Author Comment to Reviewer 1

We understand that the greatest concern is the presentation of the silica cycle. Following the reviewer's suggestion, we will expand Figure 9 to directly compare simulated Si concentration against the observed. Regarding d30Si, we thank the reviewer for pointing us to a recent paper by de Souza et al. (2012), which we found to be very useful. There are actually a number of features in our simulated d30Si that are consistent with the new paper. They include the north-south d30Si gradient in the deep Atlantic and the sharp vertical gradient near the surface. We will expand Figure 10 and modify the text to make these points. Finally, in response to the comment that "the proposed mechanisms responsible for the distribution of Si limitation are not well explored," we point out that we actually propose no new mechanisms ourselves. The mechanisms, noted in section 4.1, were presented by the data-based study of Sarmiento et al (2004). In this submission, we are presenting Si* and Si/N uptake ratio from our model to show that key features of the silica cycle described by Sarmiento et al. are reproduced by our model. In our revision, we will move the discussion of the Sarmiento et al. mechanisms from Section 4.1 to Section 1 (introduction) and expand it so that our intent is clear.

Specific comments and suggestions:

Page 3002, line 7: Maybe: ": : : allow for a land albedo feedback ..." or "allow for land albedo feedbacks : : :" would be better.

We will edit the text as suggested.

Page 3002, line 16: Surely you mean the snow-free albedo ranges from 0.2 to 0.5, although 0.5 seems very high for a snow-free albedo. There must be some kind of a land snow feedback in MESMO 1 : isn't there? Can you please clarify this.

Yes, we will clarify this point. There is actually confusion with regard to the semantics of albedo. See our response to the next comment, which addresses this point.

Pages 3002-3003, lines 22-2: It seems to me that the planetary albedo should be a diagnostic quantity not something you specify. This is partly semantics but maybe you should think of this as a change in atmospheric albedo, which is then held fixed, while your planetary albedo is free to vary with changes in snow, ice or vegetation. It is not clear at the moment if your scaling effects these "feedbacks" on planetary albedo or not. Please clarify this. Since your atmosphere is two dimensional, another way to "tune" the surface air temperature would be to make a small, uniform, constant adjustment to the outgoing longwave parameterization.

What we mean by planetary albedo is likely equivalent to what the reviewer is referring to as atmospheric albedo. Our use of the phrase "planetary albedo" is consistent with Marsh et al. (2011). As shown in their Figure 5E, planetary albedo in MESMO/GENIE modifies the incoming shortwave radiation at the top of the atmosphere and is dependent on latitude. We will make this point clear in our revision and also that surface albedo is modified by ENTS when it is coupled in MESMO 2E.

Page 3003: Line 10: Missing "." after (1989).

We will edit the text as suggested.

It makes sense to try and make the wind forcing more consistent. It would appear that a side benefit of including seasonally varying winds is that you can educe the use of an arbitrary scaling constant. Your comment that the ECMWF winds are 40% stronger than the observation-based winds (SOC) over the Southern Ocean, is curious. Both are observation based. They may just have different ways of interpolating" missing data. Is this difference in strength before or after scaling the MESMO 1 wind stress? Please clarify. Perhaps some of the differences are from comparing monthly to annual mean strength? Monthly average winds may well be stronger than annual average winds (daily, stronger still) - especially in areas of high variability. In a similar vein, do you use the monthly mean of daily wind speed from ECMWF? Averaging daily or hourly wind speed would be preferable as it would be much stronger (more realistic) than obtaining wind speed from monthly averaged winds. It is not clear what you did here.

Josey et al. (2002) have a good explanation of the 40% discrepancy. We remind the reviewer that SOC and ECMWF are not equally based on observations. ECMWF is a reanalysis, making use of a dynamical model. So, where there is little data constraint, the model dynamics will have more to say about the ECMWF winds than data, so it is not a simple matter of interpolation. Also, we would choose not to use daily winds for computational efficiency, which allows for large model time steps measured in days.

Page 3003, lines 14-18 and Table 1. You should define the domains over which you calculate NADW, CDW and NPDW delta 14C.

Thanks for pointing this out. We will define the domains and explain that they are from a GRL paper summarizing OCMIP results (Matsumoto et al., 2004).

Page 3003, line 23: I think you should have a comma after "completely".

We will edit the text as suggested.

It is unfortunate that MESMO must jump through such hoops in order to maintain a reasonable Atlantic meridional overturning circulation. Having to scale the winds in the North Atlantic and have different fresh water corrections in different basins and hemispheres makes me concerned that the response of the overturning to climate change is not going to be realistic. If you can not simulate the mechanism that creates the overturning correctly, how can you believe the response? Would it not just be best to remove all the wind scaling and just use a single fresh water adjustment? - it is simpler (more transparent) at least. Do you know how much more of a fresh water adjustment would be required if you removed wind scaling altogether? Would this have other consequences? Having said that, it is disturbing that this fresh water adjustment appears to be getting larger rather than smaller (in MESMO 2 compared to MESMO 1). Can you explain why the newer land surface scheme (in 2E) has a higher overturning? Is this heat (lower albedo => colder NA?) or fresh water related? Changing these flux corrections is perhaps beyond the scope of this paper but this is something that should be carefully considered in terms of future development.

Unfortunately the freshwater flux in question is necessary in order to have a reasonable MOC with this model architecture. In a more detailed analysis of GENIE physics, Marsh et al. (2011) note that the typical MOC does not develop without the flux in all their model configurations Also, that MESMO 2E has a stronger MOC than MESMO 2 for the same physical parameters clearly indicate that the difference has to do with the presence of the land model ENTS.

Page 3004, line 20: It is not clear where the values for the new "Kx's (for diatoms) come from? The values seem a bit arbitrary. Any justification seems hand-wavy. How well can you constrain any "tuning" of these parameters?

In our revision, we will make clear that most of the values for MESMO 2/E were chosen on the basis of the values used in MESMO 1 so as to maintain some sense of continuity. In MESMO 1, there was only one class of phytoplankton, which was split into two in MESMO 2/E, so the Kx's for the small and large in 2/E bound the single Kx in version 1. For example, Kno3=3.4 in version 1 is intermediate between 5.0 in large and 0.5 in small in version 2. Table 2 will be modified to include the MESMO 1 values for clarity.

Page 3004, line 22: "nutreint" should be "nutrient".

We will edit the text as suggested.

Page 3005, line 16-18. Not sure I see the justification for this in Sarmiento et al 2004. Please explain this more clearly.

The justification is clearly provided by Figure 2c of Sarmiento et al. in which the minimum Si/N is 0.30 and found in the North Atlantic. We will note this explicitly in our revision.

Page 3006, line 11: Maybe "as the soluble" would be better. In the discussion of how iron is implemented, it is not clear what is new and what is in the original GENIE "framework". Can you make this more explicit?

The suggested phrase will be used in revision. Also, the way the Fe code was described in our submission reflects a conversation we had with the author of Genie Biogem Fe code and may be somewhat awkward because the original Fe code had not yet been described in the literature but was very close to being described. We will confer with the author of the original code about the status of his description. We will also provide two essential equations of Fe cycle, those for scavenging and ligand binding.

Page 3007, line 2: I think you mean "nM". "nm" is reserved for nanometre rather than nanomolar. Maybe using the equivalent nanomole per Litre [nmol L-1] would be less confusing. You could also use nmol kg-1 since most of your other concentrations are in mol kg-1 (although the SI unit for concentration is mol m-3).

This was a typo. We will consistently use nmol kg-1.

Page 3008. line 2" "phytosynthesis" should be photosynthesis".

We will edit the text as suggested.

Page 3008; lines 4-5: Maybe provide a reference here.

We will cite the O'Leary (1988) paper here.

Page 3008, lines 12-13: Can you explain why these are so much lower than in Williamson et al. (2006)?

These are freely adjustable parameters and there is no *a priori* no correct value for any of these.

Page 3008, lines, 14-18: 5.5% is quite large. Again I am not sure adjusting planetary albedo is the best way to tune global average absolute SAT. If you continuously scale the overall planetary albedo you will also be modifying the effect of surface changes. Planetary albedo is also not very uniform and so changes in SAT will be concentrated in certain locations. Since the source of the absolute SAT error can not be easily identified, I still think a single, uniform constant added to the (relatively uniform) outgoing longwave would be a better way to adjust this. I suspect any adjustment would be very small compared to any uncertainty in the observed outgoing longwave.

See our response to the albedo semantics above. What we did is essentially the same as suggested (i.e., adding a single uniform constant).

The "equilibrium" simulations are not well described. What is the forcing year and how long is the spin-up?

We will describe that the forcings are indeed preindustrial. The simulations are several thousand years long. Equilibrium is judged on the basis of steady state deep ocean radiocarbon content.

The increase in overturning in MESMO 2E is clearly beneficial to the simulation of relatively young delta 14 C in NADW and this helps increase your NPDW and NADW contrast. It would be better yet if the overturning could be increased further. I am not sure I buy the argument that it is poorly resolved shelf processes that are the problem in not maintaining old NPDW. I do agree that AABW production and its transport may be to blame but, in general, your CDW is about the right age. It may be that your sea ice is too far north, causing AABW formation to form too far north (reducing its isolation). It is possible that your deep horizontal mixing is too high. Your ventilation of the NPDW may be too strong. You may also have inadequate topographic resolution to slow the invasion of younger AABW into the North Pacific. There are many possibilities but I think missing shelf processes is probably a minor one.

In our revision, we will mention some of these other possibilities as well as the missing shelf processes, which do control the residence time of surface water and thus gas equilibration.

Page 3009, lines 8-12: Are the reanalysis winds really stronger - even after scaling the MESMO 1 winds by 2?

We will provide a new figure that compares the old and new winds.

Page 3009, lines 20-22: Is your SAT being compared over the 1960-1990 average of Jones et al. or is your SAT at some pre-industrial equilibrium? You should be more specific as to the times in your comparisons.

Yes, thanks for this point. Our SAT was indeed preindustrial equilibrium, while the Jones average is for the 1960-1990 period. We will make this point clear in our revision and Table 1 for SAT as well as for other relevant variables.

Page 3010, lines 1-5: Perhaps you should say something more like "contributes" rather than "leads" here. Certainly some NADW is formed in the GIN seas but much of it is also formed in the Irminger and Labrador seas. "Iceland" is also one word rather than two.

We will edit the text as suggested.

Page 3010, lines 7-9: How was this tuning done? It seems to me that optimal solutions might not be unique given the uncertainly in the parameters and the data you are comparing to. Was this objectively tuned? If so, to what?

MESMO 1 was objectively tuned. The tuning of MESMO 2/E started from MESMO 1 and was based on the experimental experience, which is still the primary way of tuning (e.g., Marsh et al., 2011). Lenton et al. (2006) even state that their subjective tuning was superior to objective tuning. We will make this point clear in the revision.

Page 3010, lines 13-16: It seems like you have tuned the model to show Si limitation for LP over much of the globe (your figure 8), justified with Sarmiento et al. 2004. However, looking at Figure 8 from Moore et al. 2002 (see figure below) it would seem that they suggest silica is limiting over a very small area while iron is the dominant limitation for diatoms over much of the ocean (nitrogen being the other major limitation). I realize that this is just another model but can you reconcile your figure 8 with figure 8 from Moore et al.?

To the extent that the Moore and Sarmiento results are incompatible, we obviously cannot be consistent with both at the same time. While the Moore model is arguably one of the best ecosystem models around today, it is a model. On the other hand, the Sarmiento results are based on analysis of observations. In this submission, we have consciously made the decision to target the Sarmiento results. We will discuss this point in our revision.

Page 3010: The model seems to have about 73% of total production from diatoms, which seems a bit high. Are estimates not closer to 50%? Do you have a good reference for this (maybe Nelson et al., 1995)? Your high values of opal production (upper limits of estimates) seem to support the idea that the model has excessive diatom production.

There is in fact a wide range in the literature. In their textbook summarizing the state of knowledge, Sarmiento and Gruber put the range at 20-90%. We will note this in the revision.

Page 3011, lines 20-23: Is the improved upwelling (compared to MESMO 1) from increased wind stress?

It is most likely related to the new wind stress fields, which are stronger in part of the year than the annual mean winds used in MESMO 1.

Page 3012, lines 1-9: It seems that Si(OH)4 in the North Atlantic is pretty similar to the North Pacific (Figure 9a). I think it should be lower, as in Figure 1 of Horn et al. (2011). Is there a reason for this? You should show the observed values in Figure 9. The contrast between the northern Atlantic and Pacific basins does not show up in modelled Si(OH)4 (as observed) but it does in modelled Si:N uptake. Is the low Si:N uptake just due to high iron from dust in the North Atlantic? Perhaps a bit more detailed explanation in the text would be helpful.

As mentioned above, we will expand Figure 9 to include panels for observations. The difference in Si/N uptake between the two basins is indeed entirely due to the difference in iron input. The high iron input in the Atlantic lowers the Si/N take ratio. We will make this clearer in our revision.

Page 3013, lines 22-28: Considering that the modelled vegetation stocks do not include land use change, these seem a bit low. It looks as if the modelled boreal forest in Northern Asia is under represented. Is this due to a poor climate simulation there? Again, I am not suggesting the vegetation distribution needs to be changed - only that it should be considered when making further improvements.

The modeled stocks may be a bit low compared to observations but represent a slight improvement over Williamson et al. (2006), who first described ENTS. As ENTS is a (simple) model, not everything will be right. We can discuss the points made by Marsh et al. (2011) that precipitation in simple atmospheric models like the EMBM is poorly represented and continental interiors are too dry. This may explain the modeled boreal forest noted by the reviewer.

Page 3015, lines 1-3: This sounds slightly awkward to me. Maybe "By implementing the existing Fe code, two classes of phytoplankton, and a dependence of the Si(OH)4 utilization on Fe availability, the model is able to simulate key features of the marine silica cycle. These features include extensive Si(OH)4 limitation : : :" would be better. I think that you have not really shown that the mechanism behind Si depletion is due to low values being exported via AAIW. This still seems speculative, and while this might be something you could show in a model, I see no evidence that this is the case. Perhaps this is suggested by Figure 9c, but it is not very convincing. Consider revising the wording of the second last sentence in the summary.

As we note in the first paragraph, the reviewer seems to have misread this part of the text. We do not propose new mechanisms of the marine silica cycle. The mechanisms have been proposed by Sarmiento et al. (2004) based on their analysis of data. So there is evidence for the mechanisms discussed, and in fact they seem strongly supported by data at this point. We will discuss this point more clearly in the Introduction.

Page 3015, lines 7-9: Maybe something like ", which have recently received attention," would be better. Even that sounds a bit odd - maybe also consider revising your last sentence.

We will edit the text as suggested.