Response to comments by Referee 3 (Lambert Caron).

We are thankful to Dr. Caron for his positive comments about our study. For a general discussion of the main changes that we have made to the original work, in particular with respect to uncertainty assessment, we refer to the first part of our rebuttal to Reviewer 2. In the following, we address his point-by-point comments (reproduced in italics).

Major points:

1- *l39: I know that other authors have used these terms in previous papers in a somewhat interchangeable manner, but I think it is important to distinguish data-driven from empirical. GIA models derived from partial differential equations (e.g. using love numbers) are not empirical (they are based on a physical theory) but they can be data-driven if their parameters are inverted from a dataset. Among such models are for example Peltier et al. (2015), Lambeck et al. (2014) or Caron et al (2018). Because the authors use such theory to generate their fingerprints, I would argue that their approach is not empirical (and I believe the title should reflect that), and in fact amounts to rescaling the loading history via the least square coefficients as was done in the aforementioned papers, and others before them. In my opinion, that is something the authors could put forward as an advantage of their approach, as it means it is consistent with how we otherwise model and understand the physics behind surface loading and deformation of the Earth interior.*

We have indeed interchangeably used the terms "empirical" and "data-driven" in order to differentiate our approach from previous studies. We also agree that, since our fingerprints are based on an analytical relation between surface load changes and earth mechanical properties, the term empirical is probably too strong.
However, since the fingerprints represent the combined effect of ice mass changes through time and earth rheology (plus a pseudo-spectral solution of the sea level equation), we disagree with the statement that our approach is equivalent to rescaling the loading history as done in some of the cited papers. In other words, even if our results could be used to guide a revision of the input ice histories (see rebuttal to comment #17 by Reviewer 2), we cannot quantify by how much, nor decide whether this revision should also include a change in the assumed mantle viscosity.
Hence, we are going to change the term "empirical" into "semi-empirical", and expand the discussion about the difference between this and other approaches. As suggested, we will emphasize more fact that the fingerprints are based on fundamental physics.

2- *l39 (cont'd) In particular, it means we could compare the GIA scaling coefficients (here the inverted coefficients of the fingerprints) with the values found in the literature and that are based on inverting RSL, and other datasets. That exercise cannot easily be done with true empirical models as they are not built on comparable basis functions. An important question this paper could (begin to) illuminate by showing these coefficients is therefore: are GIA models preferred by GRACE statistically different from the ones constrained with traditional datasets?*

Considering that we somehow disagree on the physical interpretability of our model results, we are not sure how the reviewer meant to realise such a comparison. Nonetheless, we will add a specific comparison against two available global models: Peltier et a. (2015), and Caron et al. (2018). Geoid trend differences are reproduced in Figure R5 and R6 below. We note here that our results are generally smaller than both cited models, though closer to ICE-6G(VM5a), and that the residual patterns are very different between the two cases.

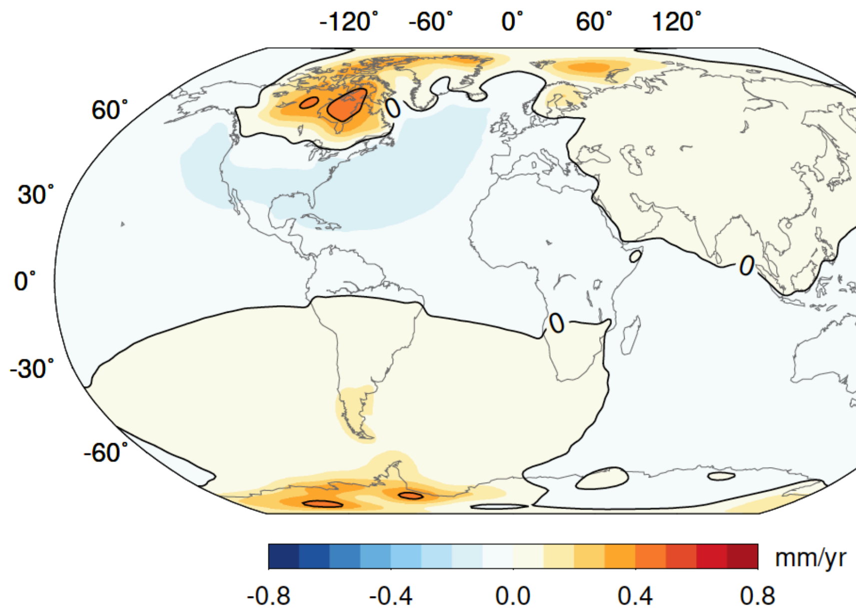

Figure R5: geoid height trend resulting from the difference between ICE-6G(VM5a) and the updated ensemble from this study.

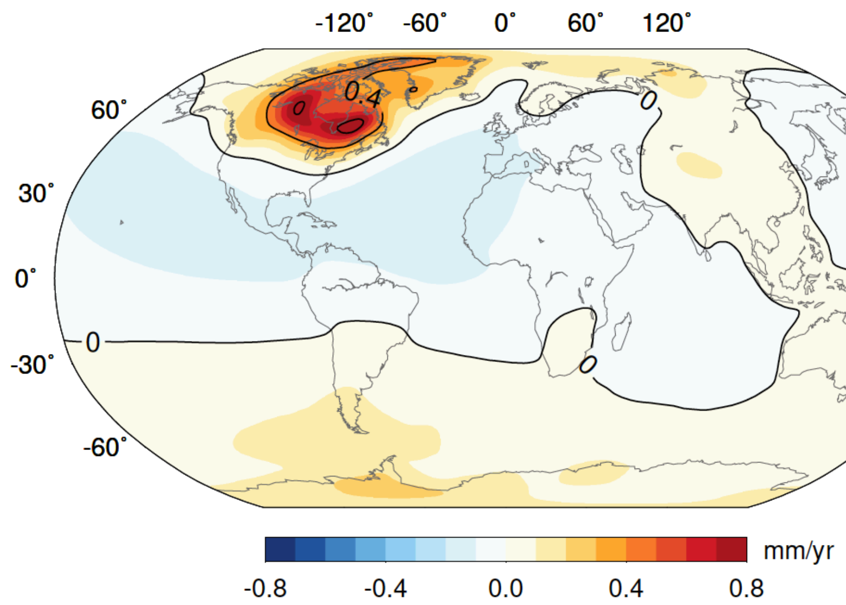

Figure R6: geoid height trend resulting from the difference between Caron et al. (2018) and the updated ensemble from this study.

3- *l81: What is the impact of the number of evaluated cases (here 4) on this statistical analysis? Would the authors expect a lot of differences from a more comprehensive exploration of the parameter space (particularly the viscosity profile)? How much does this limit the applicability of these results to correct GRACE?*

This comment is not applicable anymore, since the new ensemble is based on 132 cases. We would like to note that the new solution is fairly similar (the largest difference being in polar motion, mostly due to the improved approach used for the new solution), but the uncertainties are significantly different.

4- *l88: Ice histories such as that of ANU and ICE-6G_C have been crafted such that when combining all of their regional components, they are able to explain paleo RSL data (especially through the eustatic sea level curve). By recombining regional components of different models, is the solution still consistent with what we know about past RSL - and therefore part of what validates these ice histories in the first place?*

We do not think that we are able to answer this question, nor should we be, and that is why we consider this model to be semi-empirical. A manner to answer this question would be to repeat the inversion that has led to the input ice histories, using our solution as an additional and strong constraint on present-day geoid rates. Such an effort is beyond the scope of this paper.

5- *l113: The authors unfortunately do not really elaborate on their uncertainty quantification approach, and only state that they combine all 4 solutions into an average. How did the author calculate their standard deviation map? Did they: a) take the least-square optimized signal of each of the 4 cases, and then calculated the standard deviation between them (which the first sentence at l119 seems to point to), b) calculated the variance/covariance matrix of the coefficients for each case from the least-square system, which using the notations of Yun et al. (2019) should be a term with a form along the lines of (F'T'PTF)ˆ-1, and then averaged that covariance matrix between the 4 cases, c) a method similar to b), with a weight associated with each of the 4 cases in the averaging process to take into account that some of them allow smaller residuals than others, d) use yet another method? Out of these possibilities, a) is not an appropriate estimator, it would underestimate the uncertainty as it neglects the level of constraint of each least-square inversion. One could imagine a situation where all 4 best fit produce a similar signal for a given grid point or Stokes coefficient, but with a high variance/low confidence for that value. b) assumes that all 4 cases should have the same weight, which would be acceptable if they yield a similar sum of the residuals, c) being be more indicated otherwise. As this explanation is missing, it is difficult for me to understand and critically examine the results section of the manuscript, and going back to Yun et al. (2019) which details the method, I could not find the information related to uncertainty quantification either. I would add that if the authors mean to provide their model to the GRACE community for correcting GIA, it is very important that the treatment of uncertainty quantification be transparent*

Indeed, we had computed the model uncertainties based on option a). That is still what we do, in the sense that we compute mean and standard deviation of the ensemble. However, we think that the new uncertainties (shown in the reply to Reviewer 2, Figure R2) are more realistic, considering that the new solution is based on a much larger ensemble, in turn generated by using a rather wide spectrum of viscosity values for the upper and lower mantle.

6- *l179-182: An additional benefit of showing the covariance matrix of the least-square coefficients is that one can verify the degree of independence (or some measure of it, at least) between the different fingerprints by transforming it into a correlation matrix. This way sufficient orthogonality does not have to remain an assumption.*

As also suggested by Referee 1, we have prepared a correlation matrix (see Figure R1 in the rebuttal to Referee 1). We believe that this new figure shows that the fingerprints are quite orthogonal.

Minor points:

7- *l8: if the authors are referring to RSL indicators, they only point to a local level, not global*

True. We meant "globally distributed RSL indicators".

8 - l35: This reference should be Caron et al. 2018, not 2017

Corrected.

9- *l78: Why do the authors assume the Earth to be compressible for the fingerprints of the previous section but incompressible for GIA deformation? Is this not inconsistent?*

Elastic fingerprints need to be compressible, otherwise signals in the near field of load changes are unrealistically small. Since present-day GIA fingerprints only reflect mantle relaxation (no near-field load changes, apart from slowly changing ocean loading), the difference between compressible and incompressible solutions on gravity changes is actually quite small (Tanaka et al., 2011, GJI 184). At the same time, incompressible models of viscoelastic relaxation are numerically more stable. Hence, we prefer the possible presence of a small inconsistency against the risk of larger systematic errors in model output. Besides, considering that we now use a larger ensemble, consistency between the two classes of fingerprints (GIA and present-day) is not a major issue: if any is present, it will contribute to the ensemble error.

10- *l81: It was not clear for me at first read whether the authors were combining ICE-6G in one region with another model in the other region, despite the previous sentence. I suggest rewording along the lines of: "we use either GLAC1D (Tarasov et al. 2012) and ANU (Lambeck et al. 2010) in North America and Northern Europe, respectively, or ICE-6G_C (Peltier et al. 2015) in both regions."*

Indeed, that is what we meat. The sentence will be modified as suggested.

11- l81: "of" should read "or"

Corrected.

12- l88: The authors reference Ivins & James (2005) for the IJ05 model, which had an updated version (IJ05_R2) released in 2013 (Ivins, E. R., T. S. James, J. Wahr, O. Schrama, J. Ernst, F. W. Landerer, and K. M. Simon (2013), Antarctic contribution to sea level rise observed by GRACE with improved GIA correction, Journal of Geophysical Research: Solid Earth, 118(6), 3126–3141). If the authors used the updated version, this is simply a matter of updating the reference, but if not I would be curious to know why they chose the old version and if they expect a significant change from this choice. The volume of the Antarctic ice sheet at the LGM is different by about a factor 2 for example.

We were actually not aware of such a large difference in ice volume at LGM between IJ05 and IJ05_R2. Nonetheless, our GRACE-only approach cannot properly solve for Antarctic GIA, due to the large spatial extent and rather linear temporal evolution of ice sheet mass change, as well as its spatial overlap with GIA signals.
Hence, we have decided to stick to our initial approach, which is to use a single Antarctic fingerprint that had been roughly calibrated against the GRACE-ICESat combination by Riva et al. (2009). In that sense, it does not matter what the ice history actually is, since it is the GIA geoid signature that is used here (i.e., the combined effect of ice evolution and viscosity structure). We note that we still allow this fingerprint to be scaled within the inversion, but that only affects the actual magnitude of the signal, not its spatial pattern.

13- l116: "rounder": do you mean smoother or with a more circular shape?

We meant of a more circular shape, but this has actually changed in the new ensemble solution (the largest differences with respect to ICE-6G in North America are now over and West of Hudson Bay, see Figure R5), so the whole sentence will be modified.

Kind regards,
Riccardo Riva and Yu Sun