Clim. Past Discuss.,
https://doi.org/10.5194/cp-2018-120-AC3, 2019

Climate of the Past Discussions

Interactive comment on "Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling" *by* Robert Tardif et al.

Robert Tardif et al.

rtardif@atmos.washington.edu

Received and published: 18 January 2019

Authors' Responses to Anonymous Reviewer 3

We thank the referee for helpful comments on the manuscript.

I have two main concerns about the paper, which together require that the manuscript undergo major revisions before it is acceptable for publication. The first is the character of the derived reconstruction and the unsatisfactory verification of the product using only observational data. The second is the use of

Printer-friendly version

Discussion paper

multiple ad hoc methodological choices, none of which are reasonably justified or widely tested.

The first concern, or two concerns really, were also raised by other referees. We have undertaken a revision of the manuscript that will include a more complete discussion on the character of the temperature reconstruction, compared and contrasted with other reconstructions (see response to next comment). We also plan to expand the verification performed in proxy space and move the discussion from the supplementary material into the main body of the paper.

The referee's concern on the justification of our selection of the data assimilation parameters will also be addressed. A brief mention on how those choices were made will be included. We believe however that a lengthy discussion on these is not warranted in the present paper, as these choices have been discussed in prior publications. However we agree that some clarification on how these choices were made should be included to convince the reader that careful consideration was involved.

I am struck by the comparison in Figure 2a and the little attention the authors give to the differences between the previous LMR product and the newer version (not to mention the complete lack of comparison between either of these results and other temperature reconstructions).

We agree that our results should be better framed with respect to other reconstructions, as also echoed by other reviews. We have compared LMR results of Northern Hemisphere (NH) temperature and other reconstructions included in the IPCC AR4 and AR5 reports; see Figure 1 in this document as an example. A low-pass filter has been applied on all results to focus on the lower frequencies. Such comparisons will be the basis of a discussion contrasting our results with other efforts in temperature

reconstructions in a revised Section 3.

The GMT from the newer product looks almost like white noise and has lost not only the multi-decadal to centennial variability in the first product, it is also likely at odds with the now large collection of global and hemispheric temperature reconstructions spanning the last millennium or more. The authors not only need to spend more time discussing this issue, they also need to compare their results to the collection of large-scale temperature and hydroclimate reconstructions currently available.

The perspective gained from the comparison shown in Figure 1 suggests that indeed there are differences between the updated and prototype LMR reconstructions, as well as with other reconstructions, but also provides some evidence that a claim of "white noise" is an overstatement. The differences between the other reconstructions themselves serve as a testament of the uncertainties characterizing climate reconstructions. In that context, we should to point out that most other reconstructions are found within the bounds of the LMR ensemble most of the time, indicating some level of agreement between the different products, when framed within the context of the uncertainty defined by our reconstructions.

We do acknowledge that a loss of low frequency variability characterizes our updated reconstruction. As pointed out by the referee about global-mean temperature, also reflected in the NH-mean temperature results in Fig. 1, the perceived loss of variability is mainly the result of differences limited to three distinct periods. These are: a colder medieval period, most notably during the 875–1050CE period, and warmer temperatures during the 1600–1700CE and 1810–1920CE periods during the Little Ice Age (LIA). We further note that the updated LMR is among the warm outliers during these cold periods compared to the prototype and other reconstructions. An analysis of results from a large number of reconstruction experiments has allowed us to conclude

that colder temperatures during the medieval period, compared to the prototype LMR, are related to the change from an earlier proxy dataset PAGES 2k Consortium (2013) to a recently published collection PAGES 2k Consortium (2017). The global temperature composites presented in PAGES 2k Consortium (2017) shows that a distinctly warmer medieval period isn't a prominent feature of the new collection, and does not result from other updates to our data assimilation system. As this dataset reflects the community's most recent and rigorous identification of proxy records suitable for temperature reconstructions, we believe that the lack of a "classic" medieval warm period in our updated reconstructions of global mean surface temperature (GMST) and NH-mean temperature should not necessarily be considered an outstanding shortcoming of our updated reanalysis.

On the other end, beneficial further refinements to our DA system have been identified as a result of issues raised by the referees. A notable sensitivity of reconstructed LIA NH-mean temperatures to the set of assimilated tree ring width chronologies has been identified, underlining possible weaknesses in observation error variance estimation (i.e. the R_k terms in equations 4 in the manuscript) for these records. Additional reconstruction experiments have shown that colder LIA temperatures are obtained when some of the large number of tree ring width records from the Breitenmoser et al (2014) collection are not assimilated (see Fig. 2). A more important low frequency variability is recovered, in better agreement with other reconstructions.

As suggested by the referee, a revised Section 3 will include a more complete discussion of these updated results, including comparisons with other available reconstructions.

...it is essential for them to do more to verify their results beyond the comparisons they make to observational data. While the latter is important and useful, it is not enough. Incidentally, the authors do perform validation exercises on a withheld period of observational data and using withheld proxy series, but that

work is buried in the supplemental and not adequately discussed in this context. More should be made of those efforts, which strengthen the authors results with truly out-of-sample validation experiments.

We agree. We will include an expanded discussion of proxy verification results in the main body of the text.

I do not think the use of CE is the same as it is traditionally used in the paleo literature, given that the latter approach requires a true cross-validation period. The authors should clarify this point.

The formulation has been used in numerous other published work. We have found that CE is a valuable complementary metric to correlation, due to its sensitivity to mean errors and representation of variance in the evaluated fields. We will contrast our implementation of the metric cmoapred to its traditional use in the literature.

I think it is further important for the authors to derive validation experiments for the sparsely sampled periods early in the proxy network (e.g. deriving reconstructions using only subsets of the proxies that extend back to specific time intervals). This would go a long way toward helping to better understand the loss of proxy information back in time. This is partially addressed by the variance exercise the authors perform, but more can be done. Recons for temporal subsets of the proxy network would in fact be more useful than the MC sampling of the proxy network that the authors perform, given that it would be systematic and inform a direct question about the influence of the declining proxy network.

This is a very good suggestion. Experiments similar to what is suggested here have

been conducted in support of another publication, currently under review. This issue can also be investigated through an assessment of reanalysis performance performed in proxy space, with a focus on different time intervals. Proxy verification results will be included in the revised manuscript. This characterization should also be framed in the context of the uncertainties characterizing reconstructions. This capability is enabled in LMR through the availability of ensemble-member information. This complementary perspective will also be incorporated in the revised version of the paper.

Here is a ... list of choices the authors have adopted that are not accompanied by any justifications or sensitivity discussion: 1-Use of 100-member ensemble; 2-Use of the CCSM4 last millennium simulation as the prior; 3-Use of 51 MC realizations; 4-Use of a proxy sampling scheme based on 75% of the proxy records; 5-Degradation of the model resolution to a ∼5x5 grid. All of these choices undoubtedly influence the derived LMR product. Some of them can be justified based on discussions in the literature. Some of them require empirical demonstrations. All of them come across as ad hoc. I would also venture to guess that the LMR results are more dependent on a couple of these choices than the other dependencies that the authors more systematically test. It is therefore essential that the authors do a better job of justifying these choices and convincing the reader that they are either reasonable choices or chosen based on some methodological/logistical rationale.

Some of these choices have been discussed in prior publications. See Hakim et al (2016) and references therein for example. However we agree that text reminding the reader of how parameters were chosen should be included to convince that careful consideration was involved. To further clarify the context, we should point out that some parameter values used here were chosen to to maintain consistency with the configuration used in Hakim et al (2016), in order emphasize the impact of updates

specifically addressed in this work.

Some of the parameters were originally chosen based on practical considerations. For example, the ratio of assimilated versus withheld proxies and the number of Monte-Carlo realizations were chosen so that the random sample of proxies set aside for independent validation is representative of the overall dataset. Such considerations will be more clearly described in our revision of the manuscript.

I should specifically mention the use of the CCSM4 as the prior. The authors say nothing about how their results might depend on the model prior and whether they have tested alternative last-millennium simulations in their analysis. This is an obvious question and the authors need to address it.

This topic is addressed in Hakim et al (2016), with results from reconstructions using a wide range of calibration and prior data (see Fig. 12 of that paper). Work on this topic has since been expanded and explored in more detail. These results are, however, expected to be the subject of a separate publication.

Page 2, lines 5-6: What does "synthesizing information" mean? This is vague and I am not even sure the statement is true. There are lots of central challenges of paleoclimate science, and it is arguable that what the authors are alluding to is one of them. This strikes me as an unsupported justification for what the authors subsequently say they are attempting to do.

We will modify the text to more precisely reflect the goals of the work presented in our paper. In this part of the text, our goal is to emphasize that data assimilation is a powerful framework in which information from proxies and numerical model simulations is combined for the production of, hopefully, robust climate reconstructions.

Page 2, lines 30-32: This is a much more mundane objective than the sense given in the abstract. Are the authors attempting to release a shiny new LMR product or should this be seen as an iterative verification step toward some improved effort down the line?

We apologize if the text does not clearly convey the goals pursued here. As with any complex problem to be solved, as is the case here, improvements are generally incremental and modest. The main goal of efforts reported in this paper has been to improve our system over the LMR prototype, and to deliver an incrementally improved product, while being fully cognizant of the fact that room for improvement remains. In the revised manuscript, we will attempt to convey these goals in a more precise manner.

Page 8, line 9: The use of precipitation is not justified and concerning. First, precipitation is almost never the variable associated with moisture sensitivity in trees - some measure of soil moisture is. It is therefore not clear why the authors used precipitation and how it influences their results. Why not use a more conventional variable like PDSI? Secondly, how do the characteristics of precipitation influence the results? Does it matter that precip is likely not Gaussian and that it has limited spatial and temporal covariance structure? Is the use of precip perhaps adding to the loss of low-frequency variance in this new LMR product? My guess is that this specific choice has a large impact on the derived reconstruction and the use of precip is not justified in any way.

The referee is raising fair and important points here. We acknowledge that soil moisture is the preferred response variable for the modeling of tree-ring widths. However, given our approach, which relies on the availability of calibrated forward models, the absence, to our knowledge, of a reliable century-long soil moisture dataset is an im-

portant limiting factor. An alternative, as pointed out by the referee, would be PDSI (or other drought indices such as SPEI) instead of soil moisture. This option, is enabled within our framework with the use of the Dai et al (2004) dataset for forward model calibration. We intend to generate test reconstructions using this configuration and compare with results using precipitation. We believe these comparisons will allow the impact of the issues raised by the referee to come into greater focus.

To provide additional context on our efforts reported in the submitted manuscript, more particularly with the development of bilinear PSMs, our intent is to replicate VSlite's (Tolwinski-Ward et al, 2011) general approach of combining the influences of temperature and moisture in modeling tree-ring width variability, albeit in a simpler fashion, while avoiding the issues associated with VSlite's formulation involving thresholds (see Dee et al, 2016). We will perform a careful quantitative comparison with the more conventional approach using linear models calibrated on PDSI and discuss the results in our revised manuscript. Particular attention will be placed on the characterization of low frequency variability to assess whether the updated LMR's characteristic loss of variability is rooted in the use of precipitation to forward model tree-ring width proxies.

Figures: In general, there is a lot of small text in the figures that is hard to read and also rather confusing and messy.

We will simplify the design of the revised figures.

Figures: The many colorbars are also unnecessary in many plots when one would do.

We will streamline the design of the revised figures.

References

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N. J. and Perkins, W. A., The Last Millennium Climate Reanalysis Project: Framework and First Results, *JGR: Atmospheres*, 121, 6745–6764, doi:10.1002/2016JD024751, 2016

PAGES 2k Consortium, Continental-scale temperature variability during the past two millennia, *Nature Geoscience*, 6(5), doi:10.1038/ngeo1797, 2013

PAGES 2k Consortium, A global multiproxy database for temperature reconstructions of the Common Era, *Scientific Data*, 4(170088), 1–33, doi:10.1038/sdata.2017.88, 2017

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K. and Anchukaitis, K. J., An efficient forward model of the climate controls on interannnual variation in tree-ring width, *Climate Dyn.*, 36, 2419–2439, doi: 10.1007/s00382-011-1062-9, 2011

Dee, S., Steiger, N. J., Emile-Geay, J. and Hakim, G. J., The utility of proxy system modeling in estimating climate states over the Common Era, *J. Adv. Model. Earth Syst.*, 8, 1164–1179, doi: 10.1002/2016MS000677, 2016

Breitenmoser, P., Brönnimann, S. and Frank, D, Forward modelling of tree-ring width and comparison with a global network of tree-ring chronologies, *Climate of the Past*, 10, 437–449, doi:10.5194/cp-10-437-2014, 2014

Dai, A., Trenberth, K. E. and Qian, T., A global data set of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming, *J. Hydrometeorology*, 5, 1117–130, doi:10.1175/JHM-386.1, 2004

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120, 2018.

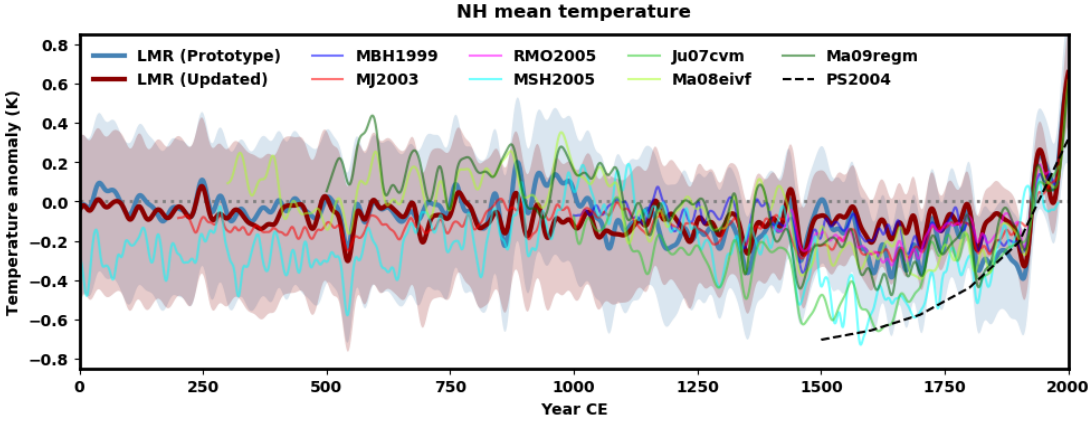

Fig. 1. Reconstructed Northern Hemisphere temperatures during the Common Era from LMR (prototype and updated reanalysis discussed in the original paper submission) and other reconstructions.

Printer-friendly version

Discussion paper

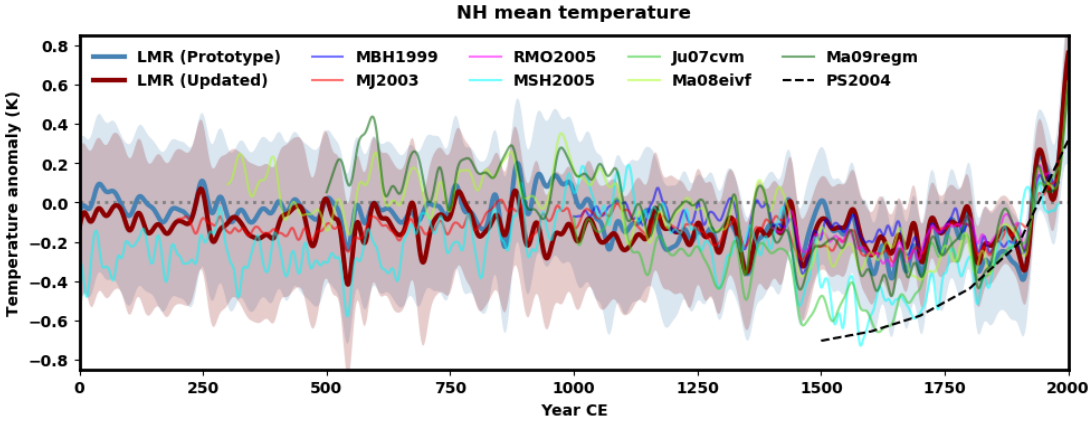

Fig. 2. As in Fig. 1, but with updated LMR results generated by applying some screening of the Breitenmoser et al. (2014) tree-ring chronologies.

Printer-friendly version

Discussion paper