS, with respect to annual mean phosphate, nitrate, and oxygen.~~ Simulated biogeochemical fluxes of RetroMOPS^{o} are generally lower than those of MOPS^{oD}, and their horizontal pattern is less pronounced (Fig. 2). This likely arises from the prescribed, constant phytoplankton concentration of RetroMOPS^{o}, which mutes biogeochemical dynamics in productive regions of the high latitudes and upwelling areas. Because RetroMOPS^{o} applies the same parameters as MOPS^{oD} for oxidant-dependent processes, its global fixed nitrogen loss and gain is comparable to that of the more complex model.

The total misfit to observed dissolved tracer concentrations of RetroMOPS^{o} is only about 4% higher than that of MOPS^{oD}, suggesting that even the simple RetroMOPS can perform almost as well as MOPS with respect to annual mean phosphate, nitrate, and oxygen. As for MOPS, optimisation of RetroMOPS against dissolved tracer concentrations ~~improves the~~ results in a good fit to global estimates of biogeochemical fluxes (Table 4), and indicates ~~,~~ that these tracers can provide means to calibrate biogeochemical model fluxes on a global scale, even - or especially - for a model as simple as RetroMOPS.

3.3 How much complexity is needed?

~~Optimisation of parameters for oxidant-dependent processesresults in~~ Current, state-of-the-art biogeochemical models address questions such as the future evolution of oxygen minimum zones, or uptake of anthropogenic carbon by the ocean (e.g. Cocco et al., 2013; C Compared to these models MOPS and RetroMOPS presented here are of a rather low structural complexity. RetroMOPS is quite similar to early models addressing these tasks, among them the pioneering work of Ernst-Maier Reimer (e.g. Bacastow and Maier-Reimer, 1 while MOPS resembles models of intermediate complexity such as HAMOCC (e.g. Six and Maier-Reimer, 1996; Maier-Reimer et al., 200 HadOCC (Palmer and Totterdell, 2001). However, very simple models such as RetroMOPS are still being used, e.g., for inverse methods (e.g. Kwon and Primeau, 2006, 2008) or to investigate specific processes, where their computational efficiency and structural simplicity facilitates model analysis (e.g. Parekh et al., 2005; Kwon et al., 2009; Primeau et al., 2013). In contrast

to these very simple model are models that simulate different plankton groups and size classes of detritus, e.g. PISCES (Aumont et al., 2015), MEDUSA (Yool et al., 2013), or PlankTOM (Le Quere et al., 2005).

Despite this large range of structural complexity, there have been only few studies which evaluate these models against a common data set, and with a common circulation. One example is the study by Kwiatkowski et al. (2014), who compared the output of six different global biogeochemical models, coupled to a common circulation model, and simulated over 118 years, against data sets of surface pCO_2, DIC, alkalinity, DIN, Chl *a*~~*slightly better fit to observed tracers, and in a much better fit to observed estimates of global fixed nitrogen turnover.*~~ and primary production. The models varied in complexity from seven to 57 compartments, and thus also in their computational demand by almost a factor of five. To assess model skill Kwiatkowski et al. (2014) ranked the models with respect to spatial correlation between, and variance of, model and observations. In general, the more complex models performed better with respect to simulated variance, but the simpler models better with respect to spatial correlation. Although no model was superior across all metrics, they concluded that "Results suggest that little evidence that higher biological complexity implies better model performance in reproducing observed global-scale bulk properties." (Kwiatkowski et al., 2014). This conclusion may be even more obvious when taking into account the ability of the different diagnostics to distinguish among the models: For example, spatial correlation of DIN in their study varied only between 0.79 to 0.94, and normalised standard deviations for DIN, alkalinity and DIC varied less than 10% around the average standard deviation. The small variation in some metrics renders model assessment by ranking difficult. Excluding these diagnostics from the model assessment may result even in an advantage for the simpler models (MEDUSA or HadOCC) with respect to spatial correlation and a reasonable performance with respect to standard deviation.

The lack of distinction between models and their ability to represent biogeochemical tracers is corroborated by the study by Galbraith et al. (2015), who evaluated three different biogeochemical ocean models within a common framework for the earth system. The models varied in complexity between one to 30 components. Following a spin-up over 100 years, Galbraith et al. (2015) analyse both a transient and preindustrial scenario with respect to the model's representation of macronutrients, oxygen, DIC, and export. All three models performed quite similar with respect to the observed tracer fields, as well as with the transient evolution of carbon uptake and oxygen concentrations. Therefore, in the presence of noise inherent in observations, and given the sparsity of biological data sets, so far it seems unresolved if more complexity is indeed beneficial - at least if the model is supposed to represent mostly biogeochemical processes, instead of biological interactions, and is compared against bulk biogeochemical properties.

4 Conclusions

Based on a global metric for biogeochemical tracers this study assessed the skill of two optimised global biogeochemical ocean models, as well as the metric's capability to constrain the often uncertain model parameters.

Similar to an earlier study (Kriest et al., 2017) that targeted as parameters relevant for biogeochemical processes at the sea surface, parameters for oxidant-dependent processes in the mesopelagial could only be determined with a wide range of uncertainty. The reason for this lack of resolution can be found in the small volume occupied by either surface, or oxygen

minimum zones (where oxidant-dependency is of relevance). Omission of the eastern equatorial Pacific from the misfit function ~~results in a slight upward bias of global fixed nitrogen fluxes. It does not , however,~~ increases uncertainty in parameter estimates, but does not fundamentally alter the outcome of optimisation.

In contrast, parameters relevant for large-scale, global distributions of oxygen, such as remineralisation length scale or stoichiometry could be determined with a high fidelity; these parameters were very similar in all experiments, and point towards a shorter remineralisation length scale of $b = 1.3$ to $b = 1.4$, as compared to the canonical $b = 0.858$ suggested by Martin et al. (1987).

Despite the uncertainty in estimates of some parameters, and very small differences between models in the residual misfit, optimisation of parameters for oxidant-dependent processes results in a much better fit to observed estimates of global fixed nitrogen turnover. The remaining mismatch to observations can partly be attributed to circulation. Model optimisations with different parameterisations of circulation and the equatorial intermediate current system (e.g., using TMs extracted from the UVic model; Kvale et al., 2017) will help to examine, if a different parameterisation alters the current requirement for very high nitrate threshold of denitrification, that currently helps to prevent nitrate from depletion.

Oxygen and nitrate add important additional constraints on the estimation of biogeochemical parameters. Of particular importance is that, in addition to the spatial information they provide, their flexible inventory introduces the bias as additional information for model calibration. The different time and space scales of processes ~~relavant~~ relevant for their inventory may help to constrain parameters that govern dissolved organic matter production and decay. The effect of these tracers on parameter estimates is of particular importance for models such as RetroMOPS and MOPS, that aim at conserving all oxidants. It may be weaker for models that continue remineralisation even under suboxic and/or low nitrate conditions, thereby implicitly assuming some "hidden" oxidants. In these models it could be useful to track and examine potential oxidant deficits for model evaluation.

~~The~~ DOP recycling rate affects surface DOP and phosphate concentrations conversely: either the model performs relatively well with respect to DOP. In this case phosphate concentrations are overestimated by the model. If the model performs well with respect to phosphate, it overestimates surface DOP. Observations of DOP as additional constraint on model parameters will help to find out if there is a model solution that fits all tracers ~~simulated in RetroMOPS~~equally well.

With respect to annual mean tracer concentrations the simple model RetroMOPS can perform almost as well as the more complex model MOPS~~. Although spatial~~ , the residual misfit being only 5% larger. Spatial patterns of fluxes in RetroMOPS are less pronounced, but global tracer concentrations, inventories and fluxes are comparable to that of MOPS, and in agreement with observed estimates.

Although it is obvious that low to intermediate complexity models such as the models presented here cannot represent the level of detail embedded in models with e.g., several plankton size classes, so far evaluation with respect to the bulk, biogeochemical observations does not seem to indicate any superiority of more complex models on a global scale. This of course may change if our scientific interest and model purpose is directed towards shorter time scales, or surface patterns, for which the misfit function applied provides little information. In this case more complex data sets, such as different plankton groups, or particle size distribution, may provide further insight about the level of model complexity required. If focusing on

large ~~scale dynamics~~scales, however, a simple model such as RetroMOPS or similarly simple models may suffice to represent and analyse much of the biogeochemical dynamics in the ocean.

Acknowledgements. I am very thankful for having met Ernst Maier-Reimer, who pioneered global biogeochemical ~~modeling~~modelling. In his thoughtful and kind way he taught me to view global ocean biogeochemistry before the background of long time and large space scales. ~~Thank you, Ernst!~~

This work is a contribution to the DFG-supported project SFB754 and to BMBF joint project PalMod (FKZ 01LP1512A). Parallel supercomputing resources have been provided by the North-German Supercomputing Alliance (HLRN). The author wishes to acknowledge use of the Ferret program of NOAA's Pacific Marine Environmental Laboratory for analysis and graphics in this paper. I thank three anonymous reviewers and Friederike Hoffmann for their constructive and helpful comments.

References

Anderson, T.: Plankton functional type modelling: running before we can walk?, J. Plankton Res., 27, 1073–1081, 2005.

Anderson, T. R.: Confronting complexity: reply to Le Quere and Flynn, Jour. Plankton Res., 28, 877–878, doi:10.1093/plankt/fbl016, 2006.

Aumont, O., Ethe, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geosci. Model Dev., 8, 2465–2513, 2015.

Bacastow, R. and Maier-Reimer, E.: Ocean-circulation model of the carbon cycle, Clim. Dynam., 4, 95–125 1990.

Bacastow, R. and Maier-Reimer, E.: Dissolved organic carbon in modeling oceanic new production, Global Biogeochem. Cy., 5, 71–85, 1991.

Breitbarth, E., Oschlies, A., and LaRoche, J.: Physiological constraints on the global distribution of *Trichodesmium* - effect of temperature on diazotrophy, Biogeosciences, 4, 53–61, 2007.

Buesseler, K., Lamborg, C., Boyd, P., Lam, P., Trull, T., Bidigare, R., Bishop, J., Casciotti, K., Dehairs, F., Elskens, M., Honda, M., Karl, D., Siegel, D., Silver, M., Steinberg, D., Valdes, J., Mooy, B. V., and Wilson, S.: Revisiting carbon flux through the ocean's twilight zone, Science, 316, 567–570, 2007.

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

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

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

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

Dietze, H. and Loeptien, U.: Revisiting "nutrient trapping" in global coupled biogeochemical ocean circulation models, Global Biogeochem. Cy., 27, 265—284, 2013.

Doney, S., Lindsay, K., Caldeira, K., Campin, J., Drange, H., Dutay, J., Follows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Madec, G., Maier-Reimer, E., Marshall, J., Matear, R., Monfray, P., Mouchet, A., Najjar, R., Orr, J., Plattner, G., Sarmiento, J., Schlitzer, R., Slater, R., Totterdell, I., Weirig, M., Yamanaka, Y., and Yool, A.: Evaluating global ocean carbon models: The importance of realistic physics, Glob. Biogeochem. Cyc., 18, doi:10.1029/2003GB002150, 2004.

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

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

Dutkiewicz, S., Follows, M., and Parekh, P.: Interactions of the iron and phosphorous cycles: A three-dimensional model study, Global Biogeochem. Cy., 19, f2, 2005.

Dutkiewicz, S., Follows, M., Heimbach, P., and Marshall, J.: Controls on ocean productivity and air-sea carbon flux: An adjoint model sensitivity study, Geophys. Res. Lett., 33, 2006.

Flynn, K. J.: Reply to Horizons Article 'Plankton functional type modelling: running before we can walk' Anderson (2005): II. Putting trophic functionality into plankton functional types, JOURNAL OF PLANKTON RESEARCH, 28, 873–875, doi:10.1093/plankt/fbl015, 2006.

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

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

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

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

Graven, H., Gruber, N., Key, R., Khatiwala, S., and Gireaud, X.: Changing controls on oceanic radiocarbon: New insights on shallow-to-deep ocean exchange and anthropogenic CO_2 uptake, J. Geophys. Res., 117, f0c, 2012.

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

Hansen, N. and Ostermeier, A.: Completely Derandomized Self-Adaptation in Evolution Strategies, Evolutionary Computation, 9, 159–195, 2001.

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

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

Hopkinson, C., Vallino, J., and Nolin, A.: Decomposition of dissolved organic matter from the continental margin, Deep-Sea Res. Pt. II, 49, 4461–4478, f2c, 2002.

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

Khatiwala, S.: A computational framework for simulation of biogeochemical tracers in the ocean, Global Biogeochem. Cy., 21, GB3001, 2007.

Khatiwala, S.: Fast spin up of ocean biogeochemical models using matrix-free Newton-Krylov, Ocean Modell., 23, 121–129, 2008.

Khatiwala, S., Visbeck, M., and Cane, M. A.: Accelerated simulation of passive tracers in ocean circulation models, Ocean Modell., 9, 51–69, 2005.

Kriest, I. and Oschlies, A.: Numerical effects on organic matter sedimentation and remineralization in biogeochemical ocean models, Ocean Modell., 39, 275–283, 2011.

Kriest, I. and Oschlies, A.: Swept under the carpet: The effect of organic matter burial in global biogeochemical ocean models, Biogeosciences Disc., 10, 10 859–10 911, 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, doi:10.5194/gmd-8-2929-2015, www.geosci-model-dev.net/8/2929/2015/, 2015.

Kriest, I., Khatiwala, S., and Oschlies, A.: Towards an assessment of simple global marine biogeochemical models of different complexity, Prog. Oceanogr., 86, 337–360, 2010.

Kriest, I., Oschlies, A., and Khatiwala, S.: Sensitivity analysis of simple global marine biogeochemical models, Global Biogeochem. Cy., 26, 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, doi:10.5194/gmd-10-127-2017, www.geosci-model-dev.net/10/127/2017/, 2017.

Kvale, K., Khatiwala, S., Dietze, H., Kriest, I., and Oschlies, A.: Evaluation of the Transport Matrix Method for simulation of ocean biogeochemical tracers, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2017-29, 2017.

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschoen, M., Enright, C., Halloran, P. R., Le Quere, C., de Mora, L., Racault, M. F., Sinha, B., Totterdell, I. J., and Cox, P. M.: iMarNet: an ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework, Biogeosciences, 11, 7291–7304, doi: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, 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, 2009.

Landolfi, A., Dietze, H., Koeve, W., and Oschlies, A.: Overlooked runaway feedback in the marine nitrogen cycle: the vicious cycle, Biogeosciences, 10, 1351–1363, 2013.

Le Quere, C.: Reply to Horizons Article 'Plankton functional type modelling: running before we can walk' Anderson (2005): I. Abrupt changes in marine ecosystems?, Jour. Plankton Res., 28, 871–872, doi:10.1093/plankt/fbl014, 2006.

Le Quere, C., Harrison, S., Prentice, I., Buitenhuis, E., Aumont, O., Bopp, L., Claustre, H., Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K., Legendre, L., Manizza, M., Platt, T., Rivkin, R., Sathyendranath, S., Uitz, J., Watson, A., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, GLOBAL CHANGE BIOLOGY, 11, 2016–2040, doi:10.1111/j.1365-2468.2005.01004.x, 2005.

Leles, S., Valentin, J., and Figueiredo, G.: Evaluation of the complexity and performance of marine planktonic trophic models, Annals of the Brazilian Academy of Sciences, 88, 1971–1991, doi:10.1590/0001-3765201620150588, 2016.

Letscher, R. and Moore, J. K.: Preferential remineralization of dissolved organic phosphorus and non-Redfield DOM dynamics in the global ocean: Impacts on marine productivity, nitrogen fixation, and carbon export, Glob. Biogeochem. Cyc., 29, 325–340, doi:doi:10.1002/2014GB004904, 2015.

Letscher, R., Moore, J. K., Teng, Y.-C., and Primeau, F.: Variable C : N : P stoichiometry of dissolved organic matter cycling in the Community Earth System Model, Biogeosciences, 12, 209–221, doi:10.5194/bg-12-209-2015, www.biogeosciences.net/12/209/2015, f2, 2015.

Li, X. and Primeau, F.: A fast Newton–Krylov solver for seasonally varying global ocean biogeochemistry models, Ocean Modell., 23, 13–20, 2008.

Löptien, U. and Dietze, H.: Constraining parameters in marine pelagic ecosystem models – is it actually feasible with typical observations of standing stocks?, Ocean Sci., 11, 573–590, 2015.

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

Maier-Reimer, E.: Geochemical cycles in an ocean general circulation model. Preindustrial tracer distributions, Global Biogeochem. Cy., 7, 645–677, 1993.

Maier-Reimer, E., Kriest, I., Segschneider, J., and Wetzel, P.: The HAMburg Ocean Carbon Cycle Model HAMOCC 5.1 - Technical Description Release 1.1, Reports on Earth System Science 14, Max-Planck-Institute for Meteorology, Hamburg, 2005.

Marchal, O., Stocker, T. F., and Joos, F.: A latitude-depth, circulation-biogeochemical ocean model for paleoclimate studies. Development and sensitivities, Tellus Series B-Chemical And Physical Meteorology, 50, 290–316, 1998.

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

Martin, J. H., Knauer, G. A., Karl, D. M., and Broenkow, W. W.: VERTEX: carbon cycling in the Northeast Pacific, Deep-Sea Res., 34, 267–285, 1987.

Matear, R. J. and Hirst, A. C.: Long-term changes in dissolved oxygen concentrations in the ocean caused by protracted global warming, Global Biogeochem. Cy., 17, 2003.

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

Moore, J. K. and Doney, S. C.: Iron availability limits the ocean nitrogen inventory stabilizing feedbacks between marine denitrification and nitrogen fixation, Global Biogeochem. Cy., 21, GB2001, 2007.

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, 2007.

Orr, J., Najjar, R., Sabine, C. L., and Joos, F.: Abiotic- HOWTO. Internal OCMIP Report, Tech. Rep. revision: 1.16, 25 pp., LSCE/CEA, Saclay, Gif-sur-Yvette, France, ocmip5.ipsl.jussieu.fr/OCMIP/phase2/simulations/Abiotic/HOWTO-Abiotic.html(lastaccess:28November2013),2000, 2000.

Palmer, J. and Totterdell, I.: Production and export in a global ocean ecosystem model, DEEP-SEA RESEARCH PART I-OCEANOGRAPHIC RESEARCH PAPERS, 48, 1169–1198, doi:10.1016/S0967-0637(00)00080-7, 2001.

Parekh, P., Follows, M. J., and Boyle, E. A.: Decoupling of iron and phosphate in the global ocean, Global Biogeochem. Cy., 19, 2005.

Paulmier, A., Kriest, I., and Oschlies, A.: Stoichiometries of remineralisation and denitrification in global biogeochemical ocean model, Biogeosciences, 6, 923–935, 2009.

Piwonski, J. and Slawig, T.: Metos3D: the Marine Ecosystem Toolkit for Optimization and Simulation in 3-D – Part 1: Simulation Package v0.3.2, Geosci. Model Dev., 9, 3729–3750, doi:10.5194/gmd-9-3729-2016, www.geosci-model-dev.net/9/3729/2016/, 2016.

Priess, M., Koziel, S., and Slawig, T.: Marine ecosystem model calibration with real data using enhanced surrogate-based optimization, Journal of Computational Science, 4, 423–437, 2013.

Primeau, F. and Deleersnijder, E.: On the time to tracer equilibrium in the global ocean, Ocean Sci., 5, 13–28, 2009.

Primeau, F., Holzer, M., and DeVries, T.: Southern Ocean nutrient trapping and the efficiency of the biological pump, J. Geophys. Res., 118, 2547–2564, f0c, 2013.

Raimbault, P., Garcia, N., and Cerutti, F.: Distribution of inorganic and organic nutrients in the South Pacific Ocean - evidence for long-term accumulation of organic matter in nitrogen-depleted waters, Biogeosciences, 5, 281–298, 2008.

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B., Slawig, T., and Oschlies, A.: Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling, Biogeosciences, 14, 1647–1701, doi:10.5194/bg-14-1647-2017, 2017.

Schmoker, C., Hernandez-Leon, S., and Calbet, A.: Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions, Jour. Plank. Res., 35, 691–706, 2013.

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

Shimoda, Y. and Arhonditsis, G.: Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge, Ecological Modelling, 320, 29–43, doi:10.1016/j.ecolmodel.2015.08.029, 2016.

Siberlin, C. and Wunsch, C.: Oceanic tracer and proxy time scales revisited, Clim. Past, 7, 27–39, 2011.

Six, K. D. and Maier-Reimer, E.: Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model, Global Biogeochem. Cy., 10, 559–583, 1996.

Torres-Valdes, S., Roussenov, V., Sanders, R., Reynolds, S., Pan, X., Mather, R., Landolfi, A., Wolff, G., Achterberg, E., and Williams, R.: Distribution of dissolved organic nutrients and their effect on export production over the Atlantic Ocean Distribution of dissolved organic nutrients and their effect on export production over the Atlantic Ocean, Glob. Biogeochem. Cyc., 23, doi:10.1029/2008GB003389, 2009.

Van Mooy, B., Keil, R., and Devol, A.: Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrificiation, Geochim. Cosmochim. Ac., 66, 457–465, 2002.

Wallmann, K.: Phosphorus imbalance in the global ocean?, Global Biogeochem. Cy., 24, 2010.

Ward, B., Friedrichs, M. A. M., Anderson, T., and Oschlies, A.: Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, Jour. Mar. Systems, 81, 34–43, 2010.

Wunsch, C. and Heimbach, P.: How long to oceanic tracer and proxy equilibrium?, Quaternary Science Reviews, 27, 637–651, 2008.

Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-1.0: a new intermediate complexity plankton ecosystem model for the global domain, Geosci. Model Dev., 4, 381—417, 2011.

Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies, Geosci. Model Dev., 6, 1767–1811, 2013.

Yoshimura, T., Nishioka, J., Saito, H., Takeda, S., Tsuda, A., and Wells, M. L.: Distributions of particulate and dissolved organic and inorganic phosphorus in North Pacific surface waters, Mar. Chem., 103, 112–121, 2007.

Table 1. Experimental setup of optimisation. Parameters that stay fixed are highlighted. For parameters subject to ~~be optimised~~ optimisation we indicate the assigned, a priori lower and upper parameter boundary (parameter range, R^{A}_{Θ}) for optimisation in square brackets. "-": not applicable for this model.

Experiment	MOPS^{r}	MOPS^{oS}	MOPS^{oD}	RetroMOPS^{r}	RetroMOPS^{o}	unit
σ_{DOP} σ	-	-	-	**0.67**	[0.4 - 0.8]	
λ_{sDOP} λ_{s}	-	-	-	**0**	[0.0 - 3.6]	y^{-1}
λ_{DOP} λ	**0.17**	**0.17**	**0.17**	**0.36**	[0.036 - 3.6]	y^{-1}
I_{c}	**24**	[4 - 48]	**9.65**	**9.65**	**9.65**	$\mathrm{W\ m^{-2}}$
K_{PHY}	**0.03125**	[0.001 - 0.5]	**0.5**	**0.5**	**0.5**	$\mathrm{mmol\ P\ m^{-3}}$
μ_{ZOO}	**2**	[1 - 3]	**1.89**	-	-	$\mathrm{d^{-1}}$
κ_{ZOO}	**3.2**	[1.6 - 4.8]	**4.55**	-	-	$(\mathrm{d\ mmol\ P\ m^{-3}})^{-1}$
b^{*}	**0.858**	[0.4 - 1.8]	[0.4 - 1.8]	**1.0725**	[0.4 - 1.8]	
$R_{-\mathrm{O2:P}}$	**170**	[150 - 200]	[150 - 200]	**171.7**	**171.7**	mmol O_2:mmol P
μ_{NFix}	**2**	**2**	[1 - 3]	**1.19**	**1.19**	$\mathrm{nmol\ N\ d^{-1}}$
DIN_{min}	**4**	**4**	[1 - 16]	~~**15.79**~~**15.80**	**15.80**	$\mathrm{mmol\ N\ m^{-3}}$
K_{O2}	**2**	**2**	[1 - 16]	**1.0**	**1.0**	$\mathrm{mmol\ O_2\ m^{-3}}$
K_{DIN}	**8**	**8**	[2 - 32]	**31.97**	**31.97**	$\mathrm{mmol\ N\ m^{-3}}$

* Note that from b (the optimised parameter) in MOPS we calculate the rate of vertical increase in sinking speed a of $w = a\,z$, via $a = r/b$. For r we assume nominal detrital remineralisation of $r = 0.05\ \mathrm{d}^{-1}$. The resulting values for a are: 0.058275 ($b = 0.858$), 0.0278 (lower boundary) and 0.125 (upper boundary).

Table 2. Results (misfit J) of sensitivity experiments with model RetroMOPS, regarding parameters ~~λ_{sDOP}~~ λ_{s} and ~~λ_{DOP}~~ λ for DOP decay rate. The misfit of the reference scenario RetroMOPS^{r} is indicated in bold.

	~~$\lambda_{\mathrm{sDOP}} = 0$~~ $\lambda_{\mathrm{s}} = 0$	~~$\lambda_{\mathrm{sDOP}} = 0.18$~~ $\lambda_{\mathrm{s}} = 0.18$	~~$\lambda_{\mathrm{sDOP}} = 0.36$~~ $\lambda_{\mathrm{s}} = 0.36$
~~$\lambda_{\mathrm{DOP}} = 0.18$~~ $\lambda = 0.18$	0.502	0.480	0.480
~~$\lambda_{\mathrm{DOP}} = 0.36$~~ $\lambda = 0.36$	**0.466**	0.476	0.493
~~$\lambda_{\mathrm{DOP}} = 0.72$~~ $\lambda = 0.72$	0.503	0.522	0.539

Table 3. Optimisation results: minimum misfit J^*, optimum parameters and their uncertainties. To determine parameter uncertainty, we selected a group Ω of the 1‰ best individuals, i.e. individuals defined by a misfit $J_i : J_i/J^* - 1 \leq \Delta_J$, with $\Delta_J = 0.001$. The number of these individuals $N(\Omega)$ is also denoted as fraction $n(\Omega)$ of all individuals of the optimisation $\lambda \times N$, where N is the number of generations, and $\lambda = 10$ the population size. For each parameter Θ the first column gives the optimal parameter Θ^* (i.e., the average parameter of the last generation). The second and third column present the parameter range of all individuals of Ω, expressed as absolute value ($R_\Theta(\Omega)$), and normalised by the a priori range of parameters (R^{A}_Θ; see Table 1): $r_\Theta(\Omega) = R_\Theta(\Omega)/R^{\mathrm{A}}_\Theta$ value.

Experiment:	MOPS^{oS}			MOPS^{oD}			MOPS$^{\mathrm{oD}}_*$			RetroMOPS^{o}		
Parameter	Θ^*	$R_\Theta(\Omega)$	$r_\Theta(\Omega)$	Θ^*	$R_\Theta(\Omega)$	$r_\Theta(\Omega)$	Θ^*	$R_\Theta(\Omega)$	$r_\Theta(\Omega)$	Θ^*	$R_\Theta(\Omega)$	$r_\Theta(\Omega)$
$\cancel{\sigma_{\mathrm{DOP}}}\ \sigma$	-	-	-	-	-	-	-	-	-	0.73	[0.7-0.7]	6
$\cancel{\lambda_{\mathrm{sDOP}}}\ \lambda_{\mathrm{s}}$	-	-	-	-	-	-	-	-	-	0.02	[-0.1-0.2]	8
$\cancel{\lambda_{\mathrm{DOP}}}\ \lambda$	-	-	-	-	-	-	-	-	-	0.47	[0.4-0.5]	4
I_{c}	9.66	[8.9-10.3]	3									
K_{PHY}	0.50	[0.4-0.5]	28									
μ_{ZOO}	1.89	[1.6-2.0]	22							-	-	-
κ_{ZOO}	4.57	[3.0-4.7]	53							-	-	-
$b^{\S}$	1.34	[1.3-1.4]	4	1.39	[1.4-1.4]	3	1.41	[1.4-1.4]	2	0.98	[1.0-1.0]	2
$R_{-\mathrm{O2:P}}$	167.0	[165-170]	9	171.7	[170-173]	6	174.9	[174-176]	5			
μ_{NFix}				1.19	[1.1-1.4]	13	1.47	[1.4-1.6]	10			
DIN_{min}				15.80	[13-16]	20	12.96	[12-16]	25			
K_{O2}				1.00	[0.3-1.8]	10	1.00	[0.5-1.4]	6			
K_{DIN}				31.97	[30-34]	12	31.97	[22-33]	35			
J^*		0.450			0.439			0.427			0.458	
$\lambda \times N$		1820			1190			2000			660	
$N(\Omega)$		718			514			1285			262	
$n(\Omega)$		39			43			64			40	

Table 4. Global annual fluxes of primary production (P), grazing (GRAZ), fixed nitrogen loss through pelagic denitrification (NLOSS), export production (F120, flux through 120 m), flux through 2250 m (F2250), and benthic burial (BUR), in Pg N y^{-1}, for the reference experiment of $MOPS^{r}$, $MOPS^{oS}$, $MOPS^{oD}$, $MOPS_{*}^{oD}$ and RetroMOPS, for which we show the fluxes of the (best) reference experiment, $RetroMOPS^{r}$, the range of all sensitivity experiments, and the optimised run, $RetroMOPS^{o}$. Also shown are some globally derived, observed estimates. Conversion between different elements was carried out via N:P=16, and C:P=122.

Experiment	P	GRAZ	NLOSS	F120	F2250	BUR
$MOPS^{r}$	5.44	3.52	0.098	0.918	0.107	0.051
$MOPS^{oS}$	7.52	4.74	0.117	1.102	0.056	0.018
$MOPS^{oD}$	7.70	4.97	0.068	1.080	0.055	0.022
$MOPS_{*}^{oD}$	7.80	5.06	0.083	1.081	0.053	0.021
$RetroMOPS^{r}$	5.56	-	0.078	1.194	0.043	0.010
RetroMOPS (range)	4.88-6.21	-	0.076-0.084	1.076-1.286	0.039-0.047	0.008-0.014
$RetroMOPS^{o}$	6.31	-	0.071	1.12	0.052	0.009
Observed[§]	7.68-8.09	4.79-5.71	0.05-0.08	0.29-1.53	0.03-0.07	0.02

[§] Observed fluxes are from Carr et al. (2006, primary production), Honjo et al. (2008, particle flux), Lutz et al. (2007, particle flux), Dunne et al. (2007, particle flux), Schmoker et al. (2013, primary production, zooplankton grazing excluding/including mesozooplankton grazing), Wallmann (2010, burial; without shelf and slope region), and Kriest and Oschlies (2015, fixed nitrogen loss).

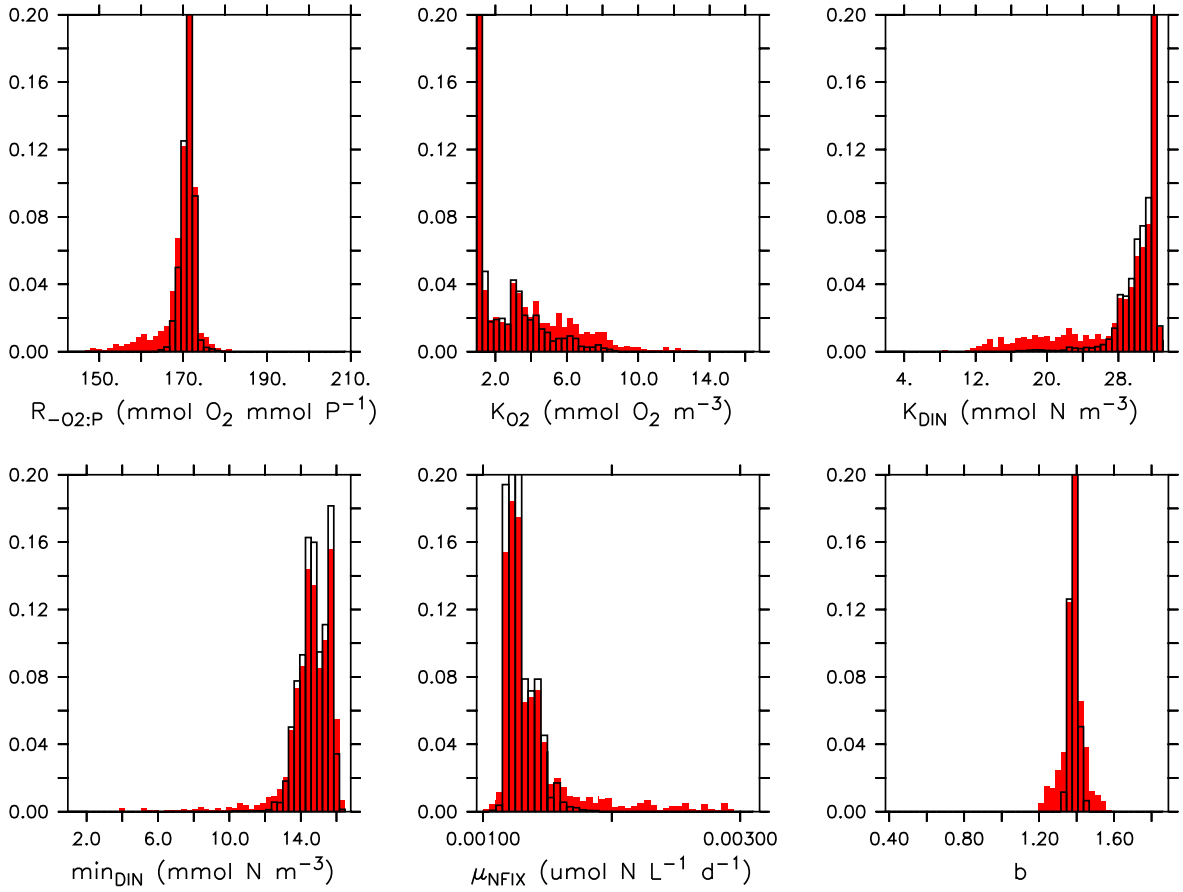

Figure 1. Parameter distribution of model simulations obtained during the optimisation of MOPS^{oD}, whose misfit do not exceed a threshold limit of $\Delta J = 1.1\,J^*$ (10%, red bars) or $\Delta J = 1.01\,J^*$ (1%, open bars) of the minimum misfit J^*. For the projection parameters of all model simulations in the optimisation trajectory were grouped into 50 classes.

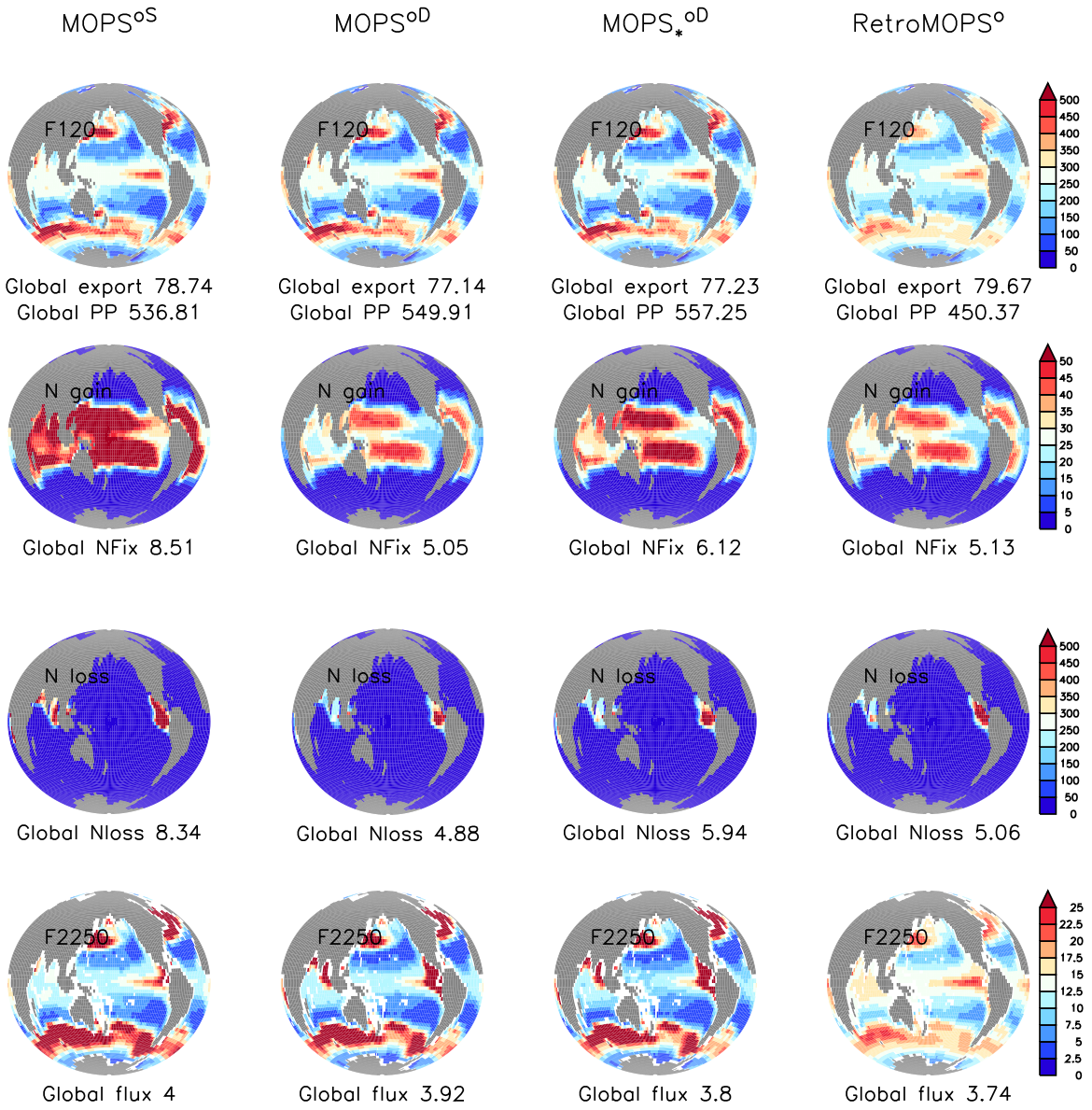

Figure 2. Biogeochemical fluxes of MOPS^{oS}, MOPS^{oD}, MOPS$_*^{oD}$ and RetroMOPS^o. Top: Export production (here: sedimentation at 120 m). Second row from top: nitrogen fixation. Third row from top: fixed nitrogen loss through pelagic denitrification. Bottom: sedimentation at 2250 m. All fluxes in mmol N m^{-2} y^{-1}. Each subpanel also gives the global flux in Tmol N y^{-1}.

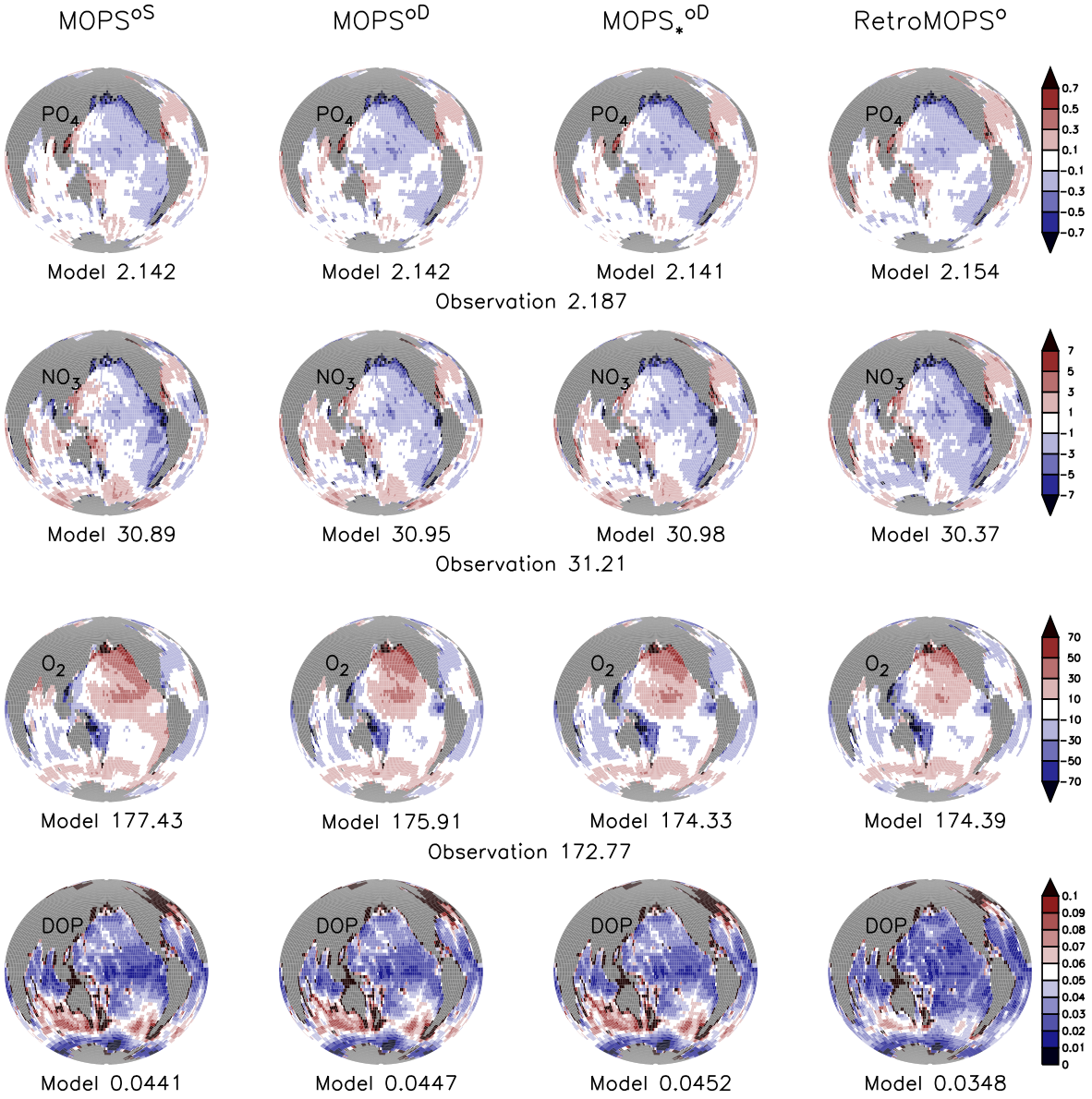

Figure 3. Vertically averaged tracers of MOPS^{oS}, MOPS^{oD}, MOPS$_*^{oD}$ and RetroMOPS^o. Top: phosphate. Second row from top: nitrate. Third row from top: oxygen. Bottom: DOP. Phosphate (mmol P m^{-3}), nitrate (mmol N m^{-3}) and oxygen (mmol O_2 m^{-3}) are expressed as deviation from observations (Garcia et al., 2006a, b), DOP is given in absolute concentrations (mmol P m^{-3}). Each subpanel also gives the global average tracer concentration in mmol m^{-3}.

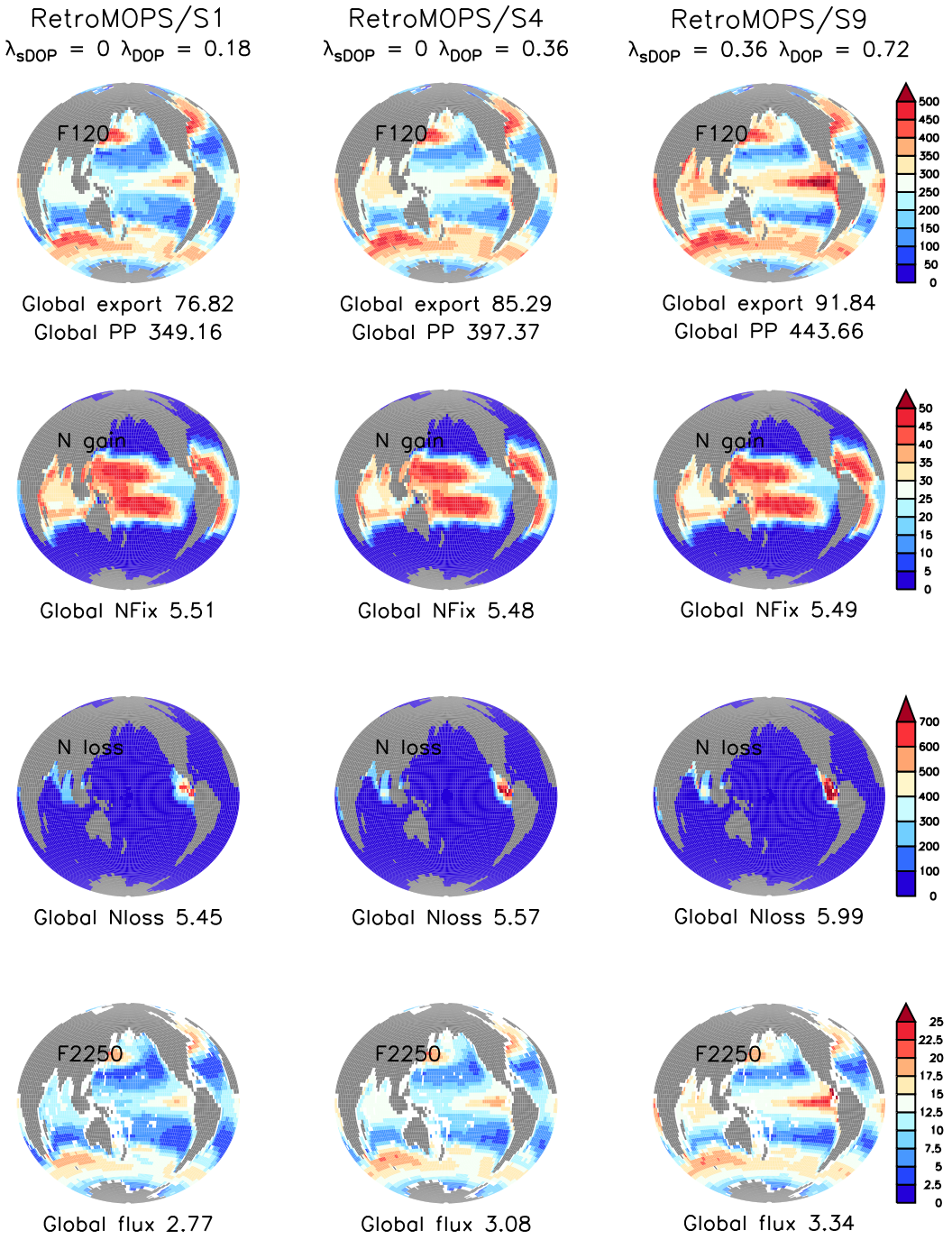

Figure 4. As Fig. 2, but for three sensitivity experiments with model RetroMOPS.

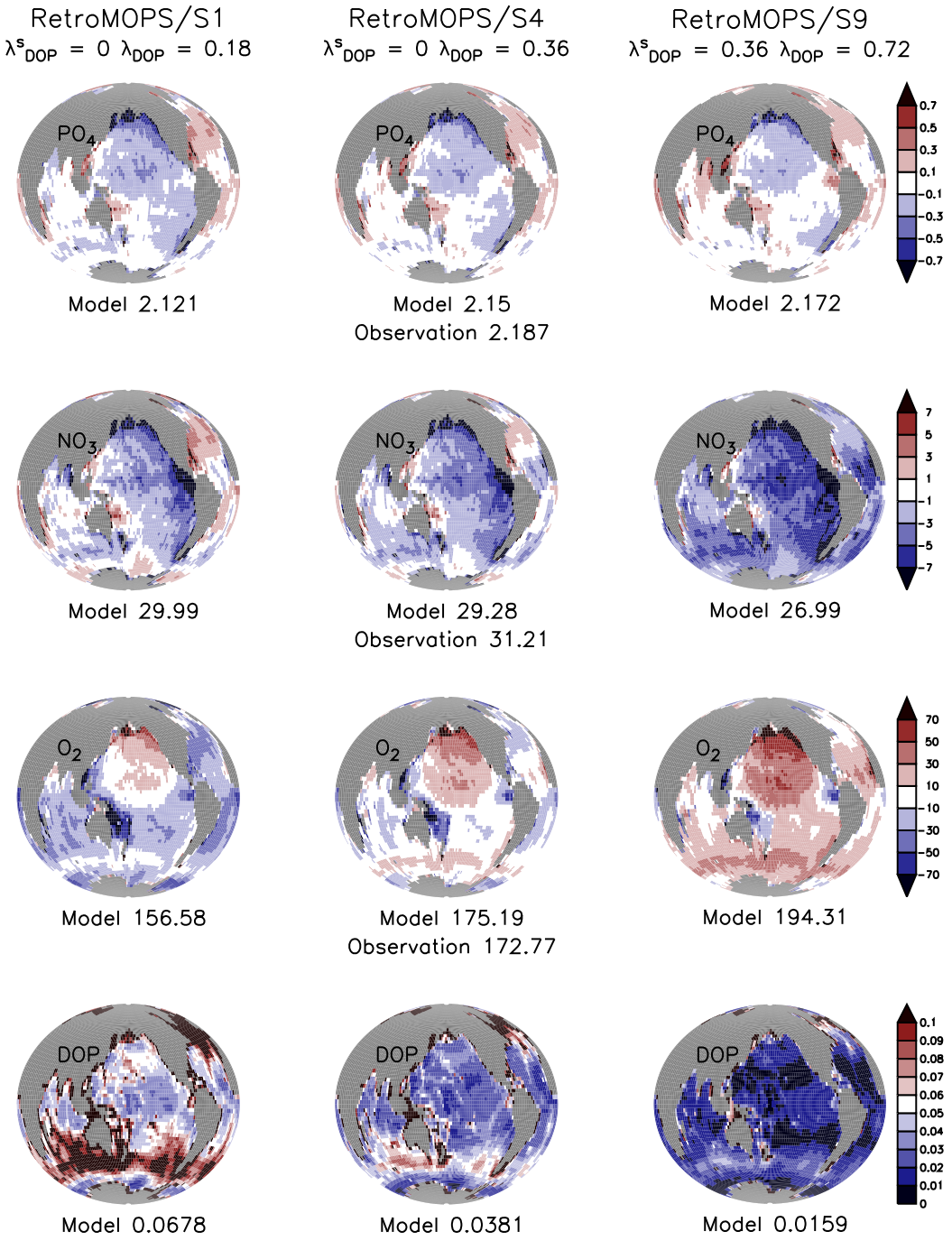

Figure 5. As Fig. 3, but for three sensitivity experiments with model RetroMOPS.

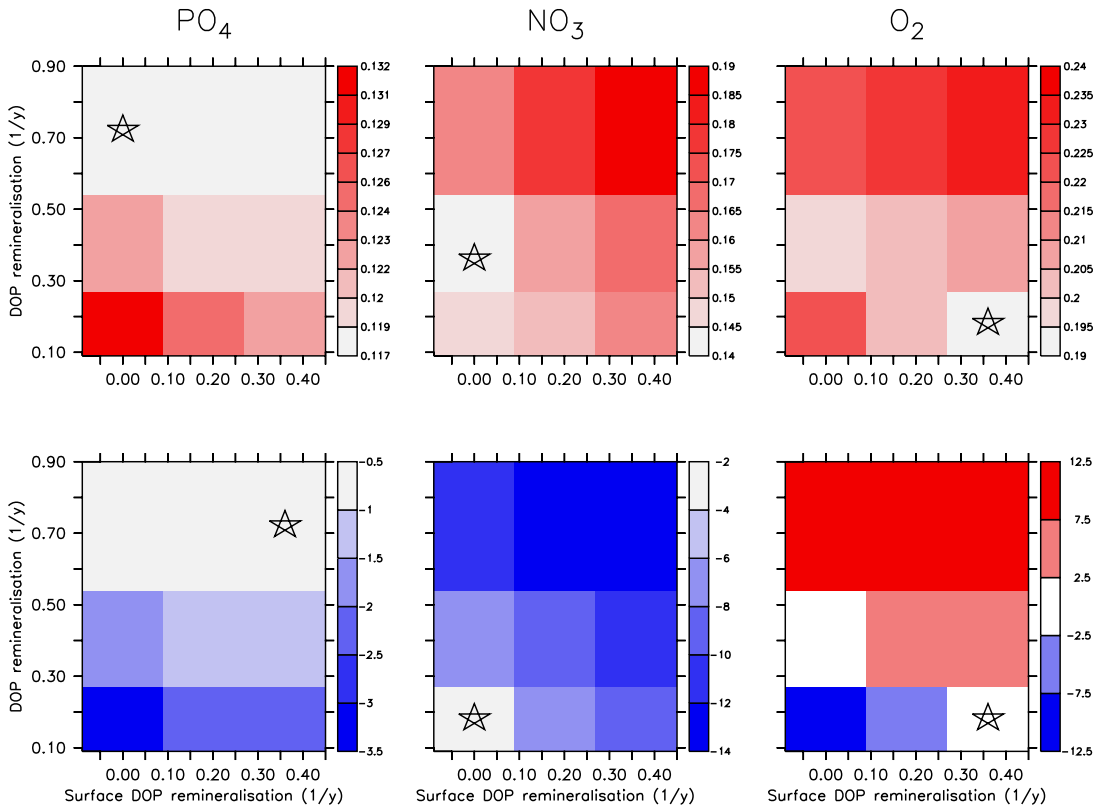

Figure 6. Components of the misfit function ($J(j)$ of Eqn. 11; upper panels) and model bias (lower panels), projected onto ~~λ_{sDOP}~~ λ_{s} and ~~λ_{DOP}~~ λ. Bias is expressed as $(\overline{m_j}/\overline{o_j} - 1) \times 100$, where $\overline{m_j}$ is the global average model tracer, and $\overline{o_j}$ the average observed tracer, for the three tracers phosphate ($j = 1$; left panels), nitrate ($j = 2$; mid panels) and oxygen ($j = 3$; right panels). An open star indicates the respective lowest misfit or bias.

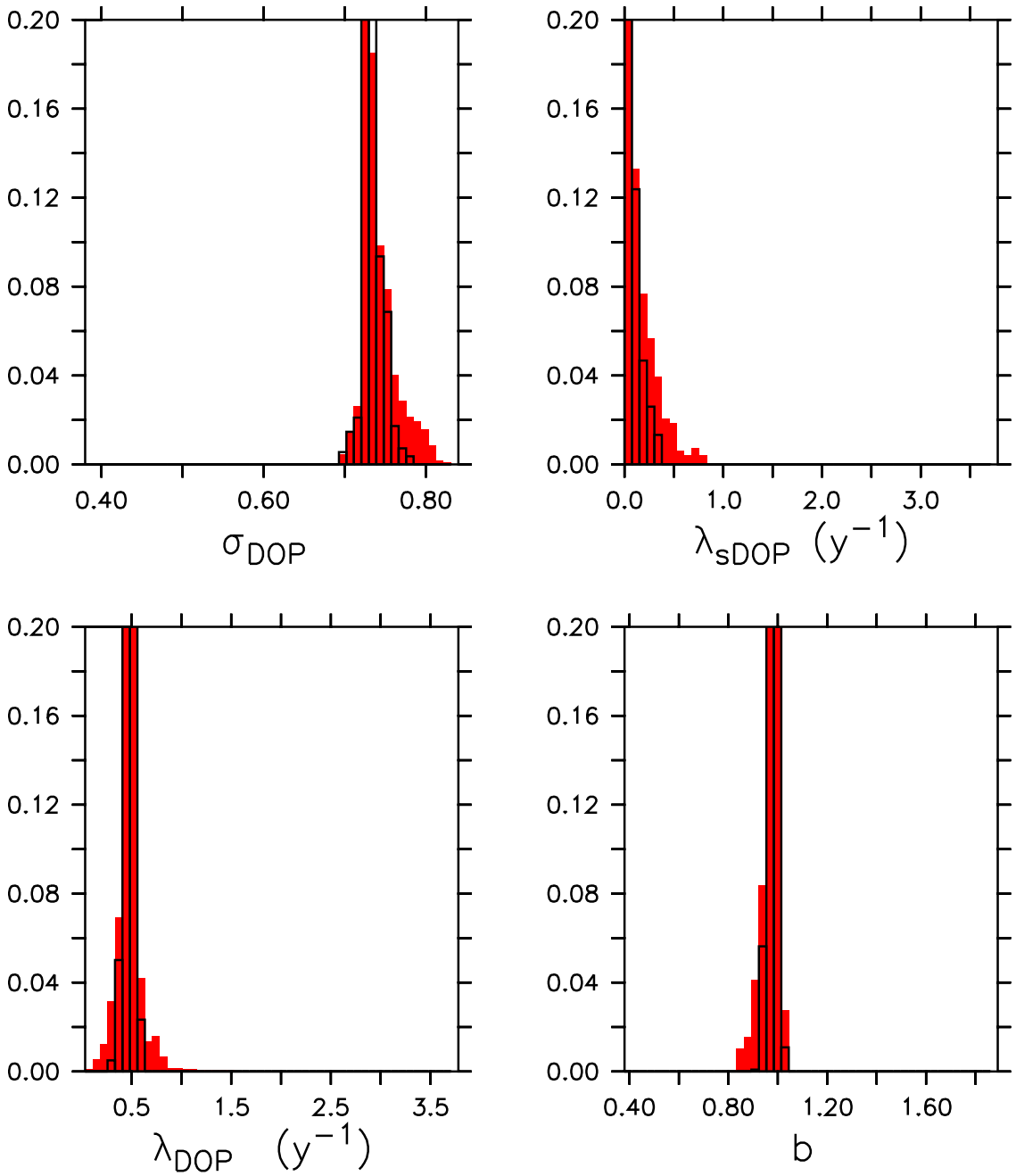

Figure 7. As Fig. 1, but for optimisation RetroMOPS°.

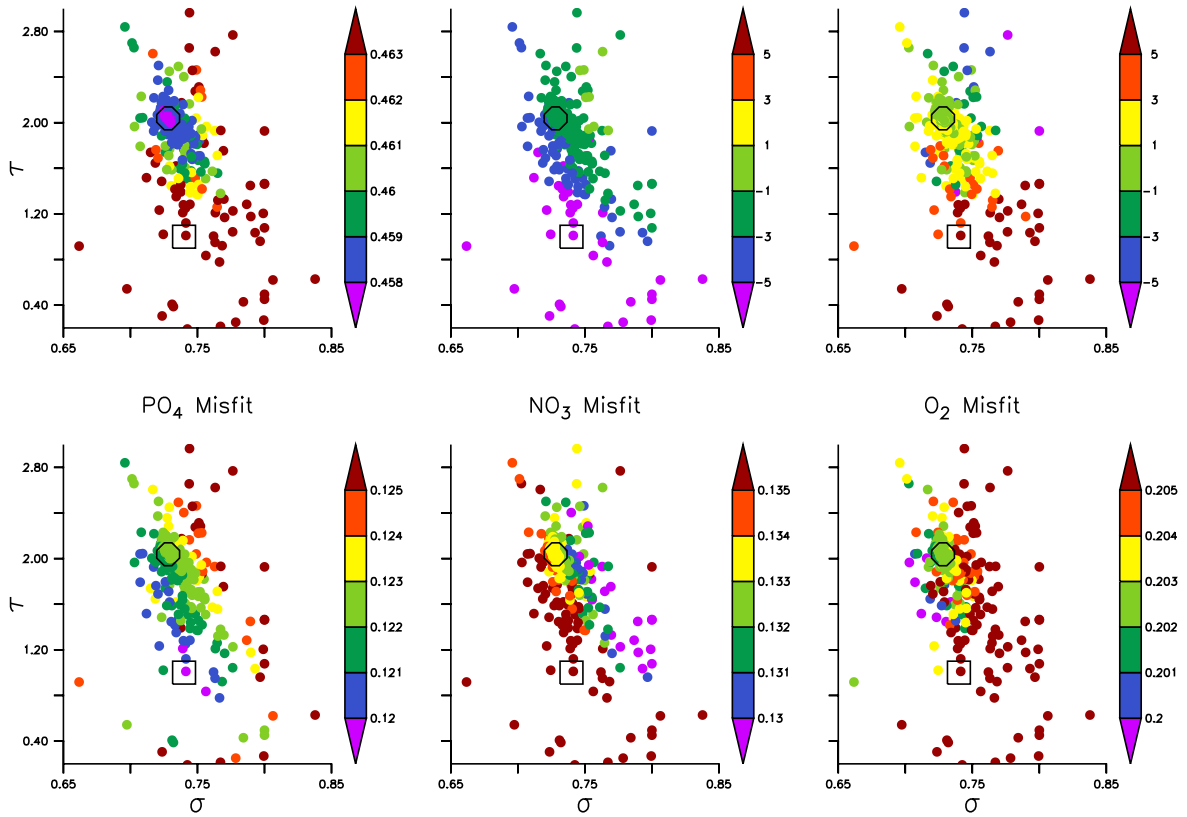

Figure 8. Model misfit and relative bias b_j of RetroMOPS°, plotted for parameter combinations of ~~σ_{DOP}~~ σ and DOP decay timescale τ, where ~~$\tau = 1/(\lambda_{\mathrm{DOP}} + \lambda_{\mathrm{sDOP}})$~~ $\tau = 1/(\lambda + \lambda_{\mathrm{s}})$. Relative bias is evaluated by $b_j = (\overline{m_j}/\overline{o_j} - 1) \times 100$, where $\overline{m_j}$ denotes the global mean model concentration of tracer j, and $\overline{o_j}$ the observed mean. Model misfit is shown as total misfit (J of Eqn. 11; upper left), and ~~seperated~~ separated into it components, normalised by $\overline{o_j}$ ($J(j)$ of Eqn. 11; lower panels). The analysis is restricted to all individuals i whose b differs less than 5% from optimal b^*, i.e. $|b_i/b^* - 1| < 0.05$. For better visibility some model solutions (≈ 10), that are outside the range $0.65 \leq \sigma \leq 0.85$ and $0.2 \leq \tau \leq 3$ have been ~~ommitted~~ omitted from the plot. Open squares denote optimal estimates by Kwon and Primeau (2006, total phosphate constraint), open circles the optimal parameter from this study.

Supplement:

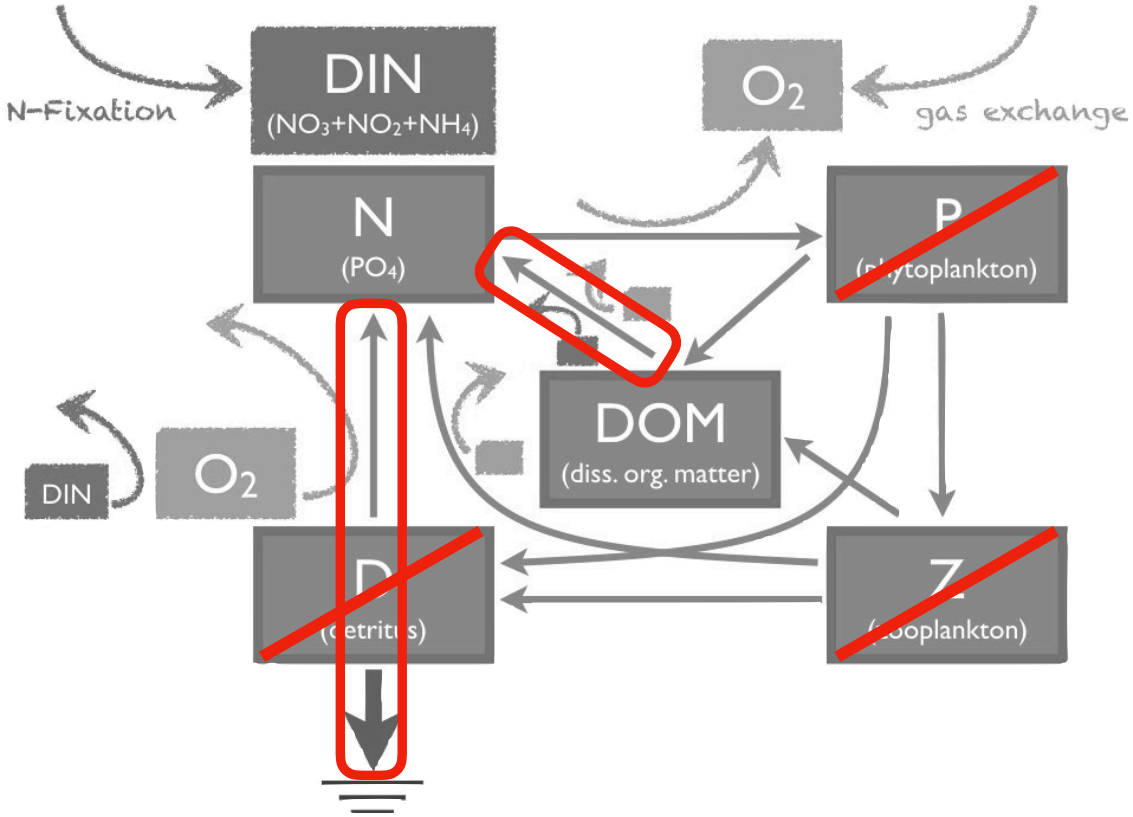

Figure S1. Diagram illustrating the changes for the downscaling of model MOPS (gray) to RetroMOPS. Omitted compartments are indicated by red diagonal bars. Structural changes for fluxes between the compartments are indicated by red borders.

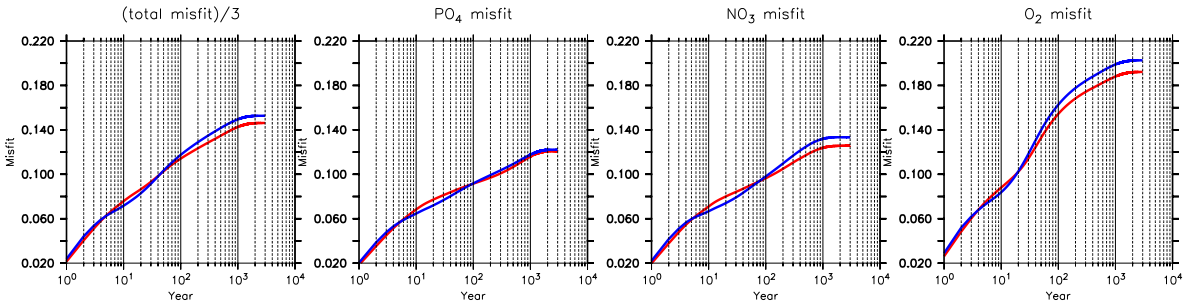

Figure S2. Model misfit ~~, plotted for each pair~~ of ~~parameter combinations~~ the best (final) candidate of MOPS^{oD} ~~. Color indicates misfit~~ (~~see the color bars on the right~~red) ~~. A circle indicates~~ and RetroMOPS^{o} (blue), plotted over the ~~parameter~~ entire spin-up of ~~one individual of the last generation~~3000 years (as log scale). ~~For better visibility the parameter range~~ Left panel: total misfit J, divided by three. Left to ~~its boundaries~~ right panels: components (~~see Table 1~~phosphate, nitrate, oxygen) of the misfit function. Note that, depending on simulation time, a different model type may exhibit the lowest misfit.

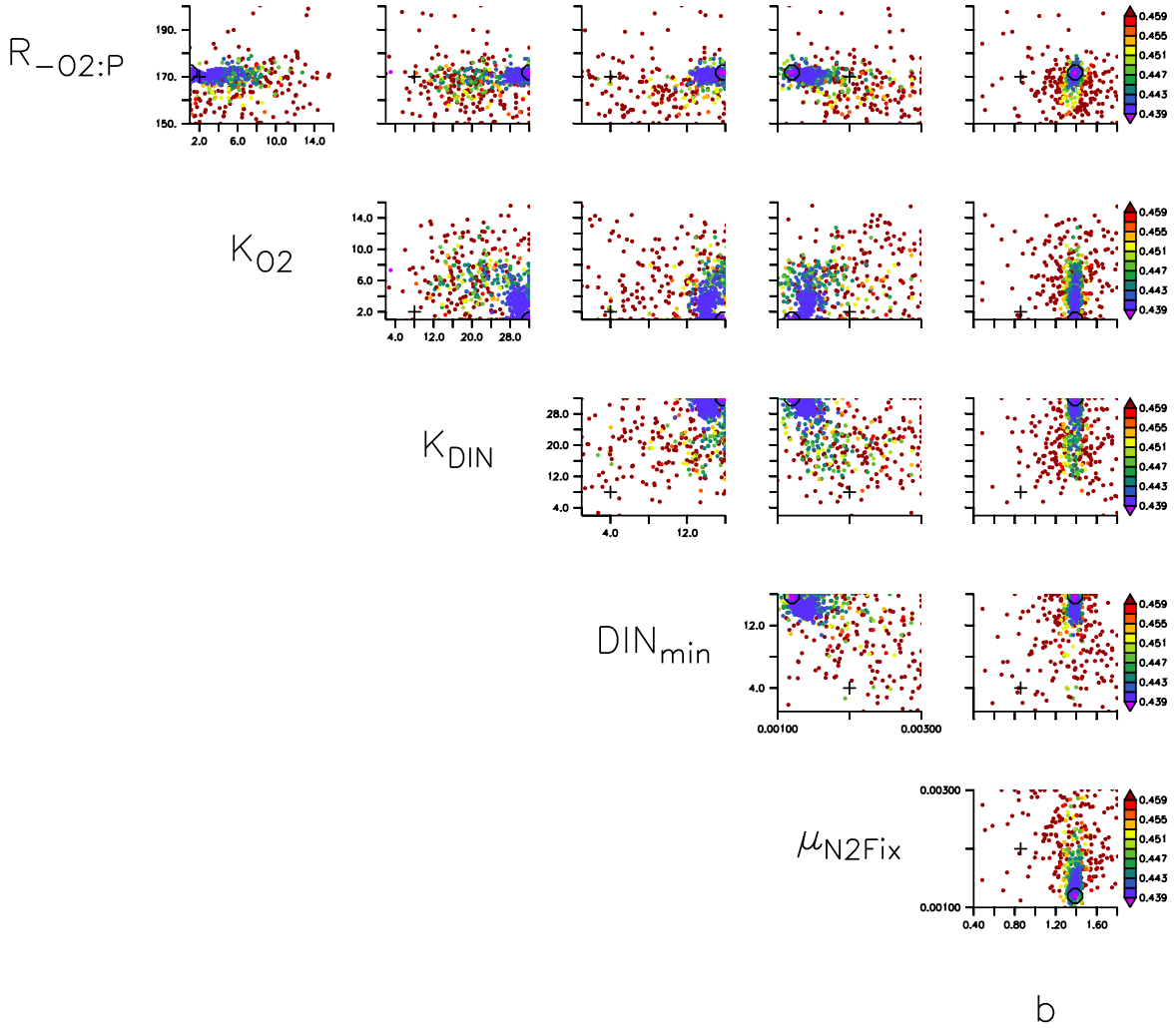

Figure S3. Model misfit, plotted for each pair of parameter combinations of MOPS^{oD}. Colour indicates misfit (see the colour bars on the right). A circle indicates the parameter of one individual of the last generation. For better visibility the parameter range to its boundaries (see Table 1).

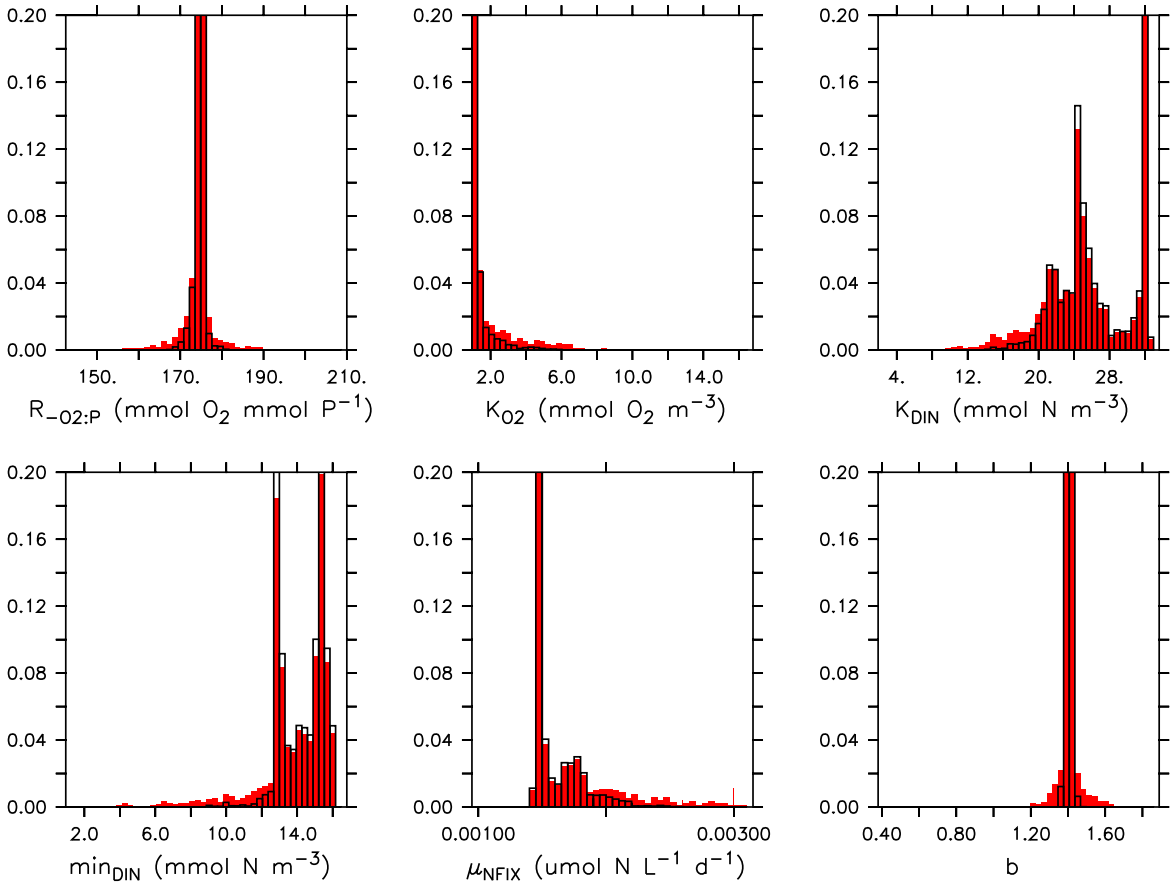

Figure S4. Parameter distribution of model simulations obtained during the optimisation of $\mathrm{MOPS}_*^{\mathrm{oD}}$, whose misfit do not exceed a threshold limit of $\Delta J = 1.1\,J^*$ (10%, red bars) or $\Delta J = 1.01\,J^*$ (1%, open bars) of the minimum misfit J^*. For the projection parameters of all model simulations in the optimisation trajectory were grouped into 50 classes.

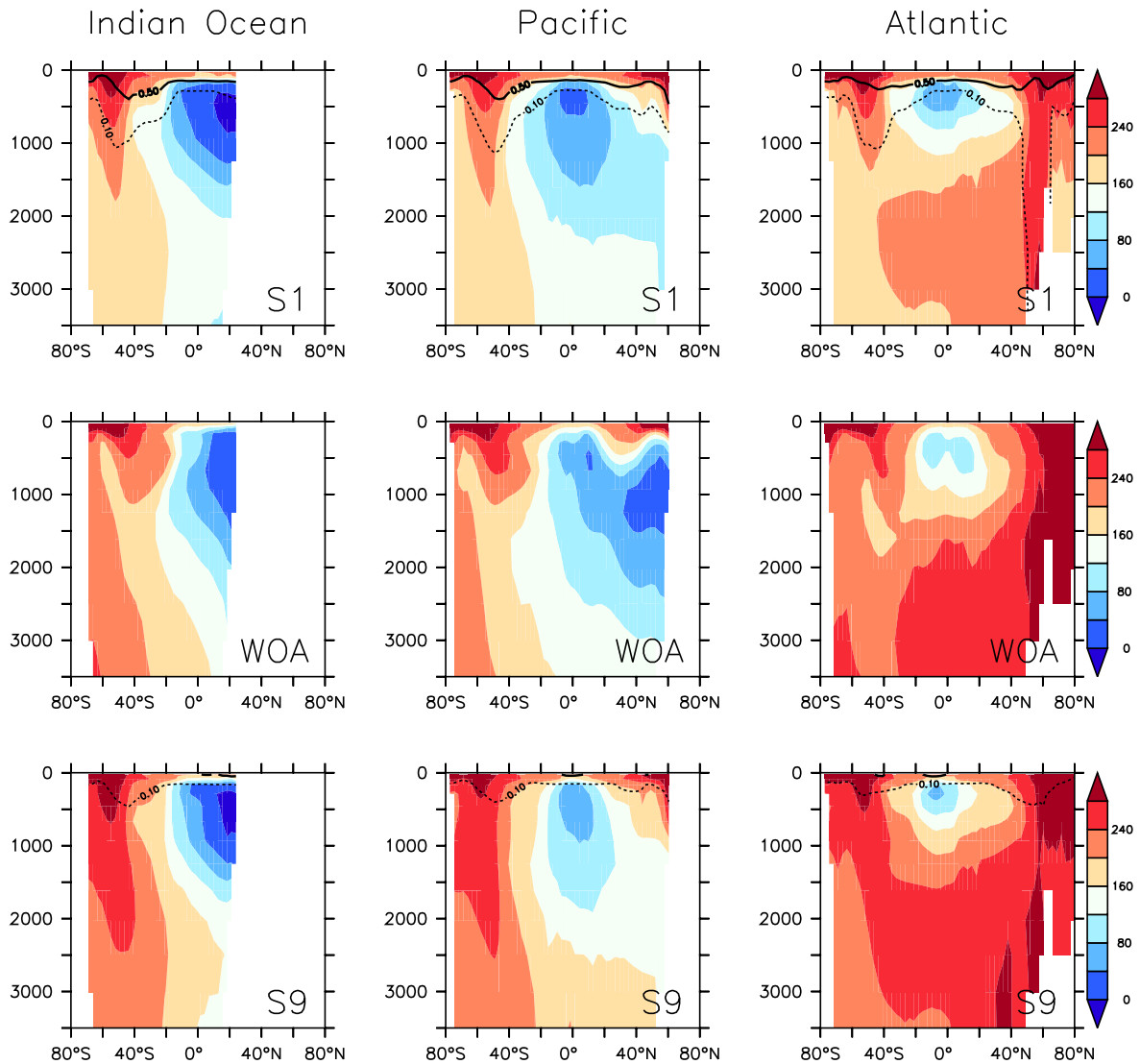

Figure S5. Zonal mean oxygen for three different basins of two sensitivity experiments with RetroMOPS and observations. Upper panels: RetroMOPS with low DOP recycling (~~$\lambda_{sDOP} = 0$~~$\lambda_s = 0$, ~~$\lambda_{DOP} = 0.18$~~$\lambda = 0.18$). Lower panels: RetroMOPS with high DOP recycling (~~$\lambda_{sDOP} = 0.36$~~$\lambda_s = 0.36$, ~~$\lambda_{DOP} = 0.72$~~$\lambda = 0.72$). Mid panels: observations. Contour lines show simulated zonal average DOP (dashed: 0.1 mmol P m^{-3}. thick: 0.5 mmol P m^{-3}).

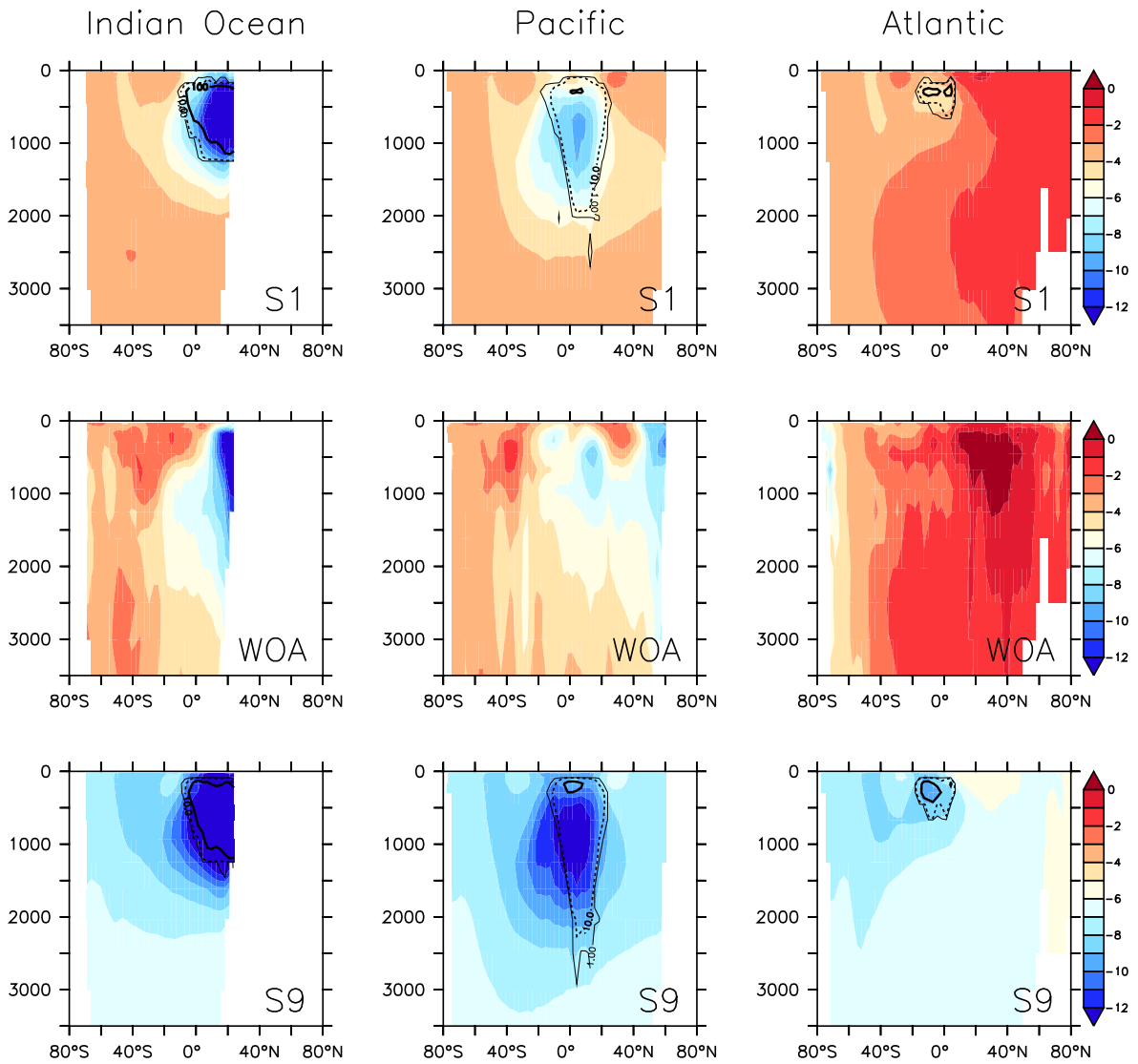

Figure S6. Zonal mean N* (NO_3-$PO_4 \times 16$ for three different basins of two sensitivity experiments with RetroMOPS and observations. Upper panels: RetroMOPS with low DOP recycling (~~$\lambda_{sDOP} = 0$~~$\lambda_s = 0$, ~~$\lambda_{DOP} = 0.18$~~$\lambda = 0.18$). Lower panels: RetroMOPS with high DOP recycling (~~$\lambda_{sDOP} = 0.36$~~$\lambda_s = 0.36$, ~~$\lambda_{DOP} = 0.72$~~$\lambda = 0.72$). Mid panels: observations. Contour lines show simulated zonal average annual fixed nitrogen loss through denitrification (thin: 1 μmol N m^{-3}y^{-1}, dashed: 10 μmol N m^{-3}y^{-1} thick: 100 μmol N m^{-3}y^{-1}).

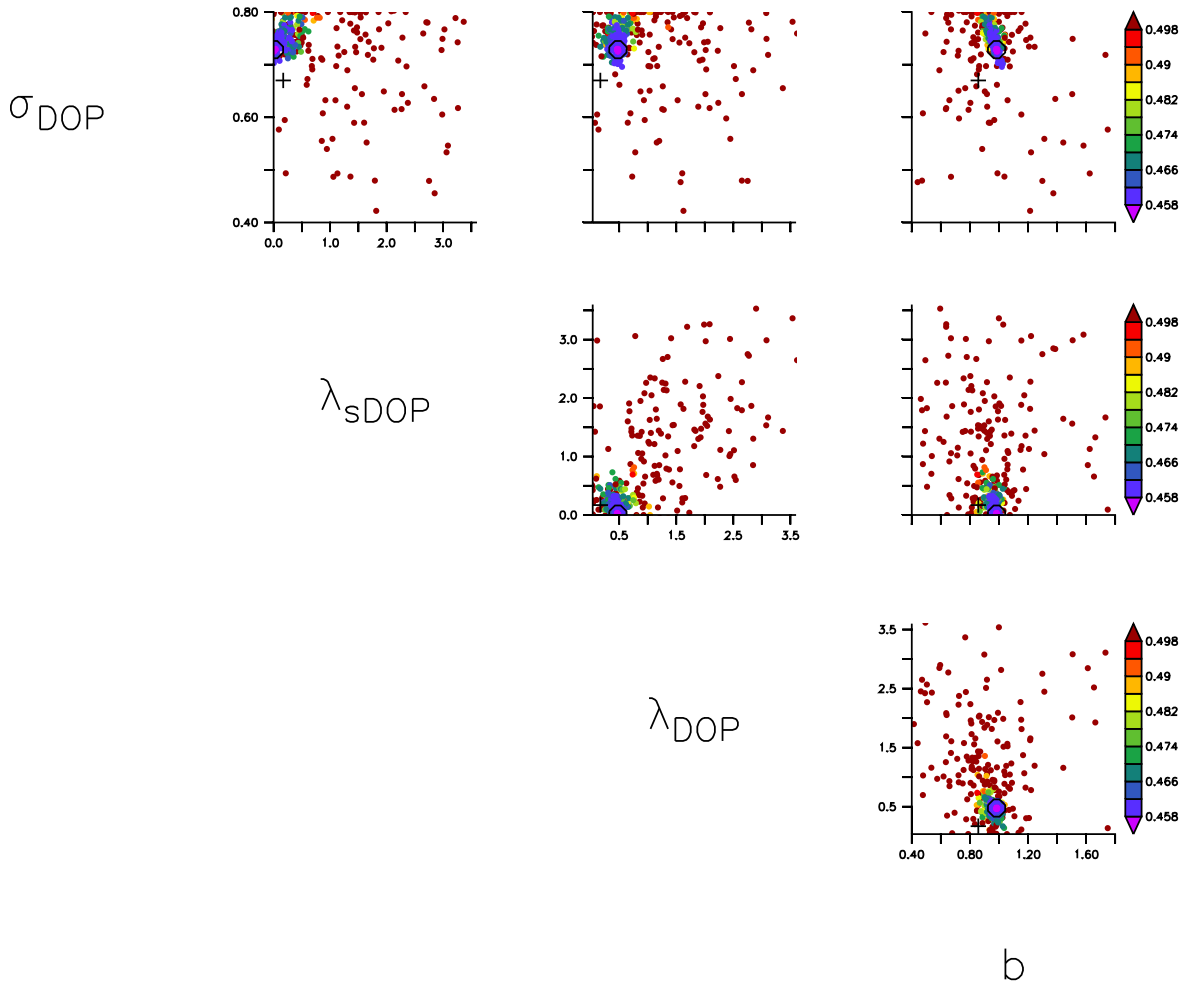

Figure S7. Model misfit, plotted for each pair of parameter combinations of RetroMOPS°. ~~Color~~ Colour indicates misfit (see the ~~color~~ colour bars on the right). A circle indicates the parameter of one individual of the last generation. For better visibility the parameter range to its boundaries (see Table 1).