Response to Reviewers Comments (essd-2019-249)

REVIEWER #2

This dataset will potentially be valuable to other paleoclimate researchers and is well suited to be published in ESSD. However, I think the database would greatly benefit from more thorough presentation of the data in terms of their quality and their limitations (i.e. uncertainties therein and potential biases). For example, the data density (temporal resolution) for each record is not given or accounted for (n=? in each average SST value), the influence of outlier SST values on the LIG averages is not adequately addressed, and the spatial biases due to latitudinal/ longitudinal binning and/or lack of spatial resolution are not explored. The criteria for including records in the database need to be more rigorously and explicitly defined (were datasets rejected? how different is this compilation from the recent Hoffman 2017 compilation?).

Because we are not investigating centennial and millennial-scale variability, we were able to expand the number of records to that reported by Hoffman et al. The key criteria was that there was a minimum of three SST estimates across the LIG. In contrast, Hoffman et al. was focussed on time series data that required: 'The sample resolution ranges from centennial to <4000 years on their published age models, with a median resolution of 1100 years.' We are therefore able to report almost double the number of records to that presented by Hoffman et al. (189 vs 104 mean annual SSTs). Inevitably there are differences in the number of analyses undertaken through the different records which is dependent on the accumulation rate. In addition to the large database of temperature reconstructions, in response to the reviewer's suggestion, we now include the temporal density of the SST observations. For the error calculated for the regional and global SST anomalies, we incorporate the errors from the SST proxies (reported in the database), and the error associated with estimating regional and global SST from limited spatial coverage. To achieve this we propagated the SST errors for each measurement through each of the averaging steps (i.e. temporal to grid cell to zonal to area-weighted global) in our ocean-area-weighted average, as described by McKay et al. (2011). We used quoted error estimates for each study where reported. If not available, we applied proxy-specific error estimates. Although the impact of the spatial coverage was not explored in this study, it has been previously estimated in McKay et al., 2011. In that study, the error associated with the limited spatial range of the oceanographic proxies was estimated by calculating 1000 random 1-year global SST anomalies over the twentieth century, and comparing that to averages derived using only the paleoceanographic network available to that study. With that approach, they found no systematic biases associated with spatial network, and a 1 sigma uncertainty estimate of <0.1 degree. In this study, we've expanded the spatial network, and so it's reasonable to to consider ±0.1 degrees Celsius a reasonable, high-end estimate, making the contribution of spatial uncertainty modest in comparison to the other uncertainties in the study.

Furthermore, the uncertainties acquired by applying the ocean drift correction are not addressed, nor are other models explored or tested to demonstrate model sensitivity.

The full method of the ocean drift is provided by van Sebille et al. (2015). This approach tracks virtual particles in an eddy-resolving ocean model, the Japanese Ocean model For the Earth Simulator or OFES. In future work we would like to explore other models. In our previous work, however, we utilised the INALT01 model and found the ±1 s.d. of the INALT01, OFES and proxy distributions overlap. See figure below using two examples.   We therefore consider the OFES to provide a robust estimate of possible drift in this early study.

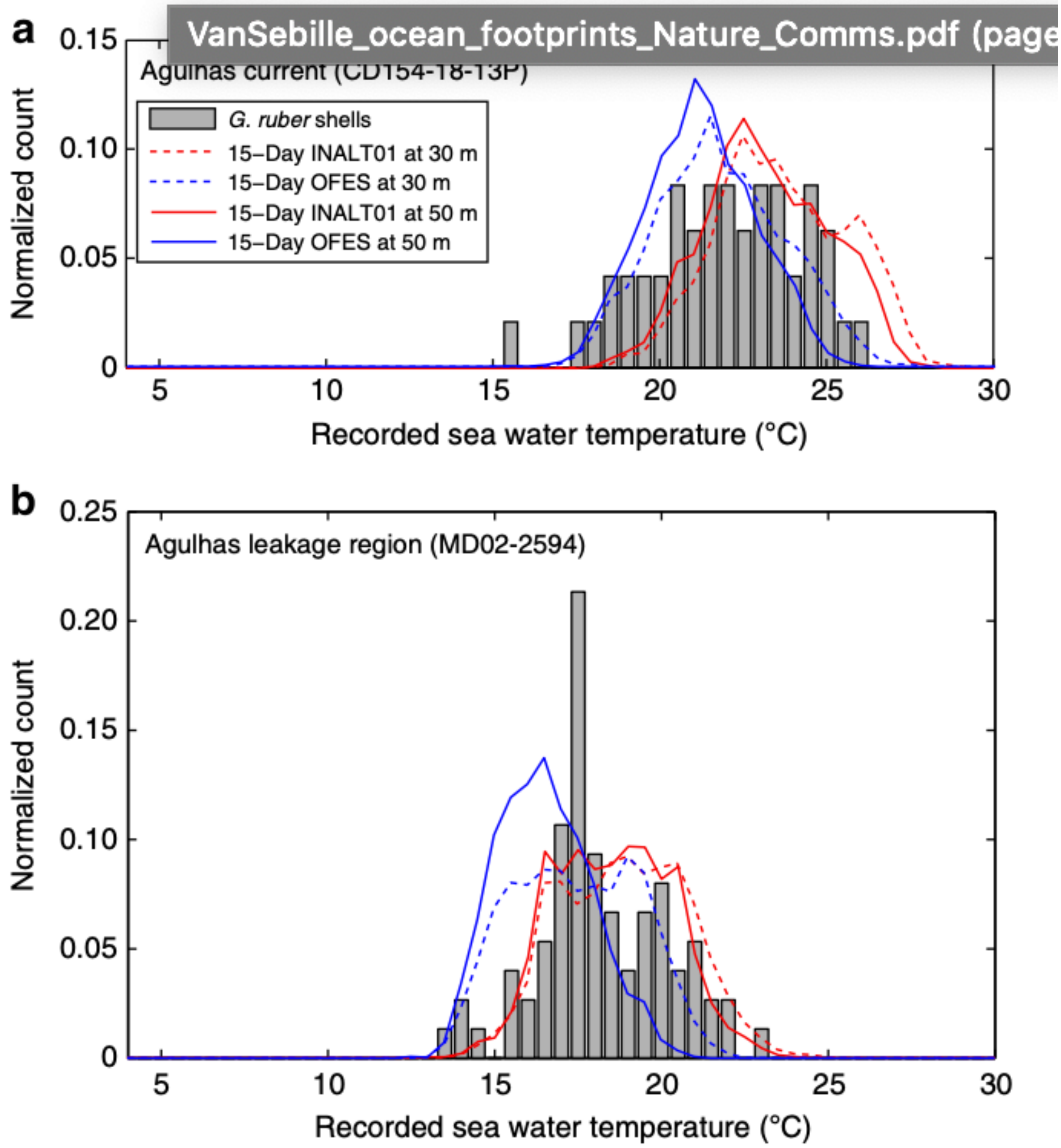

Figure from van Sebille et al. (2015): Distributions of temperature at two cores in the Agulhas region. The observed proxy temperatures (grey bars) at (a) the Agulhas Current core and (b) the Agulhas leakage core are compared with the temperature distributions for the virtual foraminifera experiments in the INALT01 model (red) and the OFES model (blue).

Reference: van Sebille, E., Scussolini, P., Durgadoo, J.V., Peeters, F.J.C., Biastoch, A., Weijer, W., Turney, C., Paris, C.B., Zahn, R., 2015. Ocean currents generate large footprints in marine palaeoclimate proxies. Nature Communications 6, 6521.

Additionally, the authors attempted to avoid complications arising from chronological alignment of proxy records by averaging over the entire LIG period; however, there is zero discussion of how the δ18O minimum was defined in each record, how well this minimum

was expressed in their 203 different sites, or to what degree errors were inherited due to local variations in benthic δ18O (even though the authors admit that such variations may temporally offset marine records by up to several millennia). In some cases, the SST records relied on proxies other than benthic δ18O to define the LIG time period, but it is nowhere explained what alternative proxies were used, how many records for which this was the case, or to what extent it might have influenced the results. The authors also do not address to what extent aligning the δ18O minima (because that is effectively what they are doing) warps the original age scales in the 203 records, except to show a very limited number of datasets (4) in Figure 2 – and there it is evident that the differences from the original age scales are substantial in some cases. Put another way, the authors need to address to what extent local variations in benthic δ18O might cause them to falsely identify the LIG time period and ultimately bias their LIG average temperature.
As the reviewer correctly identifies no one method provides an absolute age model for the last Interglacial. Even the use of d18O to define the LIG has an age uncertainty of 1-2 millennia. In some records where d18O was unavailable, other proxies used by the original authors have been used to identify the placement of the LIG; for instance, the CaCO3 content of the sediments as a measure of glacial-interglacial variability. However, it is important to note that we are not aiming to resolve centennial and millennial-scale variability through the interglacial and while we acknowledge that some individual SST estimates may not fall within the LIG or have been excluded (due to the uncertainties in the d18O for defining the interglacial) we consider the averaging of values across the full interglacial provides a robust value for each record and ultimately the regional and global reconstructions.

Finally, the manuscript would benefit from a comparison to other published LIG SST compilations (and estimates of thermosteric sea level rise) so that the reader either has some context for whether the new LIG reconstructions are reasonable, and/or why the new data are novel or represent an improvement on preexisting work. The authors also need to clarify what portion of the ocean volume their thermosteric sea level rise applies to (only surface 700 m?). It is confusing in the text as most of the authors' statements make it sound like whole ocean thermosteric sea level rise was calculated (I am still not 100 % certain).
We have now expanded the discussion of how we calculated the thermosteric sea level rise. As the reviewer correctly surmised we had originally determined this for the uppermost 700 m of the ocean. But we have now expanded the analysis to include the uppermost 2000 metres (approximately half the world's ocean) and 3500 metres. The 2000 metre depth warming provides comparable results to those reported by Shackleton et al (2020) and Hoffman et al (2017) which we have now discussed in the text.

If these comments can be sufficiently addressed, I see no reason not to publish this useful database.
We thank the reviewer for their support.

Specific comments (main text):
Line 109-110 – I cannot grasp how reliable this method was for selecting the LIG time period from the various proxy records based on what is presented in the manuscript. Were there any objective criteria for selecting δ18O minima? The authors must describe what they mean by "other complimentary proxy values," and state for how many records in the database this

applies. The authors also must state what they mean by “such a δ18O plateau is not obvious.” Were there objective criteria for electing to use alternative proxies rather than δ18O? The authors seem to think spatial variations in δ18O are not an important source of error in their approach, though they admit below that local variations can cause offsets of several millennia. Please provide more convincing arguments for this method and demonstrate to what extent these local δ18O variations are important for your analyses.
We have addressed this issue in the main manuscript by explicitly recognising the uncertainties in the recognition of the d18O minima (and other proxies such as $CaCO_3$) in each record, stating the uncertainty in this method and emphasizing the averaging of values across the full interglacial provides a robust value for each record and ultimately the regional and global reconstructions (see above).

Line 159-164 – The wording in this section is a bit too sleight of hand in my opinion. I disagree that the strategy is better than aligning records to a common temporal framework, or that it somehow circumvents the problem of generating time series data. While I agree that the authors do not interpret temporal trends (though they do distinguish the first 5 kyr from the rest of the LIG), by averaging over the selected periods with minimum δ18O the authors are in essence still aligning records to a common chronology because their analysis assumes the periods were coeval. I also disagree that this strategy is better than the example of aligning North Pacific data with EDC δD (which they state could be off by 1-2 millennia) because Figure 2 shows even larger temporal offsets of up to ~ 6 kyr (for example the end of the LIG in MD06-2986). The authors still need to present a convincing argument that aligning benthic δ18O is robust against the spatio-temporal variability between sediment cores, and then please state some estimate of the uncertainty and inherited SST error.
The age models reported in Figure 2 are from the original studies. We have not attempted to generate new age models. We are simply recognising the LIG in each record and then averaging the SST estimates over what we consider to be a common time period. The statement about the alignment of North Pacific data with the Antarctic EDC δD was to emphasise the challenges of identifying asynchronous changes between the hemispheres. Here we take a different approach to derive a first-order estimate of the temperature through the Last Interglacial, bypassing such issues.

Line 188-197 – Could you show some sensitivity analysis by running the model with different circulation? Just bracketing a plausible range would be enough to demonstrate the sensitivity. Also, I am very keen to see how the core top calibrations may change due to the ocean drift. I know the full analysis is beyond the scope of this paper, but perhaps selecting only a few core top measurements and examining how impacted they are by ocean drift would be useful for demonstrating the concept?

Unfortunately, recent work by EvS and colleagues (Nooteboom *et al*., 2020, *PlosOne*), has demonstrated that palaeoclimate modelling simulations have insufficient spatial resolution to capture mesoscale features that are critical for modelling particle drift. We hope future modelling outputs will enable this work to be undertaken. As a result, in the revised manuscript, we have acknowledged that the drift is estimated by contemporary ocean circulation which we consider to be a reasonable first-order approximation of the Last Interglacial. In future work we would like to undertake a detailed study of the impact of drift on the calibration but such a study would be beyond the scope of this database. We hope by highlighting the potentially substantial impact of drift (particularly in some key locations) this may be a focus for future research for others in the community as well. Reference: Nooteboom, P.D., Delandmeter, P., van Sebille, E., Bijl, P.K., Dijkstra, H.A., von der Heydt, A.S., 2020. Resolution dependency of sinking Lagrangian particles in ocean general circulation models. *PLoS ONE* 15, e0238650.

Line 203 – How is the uncertainty determined? If most proxies have uncertainties of 1-2 ◦C, it seems like the uncertainty on the mean should be larger than 0.1 ◦C.
We have described this more fully in the revised manuscript.

Line 213 – So far I did not realize that you were just calculating the thermal expansion of the upper 700 m of the ocean. I highly recommend saying this in the text prior when stating your results (e.g. in the abstract and also in the introduction when discussing previous sea level work). Otherwise, the reader may think you mean thermosteric sea level due to whole ocean thermal expansion (deep-water and surface).
Done. We apologise for the confusion.

Line 296-305 – Please specify here that the authors mean thermal expansion of the top 700 m of the ocean (which I think is what they mean, though it needs to be clarified more explicitly in the text). The authors should compare their result to other estimates of the thermosteric component of LIG sea level in addition to the McKay result (Hoffman et al., 2017;Shackleton et al., 2020).

Done.

Line 303-305 – This statement is too strong without explicitly stating that the deep ocean was not considered. Readers will misinterpret it to mean whole ocean thermosteric. Or, if the deep ocean was considered (I am still unclear about whether the authors did this or not), it must be justified why SST estimates alone were used to estimate whole ocean thermosteric sea level rise and why the estimates were so low compared to other work (e.g. Shackleton 2020).

Done. We apologise for the confusion.

Figure 2 – Showing the alignment of only four marine cores is much too limited to give readers any sense for how much the 203 chronologies were distorted when the authors picked δ18O minima to delineate the LIG time period, over which they averaged the SST results. Figure 2 demonstrates that for none of the four cores shown did the LIG actually

occur during the period 129-116 kyr (on their respective age models), and in core MD06-2986 the LIG notably occurred during a span of only about 5 kyr. Can you say with confidence (or even better, demonstrate for readers) that the cores in Figure 2 represent the full range of chronological differences in the δ18O minima between all of the records? Additionally, please improve the figure resolution so that the text and traces are not blurry.

We apologise for the blurriness of the figure. We have now resolved this. The figure is for illustrative purposes and reports the chronologies for the original studies. We have not developed new chronologies for the records (as undertaken by Hoffman et al and Capron et al). Instead, we have used the d18O minima to define a common period to derive a mean temperature.

Figure 3 – This is confusing. It looks like only the modern data were run through the drift correction. I thought the correction was applied to each LIG average.

The drift correction was undertaken using a modern ocean configuration and the temperature offset applied to the average LIG estimate for each site.

Figure 4 – I recommend plotting a third panel showing the residual between the original SST and the drift-corrected SST.

We can provide this panel if the editor would like.

Table 1 – It strikes me as odd that the DJF and JJA global SST values are both negative, whereas the mean global SST value is positive. What delineated a DJF and JJA record from the other 189 records? How much overlap is there between the 92 + 99 seasonal records and the 189 annual records?

The seasonal estimates are provided in the database. Seasonal temperature estimates are challenging to provide with confidence given the seasonal biases of proxies which are likely latitudinally-dependent. As a result we consider the annual estimates to be more reliable.

Table 2 – Similar comment as above. Specific comments (regarding the Excel file):

Sheet 1 – The spatial delineations are confusing. Why do you average > 45◦ and then also > 50◦ with only 5◦ difference? Please justify.

These estimates are to provide a measure of changes in the polar latitudes. There are considerably more records polewards of 45˚ so we included both to provide a measure of the robustness of the zonal reconstructions. This is now given in the revised manuscript.

Column H - By “Jan-Dec” do you mean annual? Just say “annual” so as not to be confused with “DJF.”

Done. We apologise for the confusion.

Technical corrections:

Line 42 – "The timing and impacts. . . remain. . ." instead of "remains."

Done.

Line 47 – Better references exist for "multi-millennial duration shifts in the Earth system took place in the past." The ones used here appear to mostly be about Anthropocene/ future tipping points.

Done. We have replaced with more appropriate references.

Line 51 – Can you provide a reference for 129,000-116,000 years ago, if it is elsewhere defined? Otherwise state it is the authors' definition.

Done. The reference is from Dutton et al. (2015, Science).

Line 56 – Global Mean Sea Level should not be capitalized.

Done.

Line 57 – There are better references for the observation of abrupt shifts in regional hydroclimate during the last interglacial than Thomas et al. 2015. Why not just cite cave record papers (Wang et al., 2008;Cheng et al., 2016), for example?

Done.

Line 58 – Buizert 2014 is not about CO2. Kohler 2017 is partly, but why not cite the original data? (Petit et al., 1999;Barnola et al., 1987) or (Bereiter et al., 2015) for the most recent compilation of CO2 ice core data.

This is correct but Buizert et al. do report CO2 measurements from Taylor Dome. However, we have included these other references.

Line 61 – Provide references for "considerable debate" about the contribution of sources to sea level rise.

Done.

Line 74 – Cite also (Hoffman et al., 2017).

Done.

Line 80 – Sea-Surface Temperature should not be capitalized.

Done.

Line 83 – Can you move the Mercer 1978 reference to somewhere in the middle of the sentence? At the end of the sentence it looks like it is a reference for the Paris Climate Agreement.

Done.

Lilne 117 – Does “maximum” refer to the average of the first 5kyr? I recommend changing the wording because “maximum” can be interpreted here that your means are upper limits.

This is a fair point and we have changed.

Line 121-123 – I don’t think Figure 3 should be referenced here, as it doesn’t really relate to what is said in the sentence.

Done.

Line 125-129 – Again the use of the word “maximum” could be misunderstood to mean you only used the highest values in the datasets, especially on line 126.

Done.