We thank the reviewer for the positive evaluation and valuable comments. We provide a point-by-point response to each comment, with reviewer's comments in black, and author's responses in purple. The modified manuscript text is shown in quotation marks and italic, with additions in bold.

This work compares different estimates of ocean mass contributions to sea level rise. These contributions are themselves derived from different estimates of contributions to ocean mass (ice sheets, glaciers, land water storage) which are propagated to SL fingerprints using the SLE. This work in itself, and a comprehensive documentation of resulting ocean mass trends and discrepancies between estimates at the regional level, would deserve a publication. The authors also derive uncertainties on ocean mass trends which they separate into temporal uncertainty (the amount of uncertainty coming from the natural variability in records), spatial-structural (coming from the fact that the position of sources is not exactly known) and intrinsic (uncertainty in the data itself, the way we measure it for example). This represents a large part of the paper (methods are well documented and the results well presented).

Thank you for these positive comments.

My main concern about this paper is that the representativity (or accuracy) of these uncertainties is not discussed, despite what appears (to me at least) to be inconsistencies across datasets.

Thank you for calling attention to the lack of discussion about the uncertainties representativity. We have incorporated your remarks below, leading to a better discussion of the uncertainties.

I'll try to give a few examples below:

* Section 3.1 and Figure 2 are dedicated to the noise model selection. No information is provided about the goodness of fit of the selected (optimal) noise model. As far as we are told, all models could largely fail at representing the variability in the records (I'm pretty sure this is not the case, but please provide a metric). This could help with the interpretation of Figure 2 where discrepancies between noise models fitted to different datasets are striking: to me this means that the datasets are unable to observe the same processes, even between two GRACE based datasets.

The goodness of the fit was assessed by using the modified Bayesian Information Criterion, as explained in lines 151-153. We have clarified this by adding the following information to the methods section 2.2.2 and the histogram as Figure A2 in the Appendix:

"Since these criteria are relative values, they can not be compared between different data sets. Thus, we compare the criteria of different noise models for each data set and each grid point separately. To select the best noise model, we compute the relative likelihood of BIC_tp, and select the model with values smaller than 2 (Burnham & Anderson, 2002; Camargo et al., 2020). Note that all noise models reasonably capture the variability of the time series (Figure A2), as their scores are always within a similar range. "

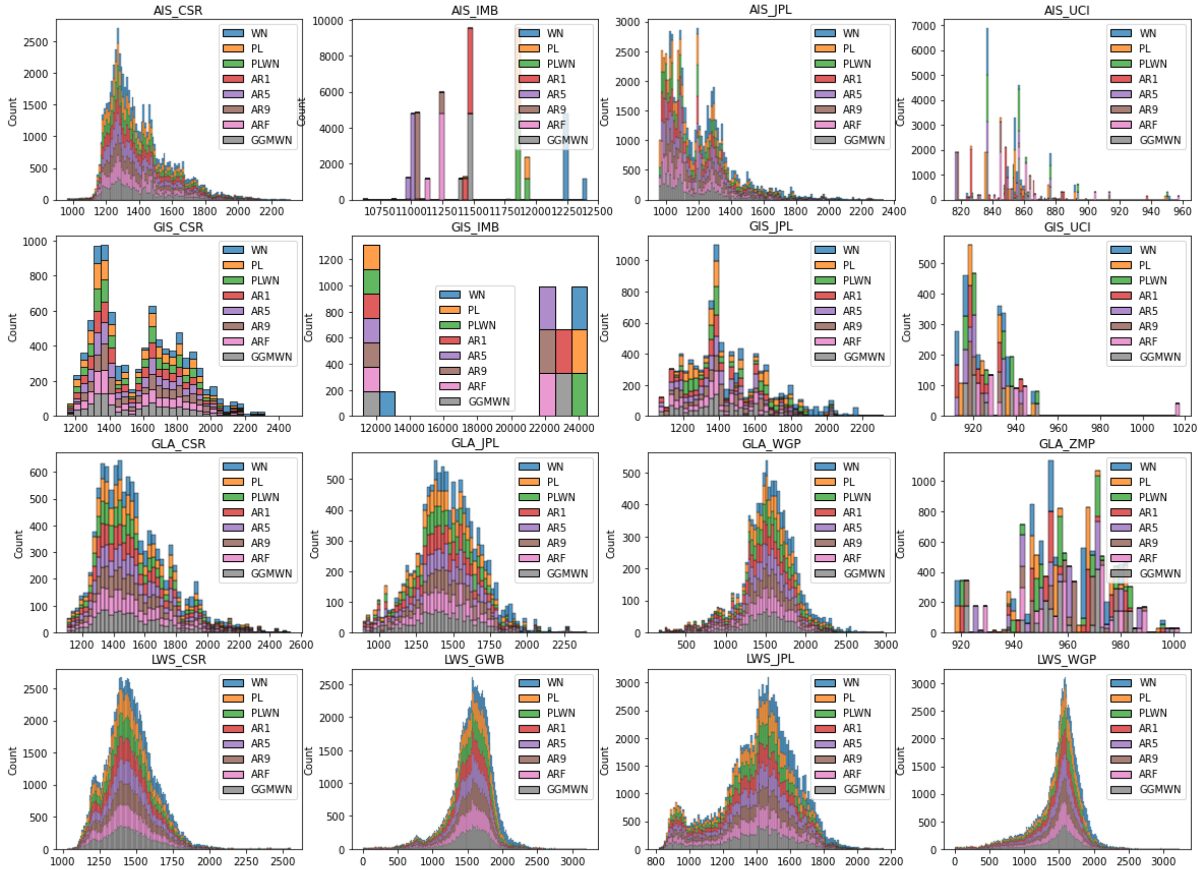

Figure 1. Histograms of the BIC_tp score for each dataset, used to select the optimal noise models. x-axis shows the BIC score, and y-axis the number of grid points (count). Note that all models have scores within the same range, showing that no model fails in capturing the signal of the observation. However, some models have a slighter lower score, and those will be the ones selected as the optimal noise model. (included in Appendix as Figure A1)

A synthetic table of datasets resolution/content could be useful here (rather than having the refer to Appendix A).

We have added an extra column to Table 1 in the main text, which now includes the information of the spatial resolution in the dataset (see table below). However, we prefer to keep the more detailed information in Appendix A, for clarity.

* Section 2.2.2 describes the uncertainty propagation rules used in the paper. I'm concerned by two points here:

1. When considering spatial-structural uncertainties the authors scale the fingerprints to 1 mm/yr amplitude to avoid too much spread across an ensemble of only 4 members. How much uncertainty reduction does this scaling provide? This should be documented.

We perform the normalization to isolate the uncertainty due to the spatial pattern of the mass change which is fed into the sea-level equation. If we would take the unscaled standard deviation of the datasets instead, the result would represent more than just uncertainty due to the spatial pattern because each data set has a different magnitude.

We illustrate the difference between normalizing and not normalizing in the figure below (right panel is Figure 4 of the manuscript). This shows that the normalization highlights the differences close to the mass change source, while without the normalization the differences in the far field are highlighted, as a result of the different global mean values in each dataset.

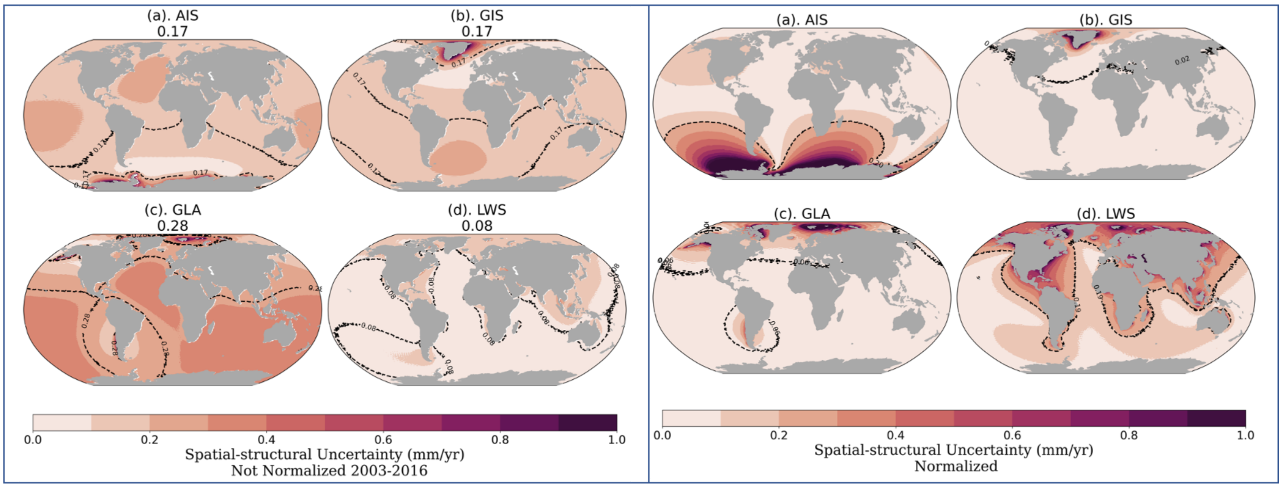

Figure 2. Structural uncertainty without the normalization (left) and with normalization (right, figure 5 in the revised manuscript).

We have rewritten the Methods Section as follows:

"The second uncertainty we consider is the spatial-structural uncertainty (Figure 1b, right column). Studies that combine a large number of datasets often base the structural uncertainty of an estimate on the standard deviation over the individual datasets in relation to the ensemble mean (Palmer et al., 2021; Cazenave et al., 2018}. However, the small number of samples in our study (4 estimates for each contribution) could lead to unrealistic structural uncertainties when simply based on the standard deviation, as individual outliers could bias the ensemble mean. Instead, we compute the spatial-structural uncertainty by estimating the standard deviation based on the normalized fingerprint for each contribution. First, we use the trend of each contribution to compute sea-level fingerprints normalized to 1 mm/year of global mean SLC. ***The aim of the normalization is to isolate the effect that the mass source distribution has on the fingerprint pattern. In other words, it removes the influence that the different central estimates (mean) have on the spatial standard deviation.*** *We then take the standard deviation across the four (normalized) datasets for each mass source contribution, which leads to four normalized spatial-structural uncertainties reflecting the uncertainty associated with the different spatial resolutions of the datasets."*

2. Regarding the intrinsic uncertainty propagation, the authors use a no-covariance hypothesis. There are many ways error covariances could affect GRACE/GRACE-FO measurements (instrument ageing, operation mode switches). I understand the scope of the paper is not to revisit GRACE error characterization but this could be mentioned ?

Indeed, instrument ageing and switches on operation mode will influence the signal of GRACE/GRACE-Fo measurements. However, the intrinsic uncertainties discussed here represent the errors in the monthly GRACE gravity field solutions, arising from measurement, processing and aliasing errors. We have added a comment to clarify what the intrinsic uncertainties represent.

These errors will exhibit covariances, which were not included previously, as pointed out by both reviewers. We have therefore modified the intrinsic uncertainty computation to now include error covariance. The intrinsic uncertainty is propagated by perturbing the time series with random noise multiplied by the uncertainty for 1000 times. We then compute the trend for each time series, and use the 95% confidence interval of the distribution as the intrinsic uncertainty. We have adapted the method section accordingly.

"The final type of uncertainty considered in our assessment is the intrinsic uncertainty, which represents the formal errors and sensitivities in the measurement system and needs to be provided with the observations/models by the data processor/distribution center. The intrinsic uncertainty was only provided with the JPL and IMBIE datasets. For all other cases, the uncertainty budget does not include the intrinsic uncertainty. ***The uncertainties provided with the JPL Mascons represent the scaling and leakage errors from the mascon approach (Wiese et al., 2016), and, over land, are scaled to approximate the formal GRACE uncertainty of Wahr et al. (2006). The latter represent errors in monthly GRACE gravity solutions, encompassing measurement, processing and aliasing errors (Wahr et al., 2006).*** *Note that, while the mascons have been corrected for mass changes due to the glacial isostatic adjustment (GIA) with the ICE6G-D model (Peltier et al., 2018), the intrinsic uncertainties of JPL mascons do not represent the uncertainties from the GIA correction, which can be large depending on the region (Reager et al., 2016; Wouters et al., 2019). For example, the choice of the GIA model used for the correction could lead to uncertainties representing 19 percent of the signal in Antarctica, but less than one percent in Greenland (Blazquez et al., 2018). Since estimating GIA uncertainties is in itself an open issue (Caron et al., 2018; Simon and Riva, 2020), we did not attempt to propagate full GIA uncertainties into the fingerprints.* ***Since the intrinsic uncertainty represents systematic errors and instrumental noise, which might be serially correlated, we assume that the errors can be approximated by a random walk. We therefore generate an ensemble of 1,000 time series by perturbing the original rate with random normal noise multiplied by the uncertainty time series. We then compute the trend for each ensemble member. We use the half width of the 95% CI as input in the SLE model, to show how the mass associated with the intrinsic uncertainty is distributed over the oceans"***

* It is unclear to me if all datasets are consistent within the estimated uncertainties of if there remains regions where this is not the case ?

We are not sure what the reviewer means with this comment. For a given contribution, the estimated trends based on different datasets agree within their respective uncertainties. Regarding the temporal uncertainties, even when using different noise models (Figure 2), the uncertainties are consistent for each dataset When comparing the different contributions (datasets), then the LWS was much higher than the uncertainties of AIS, GIS and GLA, for the 3 types of uncertainties considered here. But within a contribution (e.g., LWS), the uncertainties of the different datasets (CSR, JPL, WGP, GWB) were consistent. We hope this answers the reviewer's question.

* I have the feeling that the uncertainty estimation presented here is likely a lower bound to the true uncertainty on the ocean mass contribution to SL. Could you comment on that ?

We would actually argue that the uncertainties presented here are an upper bound, although there is no straightforward way of knowing the true uncertainty of ocean mass contribution to sea-level change. Most studies include only one type of uncertainty (intrinsic, temporal or structural). The objective of our work was to characterize all of these types of uncertainties, and thus we find larger uncertainties than other studies. Nonetheless, we did assume independence of the different types of uncertainty, and did not propagate GIA uncertainties into our fingerprints. Hence, while we do not know the true barystatic uncertainty, and have in general higher estimates than the ones published so far, it is possible that we are still underestimating the true uncertainty on the ocean mass contribution to sea-level change.

We have added a comment about it in the discussion:

"In this study we assessed the uncertainties related to the barystatic-GRD contribution to regional SLC, in particular the spatial distribution of the uncertainties. ***There is no straightforward way of knowing the true uncertainty of ocean mass contribution to sea-level change. Compared to previous studies (e.g., Horwarth et al., 2021), we tend to find larger uncertainties, thus our approach seems to be conservative. Nonetheless, we did assume independence of the different types of uncertainty, and did not propagate GIA uncertainties into our fingerprints, which could lead to even larger uncertainties.*** *Our results highlight that improving the spatial detail of land ice mass loss products, as well as determining more accurate land water storage trends, would lead to better SLC estimates. In addition, our findings can be used to inform projection frameworks. For example, we show that the distribution of ice in the Antarctic Ice*

https://doi.org/10.5194/esd-2021-80-RC1

Sheet has a significant impact on regional SLC, even in locations far from the ice sheets, such as The Netherlands. This means that, depending on the region of a collapse in the Antarctic Ice Sheet, the sea-level rise projections, which are often based on uniform ice sheet distributions and static fingerprints (e.g., Slangen et al., 2012, Jevrejeva et al, 2019) , may have large regional deviations due to spatial differences in the mass source. Incorporating the insights of uncertainty assessments in sea-level frameworks (as in Larour et al., 2020) should eventually lead to better sea-level projections."

https://doi.org/10.5194/esd-2021-80-RC1