SP-2023-28 Submitted on 02 Aug 2023

Reply to the reviewers.

"Variability of manometric sea level from reanalyses and observation-based products over the Arctic and North Atlantic Oceans and the Mediterranean Sea" by Andrea Storto, et al.

We thank the reviewers for the careful reading, the encouraging comments, and the many suggestions to improve the quality of the manuscript. We report below a point-by-point answer to the Reviewer's comments (Reviewer in bold font, our reply in regular font). Additionally, please note the limits in terms of the number of words and the number of figures for this type of submission, which we already exceeded in the original version; therefore, we cannot add new figures/text but only improve/change/rearrange the original manuscript.
Finally, the GRACE dataset went through reprocessing, and we have replaced the previous dataset with the most recent one, which shows some non-negligible changes in the North Atlantic and Arctic regions. Additionally, the SLB dataset is also adjusted to account for the deep ocean correction that is not used any longer at the regional scale. With such reprocessing, the consistency of the three datasets has even improved, and a few critical issues (e.g. large Arctic trend) have been mitigated.

Reply to Anonymous Reviewer 1

Main comments

This is a short and to the point paper, but perhaps it is a bit too short: I am missing some details and information that I think would make it stronger and more informative/appealing, as at the end, I'm left with the question: is there one method really better than the others? The authors write that 'The results are intended to (..) guide users in the choice of the specific product, depending on the region of interest' (L282-283), but to me it is not clear what the choice should then be based on, as with this information it is not possible to pick a 'best' approach, or is there something I've missed?

We added a sentence on this in the last section (lines 300-305), also incorporating the comments from the second reviewer about the reliability of the data in the Arctic region, observational sampling, etc.

Uncertainties. There is very little attention to the spread in the results, and uncertainties are only sparingly mentioned or shown. For instance, Figure 1 (or any of the figures and most of the tables) shows no uncertainties, while this should be possible (?), given that for instance the GRACE dataset is an ensemble of 120 solutions. Including the uncertainties is essential to get a feeling for the consequences of using different methods in the manometric sea level in the different basins, and as it stands the three methods can only be compared very qualitatively.

Adding uncertainties in Figure 1 would decrease the legibility of the figure, already very busy. Note also that Table 2 (and related discussion) already contains (last column) the time-averaged uncertainty for each dataset and basin. In the revised version of the manuscript, we added the uncertainty bars on the yearly mean lines of Figure 1.

Comparison to the global mean/total sea level change. Is it possible in figure 1 to also include (a panel showing) the global mean barystatic change for the three methods? (or at least GRACE and SLB, given the argumentation in l161?). Now showing only the global barystatic from SLB in Fig 1 feels a bit arbitrary as the reader does not know how similar (or different) these global time series are. In fact, showing the total sea level change (not only the manometric) for the global mean and the basins might be interesting too for reference, especially since for instance l.200 refers to the total change?

As mentioned above, the length of the manuscript is limited, so we cannot add any more figures. The reviewer may refer to Barnoud et al., 2023b (the reference is available in the manuscript) for comparing the barystatic sea level changes from the SLB and GRACE methods. We added a sentence on this in the revised version of the manuscript (section 2.4)

Regional differences. Is it possible to include maps: how does the manometric signal vary spatially in these basins? I understand that time series are difficult, but the authors could for instance plot the linear manometric trend (mm/yr)?

This is an interesting point; however, as we mentioned above, we already exceeded the length of the manuscript and the number of total figures. Unless the editor grants us the possibility of adding another figure, to remain within the manuscript limits we cannot add a figure. We will comment on the spatial distribution in the revised version in a new sentence, adding, however, “(not shown)”.

Figures. Please, can the figures be constructed in a colour-blind friendly way by choosing different colours (figs 1&2) and/or line styles (fig 1)? I’d suggest to change the colour bar of Fig 2 into a gradual one (choosing one colour which gets darker for higher correlation), as the colours now make it near impossible to interpret this figure. (see https://www.nature.com/articles/s41467-020-19160-7 for reasons why the rainbow scale is not a good scale to use). Alternatively: wouldn’t it make sense to provide this fig3 information in a table format, so that uncertainties can also be included? Fig 1; Would it make sense to plot the linear trends in figure 1? (it may become too busy though). Fig 2; given that these correlations are mirrored, wouldn’t it make sense to only show the half matrices, as basically one only needs the three blocks in the upper left corner of each correlation plot. Fig 3; can uncertainties whiskers be included on the bars?

Thanks for the suggestion, we have replotted the figures in a color-blind-friendly palette. Regardingthe additions: figure 1 is already too busy, especially if we add the uncertainty bars. Figure 2:we now explicitly state that the correlation matrices are symmetric by construction; however, we already tried to plot half matrix only, and the plot is less aesthetically appealing than plottingthe full matrix. Figure 3: we have added the uncertainty.

Minor comments

Is there a specific reason for focusing on these three basins? The data covers the global ocean, doesn’t it?

We chose these basins as a compromise between geographical interests (basins of interest for the European communities and, thus, the Copernicus Marine Service, excluding however too small basins - Black and Baltic Seas, etc. - which won't be enough constrained by the observing networks used, and for which the recourse to regional modeling systems would be more appropriate). We added a short sentence on this in the first paragraph of the Summary & Conclusions.

L 85. 'assessing the multi-method mean signal' – I don't think this is done in the paper?I could only find this for the separate methods?

Thanks for spotting this inconsistency. Indeed, this objective was planned in a preliminary version but has not been treated in the present manuscript and is therefore removed in the revised version.

L184 'significantly different' – is this statistical significance?

The basins show many statistically significantly different metrics, but here it was meant in a more general (not statistical) sense, so we remove "significant" for clarity.

L186-188; 'except during the first and last years'? ; is it only due to the final year that the trend is this high? How sensitive is the trend to those first and last years?

Thanks for pointing this out. Indeed, the bootstrapping technique used to quantify the trend uncertainty removes part of the timeseries, and thus exactly quantifies the sensitivity of the trend to individual years. We added a sentence on this to explicitly point it out, in section 2.4.

L190 add a cross-ref to Table 3 here

Added in the revised version.

L200 'the global barystatic signal'?

Corrected.

l203 – unclear what 'the total trend' is: is this the total barystatic trend, and is it in the basin or the global mean? How can the trend in a basin 'explain' a total trend? (the other way around sounds more logical?)

"total sea level trend" means the SSH trend (manometric plus steric) as seen by altimetry, and not the global as the reviewer imagined. We clarified this point in the revised version.

L212 – 'generally': in the NA and Medi, the correlations between GRACE and other datasets are always lower than for the SLB-GREP combo, isn't it?

Thanks, you are right. We modified the sentence accordingly, removing the adverb "generally".

Reply to Reviewer 2

Comments

I have serious concerns on the calculations over the Arctic Ocean from the SLB method (primarily from the Argo data), with slightly lesser concerns about the North Atlantic study.
How do the authors calculate steric anomalies in the Arctic when there are only small numbers of floats in this region???? As someone who is currently working on analyzing individual Argo float data for a project and not just using analyzed grids, I can assure the authors that there are not sufficient observations in the Arctic to even begin to do the calculations they are attempting. Most of the analysis grids cut-off the data at 65° because of this. If there are "values" in the grid cells, they must be from a climatology or VERY limited data and extrapolation. Unless the authors can justify that there are sufficient Argo observations in the Arctic to support their calculations, I cannot accept that any calculation of ocean mass based on altimetry - Argo data is credible in the Arctic Ocean.

We thank the reviewer for this comment; indeed, the Arctic region is affected by large uncertainties for the SLB product, due to the poor observational sampling. Our approach for the revised version was to i) explicitly mention these limitations for the SLB product (the other two being less affected, as reanalysis bears information on the meridional transports and atmospheric forcing, and GRACE is not necessarily affected besides issues with satellite footprint geometry and leakage); ii) discuss in more details the uncertainty of the products and compare it between the products, as also suggested by Reviewer 1. Please note, that the focus here is not to show only the SLB product but to discuss the advantages and disadvantages of the three state-of-the-art datasets, for the selected basins. That is why we do not want to exclude SLB in the Arctic, but better mention and quantify its limitations at high latitudes.

There may be some validity in the using the ocean state estimate, but this is also limited by the altimetry data problems over the Arctic due to inclinations of the satellites, sea ice, etc. In the manuscript between lines 185-195, the authors point out that the trend in ocean mass using the GREP method (the ocean reanalysis steric correction) is 6.2 mm/year, compared to 2.5 mm/year for the gravimetry (or only a little over the global mean rate). They comment later: "Note that the GRACE-derived trend is likely too large, as it exceeds the altimetry-based total sea level trend of 2.9 mm yr-1, although the latter is characterized by significant under-sampling at high latitudes and ice-covered regions". I have to assume this later statement is referring to the GREP estimate, not GRACE. And if it is 6.2 mm/year, that means an enormous negative steric change over the Arctic (3 mm - steric = 6.2, means steric = -3.2 mm/year. This would have to be caused by either a cooling or salt gain, which doesn't really make physical sense based on observations of the Arctic warming and freshening. This does not encourage one to trust either the GREP or the ARGO-based SLB estimate, and I suspect for the same reasons -- there simply are not enough T/S measurements there to support the calculation. IMO, the entire Arctic analysis should be stricken because the data sets being used are not adequate to measure what the authors want. If they want to include it, they need to do a much better job of describing the limitations of the various data and models in the Arctic.

Please see above for the issue of the Arctic Ocean data reliability. We clarified this point in the revised version, but we still believe there is value in assessing and comparing the manometric

sea level in the Arctic, eventually pointing to the weaknesses (these datasets are however state-of-the-science datasets that are broadly used also for Arctic studies, regardless of the observational under-sampling).
Additionally, it is important to note here that we do not expect the budget to close at the regional scale and that large errors may affect any of the components of the budget at the regional scale (altimetry, grace, and Argo), especially over the Arctic where there are known issues with sampling both for the in-situ and altimetry datasets. The new version of the GRACE dataset over the Arctic shows a trend still quite large but not as unrealistic as before; masks are not exactly the same so the comparison with altimetry is only qualitative.
The large trend is from GRACE and not GREP, and we added explicitly that the dataset is not built to close the regional or basin-scale budget.

The analysis in the North Atlantic and Med. Sea is better, because the data can support this, but I am troubled by the fact the authors did not use the salinity data in the Argo-based method. They do this because of a small residual drop post 2016 that appears in global halsosteric sea level variations. While this is a legitimate concern for global studies, it is not for regional studies. There can be large, real halosteric (salinity) fluctuations that are balanced by a compensating temperature change. This is known as "density compensation" and is a common feature in T/S data where water masses are being mixed, there are large fronts, eddies, and deep convection -- all common features of the North Atlantic. I can assure you that if you look at the T/S grids (or profiles) you will see these features all over the Atlantic and Med. Sea and that they can be quite large -- tens of cm of halosteric/thermosteric sea level change, that when added cancel so the steric change is small. Such an event happened in the late 70s called the "Great salinity Anomaly" and based on what I have been observing in the Argo profiles, something similar has been happening over the last several years. I haven't seen anypapers on it yet, but my point is that the authors should NOT exclude salinity from theircalculations on regional estimates. Yes, there may a small global drift, but by ignoringsalinity they are eliminating signals that are tens of cm that will cancel some of the thermometric variations. These will not necessarily average to zero.I suggest the authors recompute their Argo-based SLB estimate including salinity, and just note that the trend may be a little off at the end of the record because of the apparent salinity drift.

Please note, that there has been a misunderstanding here, as probably the text in the original manuscript was not clear enough. Indeed, we only neglect the halosteric contribution for the calculation of the global mean (i.e. the barystatic). In all regional time series - on which the manuscript is based - we do consider the full steric signal, including both the thermosteric and halosteric components, for the SLB product. Therefore, we thank the reviewer for his extensive comments, and we are sorry for the misunderstanding, but these do not apply to the data used in the manuscript. In the revised version of the manuscript, we have now stressed that the manometric estimate based on the SLB method is calculated as the difference between the total sea level changes from altimetry and the full steric changes from Argo, including the halosteric variations. Now clarified in section 2.2.

My final major concern is on the analysis of the relationship with various climate indices. This section is so short, that I cannot fully understand how the relationships were established or if the climate indices were smoothed in any way. For instance, the NAO has a lot of short-term (month-month) variability, while the AMO has a large 60-year oscillation which will correlate with the trend over a twenty-year period. That's not really good evidence of a relationship. But not knowing exactly what was done, and

how the percentages were computed, I cannot judge this. Unless the authors choose to expand this with a more thorough description of the analysis, I cannot support it being included.

In the revised version we add some details on the calculation and some discussions on the climate modes, although the length of the paper is very limiting, and we cannot add many details. Data used for these calculations are monthly raw means without any low-pass filtering, like many other works focusing on the climate mode fingerprints on sea level (Pfeffer et al., 2022, and references therein). We have followed this strategy without arbitrarily filtering the data, in the multivariate regression framework. Details were added in section 2.4 and 3.

There are some more minor comments, all of which should be easy to fix:

1. On the discussion of the Boussinesq approximation: authors should add “cannot represent the steric expansion…” They can (and do) measure the non-global parts quite well.

Thanks for this point, we include it in the revised version to stress that only the global steric signal.

2. It is true that reanalysis models “make barystatic and manometric terms often unrealistic” is true. But in most cases they are also non-existent! This should be added.

This point is not clear. All reanalysis models are forced by atmospheric reanalyses, and the freshwater cycle implied by this forcing is not balanced, meaning that the barystatic and manometric terms are not realistic by construction (but existent). Maybe we missed the reviewer’s point; therefore, we prefer to leave the manuscript unchanged in these regards.

3. There are some state estimates that have begun to assimilate gravimetry and do have reasonable barystatic variations: ECCO_v4r4 is one. Rui Ponte recently used it for diagnosing the freshening of the ocean in a paper in GRL.

We added the correct reference for ECCO, as an example (the only one) of data assimilation systems exploiting gravimetry data and showing reasonable barystatic variations.

4. “A linear trend of 0.12 ± 0.03 mm yr-1 is added to consider the contribution of the deep ocean to thermosteric sea level changes (Chang et al., 2019).” This is a value for the global average, and appropriate only for GLOBAL steric estimates. This is driven primarily by deep warming in the Southern Ocean, with lesser signals in the North Atlantic and NO evidence (or data) for such a trend in the Arctic. Please remove this "correction" for these data and merely comment on the potential of deep warming signals that are not accounted for and give some ranges -- and not just a global value! In fact, the authors should be able to analyze this somewhat with their ocean reanalysis output. What does it say?

Thanks for pointing this out. We have removed the deep ocean correction in the revised version of the data and manuscript, and we now acknowledge the fact that steric signals from the deep ocean are not accounted for. However,we are not able to give an order of magnitude for the potential effects of the deep ocean warming at the regional scale. The Magellium team has tried to use the ECMWF reanalysis ORAS5 to estimate the deep ocean warming, but this turned out to be unrealistic in several regions, and therefore the deep ocean warming has not been accounted for anywhere. There is currently not enough data to do so. Ocean reanalyses are poorly constrained as well in the deep ocean. Any range of values given for the steric contribution of the deep ocean at the basin scale would be affected by large uncertainties, so we just neglect its effect.

5. “Explained variance, as percent R2 coefficient, is used to quantify how much of the regional signal is explained by the global barystatic signal due to fast barotropic motion.” R2 (based on squaring the correlation) is NOT explained variance unless the data being compared have exactly the same variance. In this case, they likely do not. Better to compare variances: PVE = 1 – var(resid)/var(orig), where resid is the original time series minus the global barystatic signal. This is really variance explained.

Thanks, we have corrected the way the explained variance is defined and calculated, following the reviewer’s suggestions. The results are slightly different, although also SLB and GRACE timeseries have been reprocessed, although most conclusions hold.