We thank both reviewers for their constructive and insightful comments. Our point-by-point response is provided below (blue font).

Reply to RC1 (Detlef Stammer)

The authors aim to examine the effect of assimilating GRACE data, along with other datasets, on state estimates produced by the project for Estimating the Circulation and Climate of the Ocean (ECCO). By way of comparing the results against a reference run that does not include any data constraints the authors claim that the ECCO optimization leads to large adjustments in bottom pressure (pb) fields at monthly and longer timescales largely upon assimilation of GRACE data. Another conclusion drawn is that a mean ocean mass constraint is essential for mitigating large imbalances in freshwater fluxes derived from atmospheric reanalyses (used as prior forcing) and for producing a realistic barystatic sea level curve. Inspecting the residuals, the authors also point to problems with the GRAC data that appear to be inconsistent with other information about the ocean circulation and its variability.

The subject of the paper is important, and ultimately the manuscript should be published. However, some mayor shortcomings must be remedied first. First and foremost, the experimental setup is flawed in that only one assimilation experiment constraint simultaneously by many data sets is being used to pinpoint the influence of a specific data set – the GRACE data - by comparing the results against a reference run in which no data were assimilated at all. Obviously, this cannot work and conclusions drawn are not backed up by the results shown. This holds even more so as the GRACE data kick in at the same time when Argo data become available and so the solution will be impacted in all its aspect by both data sets and all other data as well. Ideally, exactly the same set-up should be run twice with and without GARCE data involved. At a minimum, what the authors need to do here is compare results against a previous optimization run (Release 4 described in detail by Fukumori et al. (2019) which included almost all data as constraints, but not the GRACE data. Moreover, many important details are missing in the paper regarding the approach but also the assimilation experiment itself that need to be added. More detailed comments are provided below. Once all those have been addressed satisfactorily I belief that the paper can become a significant contribution.

We have included analyses of a previous ECCO solution (Version 4 Release 1), which does include all the Argo data but not GRACE data. The results indicate no substantial impact of Argo data on bottom pressure adjustments. See new Figure 8 and related discussion in lines 248-254. As an extra check, we also carried out single iteration tests comparing cases with all data including GRACE vs. all data excluding GRACE. The results were consistent with the impact of GRACE constraints on the optimization. Further details of the revisions implemented to address the major points raised by the reviewer are provided below.

Major Comments:

1) The authors use a GRACE and GARCE-Follow on data product. But no information is provided as to what they do to this data set by themselves. This relates especially to the question if a trend was removed, with which temporal resolution the data are available and being used as constraint. Also, the source of the uncertainty shown in Fig. 1a is not specified.

Treatment of the GRACE data and the source of the errors in Fig. 1a are detailed in the references that were provided in the original version, but we have added text to clarify the points raised by the reviewer (lines 71-73, 90-93).

In Section 2.2 a different GRACE data set is being introduced that apparently is being used for comparison but differs from the data set assimilated. Would be useful to expand on this; e.g., why was the GRD aspects in the assimilated although the model cannot deal with this?

The comparative spatial analyses in the paper are based on the same GRACE dataset used to constrain ECCOv4r5, as stated in lines 105-108. The Wiese et al. (2024) barystatic sea level curve was used because the one used to constrain ECCOv4r5 was not publicly available anymore at the GRACE Tellus repository. This is noticed in lines 110-112. Finally, the product from Landerer and Wiese (2025) considers only GRD corrections. These corrections were not available at the time ECCOv4r5 was derived, thus uncorrected data was originally used in the assimilation. This is also clarified in lines 119-120.

2) What is the effective spatial resolution of the GRACE data and how is this incorporated into the cost function? Are coastal regions excluded in the cost function or is the model constrain by the GARCE data also over shallow regions?

In the ECCO estimate, the effective resolution of the GRACE data is approximately 300 km, achieved through the method described in Appendix E of Forget et al. (2015). All GRACE data, including those near coastal regions, are used to constrain the estimate. This information is provided in lines 91-93.

3) How is the Antarctic ice-shelf melt incorporated into the model. The text refers only to a heat transfer coefficient which is being adjusted. But what happens to the freshwater? Is this being use as freshwater flux as well?

The adjusted heat transfer coefficient affects the freshwater flux (and heat flux) associated with the ice-shelf melt, but not salt flux because the salinity of ice-shelves is assumed to be zero. This is clarified now in lines 96-97. The freshwater flux that is associated with melting ice-shelves is also mentioned in lines 269-271.

4) The handling of spurious changes resulting from the Boussinesq model physics mentioned only in passing in Figure caption 10 needs to be explained in detail in the text as this is an important aspect that needs to be understood to appreciate the results, e.g., with respect to the global barystatic sea level curve.

The changes resulting from the Boussinesq model physics are accounted for by applying the so-called “Greatbatch correction”, which is a temporally varying, spatially uniform sea level (mass) adjustment that ensures mass conservation. Reference to the Greatbatch correction has been added to the text (line 271).

5) The existence of earthquake imprints in the GRACE results is mentioned many time throughout the paper. I suggest discussing this early on once to avoid frequent repetition.

The effects of earthquakes impact many of the figures in the paper, starting with figure 1. It is thus necessary to mention them throughout the paper. We do introduce the issue at first mention when discussing figure 1 (lines 75-77), and we have tried changes to the text to avoid sounding repetitive (lines 143-144, 304).

Minor Comments:

1) Line 114: suggest referring to later discussions of rms difference to make this text here a bit more quantitative.

   Statement refers to differences in standard deviations, and we would rather not confuse it with the rms differences discussed one paragraph below in relation to figure 2. No changes were made.

2) Line 117: eddy variability in GARCE data? What is the effective resolution of GARCE data (see above) and which eddies are resolved in this data set? I guess what is meant here is that GRACE observes a field that also contains the effect of eddies; but that individual mesoscale eddies are not resolved in the data set. In contrast the model does not resolve eddies and does not show the effect of these unresolved eddies.

   The reviewer's guess is indeed what we meant. In lines 134-140, we have changed the wording and added more explanatory text in reference to Zhao et al. (2021), who describe at length how intrinsic processes map into GRACE effective resolutions of around 300 km. Intrinsic variability is now explicitly mentioned when discussing figure 1a (line 78-79).

3) Line 118: " … as discussed in relation to Fig. 1a …" I am not sure where this text points to; at least I could not find it.

   This referred to the discussion in lines 74-76. The discussion has been modified and expanded, also to address the comments from the other review (lines 134-140).

4) Fig. 3 and related text: the expression "cost" is being used here without introducing it. I find this term misleading and suggest using a different word, such as normalized quadratic misfit. Cost specifically points to the cost function that was not even mentioned before.

   We do introduce the cost function in section 2.1 (lines 63-69) and describe it as the reviewer notes. We have changed how the "costs" are introduced, linking them to the normalized quadratic misfits and the cost function (line 163).

5) Line 139: change "converging" to "converged"

   Done.

6) Related to Fig. 2d: How realistic is the GRACE error provided?

   GRACE errors are not involved in Figure 2d. Otherwise, the GRACE data errors are provided in Figure 1a and discussed in section 2.1 (lines 68-81) and references therein. There is no ground truth against which to assess "realism" of values in Figure 1a. In addition, those

values include representation errors and not strictly data noise. In any case, the fields in Figure 1a were the basis for the weights used in the ECCOv4r5 GRACE constraints and are used consistently here to evaluate cost values and other aspects of the ECCOv4r5 solution.

7) Line 143: “Peak value of distributions”? I suggest rephrasing the text.

   Text now reads “histograms display peaks at values of ~0.45, 0.65” (line 169)

8) Line 152: I suggest changing “observations” to GRACE data, as these are also estimates and not observations.

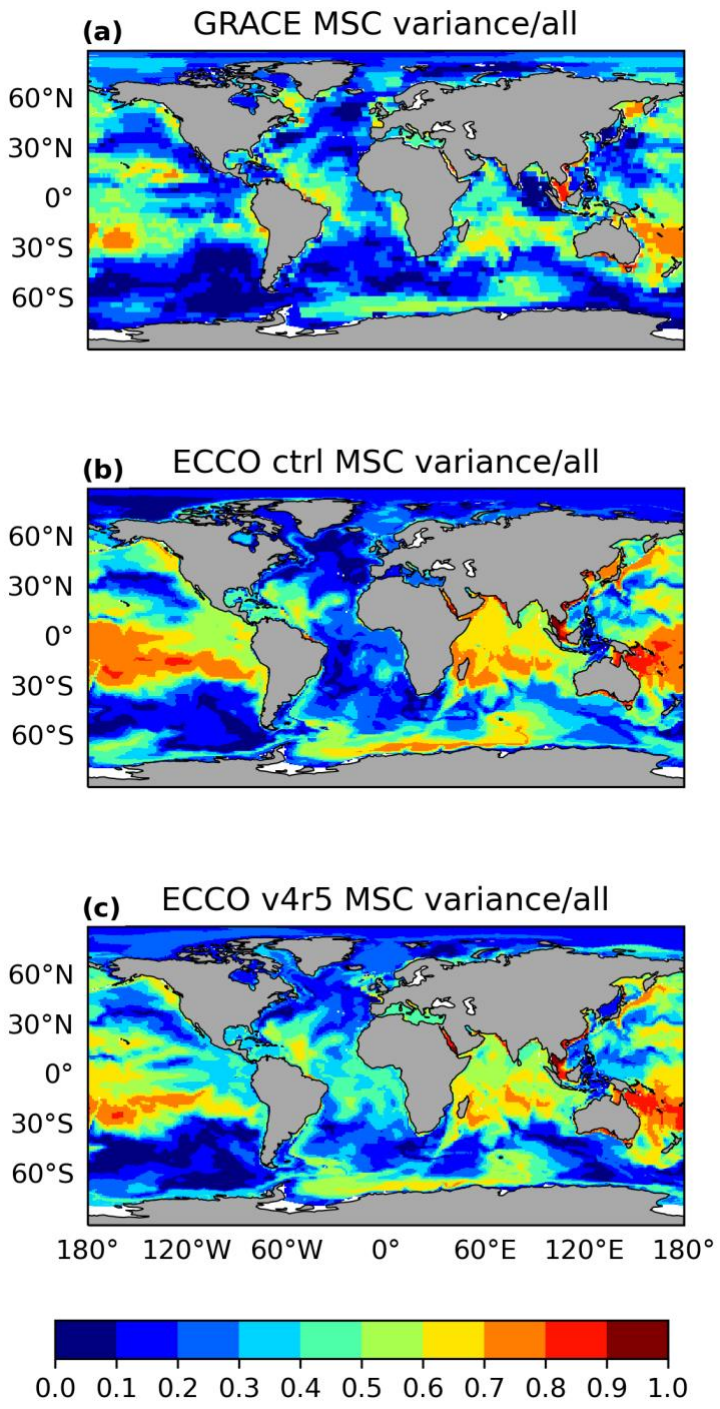
   Done.

9) Fig. 4: Details hard to see, specifically those situated in shallow regions; holds also for Fig. 6 and possibly others. Which period is considered here and in other figures?

   We have redone Figures 4, 5 and 6 as six panels in a 3 row by 2 column format to improve readability. All results pertain to the GRACE, ECCO overlapping period, as stated now in the figure captions.

10) Line 164: Seasonal cycle main component? Is this consistent with the figure shown here? What is the percentage of the seasonal cycle relative to the total variance?

    The text states the seasonal cycle is “a main component”, which of course does not imply it is THE main component. In any case, the statement is consistent with our results. Plots showing the ratio of mean seasonal variance to total variance (with trend removed) are included below for completeness.



11) Line 185: Contrary to what is stated I still see a significant seasonal cycle in these principal components 3 - 5.

Yes, there is seasonal variability in modes 3-5, but there is “more nonseasonal variability” than in the other modes, particularly after 2002. We have tried to clarify the statement in line 214.

12) Fig. 7: you detrended over a longer period; yet the GRACE trend is still prominently present in mode 3!

Yes, PC3 in Fig 7c suggests the impact of trends in GRACE data on the ECCO optimization after 2002 and this feature is made more explicit in the revised text (lines 214-216, 229).

13) Caption of Fig. 7 and others: Change ECCO fields into ECCO pb fields.

Done.

14) Fig. 8: bowls shape probably due to model drift/adjustment. Should be discussed.

The bowl-shaped behavior in Figure 8a refers to the ECCOctrl and the ECCOv4r5 minus ECCOctrl curves and is not apparent in the ECCOv4r5 curve. It is also not present for all curves in detrended results in Figure 8b. This indicates that the optimization produces considerable changes in the trends, consistent with Figure 4. The original discussion in lines 201-204 has been modified to make these points clearer (lines 234-239).

15) Fig. 9: The figure is misleading; largest adjustments happen over all shallow region and ACC, not in the Arctic as stated. All shallow regions will be affected by Argo data. So, what the impact of GRACE is as opposed to other information is entirely unclear. I suspect there in no impact of GARCE at all.

Largest values in the Arctic refer to areas near Franz Josef Land and Beaufort Sea, aside from Arctic shelves. The statement has been clarified in lines 258-260. We disagree with the assertion by the reviewer that “all shallow regions will be affected by Argo data”. There are, of course, almost no Argo data over shallow shelf regions or other regions (e.g., inside the Arctic, Japan/East Sea) with largest adjustments seen in Figure 9. Although changes in the deep ocean can affect bottom pressure in shallow regions, the possibility of Argo-induced adjustments over the deep ocean affecting the Arctic shelves, for example, is not plausible.

16) Lines 253 and Fig. 11: what about the jumps in PC1 and in PC5? I suggest putting in the times of the earthquakes as I suspect that they are the reason for those. What happens in mode 5 after 2017? Impact of GRACE-FO?

We have added lines for the two major earthquakes in Figure 11a as suggested. We also do not plot PC curves over data gaps for clarity. The text calls attention to the impact of earthquakes on modes 1 and 2 (lines 297-299) and we added mention of large change across the GRACE and GRACE-FO for mode 5 (line 301).

Reply to RC2 (Don Chambers)

Like Professor Stammer, I also feel this paper is worth publishing after some revisions are made. My major concerns (as indicated in the point-by-point comments below) concern the lack of some quantified statistics in the discussion. The discussion throughout could be improved by giving explicit values of differences (e.g., X% of the oceans have a change > 1 cm RMS).

We have added more statistics as requested. See details in the response to the point-by-point comments below.

It is a shame that the authors do not have an assimilation run on hand that was done in the same framework without assimilating GRACE data. I understand the reason why (computation costs are expensive), but I do encourage the authors to consider Dr. Stammer's suggestion on a potential way forward by considering an older run (in a slightly different assimilation framework) and comparing for that specific period around 2002 when GRACE data is introduced. This would provide the best evidence that it's not Argo profile measurements that cause the change at that period.

See the response to RC1. The comparison with Version 4 Release 1 (which uses Argo data but not GRACE data) is now discussed in the paper (see Figure 8b and lines 248-254). Results consistently point to the importance of GRACE data constraints in causing the enhanced adjustments of the bottom pressure fields after 2002 seen in ECCOv4r5.

With that said, I do find Figure 8 and 9 compelling evidence that it is GRACE that drives the improvement. However, I note the evidence that suggests this is the gap between the end of GRACE and the start of GRACE-FO, but this is not explicitly discussed in the section (see Point 9 below). I encourage the authors to take another look at this particular time-period and perhaps do some additional analysis on it to tease out the GRACE improvements further with the data they have.

We agree with the reviewer and have added more discussion of the statistics over the gap between the two gravity missions, (see response to Point 9).

Point-by-Point Comments

==========

1. Although the figure captions clarify that global means are removed before computing Pb variance, this should be discussed more in the text as to why this is done – i.e., there is a ± 1-2 cm variation in global mass (barystatic sea level), primarily with seasonal periods but also a trend. By removing monthly global means, the authors are focusing on the dynamic ocean mass variations, not geodetic. This information is glossed over in the introduction and data description but should be discussed.

   We have added more reference to these points in the introduction (lines 44-46) and at the top of section 3 (lines 122-124) as suggested.

2. Lines 115-120: "*In areas of western boundary currents (e.g., Agulhas retroflection, Gulf Stream extension, Argentine Basin), GRACE fields look more energetic than both ECCOv4r5 and ECCOctrl, consistent with the presence in GRACE of eddy variability that cannot be represented in either ECCOv4r5 or ECCOctrl, as discussed in relation to Figure 1a".* Some more discussion is warranted here. For example, the authors should reference previous works (e.g., Fu's studies in the Argentine Basin) indicating there are known and large variation is barotropic sea level (Pb) at mesoscales in these regions. While the ECCO resolution can't capture the magnitude of Pb variance due to the resolution, GRACE does capture more of it. I doubt GRACE is capturing the full magnitude or resolving the mesoscale signals, but it does likely have information there that ECCO is not capturing. It should also be noted that the GRACE data is explicitly down-weighted in these regions (see Figure 1a), which may explain why it has limited influence on ECCO there.

   We added more discussion with citation of a few relevant papers describing effects of eddies on barotropic sea level and noting the relation to the lower weights (larger errors) in Figure 1a (lines 134-140). Some of the changes also address the comments from the other review.

3. Some quantification should be added in the discussion around Figure 2, especially Figure 2d. For example, what are the values of RMSD in certain areas (e.g., the Weddell Sea). The color bars make it difficult to see a difference between 2.5 and 1.5 cm, for instance. It may be that even though the change looks small, it still may be significant. This is shown to some extent in Figure 2d, but the authors need to guide the reader a little more on what the negative vs. positive values mean – e.g., negative means the assimilation pulled it closer to GRACE while positive pulled it away. Also, please add some statistics: what percentage of the ocean improved? By how much? What was the average RMSD increase in smaller areas where there is a larger change.

   We added more guidance on the meaning of the negative vs. positive values in Figure 2d and provided more information on the statistics of those values (lines 151-154 and Figure 2 caption).

4. I understand the "Model Cost" concept, but I wonder if expressing it int erms of percent variance reduced/explained (or a similar metric) would be understood by a broader audience. Remember, this is not just being written for modelers or experts on data assimilation.

   We understand the concern by the reviewer, but we chose to stay with presentation of cost values as these are directly relevant to how the optimization works and provide a clear interpretation of its results, in relation to how the data is weighted. Percent variance reduction is a valid but broader metric. In any case, information on how much more data variability is explained by ECCOv4r5 vs. ECCOctrl is implicit in Figure 2d, and we think including percent variance explained statistics would be somewhat redundant.

5. I notice there is no discussion of correlation. My first thought reading this is does the temporal correlation in each grid improve with assimilation? And by how much? I suggest adding some analysis on this.

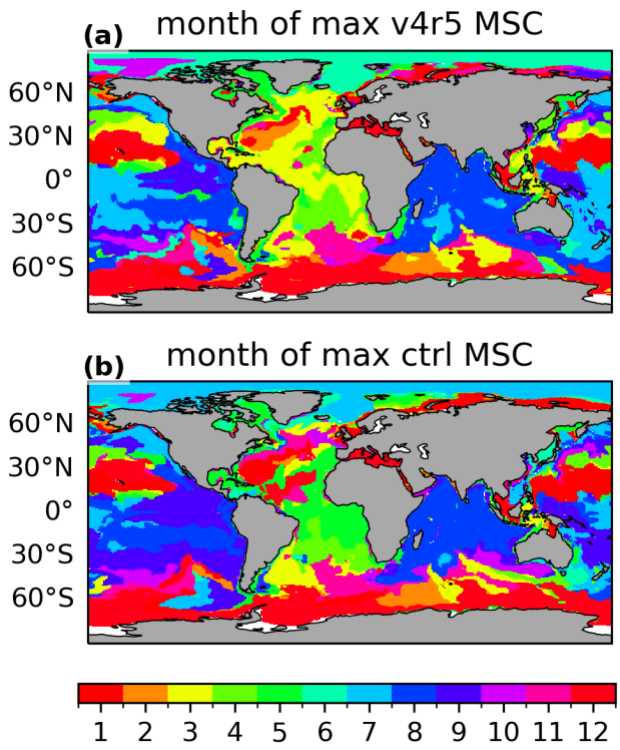
Yes, assimilation improves the temporal correlations by ~ 0.1 on average. This is now stated in lines 154-155.

6. One aspect of the trends in the control run that are not mentioned but are quite obvious to me is the negative trend throughout the Atlantic much of the Southern Ocean and positive trends throughout the Indian and Pacific Oceans north of north of about 60°S. This indicates a relatively large exchange of mass between those basins, enough to cause a change of ± 6 cm over the 20+ year period! This is indicative of a large-scale drift in the control run which is fixed via the assimilation. While there are some mentions of changes in the Southern Ocean, I feel this should be more explicitly discussed. In some respects, it's a bigger change than some of the more localized differences. That "mode" does seem to be captured in EOF3 (Figure 7). One can especially see the change in the system in that mode when GRACE is introduced. I don't see this is even discussed in that section! I think it is one of the more obvious impacts starting around 2002!

   Looking at Figure 4b and 4c, generally the positive and negative patterns noted by the reviewer in the Atlantic/Pacific/Indian Oceans for ECCOctrl (actually mostly blue, i.e., negative in the Indian Ocean) remain similar in ECCOv4r5. Aside from the Arctic, the most conspicuous differences between ECCOctrl and ECCOv4r5 occur at the high to mid latitudes of the Atlantic sector of the Southern Ocean. We have modified the text to better highlight these changes (lines 182-184) and bring up the connection with GRACE data when discussing Figure 7 (lines 214-216).

7. For the annual analysis, I would be interested in seeing a phase analysis. I know the authors use a climatology, so a phase value is not a direct estimate (unlike with a sinusoid fit). However, they could look at the month of the peak value and see if there is any evidence of a phase shift from the control to the assimilation run. Looking at the EOF analysis, I think there will be and it would be an interesting thing to show.

   As the reviewer recognizes, harmonic analyses are really needed to address phase issues properly. The alternative examination of the month of maximum value can be ambiguous, but we have produced such a plot for the ECCOctrl and ECCOv4r5 fields (see below). The results suggest that the optimization can lead to changes in phase of the mean seasonal cycle. Given the tentative nature of the analysis, we chose not to include a new figure in the paper and just mention it in a footnote in line 197 (see bottom of page 9).



8. For Figure 8. Please add some statistics of the actual values with GRACE obs and without. It will be a small change, maybe 1 cm RMS, but worth noting with a quantified metric. Same for Figure 9 – What's the percentage of the ocean where changes are > 0.5 cm? Between 0 and 0.5 cm? etc.

   Relevant average statistics for Figure 8b and Figure 9 have been added to the text (lines 241-243, 257).

9. One of the important periods for assessing the impact of GRACE is for the year when GRACE observations ended and before GRACE-FO came on-line. No other change in the observing system occurred then (unlike before 2002 with the introduction of Argo profiles into the assimilation as well). It is clear from Figure 8 that the updates to the control run Pb return to the pre-2002 levels, then move back up in 2018 when GRACE-FO is used. I see no discussion of this at all, but to me, it is the clearest indicator in the analysis that GRACE is responsible and not another data set.

   We have specifically mentioned statistics of bottom pressure adjustments over the gap period between GRACE and GRACE-FO missions (lines 241-243) and highlight their relevance in terms of supporting the impact of GRACE data in ECCOv4r5 (lines 244-245).