Author response to reviewer 2.

We thank the reviewer for the careful review comments. We understand that a primary concern has to do with the extent of the model description. The reviewer notes that "GMD is a forum for accurately and completely describing models" and views as problematic our presumption "that earlier papers should be consulted." We did indeed make this presumption especially as it relates to our earlier description of MESMO in GMD, because the journal allows a collection of related papers published in GMD to be bundled into a virtual volume. In our revision, we will make this point but also, in direct response to the review, enhance our model description with regard to "little reference to the physical framework," "cursory treatment of the marine biogeochemistry," and "nothing on simulation design." As part of this enhancement, we will introduce a couple equations on iron cycle and stable silicon isotope fractionation as requested.

The other concern is that "a step-by-step upgrade and description would have been preferable, but here each step is more of a jump." We believe that the basic structure of the paper actually already contains a step-by-step approach. For example, in section 2.1 on the new physical features, we first describe changes to albedo. Then we take a <u>step</u> and note that "the second important change made in MESMO 2" is the introduction of the new monthly ECMWF winds. We note the impact of this change as "The use of the ECMWF winds alone increases them to −80 ‰, −104 ‰, and −129 ‰, respectively." Then in the next paragraph and in another incremental <u>step</u>, we note that this "this excessive ventilation" of the deep ocean is corrected by adjusting the wind stress scaling factor. We describe the impact of changing this factor first globally then regionally. In our revision, we will reorganize, rephrase, and add new descriptions to make the step-by-step improvements more clear. It should be recognized however, that nonlinearity and feedbacks in the model would not necessarily allow the step-by-step changes to achieve the final model state in a linear fashion.

For GMD's 15 review criteria, the reviewer responded positively to all but two, which are:

6. Is the description sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? In the case of model description papers, it should in theory be possible for an independent scientist to construct a model that, while not necessarily numerically identical, will produce scientifically equivalent results. Model development papers should be similarly reproducible. For MIP and benchmarking papers it should be possible for the protocol to be precisely reproduced for an independent model. Descriptions of numerical advances should be precisely reproducible.

No. The presumption of the authors would appear to be that earlier papers should also be consulted. This is acceptable for the underpinning climate model, but the developments in marine biogeochemistry here need to be better described.

As we commented above, we do make this presumption, which reflects our view of the uniqueness of GMD, but we will also increase model description in our revision.

15. Is the amount and quality of supplementary material appropriate? For model description papers, authors are strongly encouraged to submit supplementary material containing the model code and a user manual. For development, technical and benchmarking papers, the submission of code to perform calculations described in the text is strongly encouraged.

Unfortunately I have been unable to access the supplementary material. Given my concerns with the manuscript, I judge this a secondary consideration at this point.

Our supplemental material is tar-zipped model code including all the forcing files. We have independent knowledge that someone (L. Gregoire) has successfully downloaded and opened the material, so the complaint here does not seem justified. In any case, GMD should take the lead on the accessibility of supplemental materials.

Specific comments
- Pg. 3000, ln. 17-23: while readers familiar with GENIE may understand the version used by the authors, there should at least be some mention of Earth system components, spatial representation (horizontal and vertical), time-step, etc.

Yes, we will mention this in our revision. For the record, the basic information is in the MESMO 1 description in the earlier GMD paper.

- Pg. 3001, ln. 25: the text states that diatoms are the most important biological group when it comes to export production, and cites Armstrong et al. (2002) in support; this and subsequent papers also emphasise the role of CaCO3, and even nearcontemporaneous ones, such as Klaas & Archer (2002), put the emphasis on this latter biomineral ahead of opal; as MESMO includes a parameterisation of CaCO3, the relationship of the two biominerals should be made clear

Klaas and Archer and other papers indeed discuss the importance of CaCO3 as a dense ballast material for vertical transport of organic matter. Density is not the only factor that determines global significance of the ballast materials however. The particular occurrence of opal in time and space (huge blooms in short time) and manner (ability to form mats) gives opal its great global significance. As we responded to reviewer 1, we will cite Sarmiento and Gruber's textbook to make this point.

- Pg. 3002, ln. 16: the text would be much clearer on albedo if the authors added a sentence explaining why it varies with latitude; at present readers not versed in climate might suspect everything is simply more white at higher latitudes (which, unhelpfully, is not entirely wrong)

We will add a brief explanation that insolation at high latitudes will have to pass through a greater column of atmosphere and strike the surface at oblique angle…and more white.

- Pg. 3002, ln. 16: I would be grateful for some explanation for why the authors improve MESMO 1 in several directions (physical climate, marine biogeochemistry) simultaneously rather than incrementally; not least so that the later intercomparisons allow separation of the contributions of the new components; at present, it is difficult to ascertain which change to MESMO 1 now present in MESMO 2 is most significant

Again, we understand that this is an important concern for the reviewer, and we will revise to address this point – see the second paragraph of our general response above.

- Pg. 3002: more generally, it would be helpful if the authors included figures here like those in the MESMO 1 description that give a flow chart overview of the development process

We don't imagine that a flow chart is a good use of space, but we will add a paragraph at the beginning of the model description that will serve as a road map.

- Pg. 3003, ln. 16: could the C-14 numbers here be converted to ventilation timescales using C-14 half-life?

We could, but C14 in data-model comparison (as in OCMIP) is typically expressed in abundance rather than years.

- Pg. 3004, ln. 2: given the apparent flexibility of the physical model's ventilation, why did the authors not just choose values for Atlantic-to-Pacific FW flux that give MESMO 2 and MESMO 2E MOC strengths that were the same?; instead, the authors seem happy to settle on different base physical climates in the models which (again) makes it difficult to establish the source of model improvement

We could have possibly done this in two ways: (a) as we did, keep the same physical parameter values which give different MOCs in MESMO 2 and 2E; or (b) as suggested here, further modify

the physical parameters of 2E relative to 2, so that the MOC would be similar. We don't believe either of the two ways is inherently superior. Had we done it the other way (b), it is conceivable to have received a criticism saying we should not have done so. Also, we did as we did, because 2E has a somewhat stronger and more reasonable MOC than 2; it would not make sense to "modify" (but actually degrade) 2E to the same level of MOC as 2.

- Pg. 3004, ln. 2: the authors focus here on the Atlantic MOC, but it would be helpful if they said more about the wider circulation; for instance, the strength of the ACC or the Pacific overturning cell in the various model versions

Actually we don't believe we focus solely on the Atlantic when it comes to deep ventilation. As we are primarily biogeochemists, we present deep D14C as measure of deep ventilation not just in the Atlantic but in the Southern Ocean and Pacific as well. In contrast to previous studies like Marsh et al. (2011), who do not present deep D14C, we would argue that D14C has distinct advantage as a deep ventilation metric in that it provides an observational constraint. We presented the Atlantic MOC because that is a commonly used metric.

- Pg. 3004, ln. 11: while I understand the reflex to avoid repetition in papers, this manuscript would be a lot more helpful if it added a paragraph to explain how, exactly, SP and LP are modelled; are they standing stocks of biomass, or do they simply represent fluxes of material?; the manuscript – which is ostensibly about model description – is pretty opaque on this point

Yes, thanks for making this point. We will note explicitly that SP and LP are export production, so they represent fluxes of materials across the critical depth of 100 m water depth.

- Pg. 3004, ln. 12: the authors tie primary production to the availability of CO2 rather than DIC – this is quite an unusual formulation, and is not justified here; most biogeochemical models instead view DIC as an effectively unlimited resource that phytoplankton can draw on; as well as justifying the choice here, it would be helpful to know how significant it is for model behaviour – i.e. are phytoplankton limited by CO2 anywhere?

DIC is of course abundantly available, but Riebesell et al. (1993) has shown that there can be CO2 limitation if CO2 is diffusion-limited. This is noted in our original description of MESMO 1 in GMD, which we again assume will be consulted. With the chosen half saturation constant, which is informed by Riebesell's work, CO2 does not become limiting. However we keep this mechanism in place because it is something that we could envision using under glacial or other runs.

- Pg. 3005, ln. 6: I may be misunderstanding the description here, but does MESMO 2 have a fixed C:Si ratio?; this seems very much at odds with many observations of diatom productivity; conventionally, the ratio varies with growth rate such that diatoms limited by nutrient availability (e.g. in HNLC regions) accumulate Si and shift their C:Si ratio downwards; this stems from a degree of decoupling of silicon uptake from the cell division cycle.

No, C/Si is not fixed. As noted as one of the new features of MESMO 2/E, Si/N ratio changes and depends on Fe, while C is tied to N by stoichiometry. This means that the model has variable C/Si ratio. We will explicitly note this in our revision.

- Pg. 3005, ln. 20: the text implies that opal is supersaturated in some locations in the modern ocean; I do not believe this to be the case, but can the authors clarify here?; it's certainly true to CaCO3, but I was not aware that opal was anything other than undersaturated everywhere (something which is not true at earlier points of Earth history)

No, we made no such implication. It merely states that the formulation of opal dissolution depends on a number of factors including the degree of saturation. But we will rephrase the section to avoid potential confusion.

- Pg. 3006, ln. 2: I would expect a model description paper to be far less circumspect in describing components than here; the authors refer to a version of the GENIE model at a particular time, but neither give a published source for this specific model nor attempt to formally describe the iron cycle equations used here

As we responded to Reviewer 1, the particular way we described the Fe code represents our effort to not take credit for work by the original author of Genie Biogem Fe code, which has not yet been described in the literature. We will provide equations for scavenging and ligand binding, two essential processes of the iron cycle.

- Pg. 3006, ln. 21: the relationship between remineralisation and scavenging would be much clearer if the authors simply wrote out the equations.

Yes, see above.

- Pg. 3007, ln. 9-11: I'm struggling to grasp what the authors mean here; perhaps explaining what used to happen – and the effective consequences of this – would make it clearer

This was a small bug fix. We will rephrase.

- Pg. 3008, ln. 9: "are adjusted" – how was this done?; for instance, with three parameters it's not at all obvious how an optimal solution was reached (assuming that it was!); did the authors just engage in separate and sequential tuning by eye?; or was there more to the process?

We will explain that a parameter space was crudely swept for given ranges in K18, k24, and k29.

- Pg. 3008, ln. 20: the authors jump immediately into discussing model performance at equilibrium in this section without making it in the least clear how simulations were performed; what was the initial physical state?; what was the initial biogeochemical state?; do these matter?; how long were the simulations?; is this long enough to ensure "equilibrium"?; how is this "equilibrium" defined?; the authors mention OCMIP, but are they using the sort of equilibrium used there?; is the carbon cycle closed, or is a component of it (atmospheric pCO2) held constant while the rest of the cycle comes into balance?; there are a long list of considerations here, but the authors seem to think they don't matter; again – this is inappropriate for a model description paper

As in the usual sense of carbon cycle equilibrium, it represents a model steady state using preindustrial boundary conditions. The equilibrium is reached when the deep ocean's radiocarbon content has leveled off, which typically takes several thousand years. We will explain the details of all this.

- Pg. 3009, ln. 8-10: "The stronger reanalysis winds ..." – this is stated rather baldly without any explanation; could the authors expand a little to be clear why this is the case

Yes, we will note how the stronger southern hemisphere winds enhance deep upwelling in the south, makes the deep Southern Ocean radiocarbon too young, and as a consequence reduces the NPDW-NADW difference in D14C.

- Pg. 3009, ln. 12: we segue from equilibrium simulations to transient simulations with no commentary from the authors here; I presume that OCMIP protocols are being followed, but there's nothing in the manuscript to explain.

We will explain that natural radiocarbon from equilibrium simulation is used to assess the deep ocean ventilation. In contrast, bomb radiocarbon is a transient tracer and useful as metric of decadal ventilation.

- Pg. 3009, ln. 19: how does penetration depth of anthro CO2 and CFC-11 compare with observations?

We know from previous studies that anthro carbon and CFCs are transported into the interior primarily by the intermediate waters and the Atlantic MOC. As shown, for example by Doney et al. (2004), NADW is shallower in models than in observations, so the transient tracer penetration depth would be shallower. Our model should be no exception to the rule. We will confirm this and note it in the revision.

- Pg. 3009, ln. 20: just an idle observation: would the average air-sea gradient in temperature be an instructive metric?

Generally speaking, it could be. But it would also be redundant if SAT and SST are already used as metrics.

- Pg. 3009, ln. 27: how about sea-ice volume?; or seasonal extent?

We list the maximum seasonal sea ice cover in Table 1 and show it in Figure 5. We will discuss the thickness as well in our revision.

- Pg. 3010, ln. 7: “values ... are tuned” – how?; is this done formally via a objective technique, or informally by hunch and by eye?

Tuning is by “hunch and eye,” although we will phrase it in our revision as subjective tuning.

- Pg. 3011, ln. 25: the authors attribute high North Atlantic Si to the deposition of Fe that decreases Si demand; however, Si reaches its lowest high latitude concentrations in this basin in observations; furthermore, in part this is due to stripping out of Si as water moves northward in the Atlantic basin – what’s happening here?

As pointed out by the reviewer and as we will clarify, the North Atlantic is one place where the simulated Si concentration does not compare well with observations. We will expand Figure 9 to include observations to make this point. It is nevertheless well supported by data that Si demand (relative to N) is reduced there, most likely because of the Sahara-derived Fe deposition. The reduced demand in terms of low Si/N ratios can be readily confirmed in Figure 2c of Sarmiento et al. (2004). We will suggest in our revision that the North Atlantic has a stronger opal export in the real ocean compared to our model even though Si/N is low in both, because diatom production itself is greater in reality, possibly because of the Fe deposition. In the model the North Atlantic becomes N-limited (Figure 8), so that high Fe deposition does not lead to greater diatom production. This discrepancy likely indicates the limitation of the simple ecosystem we have in the model, and it cannot capture the full mechanisms of the spring bloom.

- Pg. 3013, ln. 27: the authors state that MESMO 2E is simulating the pre-industrial vegetation state but given that the simulations include anthro CO2, is this true?; or are the authors showing vegetation at the end of the equilibrium phase?

No, this is not true. The 2E simulation is a preindustrial equilibrium run, so there is no anthro CO2. We will make this clear.

- Table 0: perhaps a table listing all of the changes to create the models would be good?

We already have Table 2, which lists all the changes. Also as noted in our response to Reviewer 1, we will expand Table 2 to include the half saturation constants from MESMO 1 to facilitate comparison of the new model with version 1.

- Figure 1: why no colour?; these are ugly and difficult to interpret

Our general philosophy with respect to the use of color is to use them sparingly. This is in consideration of those who have difficulty distinguishing certain colors, minimizing publication cost for us, and minimizing printing costs for those who may not have ready access to color printers. We actually think that the labeled contour plots in Figure 1 adequately show the strength of MOC. We will do grey scale here, as there would be only one color really (see Fig. 21 of Marsh et al. 2011 for example).

- Figure 4: is this really observed sea-surface temperature, or is it observed nearsurface temperature averaged vertically to a depth appropriate for comparison with the MESMOs?; the latter would be more appropriate (and appear to be done for “surface” nutrient)

Thanks for pointing this out. We will remake the figures for the equivalent depth range.

- Figure 5: why is this plotted so differently from the other model fields? I can see no good reason not to stick to the same format – not least because it would allow presentation of sea-ice cover or height, and would do so far more clearly than this plot

We will try the same color format as others to see if it improves the presentation. But the intent here was, again, to use less color and to more directly compare in a single figure the extent of sea ice in different models. The difference is fairly subtle that separate panels may not be effective. In any case, we will evaluate the reviewer suggestion.

- Figure 7: any chance of a corresponding plot (except MESMO 1, obviously) for Si?; also, why use Levitus (1993) when there have been a succession of World Ocean Atlas fields since then? (1994, 1998, 2001, 2005, 2009)

For Si, indeed Figure 9 will be expanded to include data. As for the date of the Atlas, we will remake the figures using the more recent atlas fields.

- Figure 8: why is P-limitation in white?; I can’t see any cells that are P-limited; also, would a parallel pair of plots (for SP and LP) showing Fn (on Pg. 3004, ln. 18) be instructive?

We will fix the color scale. The limitation shown is indeed Fn.

- Figure 11: why no comparison to observations?; I’ve seen comparable fields plotted before.

We will seek data equivalent for the land model carbon outputs for display in Figure 11.