We thank the reviewer for their constructive comments on the manuscript. We have addressed all their points in a revised manuscript and in responses below. Reviewer comments are shown in bold text and our responses in standard text, with actioned responses in green.

The major changes mostly involve providing further justification for methodology choices (i.e. use of datasets), clarification of language choice and expanded explanation for discussed results.

We have also addressed minor points attributed to typos and referencing errors.

We believe these additions have improved the manuscript's clarity and hope the editor is in agreement.

**RC2**: Anonymous Referee #2, 19 Sep 2023

**Referee**/Response/Actioned Response

**This manuscript presents a useful addition to the literature, and should in my view be published after addressing the relatively minor issues identified below.**

**Line 142: "The only differences with respect to the iron cycle parameterization used in Ward et al. (2018) are then: (1) the dust field of Albani et al. (2016) rather than Mahowald et al. (1999), (2) a mean global solubility of dust-delivered iron of 0.244 % as opposed to 0.201 % (partly due to the overall lower dust fluxes of Albani et al. (2016) vs Mahowald et al. (1999), and (3) a small reduction in the scavenging rate scaling (0.225 vs. 0.344 in Ward et al. (2018)." This is helpful information, but requires more justification. Presumably these numbers resulted from model turning, in which case this should be stated, and the objectives of that tuning described.**

The reviewer is correct, these numbers arise from model tuning. The objective of this was to implement a more up-to-date dust field into the model, as the original rendition was tuned to data from 1999. The numbers presented are a result of this change in dataset, this has now been clarified in the manuscript.

**Line 151: "we do not attempt to calculate the fractional preservation of opal in accumulating sediments at the seafloor, but instead impose a simple benthic 'closure' term and reflect biogenic matter reaching the bottom of the ocean." What do you mean by 'reflect' here? Please describe this more completely, what is the fate of this silica?**

We thank the reviewer for pointing out this potentially unclear definition. The silica reaching the seafloor is entirely dissolved, thus the global ocean silica inventory in the ocean remains unchanged and is thus "reflected" from the seafloor instead of being buried. We have changed this word choice in the manuscript for clarification.

**Line 235: What are you using old WOA data? Would updating this make a difference to your tuning, for example with increased data at high latitudes?**

Yes we are using WOA data from 2013, this enables direct comparison to the original EcoGEnIE rendition (now added in the updated manuscript). When testing with more recent WOA datasets we found minimal difference in model performance.

**Line 245: "little further change occurred in biogeochemical indicators (oxygen, phosphate <1% change etc.)" Be more specific - max surface ocean concentrations, global ocean inventory….**

We thank the reviewer, this has been clarified to M-scores i.e. those plotted in Figure 1.

**Line 281: "Mean oxygen concentration produced by this iteration is also acceptable at 156 μmol kg-1, close to the 162 estimates" - please provide a citation for this estimate.**

The average for mean oxygen concentration has been taken from the used WOA dataset. We now mention this in the manuscript.

**Line 286 I don't think that "Overall, EcoGEnIE 1.1 captures the zonal contrast in phosphate concentrations between low and high latitudes" is a very good description. The contrast is in fact between the polar and sub polar regions, excluding the Arctic, and the the rest of the oceans. Low latitude implies latitudes around the equator, where the model performance is poor.**

The reviewer is correct, it has now been clarified here that the contrast is between the polar and the subpolar, and that the contrast at low latitudes (i.e. equator) is due to poor model performance.

**"The model-data comparison is also not strictly like-for-like, because in re-gridding higher vertical resolution WOA to the model grid, elevated subsurface concentrations become averaged into the re-gridded 'surface' layer.". Does the physics of GENIE allow for a meaningful mixed layer to form? If not, there will not be the barrier to nutrients being entrained into the ~80m top level and the comparison seems reasonable? Further, can this argument apply to phosphate, but not to silica, where the results are pretty good, and one needs to instead justify why WOA is lower than your simulation in much of the ocean? Expanding your argument would help people like me who are not very familiar with the consequences of GENIE's highly simplified physics.**

The gridded WOA observation data include observations from the top 80.8m of the ocean, so will include observations from below the true mixed layer when it is shallower (thus shallow regions such as the equatorial Pacific appear elevated in phosphate in the observations), thus the issue is likely a source of bias. This has been addressed in the manuscript.

The mixed layer is only represented in the model in terms of how the light available to the phytoplankton is calculated (see section 3.2.6 of the EcoGEnIE 1.0 model description paper). EcoGEnIE then works out what the chlorophyll concentration would be if it were mixed evenly across this depth, and then works out what the average light level should be across that depth with that level of chlorophyll. This scheme is not used for nutrients or biomass, thus the model does not have a "meaningful" mixed layer in that sense.

Both phosphate and silica have very good M-scores but indeed silica is visibly better. This can be attributed to the mechanisms in which they are regulated in the model, notably during deep ocean cycling. Phosphate is regulated by decomposition of sinking material (traded off with $O_2$, see response to trade-off comment below), whereas silica is regulated by the dissolution of sinking opal. As decomposition requires respiration ($O_2$ consumption), improving model performance of phosphate (e.g. intermediate Pacific) would likely push oxygen to less realistic concentrations. Silica is not influenced by this trade-off thus higher M-scores are achievable.

**"The spatial distribution of diatoms (all size classes combined) in EcoGEnIE 1.1 agrees with previous estimates (Tréguer et al., 2018)…" This is a key criteria for this manuscript, and needs to be expanded on. Currently I believe you only present results from your model here (please make figure captions more descriptive so this is not ambiguous) - this needs to be contrasted with other estimates or datasets in a more robust way, acknowledging the challenges around data availability. At present the verification of the distribution of the key new PFT**

**you have added to the model is "… agrees with previous estimates (Tréguer et al., 2018), with high concentration in the productive regions (e.g. equatorial upwellings, subpolar regions) and peakings in the Southern Ocean at ~ 1 mmol C m" within the results section, and a comment on the relative size distribution in the discussion. Where observation based comparison is simply not possible because of limitations to available observations, explain this to the reader.**

The reviewer is correct that observation-based comparison is simply not possible because of limitations to available observations, which is also referred to in the 5.2 section when discussing plankton recording techniques. We do indeed only present diatom biogeography from the modelling and have made this clearer in the figure captions (e.g. Figure 16's caption has been changed to clarify these results are all model-borne). We have also added further explanation with the paragraph discussing spatial distribution of diatoms that alludes to the difficulties associated with ecological datasets (e.g. Maredat) and that the more robust verification methods are through biogeochemical tracers i.e M-scoring.

**Line 377: "peakings" should be "peaking"**

Changed.

**What hypotheses are stimulated by the trade-offs seen in your M-scores? This behaviour is telling you something about the system or the limitation to the modelling approach - can you propose any suggestions?**

We hypothesise that the M-score trade-off between oxygen and phosphate is due to the limitation of the model. Phosphate concentrations are regulated by decomposition of sinking matter, which influences the extent of respiration and thus the amount of oxygen consumed. As previously mentioned, achieving more accurate phosphate concentrations would lead to elevated oxygen levels and would require modification of the $O_2$:P stoichimetry. It is also likely to be an issue due to insufficient ocean mixing with, $O_2$ not being replaced to counteract diminishing concentrations. We have added a sentence in section 4.1 and the supplemental to suggest the reason for this trade-off. It is worth noting that achieving accurate oxygen performance is a common issue amongst global biogeochemical models.

**General:**

**- ensure consistent formatting of references.**

We have used the EndNote output style file for referencing and ensured they are consistent with GMD's submission requirements.

**Minor specific points:**

**Line 51 "ration" to "ratio"**

Done.

**Line 598 sort out the reference and its citation in the text.**

The citation has been removed as this sentence was removed in a re-edit.

**Table 1. Define ESD in the caption.**

This is equivalent spherical diameter and has now been defined in the caption.