**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.

Anonymous Referee #3

Review of Tardif et al.

Summary: This paper presents various sensitivity and verification tests of an updated LMR product targeting multiple climate state variables over the Common Era. The updates to the product include both methodological choices (seasonal assessments of the proxy sensitivities and tests of multiple statistical PSM formulations for univariate and bivariate dependencies) and increases in the proxies included in the network. While the abstract gives the impression that the paper serves as a the basis for the public release of a new LMR product, the paper is generally as I have stated: a large number of sensitivity and verification tests that feel more exploratory than definitive.

General remarks: 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 multiple ad hoc methodological choices, none of which are reasonably justified or widely tested (which strikes me as strange given the extensive number of sensitivity tests that the authors have performed). I expand on each of these points below.

1. 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). 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. This is not only to place their results within the now lengthy body of work in this area, but 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. Incidentally, 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.

Similar to the above suggestions, 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.

2. Here is a quick list of choices the authors have adopted that are not accompanied by any justifications or sensitivity discussion:

a. Use of 100-member ensemble

b. Use of the CCSM4 last millennium simulation as the prior

c. Use of 51 MC realizations

d. Use of a proxy sampling scheme based on 75% of the proxy records

e. 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.

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.

A few small details are also worth noting:

Pg. 2, Ln. 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.

Pg. 2, Ln. 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?

Pg. 8, Ln. 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.

Figures: In general, there is a lot of small text in the figures that is hard to read and also rather confusing and messy. This could be cleaned up a lot and the digestibility of the figures could be improved. The many colorbars are also unnecessary in many plots when one would do.