Climate of the Past Discussions

Clim. Past Discuss., referee comment RC2
https://doi.org/10.5194/cp-2022-12-RC2, 2022

Comment on cp-2022-12, by Jess Tierney

Jessica Tierney (Referee)

Referee comment on "A new global surface temperature reconstruction for the Last Glacial Maximum" by James D. Annan et al., Clim. Past Discuss., https://doi.org/10.5194/cp-2022-12-RC2, 2022

In this work, Annan and Hargreaves produce a new reconstruction of LGM temperature using the MARGO, Bartlein et al (2011) and Tierney et al. (2020) proxy databases and an ensemble of model priors from several generations of PMIP. They find lower global cooling than Tierney et al. (-4.5 vs -6C) and discuss some of the reasons for the discrepancy.

I think this is a useful contribution to the problem of reconstructing past climate fields and it is very interesting to see the outcome when a slightly different DA approach and multiple different model priors are used. However, the analyses in the paper as it is now don't identify *why* these new results are different than Tierney et al., or for that matter, Osman et al., (2021) which finds an even cooler LGM (-7C) with a more limited proxy network yet much wider priors (this follow-up paper isn't discussed here, but should be!). The authors suggest that the cold results we got are b/c of the CESM priors. It definitely could be, but I'm not convinced yet (in part b/c of the strong validation we have in our studies, and agreement w/ independent estimates of glacial cooling, see below). Conversely, it seems like the relatively warm priors here are influencing the result also, given that the adjusted prior mean is -4.9 p/m 1.1C. The authors also find that if they inflate the prior mean cooling to -9.7C, they get a substantially cooler posterior (Page 14), unless they do their "translation procedure", which re-centers the prior on the data a priori. But then, isn't the result basically dependent on pre-centering the prior on the data -- which means you presuppose what level of cooling the data show? I think this could be problematic. The data might show a mean cooling of say -4.5C, but the data don't sample the coolest places on Earth (like the Laurentide) so this is almost certainly an underestimate. To what extent does the translation step bake-in the posterior result?

A couple of further thoughts on things to explore to investigate why the results here are so different from Tierney20 and Osman21.

-In Tierney20, we calculated a proxy-only global mean cooling to compare with the DA and found it has a median of -5.5C. This was based only on SST, so reasonably could be too warm since it excludes land data. So first order, the -4.5C result here just does not seem cold enough. Granted the -5.5C we calculated has to assume a scaling b/t SST and SAT (we followed Snyder et al, 2016 and drew this scaling from PMIP). Nevertheless you could calculate a proxy-only value here for comparison, and also compute a proxy-only global SST (which doesn't have to be scaled). Can you calculate proxy-only global SST change and compare to the DA results (as we did in Fig. 2 of Tierney20)? We got -2.9C with the TEA network. I'm just curious whether the DA proxy posterior comes out similar to this or warmer.

-In Tierney20, for sure our prior was a bit too tight than we liked which we discussed in the paper and is evident in the U-shape of the rank histograms. However, interestingly, there was no strong mean bias in the LGM rank histogram (but there was a slight mean bias for the Late Holocene) which suggested to us that the LGM prior was not too cold; otherwise we would expect to see a skewed shape in the rank histogram. This is different than the rank histograms shown here which are much more flat. I'm not sure how to reconcile this. Our rank histograms are based on withheld validation data only (not sure if that is the case here or not). They were also calculated in proxy space, taking advantage of the forward modeling in our DA technique. So how to compare those to these?

-ERSSTv5 is used here as a preindustrial baseline instead of the Late Holocene values. What happens if you use the Late Holocene SSTs in TEA instead? Our reasoning for using Late Holocene SST was that in some locations there is a strong bias in the proxies (relative to climatological SST) so subtracting LH SST would correct for this. This said, LH SST estimates from the TEA network might suffer from non-modern coretops. In Tierney20, we did not screen the data for age control, so some of the coretops might have pre-modern (likely colder) proxy information. We speculated that this is one reason why Osman et al LGM cooling is larger. In that study, we only used proxy data with good age control so the Late Holocene data are for sure Late(est) Holocene. Perhaps it is worth running your DA using the screened proxy network of Osman et al. alone?

-Working in temperature vs. proxy space. I suspect this could produce some of the differences between the result here and Tierney et al (and Osman et al). For one, in these latter studies the forward modeling allows us to consider the seasonality of the proxies and assimilate them during different production seasons. In contrast, it seems that here all of the proxies are treated as annual mean T (?). What happens if, instead, you consider the seasonality and assimilate the proxies in their respective seasons (as we did)?

In addition, working in proxy space allows for changes in the other environmental parameters that influence the proxies, i.e. d18O of seawater for example. Whereas the derived SSTs assume a uniform change in d18O of seawater. Not sure how to test this difference, but at a minimum, it should be discussed.

-The authors can test the theory that the CESM priors are the main source of the difference by applying their approach to a CESM-only ensemble. The can apply their technique to the prior states from Osman et al. (21, 18, 16, 14, 12, 9, 6, 3, 0 ka - 21, 18; 3, 0 are same as used in Tierney20). I see now that we didn't put our priors up on the NCDC archive for Osman et al. However we can rectify this and/or directly share with the authors through a Google Drive link. As Dan Lunt suggests, this would be a very useful test.

-Finally, validation. One of the strongest pieces of evidence that the solutions of Tierney20 and Osman21 are robust is the fact that we excellent independent validation with the ice core data. If LGM cooling were only -4.5C globally, I doubt we would get as good of validation. In contrast, there is no external validation in this study. At a minimum, the authors should do a more robust k-fold style validation (leave out 20-25% of the data) rather than leave-one-out, which is not going to be a very hard test. Another approach might be to assimilate TEA, withhold/check against MARGO assemblage data, or withhold/check against Bartlein 2011.

-Finally, Seltzer et al. (2021) find, through independent analysis of noble gas data, a low-to-mid latitude cooling of 5.8C during the LGM (https://www.nature.com/articles/s41586-021-03467-6). This is largely in agreement with the deeper cooling inferred in Tierney20 and Osman21 (if not implying somewhat even more extreme cooling!). How do you reconcile these results with global or latitudinal changes inferred here?

In summary, more work has to be done to understand the origins of the differences (proxy vs T space, model priors, method assumptions). Likewise, in the future, we hope to try our DA approach on other model priors as well and understand how/whether it influences the results; our main barrier has been the lack of water isotopes in other model simulations besides CESM.

Specific Comments:

Page 5, Line 20: I agree that CESM2 is an outlier and probably needs to be removed to satisfy the condition of normality, but it is noteworthy that the resulting ensemble mean is 4.5 p/m 1.1, which is rather tight.

Page 8, line 15: The TEA database includes all of the data from MARGO that are d18O, Mg/Ca, and UK37, so if you are adding MARGO estimates double check that it is only the assemblage data, which we did not use. Otherwise the same data might be used more than once when you combine the datasets.

Page 10, line 10: Can you clarify here whether you are doing serial updates (one observation at a time) or a joint update (all observations together)? Tierney et al. and Osman et al. are joint updates. There can be some differences b/t using serial vs. joint although they are usually small.

Page 11, line 14: The authors take issue with the cold N. Pacific gyre signal in T'20, but they also show this feature in their reconstruction with a cooling of -4 to -8. Indeed, most PMIP models indicate a strong cooling here and the proxy data also suggest this. In T'20 it is indeed larger (-6 to -10), but I don't think we can say which magnitude is more accurate given the uncertainties in the proxies, and also the lack of proxy data from the center or eastern side of the gyre. Note that Gray et al. 2020 (PaleoPaleo) document gyre cooling of about -5C on average, but deeper cooling is seen in some of the raw Mg/Ca and UK37 data. It seems to me that it's hard to argue what the exact magnitude is. I would either drop this sentence or introduce more nuance here.

Page 11, line 16: With this many data points, I think leaving out a percentage of the proxy data, as was done in Tierney et al. and Osman et al., is a more robust test than leave-one-out. I suggest withholding 20%-25% of the data.

Page 12: Rank histograms. Are these calculated on the withheld validation data (they should be)? Please clarify.

Page 17, Line 9: I would not go so far as to argue that using calculated SST/SAT from the proxies is better than forward modeling, i.e. working in proxy space. This *might* be true for proxies that are univariate, but it is definitely not true for proxies that are not, which includes d18O, Mg/Ca, foram transfer functions, and pollen. To translate these to temperature, one has to make assumptions about the other environmental influences on the these proxies (i.e. pH, salinity, or pollen, moisture balance) that are going to be imperfect compared to the forward modeling scenario in which multiple environmental parameters can be accounted for. Unless you can prove otherwise, using derived temperatures will be inferior to working in proxy space where inversion is not needed and you can make fewer static assumptions.