Our response is italicized

Review of Matsumoto et al.: MESMO 3: Flexible phytoplankton stoichiometry and refractory DOM

The manuscript of Matsumoto et al. describes the third version of the MESMO model, a model of intermediate complexity dedicated to studies on ocean biogeochemistry. Major model developments include three phytoplankton functional types, a flexible phytoplankton stoichiometry based on ambient nutrient, temperature and light conditions, as well as an implementation of a refractory DOM pool, in addition to the existing semi-labile pool, also including stoichiometry of C, N, P and Fe. The model is well and comprehensively described and the substantial developments after version 2 fully justify a new version description. The paper is well structured by starting with the developments since version 1, the newly implemented processes and a discussion of the model results. It will be a useful and complete reference for the future use of this model version, and fits perfectly within the scope of the journal. Therefore, I recommend publication after the comments below are addressed:

Thank you for the overall positive and supportive assessment.

General comments:

- I think the overall improvement resulting from the changes in the new model version could be made more clear in a quantitative way. This is qualitatively described in paragraph l. 484ff, and on a 'global-mean' basis in Table 3. Adding more parameters that are difficult to constrain (such as e.g. the partitioning between DOMsl and DOMr, or uptake stoichiometry) should somehow "pay off" in terms of better agreement with observations or a more flexible representation of processes needed to model biogeochemistry in a changing ocean scenario. It is described that e.g. the nutrient pattern is better in MESMO3 than MESMO2, can you quantify this (with e.g. an RMSE, or a scatter plot/taylor diagram)? Or in Tab. 3, export C:N:P ratio seems to be less close to observations for the new version – what is the cause? Is there a better representation of the spatial pattern instead?

Following the reviewer suggestion, we modified Table 3 to include RMSE for P, N, Si, and Fe to be more quantitative. This point was also raised by the other reviewer. We modified this paragraph substantially (now paragraphs 2 and 3, section 3) to discuss where we do well (e.g., depth of O2 min and SiO2 surface depletion) and where do not to do so well (e.g., eastern equatorial Pacific is too depleted).

- In combination with the above comment: Could you describe more in detail what your calibration method was (in l. 152f you state that there was a calibration procedure, but you only state the absence of it)? Did you adjust parameters to fit a global mean like in Tab. 3 or a global mean profile of e.g. concentrations, or a specific dataset?

Continuing with the above response, our new paragraphs 2 and 3 in section 3 give more detail. We tuned MESMO 3 by hand/experience with an eye towards getting the right balance of community composition and C:N:P ratio for each PFT. As noted above, we also tried to get the O2 minimum at around the right depth, because it helps drive denitrification at approximately the right depth.

- It would be helpful if some runtime information is given for the new model in comparison to the old one. EMICs are great tools for "cost" efficient calculations, so it would be helpful for future users of the model to be able to calculate runtime.

We note that it takes 1 hour to do 1000 years on a single CPU (first paragraph, section 3).

Specific comments:

- 224: "NPP is produced" – maybe revise, e.g. "all NPP is immediately routed to…"

Changed as suggested.

- 227: do you mean "master nutrient variable"?

Changed as suggested. It was a typo originally.

- 246: If I understand correctly, your "deep POC split" pathway represents solution of DOM from POM (by e.g. exoenzymes or the like) – is there a reason why this happens only at depth and not at the surface? Or is the relative role of this process in producing DOM larger at the subsurface compared to the surface and therefore it is neglected at the surface?

No, this happens at the surface also. We have reworded to make this clear.

- 320: Does photodegradation transfer DOMr to DOMsl in your model, or does it just represent a sink for DOMr?

The latter – just a sink for DOMr. This is stated on line 321, and explicitly shown in Equations 50-52, which shows DOMr going to the dissolved inorganic nutrients instead of DOMsl.

- 463: Can you provide the units for this equation? I'm probably missing sth here, but why is density a factor in the gas-exchange equation (sticking to SI dimensions L=length, T=time, N=chemical amount/mol, M=mass unit):
  - $[N/(T*L^2)] \neq [M*N/(T*L^5)] = [L/T] * [M/L^3] * [N/L^3] * [-]$
  - flux k density delta conc. ice fraction

We appreciate that the reviewer has picked up on this correctly that the density factor is unnecessary. It is now removed from Equations 67 and 68.

- 489: Based on the changes in the code, can you pinpoint the reason for the improved nutrient fields? A sentence here would be helpful (similar to line 493, where you state why the oxygen minimum depth improved – I find this very helpful).

Please see our new paragraphs 2 and 3, section 3. Pinpointing subsurface O2 was relatively easy because there are only few parameters involved. The factors that influence the surface nutrient fields are more numerous and more difficult to pinpoint. But we believe that the faster particle sinking is important, as it would deplete surface nutrients more strongly (and ultimately reduces export production). In the process of getting the right balance of PFT community composition, we modified the optimal nutrient uptake rates and the half saturation constants. These parameters would also affect the nutrient depletion. As we noted above, we considered the strong N and P depletion of nutrients in the subtropical gyres to be improvements in MESMO 3, because there was insufficient depletion in MESMO 2. However, the RMSE is not actually lower in MESMO 3 compared to MESMO 2 (Table 3). We thus make clear that there were certain features we had targeted in our calibration (e.g., community composition, C:N:P, O2 min depth).

594: Can you give reasons for why the DOM concentration at the surface is so high compared to observations? I was wondering whether this has to do with the way the model in Hansell, 2012 is formulated, that led to estimating the turnover times: There they scale the DOM production with the square root of NPP, this potentially may lead to lower surface concentrations and – by using their turnover time fitted to this model set-up – cause the overestimated concentrations in MESMO3. (eq. 1 in Hansell, D. A., C. A. Carlson, and R. Schlitzer (2012), Net removal of major marine dissolved organic carbon fractions in the subsurface ocean, Global Biogeochem. Cycles, 26, GB1016, doi:10.1029/2011GB004069)

We believe there is a confusion here. Hansell et al (2012) determined the turnover times (1.5 yr for semi-labile and 16 kyr for refractory) "on the basis of empirical correlations of DOC with water mass age from CFC data…" Their turnover times were not "fitted to their model set up." The reviewer's question remains however, as to why our simulated DOM at the surface is so high when we used the "literature values" for key DOC model parameters. This is an important finding of our paper. While data-model agreement does not always point to concrete lessons, data-model disagreement often does. Our lesson is that: "… the data-LV run mismatch in the overall DOC_t concentration (Figure 7) and surface DOC_t pattern (Figure 9) indicates that the parameter values from the literature do not fully capture the DOC cycle and/or MESMO 3 is still lacking some important DOC process." (second to the last paragraph, section 3.2). We now elaborate that the mismatch is due to DOCr being too high. This could be a result of DOCr production being too high (e.g., fDOMr) or DOCr degradation being too slow (suggested by reviewer#2). A graduate student is currently working on this, and there is still quite a bit more to do. A point that we

make with Figure 7 is that it is possible to achieve a better match to observations than the LV model.

Figures

- Figure 1: The illustration of the sink processes of DOMr might be misleading: does the background decay happen all the time or only at the surface? If it is a true background decay happening everywhere, it might be helpful to write sth like "T_{photo} + $T_{background}$" instead of only "T_{photo}". Or does the background decay only happen at a specific depth level – then this would need to be justified why there is no first-order loss process at the surface?

Figure 1 caption explicitly notes: "the slow background degradation that occurs everywhere." We now add "ubiquitous" in the main text to be clear (Line 323)) and modified Figure 1. Also, Equation 28 has both terms (bg and photo) reflected. Practically speaking, when both bg and photo are activated in a run, DOMr degradation will be driven by photo, goes faster by 3 orders than bg.

- Figure 3: The caption beginning with "the effect of…" suggests that an effect is shown, i.e. the difference between a simulation with or without RNPG. Please either show an effect, or adjust the caption accordingly.

Caption is modified. We clarified that the comparison of panels b and c indeed shows the effect of RNPG. Both opal and CaCO3 are linked to the same PFT eukaryote, yet they are different due to RNPG.

- Figure 4: Panel b would benefit from a latitudinal mean from observations for comparison, maybe from the dataset you already cite: Martiny, A. C., Pham, C. T., Primeau, F. W., Vrugt, J. A., Moore, J. K., Levin, S. A., & Lomas, M. W. (2013). Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. Nature Geoscience, 6(4), 279-283.

We have modified Figure 4b as suggested and added observational constraints.

- Figure 7: Please define DOCt again in the caption

Caption is modified suggested.

Again, we appreciate the constructive comments.

Our response is italicized

I agree with Reviewer #1's analysis that this manuscript represents a meaningful contribution to the current and potential users of the MESMO model and is well-suited in scope for GMD. I too recommend publication after my comments below are addressed. Many of my comments ask for better documentation and/or clarity of the methods used while I also offer some recommendations on the future calibration of the optional DOM cycling module.

Thank you for the overall positive and supportive assessment.

Equation 3, Line 132-135. It is not immediately apparent how this equation is analogous to the Q10 = 2 relationship. For example, the value of FT for a temperature of 288 K is 0.973 and for a 10 degree higher value of T = 298 K the value of FT is 0.974 or only 0.1% higher.

This equation requires temp in Celsius (as noted), not Kelvin. 0C: Ft=0.2, 10C: Ft=0.6, 20C: Ft=0.73... So, it is not exactly Q10=2 but is "analogous"…higher than Q10=2 in colder waters and lower in warmer waters. As we described in the Version 1 model paper, the equation comes from Maier-Reimer (1993), a seminal work on biogeochemical modeling.

Line 141 – 144. It is unclear how Eq 1 varies as a function of zml from the statements as written. Does this mean that nutrient uptake can be confined to model level 1 if zml < zc?

Max(1,zc/zml) is meant to simulate a bloom-like condition according to the Sverdrup spring bloom model. See last paragraph of section 2.1 Thus, when zml < zc, we give it a boost by zc/zml>1. It does not change the fact that the production is confined to the upper two layers. In the Version 1 description paper, we cited Doney et al. (2006) for this formulation.

Table 2b typo: [NO3]0 is listed as [PO4]0

Corrected.

Equation 5: It is unclear how this equation produces an estimate of P:C (or C:P) that approaches the observed range of ocean C:P stoichiometry of marine phytoplankton and associated nutrient uptake to produce said plankton stoichiometry. For example: if I use Eq. 5 and the parameters from Table 2 to compute the P:C of uptake for the cyanobacteria in a low latitude oligotrophic setting with in situ [PO4] = 0.1 μM, [NO3] = 0.1 μM, T = 298K, and I = 150 W/m2 I get: 6.3 x 0.614 x 0.411 x 0.827 x 0.962 = 1.266 for P:C or 1/P:C = 0.79 for the C:P ratio. Redfield C:P is 106 or 117 to 1 and cyanobacteria should likely exhibit a C:P > 117:1. What am I missing? Is P:C expressed in per mille (‰)? There isn't any mention of this in the text or Table 2. If that is the case I compute the same example P:C = 0.001266 for C:P = 790:1. Is there any evidence from the literature for such a P-depleted uptake stoichiometry for small phytoplankton? The maximum C:P for cyanos that I am more familiar with maxes out around 350 or 400:1. Figure 4d shows the cyano C:P and its max is ~280:1. Better clarification of how Eq. 5 (and Eq. 6) in conjunction with the parameter values in Table 2 are used to compute PFT C:N:P stoichiometry is needed.

Thank you for this comment. Indeed, the P:C and N:C ratios are in permil, and we had neglected to mention this. We now make this clear in Table 2b. We also follow Equations 5 and 6 in the main text with an explanation that there are hard upper and lower limits for the C:N:P ratio following observations.

Equation 19 and the discussion on the variability of fDOM (Line 234-243). I refer the authors to Roshan & DeVries 2017 in Nature Communications who report on the latitudinal variability of this parameter from an artificial neural network model of DOC production and export. The variability is within the range ~0.01 – 0.7 which differs from the 0.28 – 0.96 values used in this study. I do not mean to infer that the one study's range is more correct than the other. I do think that the manner in which the authors vary fDOM as a function of temperature in this study to be a clever formulation and as it captures the likely environmental factors that cause fDOM to vary related to the differences in nutrient levels, light levels, and phytoplankton community structure which co-vary in the ocean.

We are aware of the RD2017 paper and believe the reviewer is referring to their NDP/Cex ratio. This ratio is actually not the same as fDOM. The denominator of their ratio is POC export (Cex), while the denominator of our fDOM is NPP. The numerator is DOM production (JprodDOM) in both.

RD2017: NDP/Ced=JprodDOM/Cex=0.71 x f_picophytoplankton – 0.101 log10(NO3)

This study: fDOM=JprodDOM/NPP

We provide more discussion about fDOM below.

Equation 31 and value of Kr in Table 2. CaCO3 is more soluble at colder temperatures unlike most solids due to the positive Gibbs free energy change at standard conditions for the dissolution reaction. The parameter Kr is listed as positive 0.69 in Table 2. Thus when computing the temperature scaling term for CaCO3 remin rate exp(Kr*T) you will get a faster remin rate at higher temperatures for CaCO3, not at colder temps. Should the value of Kr be negative 0.69? This would give a faster CaCO3 remin rate at colder temps. CaCO3 solubility is also pH dependent, through the changes of [CO32-] as a function of pH, but without a full carbonate system formulation in MESMO I think it is ok to simply vary CaCO3 remin as a function of temp since pH and temp co-vary in the ocean. The handling of the opal remin rate appears to obey the correct temperature dependence for opal with a higher opal remin rate at warmer temperatures in Eq. 33.

We stand corrected on the sign of kr in Equation 33 and Table 2d. BTW, the correct value is 0.069, not 0.69 as we had originally. It turns out that we had the positive sign for kr all this time. We have thus repeated all MESMO 3 experiments with the correct sign and remade all figures and recalculated all values presented in this work. There are some modest changes in results, but the model overall is not sensitive to this change in kr as evident from the comparison of the new version of this manuscript from the previous version. In Table 2d, we have "0.069" in MESMO 2 and "-0.069" in MESMO 3.

Line 482. Global average fDOM of 0.71 seems problematic. Other estimates that have been constrained by the available DOC concentrations across the global ocean are much lower, e.g. 20% Hansell et al., 2009; 25% Letscher et al., 2015; 17% Romera-Castillo et al., 2016; 20% Roshan & DeVries, 2017. I recommend that the authors revisit the prescribed range on fDOM in future calibration efforts of the DOM module. To be clear, in my opinion I do not believe that a fully calibrated DOM module for MESMO 3 to be a prerequisite for publication of the current manuscript.

These low numbers are not equivalent to what we call fDOM. As we discussed Roshan & DeVries 2017 above, NDP/Cex is not fDOM. Hansell et al. 2009 note: "the amount of DOC that is routed through rapid bacterial production may be as much as 50% of primary production…" This is not 20%. Letcher et al. (2015) say that their model sends 15% of modeled PP to DOM but "does not track the total production/decomposition of DOM," meaning they don't have fDOM. Romera-Castillo et al.'s ratio is DOM production to NCP.

From various independent estimates, we know that global NPP is 20-70 PgC/year, and POC export can reach 10 PgC/yr (Table 3). For simplicity, if assume that NPP is 50 PgC/year, then fDOM=DOM prod/NPP=(NPP-POC export)/NPP=(50-10)/50=0.8, which is not dissimilar to the global fDOM that we get. To give some historical context, the Ocean Carbon Cycle Model Intercomparison Project (OCMIP) adopted Yamanaka and Tajika's number of 0.67 for fDOM. This is the value that MESMO 1 and 2 used.

Line 486-489 and Figure S1. What happened to the EqPac HNLC in MESMO 3? It's gone. This suggests there is not enough Fe limitation in that HNLC to keep NO3 and PO4 in the surface or that the new variable P:C and N:C formulations allow for too much NO3 and PO4 drawdown in this HNLC, although this latter mechanism should operate in the Southern Ocean HNLC too which is retained in MESMO 3. This particular deficiency of MESMO 3, it's lack of an EqPac HNLC should be identified in the text with a possible explanation given by the authors. Also can the authors provide a more formal statistical comparison of MESMO NO3 and PO4 to World Ocean Atlas, e.g. point by point correlation and overall mean bias.

Following this suggestion, we have substantially modified this section where Figures S1-5 are introduced (new paragraphs 2 and 3, section 3). We now note and discuss the deficiency and provide RMSE. See our response to Reviewer#1.

Figure 2. Eukaryotes are limited by silica nearly everywhere in the global ocean except the Southern ocean and subarctic N Pac. How does this compare to the literature? My thoughts are that silica limitation is not as widespread in the real ocean. Perhaps the RNPG formulation for Euk growth is contributing to the high levels of Si limitation? Cyanobacteria are limited by nitrogen almost everywhere except the Southern Ocean. This is likely a result of the biased low surface NO3 concentrations simulated by MESMO 3. Most field estimates suggest the Sargasso Sea is P-limited to all PFTs with this feature only simulated for the diazotrophs in MESMO 3. I recommend a discussion of the PFT nutrient limitation patterns in comparison to the broad patterns identified by the extant literature on the topic, pointing out where MESMO 3 does a better job and where it is deficient.

We have added more description on the nutrient limitation in the second paragraph of section 3.1. One of the key features of the observed marine Si cycle that we previously reproduced in MESMO 2 and that we preserve in MESMO 3 is the negative Si (=Si-N), which originates in the Southern Ocean and spreads across the world ocean. As shown by Sarmiento et al. (2004), Si*<0 in most of the world's thermocline (except N Pac), whereas diatoms typically require Si:N of 1 or higher, implying that diatoms are generally limited by Si. Also, Ragueneau et al. (2000) indicate that 60% of surface waters have [Si]=1~10 uM where "Si uptake should be measurably limited." As far as organic C is concerned, we consider eukaryotes in MESMO 3 to be basically diatoms, which are arguably the single most important agent of C export. Thus, the large Si limitation in MESMO 3 (and seen in MESMO 2) is expected. We are not aware of any data-based global map of nutrient or Si limitation of diatoms. The global nutrient limitation map from Moore et al. (2013) is not broken down by PFTs, but it shows that N is the primary limiting nutrient even in the N Atlantic, where P is indicated as a co-limiting nutrient. Moore et al. note that the "addition of P alone does not typically result in increased autotrophic activity of biomass." We are of course aware of the extreme P depletion there and that phytoplankton use alkaline phosphatase to use DOP…we had cited Mather et al.*

Figure 4 PFT C:P stoichiometry looks great and largely matches the global patterns and variability seen in field observations.

Thanks.

Figures 7, 8, and 9. Some comments about tuning MESMO DOC. MESMO3 DOC has a positive mean bias when the authors used literature values for the DOC production and remineralization fluxes. The authors then made adjustments to the production flux (from 1% of NPP produces DOCr to 0.2%). I wish to point out to the authors that they can also make adjustments to the remineralization fluxes, i.e. increase the decay rate constants to attempt a better calibration of simulated DOC vs. observed. Neither the DOMr production or consumption fluxes are well constrained observationally at present. The MESMO 3 deep ocean DOCt gradient captures well the thermohaline circulation gradient. The simulation of DOCsl in the upper ocean is reasonable as compared to the neural network data product of Roshan & DeVries. The authors choice of tying the DOMsl production flux to temperature is a good first step to capture the behavior seen in the DOC observational data and the neural network product that DOC production peaks in the subtropical gyres where temperatures are also high. On line 647-650 the authors state that they may try to tie the fDOM parameter negatively to nutrient concentrations to try and achieve a higher DOMsl production flux in the subtropical gyres. In my expert opinion I believe that tying fDOM to both temperature and inversely to nutrients will likely capture the behavior they wish to reproduce. I refer the authors to a review of this topic on 'extracellular release' by phytoplankton within Chapter 3 of "Biogeochemistry of Marine Dissolved Organic Matter" 2nd Edition. Edited by D.A. Hansell & C.A. Carlson published by Academic Press. Rather than temperature, to be more physiologically consistent, the authors may wish to tie fDOM to PAR rather than temperature (although both co-vary in the ocean). This could be left to a future study that focuses on calibrating the DOM module in MESMO 3.

We appreciate these suggestions and insights. We currently have a graduate student working on this and these comments are very useful.

Figure S6. Why not also show the water column depth integrated denitrification flux alongside the N2 fixation flux?

We now show both denitrification and N2 fixation side by side in Figure S6.

Again, we are grateful for the careful reading of our manuscript and constructive comments.