Biogeosciences Discuss.,
doi:10.5194/bg-2017-71-AC3, 2017

Interactive comment on "Calibration of a simple and a complex model of global marine biogeochemistry" *by* Iris Kriest

I. Kriest

ikriest@geomar.de

Received and published: 30 May 2017

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 pos-

sible, 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 of time 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 want

to 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, MOPSËĘoD 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 $\sim$0.02 per day ($\sim$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 RetroMOPSˆr

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2017-71, 2017.