Clim. Past Discuss.,
https://doi.org/10.5194/cp-2018-120-AC1, 2019

Climate of the Past Discussions

CPD

Interactive comment

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 1

We thank the referee for thorough and insightful comments on the manuscript. The comments have challenged us to take a more complete look at the numerous reconstruction experiments we performed, and as a result, we have gained a more comprehensive perspective on the results.

The paleoclimatic discussion needs to go more into depth and the authors need

Printer-friendly version

Discussion paper

to be more critical about their own results. The authors compare their new data set to the previous version (Hakim et al. 2016) and conclude this new version would be an overall improvement. This conclusion is based on validation statistics in the 20th century. However, the new data set lost all multi-decadal to centennial variability in the global mean temperature and does not show a warmer medieval period nor a cooler "little ice age" anymore. The authors do not discuss this issue at all. The paper suggests that this new reanalysis version would present the more likely global mean temperature evolution of the past 2000 years although it is in contrast to what most other reconstructions and paleodata records suggest.

We agree with the referee's suggestion that results should be framed in the context of other reconstructions, and a discussion focused on the long-term perspective (not limited to the 20th century) of the updated reanalysis be included in the manuscript. To address this issue, we have prepared figures showing comparisons over the entire Common Era of LMR results (updated reanalyses and prototype) and other available reconstructions of the Northern Hemisphere (NH) temperature. Figure 1 in this document shows a comparison of LMR results (prototype and updated reanalysis) with other published reconstructions. These chosen reconstructions are presented in the IPCC AR4 and AR5 reports, restricted to those representing the entire hemisphere and extending in time well into the 20th century. Also, a low-pass filter is applied on all results to better highlight low frequencies. First, this comparison shows that most other reconstructions are found within the bounds of the LMR ensemble, indicating a general agreement between the different products, at least within the bounds of uncertainty as defined from LMR. However, as pointed out by the referee about the reconstructed global-mean temperature (GMST) presented in the submitted paper,and also reflected in the NH-mean temperature results shown here, three periods with differences between the LMR reconstructions are highlighted. These are a colder medieval period in the updated reanalysis, most notably during the 875–1050CE period, and warmer temperatures during the 1600–1700CE and 1810–1920CE periods. We also note that the updated LMR is among the warm outliers during these cold periods compared to the prototype and most other reconstructions. Additional work has been undertaken to identify the reasons behind the above-mentioned differences. Initial findings are twofold.

First, we have concluded that the reconstructed colder temperatures during the medieval period, compared to the prototype LMR, are primarily rooted in the change from PAGES 2k Consortium (2013) to the more recent PAGES 2k Consortium (2017) proxy data. Indeed, it appears that a distinctly warmer medieval period isn't a prominent feature of the new collection, as indicated by the global temperature composites presented in PAGES 2k Consortium (2017). As this dataset reflects the community's most stringent evaluation of proxy records suitable for temperature reconstructions, we believe that the lack of a "classic" medieval warm period in our updated reconstructions of GMST and NH-mean temperature is not necessarily a glaring deficiency, but rather reflects the inherent ambiguities in defining this feature, as discussed in Diaz et al (2011).

Second, the referee's comment has motivated us to revisit our numerous experiments, as well as to perform additional test reconstructions, with a focus on results during the Little Ice Age (LIA). This exercise has allowed the identification of a notable sensitivity of reconstructed NH-mean temperatures to the set of assimilated tree ring width records (see the response to the next comment).

Following the referee's suggestion, a revised Section 3 will include a more complete discussion of these results, supported by a figure similar to Fig. 1 to better contextualize the updated LMR results against previous reconstructions.

The loss of low-frequency variability is most likely a consequence of the proxy data sets used, because this is the major change in the new version. Many of

the tree-ring chronologies in Breitenmoser et al. 2014 are not climate sensitive at all or moisture sensitive. As precipitation does not show any low-frequency variability in contrast to temperature, it is a logical consequence that using covariance information from moisture sensitive trees to correct temperature data leads to a loss of low frequency variability.

The referee raises some valid points here. Key elements within the data assimilation (DA) framework related to the possible issues raised here are the forward models (here the proxy system models, or PSMs) used to estimate proxy observations from a model prior, and the observation error variance assigned to the proxies, i.e. the R_k terms in equations 4 in the submitted manuscript. Within a DA framework, there are no fundamental reasons why the inclusion of records with sensitivities other than to temperature would lead to a deterioration to reconstructions of temperature, provided that PSMs properly account for the proxy sensitivities and observation error variances representative of the observation uncertainty are specified. However, it is acknowledged that these required conditions may not always be trivial to achieve. Nonetheless, initial LMR results (i.e the prototype) suggested that our approach has the ability to delineate proxy records with weaker sensitivity to climate by assigning relatively larger observation error variances (R_k), resulting in such records only weakly influencing the reanalysis results even though they are assimilated. Also, a motivation for developing the bilinear approach to model tree-ring width (TRW) data was to gain an ability to seamlessly handle the more complex sensitivities to temperature and/or moisture of these chronologies. With this approach, in principle, TRW records can be assimilated without having to make a binary decision whether each record is dominantly sensitive to temperature or moisture, or having to screen records out a priori. However, we acknowledge that relying on simple regression-based PSMs opens up the process to the influence of spurious correlations between noisy data. With a large number of proxies considered, some records will invariably be characterized by somewhat overestimated confidence, i.e. too-small error variance, and therefore overly weighted in the update.

Additional test reconstructions, performed in response to the referee's comment, suggest that this issue is partly responsible for the warmer conditions during the LIA. Figure 2 presents results from a new experiment where a reduced number of Breitenmoser TRW records have been assimilated (updated LMR in the figure). Only the records with a bilinear calibration correlation above 0.4 have been considered. A significant sensitivity of the NH-mean temperature during the LIA is observed, primarily resulting in cooler temperatures during the 1600–1700CE and 1810–1920CE periods compared to results in Fig. 2. As a result, this new reanalysis exhibits a greater consistency with the LMR prototype and other reconstructions during the 1500–1900 CE period. The absence of a prominent Medieval Warm Period (MWP) remains however, indicating that the inclusion of unscreened records is not the sole reason for the reduced low frequency variability in the updated LMR.

A second reason may be the use of proxy data with dating uncertainties, such as ice cores, in an annual reconstruction. These proxies probably do not have age errors in the 20th century validation period but become just noise if they have an age offset of one or a few years further back in the past. The authors just conclude that using moisture sensitive data leads to improved reconstruction skill, although this is only true in the 20th century validation period but not in the pre-instrumental period, most user of this data set will be interested in.

This is an interesting suggestion. We tested this hypothesis by performing an additional experiment in which all ice cores records were withheld from assimilation. The results do not support the hypothesis put forward by the referee however. GMST and NH-mean temperatures exhibit similar multi-decadal variability compared to reconstructions which include ice core information. The main difference consists of a modified long-term trend, showing a flatter temporal evolution over most of the Common Era prior to the 20th century warming, worsening the agreement with the other reconstructions. The primary role of ice core proxy data is to rather anchor the millennial-scale cooling characterizing the pre-industrial era.

In the current version, the global mean temperature evolution is the reappearance of the famous "hockey stick" in climate science. After all the discussion, the hockey stick was rising 20 years ago, I would not publish this as a state-of-the-art temperature reconstruction, especially not without a discussion and not if it is an artefact of unscreened input data.

We agree. See responses to the comments above. In the light of results outlined above, the revised manuscript will present an alternate updated LMR, showing a greater level of low frequency variability which exhibit a better agreement with other reconstructions. The final selection of this updated reanalysis will be informed not only by the level of agreement with other reconstructions, but on the basis of objective validation of results involving verification in proxy space using independent (i.e. withheld from assimilation) records. Progress has been made in the identification of an alternate configuration of the reanalysis, and results will be described in the revised manuscript.

I see two options, the first would involve minor revisions and the second major revisions: 1. It must be stated prominently (already in the abstract) that this reanalysis should not be used or considered to have the correct multi-decadal and centennial variability and that the global mean time series over the last 2000 year potentially has serious issues. The discussion needs to include all problems of data set, too and ideas how to overcome them in the future. In general, the paper should be put more into a context of methodological improvements to achieve better products in the future instead of claiming this would be nearly the prefect reanalysis for the past 2000 years. 2. A proper screening of the data needs to be introduced that prohibits the assimilation of non-climatic information, which has

just spurious correction with observations in the short window of overlapping instrumental and proxy data (a minimum of 25 data points has been used in this study, page 5, line 2). These records will have little weight in the assimilation procedure due to large residual variance, but hundreds of little errors probably produce significant noise. Probably, precipitation limited proxies and proxies with age errors have to be removed or treated specifically, too. These are just some ideas and it will need many improvements and new experiments to find and solve the problem.

We appreciate these comments and suggestions. As outlined above, we chose to follow a path inspired by the second suggestion. A complete review of results was undertaken and new reconstructions experiments have been carried out. The newly gained perspective has led us to consider alternate reanalysis configurations which address the issues identified by the referee. Suggestions by the referee were carefully considered and integrated in the design of our latest experiments. However, we wish to underline our firm belief that an approach that seeks to simply "prohibit the assimilation of non-climatic information" and remove information from precipitation limited proxies, as suggested by the referee, is not the preferred framework in which to seek improvements in LMR reconstructions. Rather, improved forward models (PSMs), describing more accurately the relationships between climate variables and proxies, in addition to improved characterization of observation errors, are the key aspects where improvements can be achieved. We hope to convince the referee that our work on the seasonal PSMs and the development of bilinear PSMs for tree-ring proxies are worthwhile contributions to the former, while further refinements to the latter are still needed. We now believe that an approach involving some screening of the proxies available in our database, preferably less stringent than suggested by the referee, represents a compromise between the need to incorporate useful proxy information while minimizing the adverse effects of noise in the data. Our efforts will be reported in the revised version of our manuscript.

A difficulty in the review process is that the input data has not been published, yet. Hence, it is not possible to properly judge the input data base. However, it appears to be basically the Breitenmoser et al. 2014 data set with a few coral and ice core records added. Why do you not simply refer to this first publication and give citations for the additional records or wait until the Anderson et al. paper is published?

The manuscript is now available online at:

https://datascience.codata.org/article/10.5334/dsj-2019-002/

In general, the decrease of skill further back in time is not discussed sufficiently.

This is a great comment. We will address this using our framework enabling an assessment of reanalysis performance in proxy space. The revised manuscript will include a more prominent presentation and discussion of proxy verification results in the main text. This characterization should also be framed in the context of uncertainties in reconstructions. This capability is enabled by LMR through the availability of ensemble member information. This complementary perspective will also be incorporated in the revised version of the paper.

It should also be discussed why forcings are not important and what the consequences of unforced simulations ensembles are for the final product, especially further back in the past when the proxy network is sparse.

We are not sure how the referee has come to believe that forcings are not important. In fact the model simulations from which we draw prior information do include

Printer-friendly version

Discussion paper

forcings, such as pre-industrial greenhouse gas and aerosol variability, including the effects of volcanic eruptions. This point is mentioned in the manuscript. However, to bring the context into greater focus, we have found that an important characteristic of prior information within our DA framework is the amount of variance characterizing the simulations. We have found that the greater variance generally characterizing the “Last Millennium"-type simulations (which include the forcings listed above) provide for more accurate reconstructions, compared to using simulations performed without the influence of external forcings (as in the “pre-industrial” or piControl CMIP5 protocol). We also have generally refrained from using simulations which cover the 20th century warming to dispel the notion that we are “cooking the books” when reconstructing temperature trends. In our framework, temporal information (trends) come entirely from weighted information from the proxies.

I suggest to evaluate the spatial skill of the reanalysis in the 20th century but with the spatial proxy network at multiple time slices, e.g. 0 AD, 500 AD, 1000 AD, 1500 AD.

This is a good suggestion. We intend to move, and expand, the proxy verification results that are currently in the supplementary material into section 3 of the main body of the paper.

Additionally, not using forced simulation offers the potential to use them in the validation procedure. It could be checked if temporal and spatial patterns of known past events or periods are well represented in the reanalysis, e.g. spatial moisture distribution after eruptions (Iles and Hegerl, 2015).

This is also a good suggestion, however we believe that such efforts are outside the scope of the current work. The suggestion will be considered in future efforts.

Finally, it would be interesting to see a map of the regression residuals to get an idea how many paleodata records have significant influence in the assimilation procedure and which are basically ignored because they have no climate information.

We agree, however a concise presentation of this is challenging, due to the varied nature of the proxy data. We will attempt to present this information as clearly as possible in a revised supplemental material. However we believe this would be addressed in a more informative way by a formal proxy impact study, which is intended to be the subject of another paper.

Additionally, I would like to know how many records in the PAGES2Kv2 data base have expert information on seasonality? I would be interested to read how well the expert-based seasonality in the PAGES data base agrees with the objective assessment in this study. Probably, the experts did a similar search for highest correlation, maybe just including more possible combinations of growing season months.

For the 2017 publication, the PAGES 2k consortium requested that each data certifier assess the seasonality of the temperature response and report its basis. In the LiPD format McKay and Emile-Geay (2016), this information is encoded in the `climateInterpretation_seasonality` and `climateInterpretation_basis` metadata properties. All records in the database include seasonality information, either as letters (JJA) or numbers ([6 7 8]) indexing calendar months. When the basis is reported, it is either from "first principles" (e.g.

trees are known to grow in local summer, which in most cases is synonymous with June July August), or from a search for the highest correlation. When the basis is not reported, the reader is referred to the publication documenting each record, which in most cases uses a mix of first principles and search for highest correlation, similar to the approach used in this paper. A key difference between the expert assessment of seasonality and the one done in our paper lies in the choice of target datasets. Studies focusing on individual series tend to be more careful about selecting an instrumental dataset appropriate for calibration (e.g. local GHCN station, rather than GISTEMP grid box average). These choices of target datasets are likely contributing to differences in the seasonal window determined via this process.

The choice of calibration datasets in LMR is driven by the need to uniformly process a large number of globally distributed records, hence the more general, perhaps not optimal selection as is possible when one has to consider a single or few records.

We also wish to point out that the information requested by the referee on the differences between the expert-based seasonality in the PAGES data base and objective assessment in this study is already provided in the supplemental information accompanying the main manuscript (Figure S1).

I was surprised to read that the authors use an extended fall period (JJASON)? Is there any reference for trees which are limited by climatic conditions in these autumn months.

Consideration of this period has been motivated following D'Arrigo et al (2005), and the fact that some seasonal responses found in the PAGES2k metadata extend to fall months, as suggested by the results shown in Figure S1a in the supplemental material accompanying our submission. It is interesting to note however that the objectively-derived seasonal responses determined by the approach described in our paper leads

Printer-friendly version

Discussion paper

to less emphasis on those fall months compared to the PAGES2k expert data.

It would be favorable to store the data at a world data center and not at a personal homepage.

We completely agree with the referee on this point. The LMR team has engaged interactions with the project's sponsor (NOAA) to identify a suitable storage location and access point, but these have yet to be identified, hence the reference to a personal homepage at this point.

Technical corrections

Abstract: skill score increase in percent is misleading. It is easy to have a large relative increase if scores were very low in the comparison data set, e.g. in Z500 where CE improves from very negative to less negative the increase in percent is large but the skill is still negative!

Point well taken. However, our emphasis has been about quantifying differences with respect to our main benchmark, the LMR prototype, to highlight improvements. Therefore we remain convinced that the formulation used is appropriate. The fact that some skill scores remain characterized by negative values is not hidden and becomes quite clear in the core of the text.

Abstract: be more precise what is meant with "ensemble characteristics".

We agree that this concept should be more accurately defined. It will be rectified in the revised manuscript.

Introduction, Line 17: apart from paleoclimate with annual observations, the forecast model is a third important component (this is even written in the Methods section).

We agree that this statement in the Introduction could be interpreted to mean that the prior data has less importance in the DA context. A statement more accurately conveying the importance of the various components will be included in the revision.

Methods, Page 7, line 6ff: I do not understand why the calibration is done with a different gridded data set than the validation. Both data sets a based on largely overlapping instrumental observations and therefore clearly not independent.

The use of GISTEMP as the calibration dataset has largely been motivated by the fact that a larger number of proxy records could be calibrated (larger number of records with sufficient overap with valid calibration data) compared to other datasets. Therrefore, a larger number of proxies may participate to the reanalysis. We in fact perform validation using all datasets at our disposition, including the calibration dataset. Skill metrics show small differences among the various results, but remain in general agreement. Here we have chosen the Berkeley Earth dataset because it provides a greater spatial coverage, comparable to the calibration dataset, therefore providing a larger sample of spatial verification results.

Methods, Page 4 line 14: How many ensemble members?

Printer-friendly version

Discussion paper

The information is provided later in the paper, in Section 2.3, where details of the configuration are listed.

Methods, Page 4 line 14 it is written that Hakim et al. 2016 worked "with the same randomly drawn ensemble members used for every year in the reconstruction", whereas on page page 5, line 28 it is written for this study: "each using a different randomly chosen 100–member ensemble". Are both studies consistent and is each year build on different randomly chosen ensemble members?

Both statements are consistent, in that the first statement describes how the ensemble members for a given reanalysis realization are used to populate prior information in the time domain, whereas the second statement points out that different sets of states are used to populate the prior ensemble for different Monte-Carlo realizations of reconstructions. We will attempt to clarify this in the revised manuscript.

Page 5, line 1: "Only records for which a PSM can be established are shown ...". What do you mean by "shown"? There is no reference to a figure. Do you want to say that only records meeting these criteria are assimilated?

We wish to point out that a reference to Figure 1 is included in the previous sentence. We will modify the text to make sure the reader is not confused about what we are refering to in this sentence.

Methods, Page 5, line 8: Breitenmoser et al. 2014 is not screened for any climate sensitivity.

Point taken. We will modify the text accordingly.

Methods, Page 6, Proxy modeling: It should be repeated here over which period the regression coefficients are calculated. As many data points as available in the overlapping period with instrumental data but minimum 25 pairs of x and y?

This information will be clarified in the revised manuscript.

Methods, Page 7, line 30: Is there a reference that any tree-ring proxy responds to an extended fall period (JJASON)? The given references point to common growing seasons from May to August in the northern hemisphere. Why are not all combinations of growing season length tested and the optimum is chosen? In the PAGES data base there are also various different length of growing seasons defined.

For the first part of the question, see the reply provided earlier in this document. To address the second part, we concede that our objective seasonal PSM calibration methodology is a compromise between comprehensiveness and efficiency in the calculations, performed over more than 2000 proxy records. Nonetheless, we believe that while perhaps non-optimal, the resulting characterization provides realistic results, as suggested by the fact that more accurate reconstructions are obtained when this information is used in forward modeling tree-ring records. This outcome will be further supported in the revised manuscript by presenting verification results performed in proxy space using independent records.

Methods, Page 8, line 26: “local” should better be “grid box”.

The text will be modified accordingly.

Updated reanalysis, Page 9, line 5: Can you explain the localization better? Does a cut-off radius of 25000 km mean that each proxy influences basically the entire globe?

We will add additional clarifying information about the covariance localization configuration used to generate reconstructions and its influence on how proxies are informing the reanalysis.

Updated reanalysis, Page 9, line 7: Mention that the reference for the skill score is climatology.

We agree. The text of the revised manuscript will be modified accordingly.

Figures: some text is too small that it cannot be read in a print version.

We will revise the final figures to only present the most useful information on the figures.

Figures: figures should have consistent font types.

Efforts will be undertaken to use uniform font types.

References

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

McKay, N. P. and Emile-Geay, J., Technical note: The Linked Paleo Data framework – a common tongue for paleoclimatology, *Clim. Past*, 12, 1093–1100, doi:10.5194/cp-12-1093-2016, 2016.

Diaz, H. F., Trigo, R., Hughes, M. K., Mann, M. E., Xoplaki, E. and Barriopedro, D., Spatial and temporal characteristics of the climate in medieval times revisited, *Bull. Amer. Meteorol. Soc.*,99(11), 1487–1500, doi:10.1175/BAMS-D-10-05003.1, 2011

D'Arrigo, R., Mashig, E., Frank, D., Wilson, R. and Jacoby, G., Temperature variability over the past millennium inferred from Northwestern Alaska tree rings, *Clim. Dyn.*, 44, 227–236, doi:10.1007/s00382-004-0502-1, 2005

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

Printer-friendly version

Discussion paper

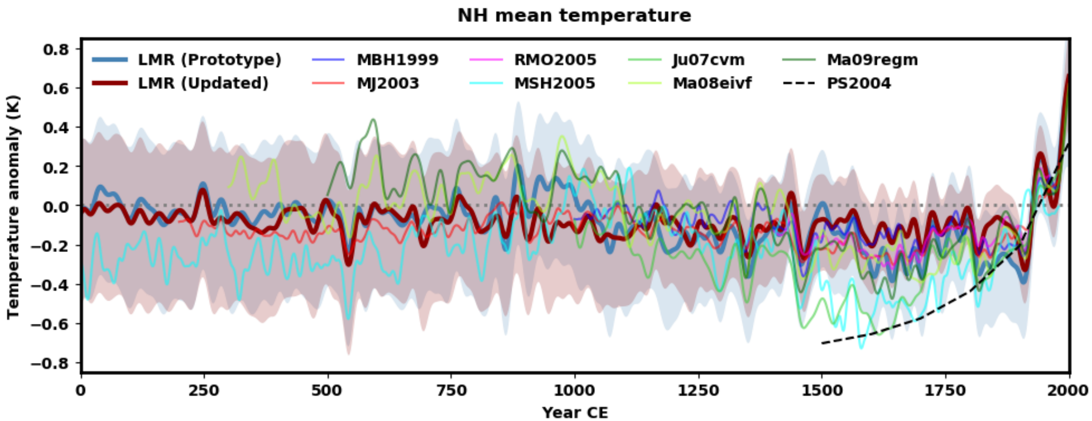

Fig. 1. Reconstructed Northern Hemisphere temperatures during the Common Era from LMR (prototype and updated reanalysis discussed in the original paper submission) and other reconstructions.

Printer-friendly version

Discussion paper

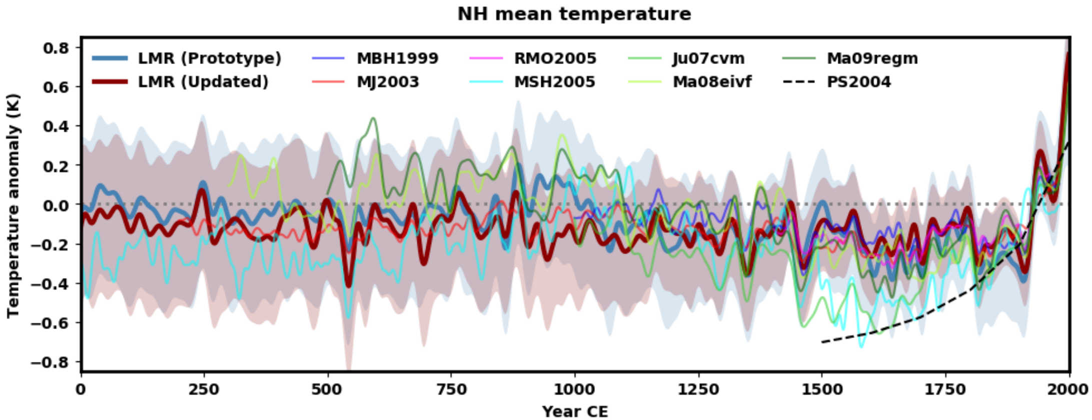

Fig. 2. As in Fig. 1, but with updated LMR results generated by applying some screening of the Breitenmoser et al. (2014) tree-ring chronologies.

Printer-friendly version

Discussion paper