General comments

1. It is great to see a study that considers how to rigorously define conditional probability distributions for RSL for paleo contexts. However, the current submission has a major flaw. The SLI residuals are not independent and this must be explicitly accounted for. The current formulation explicitly assumes independence but then contradicts this with a 1/N normalization. The consequence of SLI dependence is clear, for instance, when considering the whole Dyke RSL database for North America. The spatial−temporal density of RSL datapoints varies greatly with resultant variations in datapoint redundancy. Without taking this density variation explicitly into account, use of your scoring scheme for say deglacial ice sheet model calibration will give results with model−data fits biased to where datapoints density is highest, even if the sectors where this occurs represent just a small area fraction of the LGM North American ice complex. Until this is addressed, the statistical method is invalid.

I should also note that this flaw might have been avoided with a more careful consideration of the existing literature (which is not evident in the reference list), eg Briggs and Tarasov, 2013 and Love et al, 2016.

We fully agree that we should consider the spatial-temporal distribution of RSL datapoints. Briggs and Tarasov (2103) as well as Love et al. (2016) applied a spatial weighting algorithm to already aggregated curves in order to consider the clustering of curves in specific regions. In this study, the individual SLIs are analyzed independently. Accordingly, we apply the redundancy weighting method proposed in Caron et al. (2017) where the cross correlations of the SLIs with respect to the considered model ensemble are taken into account. Therein, for each SLI a redundancy weight $w_i = \frac{K}{\sum_{j=1}^{N_{data}} \rho_{ij}}$ is defined.

Here, *K* is a normalization constant so that $\sum_i \frac{w_i}{N_{data}} = 1$, N_{data} is the total number of data. The Pearson correlation coefficient between the ensembles of predictions *i* and *j* is represented as $\rho_{ij} = \frac{cov(i,j)}{\sigma_i \sigma_j}$ where *cov(i,j)* is the covariance between two SLIs and σ_i, σ_j are the standard deviations of the two SLIs. The redundancy weights are calculated for each SLI and are considered as prefactors in Eq. 13, that now reads:

$$\ln L = \frac{1}{N_{data}} \sum_{i=1}^{N_{data}} w_i \ln(P_{sli_i})$$

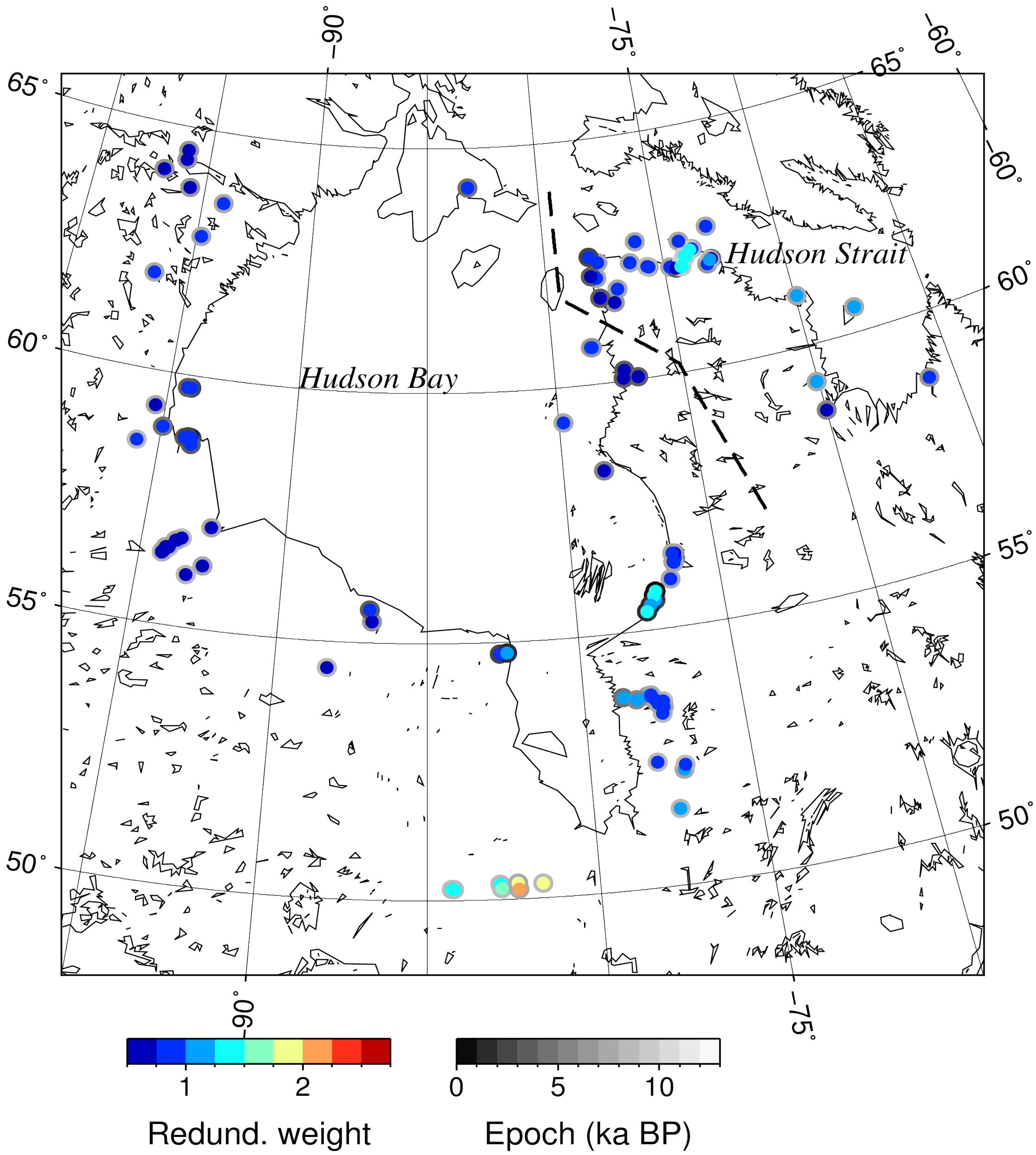

Figure x. Redundancy weights for each selected SLI. The color of the individual circle denotes considered weight, the shade of the time around denotes the calibrated age of the SLI. An overlap of SLIs could not be avoided. Dashed line separates Hudson-Strait indicators from those of the Hudson Bay.

In addition to the above procedure explanation of calculating the weights, we also included this figure into the manuscript

2. I do not understand the choice of journal. This submission would seem to me much more appropriate in GMD especially since the novelty here isn't the theory (this is standard Bayesian and probability theory) but the actual implementation. The first line in the abstract

also delineates this as a methodology paper: "In this study, we propose a statistical method to validate sea−level reconstructions using geological records known as sea−level indicators (SLIs)." Furthermore, the paper focus is on the method with the viscosity results only provided as an example : "findings are only meant to explain the method and not actually to constrain models."

The paper would also strongly benefit from more concrete details on implementation (probably best included in the supplement) to enable others to do so (especially since the software toolbox is not being made available).

Submission to GMD though requires provision of necessary code/software. This then raises an inequity between the two journals, submit to CPD and avoid the need to provide required code.... I'll defer the appropriate journal choice to the Chief Editor who should have a clearer sense of journal scope. I would like to see a statement from the editor clarifying how to resolve the scope intersection between GMD and CP with respect to software availability.

We did not consider GMD journal for submission since, as you mentioned, it requires a software, and we are not software developers. Providing the code that is user friendly would require more time. But we will, of course, respect editor's decision. We will include in the supplement a more detailed explanation of the algorithms considered and also an example of calibration with OxCal and how we extracted *pa(t).*

3. I would also like to see explicit consideration of tidal range and wave impacts, especially given the significant tides in Hudson Bay along with the well−known "storm−beach" displacement of SLIs. Once these issues (and the points below) are addressed, I would see this submission as worthy of publication in GMD (or CP if justified by the chief editor).

The tidal range for the Hudson Bay at present varies between 0 and 4 m (Webb 2014), and produces accordingly a small offset, which we decided to neglect for this study. In the outlook, we discussed the consideration of tidal ranges. In the majority of the studies from which we obtained the data (reference in the supplement of the manuscript), we could not find evidence that samples were found on the "storm-beach", apart from the local correlation suggested by the primary investigator. Therefore we did not consider the displacement of SLIs due to this effect in the study. But, since we established that Allard & Trembley (1983) related the samples from their study to "storm-beaches", as they found SLIs of 650 yr at an elevation of 4 m (Manitounuk islands), we will consider displacement in future studies. For indicators related to shore line, and not explicitly defined as picked from their living position (Klemann & Wolf, 2007), we will consider possible shift of 4 m (assumed storm beach height) .

We included this answer to the manuscript.

Specific comments

4. “For this conceptional study, we restrict ourselves to one type of indicators, shallow water shells, which are usually considered as low−grade samples giving only a lower limit of former sea level, as the depth range in which they live spreads over several tens of meters, and does not follow a normal distribution”
This statement is too sweeping. Eg Dick Peltier and myself treat certain inter−tidal species (eg Myt. Ed.) as providing more than just 1−way bounding.

Already visible from figure 4, Mytilus Edulis' living range is extending into the inter-tidal, meaning the lower limit of the depth range is smaller than 0. We merely wanted to indicate that only one of the boundaries is exactly specified close to 0 m depth.

5. “The shells’ depth range is derived from the OBIS database,”
You need to make clear whether the database only includes shells that were found in living position as well as whether the shells were living or not.

The database does not explicitly say if the samples were found in the living position, but the intention of the database is to discuss environmental conditions of species, which led us to assume that they were found in the living position. Hibbert at al (2016) uses OBIS database for living position of corals and we followed this approach.

6. “In addition to the indicative meaning, each sample’s depth is attributed to additional measurement errors which we have to account for. We assume them to be normally distributed, i.e.,”
It should be stated whether all considered SLIs were found in a living position. If not, how are the additional uncertainties addressed?

If considered SLIs are not found in a living position, our method is invalid, because we based our calculations on the specific location of each sample. Of course, it is tricky to ensure this information from the primary literature, so we decided to rely on this assumption.

7. “sums up the uncertainties derived from the leveling of the OBIS data, sigma_OBIS, and those of the SLI, sigma_SLI.”
I’m confused. Doesn’t the gamma function account for sigma_SLI?

Gamma function is only indicating depth distribution and not accounting for σ_{SLI}. Observational errors due to the leveling of the the depth in OBIS data and elevation of the SLIs are assumed as 1 m and 5 m, respectively.

8. “For the time range of considered SLIs, the IntCal13 curve”

Why wasn’t the marine calibration curve used? Furthermore there needs to be accounting of Reservoir age uncertainties (and reservoir age itself if you are using the IntCal curve). The text should also briefly describe reservoir ages uncertainties (given their non−trivial space/time variations).

From the personal correspondence with Art Dyke we found out that, while gathering data for the database, he did Marine Reservoir age correction for 440 years for those SLIs that were not already corrected in the primarily reported age. We therefore used IntCal13 atmospheric curve to avoid double correction. But, we do agree that it would be more correct to use the marine curve for this type of indicators. So, we first added back 440yr that Art Dyke accounted for, and then we applied marine curve (Marine.13) on the selected SLIs. Butzin et al. (2017) discuss spatial and temporal variability of the marine radiocarbon reservoir age during the last 50,000 years based on ocean circulation modelling. The authors did not focus on small regions like the Hudson Bay. Nevertheless their published model results (Butzin et al, 2017, data), show some variability. Therein, we find a decrease of reservoir age from about 700 years in the Hudson Strait to about 416 yr in the Hudson Bay for the last 12,000 yr. The time variability amounts to 50 yr for this time interval. In some parts near the W and SW shoreline of Hudson Bay, the basin correction reduces further to 200 yr what we do not consider in this study. In contrast, we split our data into two regions, 'Hudson Bay' and 'Hudson Strait', in which we consider basin corrections for the considered marine shells of 416 ± 50 yr and 700 ± 50 yr, respectively. The reason for this deviation is the higher sea-ice concentrations in Hudson Strait than in the central Hudson Bay. As sea ice inhibits air-sea $^{14}CO_2$ exchange, this leads to lower surface water concentrations (=> higher ^{14}C ages) in the entry of Hudson Bay than in the central bay (pers. comm. Martin Butzin).

Figure 7, after applying marine curve, now looks like the following one:

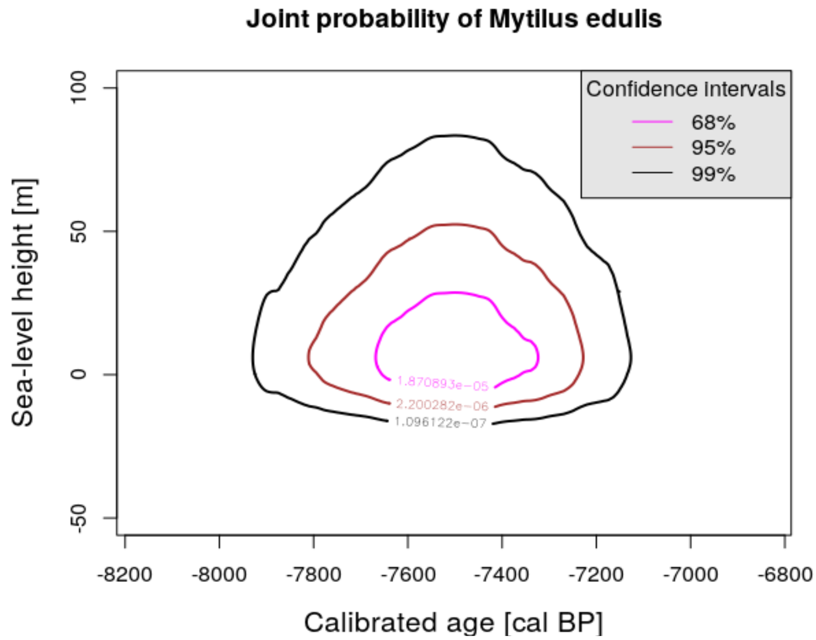

Figure 7. Joint probability density of Mytilus edulis presented as a 2d contour plot.

This explanation is included in the manuscript along with updated Figure 7. and brief description of reservoir age uncertainties.

9. # Fig 3. LGM RSL is kind of meaningless since all the SLIs are only present after local deglaciation. Better to show eg 8 ka RSL around when most of the critical Hudson Bay dates are available.

We changed the Fig 3. and replaced as you suggested.

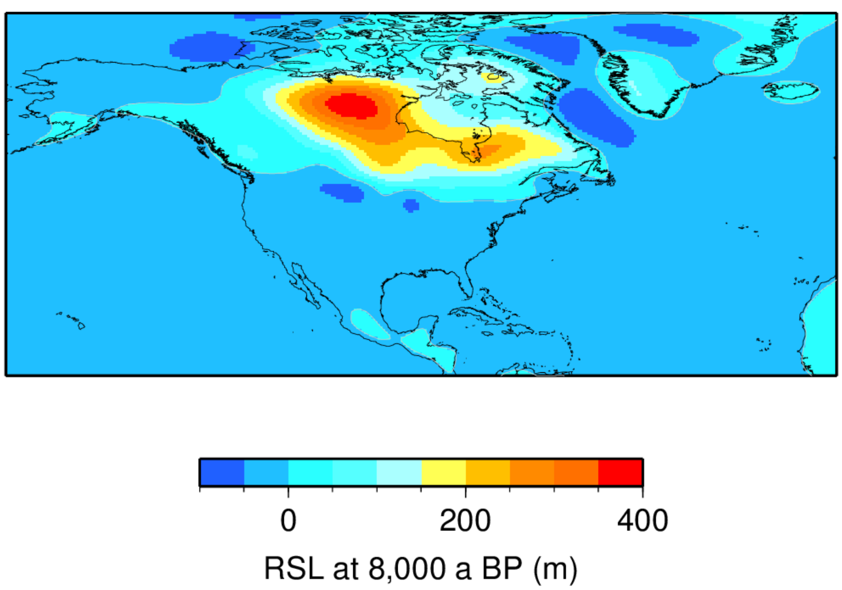

Figure 3. Relative sea level (RSL) at 8,000 years before present, in the region of Hudson Bay.

10. eq 9: here am and bm are predicted height and uplift velocity
a_m is the predicted height at t=0 only

Corrected Eq. 9: $h_m^{RSL}(t,\Omega)=a_m+b_m(t-t_m)$ where t_m is the median of calibrated age.

11. lithospehere −> lithosphere

Corrected

12. Fig 9 # please use a higher contrast colour scheme to make this easier to read

Corrected, Figure 9 is replaced with the following with recalculated fits.

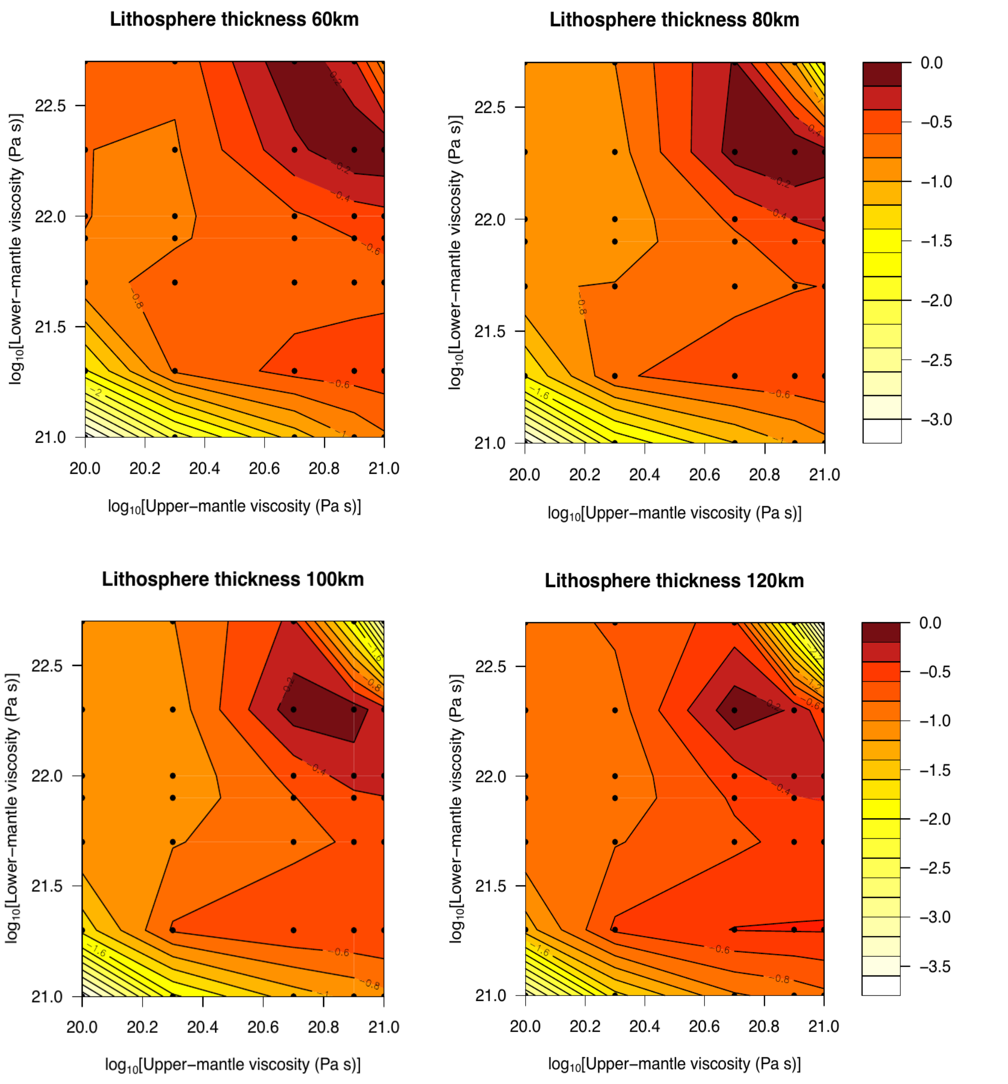

Figure 9. Model fits as function of upper- and lower-mantle viscosities for considered lithosphere thicknesses

13. Notation: equations 10−12

Use consistent notation. Eg eq 11 use P_i for conditional probability but eq 10 uses F_{h,t/m}. Best would be to use standard statistical notation for conditional probability, eg (h|x) for h conditioned on x. eq 10

We changed the notation in the Eq. 10 that now reads as follows:

$$P_{h,t|m} = \int_{-\infty}^{\infty} pa^{SLI}(t)\, ph^{SLI}(hm^{RSL}(t))\, dt$$

14. # How is this implemented? And how is pa(t) retrieved from oxcal?

oxcal is a complex enough application that a bit of guidance here would help others with their own implementation.

Based on your comment from the beginning, we included explanation about how we used OxCal to calibrate SLIs and how we retrieved *pa(t)* in the supplement.

15. Assuming that the conditional probabilities of the individual SLIs, P_i, are independent, the joint probalility eq 12

the 1/N_data normalization in eq 12 breaks the stated assumption of independent conditional probabilities. The likelihood is the joint conditional probability given by P in eq 11. ln(L) would just be SUM(ln(P_sli_i) if the residuals were truly independent. Anyway, there is no basis to assume all the SLI residuals are independent.

As discussed in the beginning of the review, we agreed that the residuals are not independent by including the spatial-temporal weights. After re-calibration with marine curve and calculation of fits with weights, we got similar results as before; 5 – 8 $x10^{20}$ Pa s for upper-mantle viscosity and values of 2 – 5 x 10^{22} Pa s for lower-mantle viscosity with lithosphere thickness of 60 and 80 km.

Literature

Allard, M. and Trembley, G. (1983), La dynamique littorale des ı̂les Manitounuk durant l'Holoce`ne, Z. Geomorphol. suppl. 47, 61–95.

Briggs R. and Tarasov L.. Evaluating model-derived deglaciation chronologies for Antarctica. Quaternary Science Reviews, 2013. http://dx.doi.org/10.1016/j.quascirev.2012.11.021

Butzin, Martin; Köhler, Peter; Lohmann, Gerrit (2017): Marine radiocarbon reservoir ages for the past 50,000 years, links to model results in NetCDF format. PANGAEA, https://doi.org/10.1594/PANGAEA.876733, Supplement to: Butzin, M et al. (2017): Marine radiocarbon reservoir age simulations for the past 50,000 years. Geophysical Research Letters, 44(16), 8473-8480, https://doi.org/10.1002/2017GL074688

Caron, L., Ivins, E. R., Larour, E., Adhikari, S., Nilsson, J., and Blewitt, G.: GIA Model Statistics for GRACE Hydrology, Cryosphere, and Ocean Science, Geophys. Res. Lett., 45, 2203–2212, https://doi.org/10.1002/2017GL076644, 2018

Hibbert, F., Rohling, E. J., Dutton, A., Williams, F., Chutcharavan, P. M., Zhao, C., and Tamisiea, M. E.: Coral indicators of past sea-level change: A global repository of U-series dated benchmarks, Quat. Sci. Rev., 145, 1–56, 2016.

Klemann V., Wolf D. (2007) Using Fuzzy Logic for the Analysis of Sea-level Indicators with Respect to Glacial-isostatic Adjustment: An Application to the Richmond-Gulf Region, Hudson Bay. In: Wolf D., Fernández J. (eds) Deformation and Gravity Change: Indicators of Isostasy, Tectonics, Volcanism, and Climate Change. Pageoph Topical Volumes. Birkhäuser Basel

Love R., Milne G., Tarasov L, Engelhart S., Hijma M., Latychev K., Horton B., and Tornqvis T. The contribution of glacial isostatic adjustment to projections of sea-level change along the Atlantic and Gulf coasts of North America, Earth's Future,4, 440-464, doi:10.1002/2016EF000363, 2016.

Webb, D. J.: On the tides and resonances of Hudson Bay and Hudson Strait, Ocean Sci., 10, 411-426, https://doi.org/10.5194/os-10-411-2014, 2014.