**Climate
of the Past
Discussions**

Interactive
Comment

# *Interactive comment on* "The extra-tropical NH temperature in the last two millennia: reconstructions of low-frequency variability" *by* B. Christiansen and F. C. Ljungqvist

**Anonymous Referee #2**

Summary

This manuscript presents two reconstructions of extratropical (30-90N) NH mean temperatures, one that spans the entire Common Era and another that extends to 1500 CE. This work is an extension of a series of papers by Christiansen and (in some cases) Ljungqvist to develop/apply the LOC method for temperature reconstructions spanning all or parts of the CE. Generally speaking, it is a well-written and organized paper, is a logical extension of their previous work, and does a thorough job of presenting the employed proxies and, to a lesser degree, in testing and presenting their derived reconstructions. I nevertheless have several major concerns that must be addressed in

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

a revised version of the manuscript. If addressed appropriately, the manuscript will be suitable for publication in Climate of the Past.

General Points

1. As mentioned above, this work is an extension of previous papers that have sought to use the LOC method to reconstruct large-scale temperature indices. This paper is nevertheless a close extension of the previous Christiansen and Ljunqvist (2011a) paper, and a critical reader might wonder whether this latest effort warrants an additional publication. I think it does, but the difference between the previous publication and the current one could be more clearly delineated. There have indeed been more proxies added and the reconstructions are now broken into full CE and 500-yr intervals (the previous work provided a 1000-yr reconstruction), but the authors also provide some additional analyses that were not done in their earlier publication. These include additional sensitivity tests on their reconstructions and some spatial analyses of the proxy reconstructions. All of this should be clearly pointed out at the onset, in addition to the fact that more proxies are used, to make clear how this effort represents an advance from the earlier publication. I would also encourage the authors to include a graphical comparison of their 1000-yr reconstruction to the new reconstructions. They describe the comparison in words, but it would be very helpful to have a direct comparison in a figure.

2. The authors' failure to provide any validation statistics is a serious flaw. They have performed sensitivity tests throughout the manuscript, and while important, amount to in-sample evaluations of the mean estimate. It is very important to compare the reconstruction to an out-of-sample estimate of the target time series, and the authors simply don't do this. There are several things that are needed. First, the target time series should be shown in all of the figures that present reconstructions (figs. 5-8). This alone will provide a visual sense of how well the reconstructions compare to the target. Note that one of the chronic problems with LOC reconstructions is a clear disagreement with the derived reconstruction and the target NH mean timeseries in their period of overlap

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

(despite good agreement in pseudoproxy tests). Secondly, a collection of validation statistics must be provide for the reconstruction. There is a host of these to choose from, and the authors need to decide which ones they believe are appropriate. I notice that this comment comes up in another review of the manuscript, and the authors have dismissed it based on the fact that their reconstruction is less reliable at high frequencies. There are several points here too, which I outline in subsections below.

2a. How unreliable are the high frequencies and what is "high frequency" in this context? The authors first need to more clearly discuss the potential problems with the high frequencies that have been highlighted in some of Christiansen's earlier pseudoproxy work. They could also show validation statistics in different frequency bands that would give a sense of how the reconstructions perform in various frequency domains. It is well understood that some of the lower frequencies are harder to validate with the short instrumental record, but this is no excuse for not looking into it. It is contingent upon the authors to convince the reader that their reconstruction carries some merit, and it is very hard to do so without some reasonable validation exercises.

2b. If the high frequencies are unreliable, why discuss them and use them in the various analyses and figures? The authors could present only a lower resolution reconstruction, which they have effectively done in their filtered time series and confidence intervals. There is precedent for this in the literature (e.g. Hegerl et al. 2007) and the reconstruction would still be of great value. I simply see no reason to present the high-frequency information if the authors themselves consider it dubious. Note that I myself am very suspicious of the high-frequency component of the reconstruction based on the demonstrated tendency of the LOC method to blow up at frequencies with periods close to the sampling interval of the time series (Christiansen 2011a).

3. Building further off of the validation exercises above, I am surprised that the authors do not plot their reconstructions with other published NH temperature reconstructions. There are several that have been published in the last few years that show increased low-frequency variability and it would be worthwhile to compare them graphically (some

comparisons are discussed verbally in the Conclusion, but of course a picture speaks 1000 words...). These reconstructions include the Esper et al. (2002), Moberg et al. (2005), Hegerl et al. (2007) and Mann et al. (2008) reconstructions (there is also the Kaufman et al. (2009) Arctic reconstruction that may be applicable here). In particular, I am quite surprised that the authors do not even cite the Hegerl et al. (2007) paper. This work is closest in methodological provenance (basically the LOC method, but using hemispheric instead of local calibration) and targets the same geographic domain and time periods (even broken up similarly as the authors do here). I also expect that many of the same proxies used in the Hegerl et al. (2007) study are included in the current work (although the present study includes more). Bottom line: the authors should graphically compare their reconstruction to some previous efforts, and spend some time specifically comparing and contrasting with the Hegerl et al. (2007) reconstruction because of its strong similarities to their own work.

4. There are two methodological issues that are not clearly resolved in the manuscript. Given the mention of different temporal resolutions for the proxies, it is not clear how the proxies with different resolutions have been blended. There is mention of interpolations within the low-resolution time series, but it is not clear how this entered in to the local calibrations. Were the proxies calibrated on degraded instrumental data first and then interpolated to combine the mean? How was this done? All of the reconstructions are presented as annual, but not all of the proxies have such high resolution. This is confusing and the details are not discussed. It would also be nice to see how much the reconstructions depend on proxies with different resolutions. An additional sensitivity test, to the degree that it is possible, would compare reconstructions using only annually resolved proxies and one derived from lower-resolution proxies. The second methodological issue involves the use of pseudoproxies to generate the confidence intervals. The pseudoproxy construction is undoubtedly more simplistic than the actual proxy noise characteristics (which likely have both multivariate and non-stationary influences). The confidence limits in this case are therefore likely to be more optimistic than reality. The authors address this tangentially by comparing reconstructions de-

rived from subsets of proxies and conclude that the effect of low-frequency noise is small. I am dubious and think that the authors should be more cautious in their characterization of the pseudoproxy estimated confidence intervals (and probably expand the description of the method on pg. 3999). They should also compare their estimate to a more traditional confidence interval estimate, e.g. one from a residual analysis.

Minor Points

Pg. 3992, Ln. 23: I think Common Era is being more widely used as a precise designation of the last two thousand years. Consider adopting this terminology throughout the manuscript instead of using late Holocene.

Pg. 3993: There is a distinction between reconstructions that target large-scale means and spatially resolved reconstructions that target temperature patterns. This distinction is not discussed and is important in the context of the discussions here (and on pg. 3994). This should be clarified.

Pf. 3993, Ln. 16: "Unfortunately, there still exist no universally accepted chronological definitions..." The reason for this is basically given in the following sentence, i.e. it is not clear that these epochs were temporally synchronous throughout the globe. It therefore is difficult to assign a specific period during which these events universally occurred. This should be pointed out more clearly.

Pg. 3993, Ln. 27: Please clarify why you have concluded that the amplitude of the LIA is the "biggest uncertainty in the climate of the millennium."

Pg. 3994, First Paragraph: The authors fail to mention several methods that have been shown to avoid underestimation in trends and low-frequency variability (Mann et al. 2008, Hegerl et al. 2007). The difference between index and field reconstruction is also relevant here. Smerdon et al. (2011) have shown that these problems may exist in field reconstructions even if they do not in the composite mean indices. A more comprehensive and balanced discussion should be included here.

Pg. 3994, Ln. 23: "...all previously have been shown to respond to temperature." I would explicitly point the reader to the references in Table 1 here.

Pg. 3994, Ln. 25: It is mentioned that the addition of more proxies is expected to reduce the confidence limits. Do they? There is no later comparison to indicate whether this is the case.

Pg. 3995, Ln. 14: The log-transforms need to be clarified and included in the aforementioned discussion about how this particular resolution was included to make annually resolved reconstructions.

Pg. 3995, Ln. 23: It is argued that well sampled regions can capture well the trend and amplitude of extratropical NH means. Why not simply compare the NH mean computed from the entire instrumental temperature field to a mean computed from only those instrumental grid cells that include proxies? This would give an explicit test, within the calibration interval, of how well the extratropical NH mean is represented from the spatial sampling of the proxies.

Pg. 3996, Ln. 15: "We have confirmed that the outliers have only marginal influence on the NH mean reconstructions." How?

Pg. 3996, Ln. 29: The authors single out a single tree-ring chronology as not available. This seems unnecessary. I also do not understand the logic behind why several reconstructions using documentary data cannot be used.

Pg. 3998, Ln. 27-29: It is ambiguous here whether detrending is common practice in the field or not. Please reword.

Pg. 3999, Ln. 8: Please provide a reference for the seasonal correlation being high on decadal and longer time scales.

Pg. 4000, Ln. 15: If the proxies are serially correlated, why not include this in the significance estimates? How many proxies would be additionally excluded if the significance estimate was adjusted for serial correlations?

Pg. 4001, Ln. 5: The authors only tangentially mention that a NH mean computed later in the 20th century would change their relative comparisons. They should address this more specifically because many of the MCA temperature comparisons consider the late 20th century for comparison. It should either be stated why the authors do not consider the later decades of the 20th century using the instrumental data or the comparison should be explicitly done.

Fig. 2: The panels are very difficult to read (particularly the headings and tick labels) and the vertical axes are all in mixed units (with no axes labels). This later point in particular should be fixed.

Fig. 9: Top panels are anomalies relative to what baseline? The calibration period?

References (not in the paper)

Hegerl, G. C., T.J. Crowley, M. Allen, W.T. Hyde, H.N. Pollack, J. Smerdon, and E. Zorita (2007), Detection of Human Influence on a New, Validated 1500-Year Temperature Reconstruction, Journal of Climate, 20(4), 650-666.

Smerdon, J.E., A. Kaplan, E. Zorita, J.F. Gonzalez-Rouco, and M.N. Evans (2011), Spatial performance of four climate field reconstruction methods targeting the Common Era, Geophysical Research Letters, 38, L11705, doi:10.1029/2011GL047372.

Interactive comment on Clim. Past Discuss., 7, 3991, 2011.