Response to Referees, 2nd round

Marco van Hulten et al.

24th January 2017

Dear editor and reviewers,

Thank you for your input. We have considered all your suggestions and adjusted our manuscript accordingly.

The manuscript comprises much originality in presenting the new very large Mn dataset of the thus far longest ocean section GA02 in combination with a novel developed first-ever world ocean Mn simulation model. This work was done by the lead author in the Ph.D. period and subsequent postdoc time in Paris, next to other projects and tasks. In those years the GA02 dataset was available, and by re-organising this data into the grid of the simulation model, a statistical comparison exercise was feasible.

Only recently the USA Atlantic GA03 dataset and even more recently the USA Pacific GP16 dataset have become available. This did allow to show qualitatively in the various graphics an indication as to how the simulation compares with these very new datasets. For a more rigorous quantitative statistical comparison there simply is right now no labour time and computation finances available to do all this extra work of converting the dataset into the model grid, and doing all the statistical calculations. Also and perhaps even more important, the simulation of OMZ regions is not a simple and straightforward task, and also would require the new findings of the Pacific GA16 section to become available in new publications, in order to be able to cleverly draft the right type of refined resolution model boxes, and the right type of model simulation equations. For example the recent articles of Hawco et al. (2017) and Ohnemus et al. (2016) do provide a wealth of new insights, that first need to be studied and contemplated upon, before beginning to make modifications to the simulation.

As a matter of fact it may well be that the only feasible approach is to make separate models only for each one of the OMZ regions, and run these off-line from the overall world ocean model. Also one should not underestimate that for a world ocean model, as this is, one cannot simply solve one problem in one ocean, without the major risk of making at the same time the simulation worse in other oceans. Also fine colleague Resing has kindly made his dataset available, but intends to first write his own paper on Mn processes in the OMZ region off Peru, before this dataset can be utilised in a more rigorous statistical intercomparison with the simulation model.

All this would require much extra work and finances for computing. The lead author now having completed the period for work on this and other projects, and starting a new job quite soon, simply cannot do this right now.

We relabelled the reviewers arbitrarily to A and B, because the 1 and 2 were interchanged in the merged report (where rather the report number was used as it were the reviewer number and Reviewer 2 happened to respond earlier this time than Reviewer 1).

Response to REVIEWER A (report 1)

General comments

> The manuscript is overall much better than the first version and, in my opinion, only requires a few minor adjustments before publications.
>
> The introduction reads substantially better than version 1 and, this time, really sets the stage to an interesting paper. One major improvement over the last version is the way the various experiments are used and presented. This is now much more logical and helpful. There are still a few sections where additional explanations would be desirable (see specific comments), but overall, the structure is good.

We have done our best in amending these issues.

> I still think that reliance on only the Atlantic data for assessing statistical model fit is a poor choice and this leads to some convoluted explanations in the main text, but this is no a major issue at this point. I respect this is the author's choice, even I believe it is a bad choice. I would simply suggest that the authors make it as clear as possible throughout the manuscript that all fits pertain to one single section in the Atlantic only.

This should now be clear throughout the manuscript. Otherwise see the above introductory response on reasons why it is right now not possible to do more statistics on data-model veracity than the GA02 intercomparison thus far.

> One remaining frustrating detail is that the authors seem to have replaced the word "realistic" with the word "accurate". Unfortunately, replacing a word doesn't fix the problem, which is to overly rely on highly subjective statements of quality. Please reduce this further by being more specific or more quantitative in your statements.

We have now realised that it is confusing to use the word accurate for both (i) the actual dataset of dissolved Mn as measured by author Rob Middag, and (ii) for the agreement of model output with this dataset. Upon ample consideration we decided to maintain the word accurate for the actual measurements, and we use the word *veracity* for the agreement of the model output with this dataset. In the first pages these concepts have been defined, *accuracy* is defined at bottom of page 4 lines 31–33, veracity is defined at the top of page 5 line 6. Obviously, these definitions also apply to the words *accurate* and *veracious*, respectively.

The precision and accuracy of the measurements of dissolved Mn of the new published GA02 section dataset are presented as exact as can be in the main text of the methods and the related Table 1 and Figures 1 and 2.

Thus far a statistical data-model intercomparison has been done only for the GA02 dataset, see Table 6 and Appendix A. The Reliability Index RI (Leggett and Williams, 1981) provides a relative comparison of the veracity between the various model runs. Here the *Reference* run with lowest Reliability Index 1.76 ± 0.08 is deemed to have the best veracity. However as mentioned deviations of two other runs *NoBio* and *OxidThreshold* from the value of the Reliability Index of the *Reference* run are insignificant. Only the two remaining runs *LowHydro* and *NoThreshold* have significantly worse Reliability Index, as indicated by bold print of these values. For the run *LowHydro* this is further confirmed by the significantly worse Pearson Correlation and RMS deviation, their values 0.64 and 0.60, respectively, also given in bold print.

Specific comments

> P5, l13, remove "on" from "on shipboard"

Done.

> P11, footnote (2). Just add that to the main text directly.

Done.

> P14, l24: "recently measurements" change to "recent"

Done.

> P15, l16: consider rephrasing this ": very soon the [Mndiss] reaches the typical background concentration", maybe " as Mndiss approaches near-constant deep background concentrations quickly such that the NADW plume is no longer discernable."

Done.

> Section 3.3, p17. Expand this section – this is interesting. Please provide an interpretation as to why Atlantic value don't change but Pacific results do.

We have added a Section 4.2.3 which discusses this more.

> P18, l15. Table 6 shows a reliability index of 1.88 for NoThresh, not 2.77.

Indeed, fixed.

> P17, l20: What is your definition of "reasonably accurate"?

We have included a definition on page 5 of the now chosen word "veracity". As it is not an exact notion, and in this context there is not such a thing as absolute veracity, we qualified this with an adverb. We tried to make it as clear as possible. The sentence has been improved as follows:

> We have shown that the *Reference* simulation gives a reasonably veracious representation of the effects of hydrothermal vents and the background concentration in the deep ocean.

> P18, l27: overstatement "only the Reference simulation is accurate". What is your definition of accurate?

We have now chosen to strictly use the word veracity for the model output, and this veracity is defined now on page 5. For the data–model intercomparison of only section GA02, the veracity is quantified by the three statistical entities Pearson correlation, RMS deviation and Reliability Index given in Table 6. For all the other data and including the GA02 section, there is qualitative data-model comparison in the various graphs of Figures 8, 9, 10, 12, 14, 15, as well as 16.

> P19, l4-5: I don't understand the logic behind that argument about "usefulness": "We think that coupling the model to the sediment would ... maybe not that useful, because the sediment source is large at shallow sediments, while most of the Mn burial occurs near the hydrothermal vents"

With “usefulness” we referred to the ideas that the model could practically be spun up and also that it would yield values close enough to the true value (or observations) that it may be useful for certain studies. We decided not to discuss coupling with the sediment in detail, and rewritten this (see Section 4.1).

> P19, l10-11: Does the model handle “small” shelf regions well? Can you make any quantitative argument about how well the model resolution handles shelves?

The sediment source was derived from Aumont et al. (2015). They note that “as a consequence of the relatively coarse resolution of ORCA2, the model bathymetry is not able to correctly represent the critical spatial scales of the ocean bathymetry. An example is the continental shelves, which typically have a width scale of 10–30 km”. An algorithm was developed to account for this, but it clearly still has shortcomings. The resulting parameterisation overestimates large shelf regions (and underestimates certain small shelf regions where oxygen is low). We removed the word “large”:

> [...] the low model resolution does not handle ~~large~~ shelf regions well.

> P19 ,l12: Why only report on Slomp’s maximum values?

Looking again at the literature, we remembered that the choice was more subtle than just taking this maximum value. We changed the text, and more clearly refer to literature values now.

> P19, l16: “out of proportion in some regions of the ocean”. Where is that specifically? Is that only in the Arctic, as alluded to in the next sentences?

It should be more clear now by mentioning the regions and adding references describing the same issue (page 20, lines 9–13 in marked up version).

> P19, l20-21: add a comma in “. . . Pacific Ocean, where. . . ”

Done (plus semicolon earlier in sentence for clarity).

> P19, l22: “In the East Pacific Ocean the California Current induces Ekman transport and hence equatorial upwelling”. Very puzzling bit of physical oceanography? Some references in support of that statement would be very welcome. I believe wind induces Ekman transport, not the California current. I’m also not aware of physical theories of equatorial upwelling that argue the California current induces it.

You are right. We repaired this.

> P19, l23: “upwelling from OMZ sediments”. Maybe “upwelling of water that has been in contact with OMZ sediments”

Yes, and we have further reformulated this part.

> P19, l24: “This is partly captured by our model”. Which part?

At the specified location (20° N, East Pacific boundary) there is an elevated value for dissolved Mn. But we rephrased this paragraph for clarity.

> P19, l25: “In the South Pacific Ocean this effect is more clear in the data of Resing et al. (2015) (Fig. 8a,b, East Pacific around 20S).”… and in the model?

Much more clear in the data than in the model, actually. We rephrased this.

> P19, l27-29: Better would be to provide a back of the envelop estimate of how much bias may come from not representing fluxes from OMZ sediments.

No measured Mn fluxes can easily be applied to a coarse model. Even if we have those fluxes, we cannot do a simple, reasonable calculation how much impact it would have on Mn concentrations, e.g. compared with other sources. The sediment source was derived from Aumont et al. (2015).

> P20, l21: “and especially at low latitudes”. Please substantiate this with a few sentences. The previous discussion was all about the Southern Ocean, not on low latitudes”

Done.

> P20, l23: What is “the most settling Mn”? Do you mean the particulate Mn fraction that contributes most to the sinking Mn flux is from biological particles?

Reformulated.

> P20, l28: include, not includes

Done.

> P22, l29: “because Mn redox does not depend on O2”. Rewrite sentence. What is “Mn redox”?

Done. “Redox” is defined at first use at page 2 in relation to the given Reaction (R1) equation. Redox is a very common word in research articles on Mn and other ‘redox’ chemical elements in the oceans, such as Fe, Co, nitrate/nitrite, sulphate/sulphide.

> P22, l31-43: “…For this reason we have not included a dependency on [O2] to the model.…” Consider rewriting these few sentences in a less convoluted way.

Done.

> P24, l6: remove “e.g.” – say what you mean in words instead.

Done.

> P25, l3: “for an accurate simulation of [Mndiss]”. What is your definition of accurate? Replacing “realistic” with “accurate” doesn’t remove the problem of relying on subjective statements.

We have defined “accurate” before.

> Appendix A, p27,l3: do you refer here as the “Pearson correlation coefficient”? please specify.

Done. It is indeed the Pearson correlation we use here, this is now stated as follows:

In addition to classical statistical indices (Pearson correlation index r, root mean square error RMS), another [...]

Appendix A – Table 6. Why are there only errors for the Reference case and not all cases? For comparison purposes, errors should be calculated on all cases.

This is not really needed but has now been added upon the request of the referee. As a matter of fact, the comparison is only with *Reference*. In Appendix A we explain that the error of only the *Reference* run is needed to decide on the sensitivity simulation's significance. We realised that we had not used the very last timestep (year 480 instead of 600) of the simulations for the statistics. Therefore we updated the statistics. The RI increased a bit for all simulations, so we decided to have another look at how close the simulations are to a steady state. Whereas at year 480 there was a decadal relative change of Mn content of 83 ppm in the surface (upper 100 m) and of 28 ppm below 100 m, this was only 24 ppm and 7 ppm, respectively, in the last decade (year 600 minus year 590). We shortly discuss this now at the beginning of Section 3.2. There are no consequences for the significances between the different simulations or any of the discussion.

Figure 7, caption. What do you mean by the word "by in "the by red" or "the by blue" lines? Probably remove this. Also, would be good to make these lines thicker on the figure. They are very thin, even when the figure is full screen.

We made the letters thicker, and remove/rephrased the colour referencing.

Figure 8: x-axis labels and sub-plot titles overlap. Fix spacing.

Done.

Figure 5 and 9. Choose a consistent name between GIPY5 or GIPY5_e.

Done. We now only use the name "GIPY5", and specify simply that it pertains the Zero Meridian part alone.

Figure 13, explicitly state in the caption if relative difference is (ref-low hydro)/low hydro or (ref-low hydro)/ref?

Done.

Figure 16, make the colored lines thicker

Done.

Response to REVIEWER B (report 2)

The manuscript by Hulten et al. has changed considerably in the revised version, with the two most important changes being that

- a biological cycling of Mn has now been implemented into the model, as a response to remarks by the reviewers. The cycling is parameterized as following the uptake and release of phosphorous from a global biogeochemical model, assuming a constant Mn:P ratio from Twining and Baines (2013). The modeled Mn distribution, as far as I can see, has no effect of phytoplankton growth in the model, i.e. Mn limitation is not included in the model. This is a reasonable first step, but should me mentioned in the model description.

We have added this.

> - mostly in response to the second reviewer, the paper now contains a much more detailed description of the manganese observations along the dutch Geotraces section GA02, including a discussion of the methods for these observations.

Yes, indeed that was our intention based on the previous reports.

> The inclusion of a biological cycling of Mn in the model is reasonable, and I think it strengthens the paper a lot. I have, however, a bit mixed feeling about the new focus of the manuscript on the GA02 section: Reviewer 2 suggested "showcasing these (unpublished?) measurements in more depth early in the manuscript would help build a sense of expectations with regard to what the model is expected to do or not to do. That could also be used as a 'roadmap' to explain how the paper is organized and why." To me this aim has not yet been reached fully, the observational results (3.1) and the modelling results, especially section 3.2 still stand side by side in a too unconnected manner.

We modified the text in order to improve the integration of model and data. In line with the recommendation of 'minor revisions', we kept changes to a minimum.

> One example for this is the elevated value of Mn in the DSOW overflow, which is visible quite clearly in Fig 7 and discussed over a few sentences in section 3.1. In the modelling part, this feature is never mentioned again, and indeed the color scale in Fig. 9 is chosen in such a way that it is not even visible in the observations anymore. My expectation is that the model does not reproduce this feature, and that is not even bad; it probably just highlights that the model is missing sediment resuspension, a locally inportant but probably globally unimportant process. In my opinion, often the most important information in model-data comparisons is where the two do not agree because here one learns about processes.

It does seem to be reproduced to some extend, even with our colour scale. Still, how well the DSOW features are reproduced is not of ultimate interest, because the resolution of our model does not describe the DSOW well.

> This is just one example, but my general impression is that the present manuscript does not integrate the observational and modelling parts enough and thus misses what the second reviewer had in mind when he suggested to focus more on observations. The model-data comparison could be made much more precise, and I would argue that the authors should try to do that in a second revision.

This issue is discussed at the beginning of our response.

> One aspect to improve is the "over-reliance on subjective statements to describe the simulations and on imprecise/loose statements" that was already mentioned by the second reviewer. I think the present manuscript still contains many too colloquial statements, but now also in the description of the observations. A few examples are that "very soon the Mndiss reaches the typical background concentration" when talking about NADW or "slightly elevated concentrations in the subsurface are also observed in ... AAIW .. but once again concentrations reach the typical background concentration". Both statements in section 3.1 would be much more precise with some indication where something happens. Likewise, "for a small part" later on the same page is not a helpful description. When comparing the NoBio with the reference run, it is stated that "NoBio generaly compares better with the observations in the Pacific Ocean especially with the US EPZT transect" Again, it would be helpful if the authors could be a bit more specific: Where is the improvement, and in what aspect? At the surface or at 500m depth?

We have improved a lot of these subjective statements. However, sometimes this was quite difficult for reasons mentioned above.

Minor comments

> Section 2.2.7: the value of 0.4 'derived by Middag et al. (2013)' should probably be $0.4 \cdot 10^{-3}$.

Corrected.

> Section 3.1: "Whereas: on line 13 should probably be "As"

Corrected.

> Page 16, line 17: maybe include "hydrothermal" before "forcing field"

Corrected.

> Page 20, line 23ff: To me this is one of the most interesting results of the new model runs including biology; but why just state this here without giving any numbers? In my first review I already suggested that it would be informative to have an idea on the relative magnitude of the sinking fluxes of authigenic and ofbiologically incorporated Mn; The authors would make that point much stronger when calculating e.g. the globally integrated fluxes of Mn from Mnox and from the biological compartment, maybe at 100m depth and at some depth deeper in the water column. Just a suggestion..

We have included 100 m flux plots, and discuss them in the paper.

> The authors might consider using the same software for plotting observational and modelling results; Fig 7 is made using ODV, Fig 9 using ferret, I believe.

We did consider this, and acknowledge that ODV and Ferret present different-looking plots. We do still use the two different programs, because there are different communities working on this study (modellers and observationalist) who use different tools. Furthermore, possibly, ODV is the best tool for interpolated observational data and Ferret is the best tool to make these kind of model–data comparison plots (see Appendix A for the description). In addition, Ferret handles the irregular ORCA grid. In short, it is a practical consideration why we use different software for plotting.