



Article being discussed

Title: Global temperature changes of the last millennium

Author(s): Ronan Connolly and Michael Connolly

Version number: 0.1

Comment/review being discussed

Craig Loehle

FEBRUARY 27, 2014 AT 3:23 PM

This is a thorough and comprehensive review and commentary. I would just add a few things. The Loehle work is one of the few which does not heavily weight certain series, and therefore avoids the data mining/spurious correlation problem. A little more emphasis could be given to the failure of the various reconstructions to statistically match each other (and therefore they do not corroborate each other).

Author response

Authors: Ronan Connolly and Michael Connolly

Craig Loehle points out that some of the global temperature proxy reconstructions use reconstruction methods which heavily weight certain proxy series relative to other series, and that this could introduce data mining/spurious correlation problems.

There are certainly merits to the approach of applying a simple “composite-plus-scale” averaging method. If the “temperature signal” in the various temperature proxy series is as strong as is sometimes implied, then this presumably should be sufficient to at least give a *reasonable approximation* of past temperature trends.

Much of the difference between some reconstructions does indeed seem to be related to the averaging method used, e.g., in Figures 10-12 of the article, compare the three Shi et al., 2013 reconstructions which all used the same proxy dataset; or compare the two Mann et al., 2008 and McShane & Wyner, 2011 reconstructions which all used the same dataset; or compare the Ljungqvist, 2010a; Christiansen & Ljungqvist, 2011; Christiansen & Ljungqvist, 2012 reconstructions which used fairly similar datasets but quite different reconstruction methods.

This indicates that:

- (a) Whatever temperature signals might be in the various proxy series, they are apparently not very strong and/or very consistent.
- (b) The choice of reconstruction method can substantially alter the estimates.

However, if the temperature signals are as weak as this implies, then it is not clear that the simplest reconstruction method is necessarily the most reliable. If the temperature signals are very weak, then it may well be that relatively complex methods are required to extract these signals:

“Everything should be made as simple as possible, but no simpler” –attributed to Albert Einstein by Roger Sessions [1].

Unfortunately, the weaker the signal and the more complex the method, the more likely it is that the method will falsely report spurious statistical artefacts and/or noise as “signal”. That is, as Craig Loehle suggests, inappropriate weighting might be leading to data mining/spurious correlation problems.

With this in mind, we propose including a new short subsection outlining and comparing the different reconstruction methods used by the studies – perhaps a new subsection between the current subsections 2.1 and 2.2.

Having said that, while the reconstruction method does seem to have a substantial influence on the resulting estimates, we do not believe it is the most important factor. For instance, just like the Loehle, 2007 estimate, the Jones et al., 1998 and Briffa, 2000 also use a simple composite-plus-scale reconstruction method which gives equal weighting to all series. Yet, it can be seen from Figures 10 and 12 of the article that each of these estimates is quite different. This indicates that this approach is highly dependent on proxy series choice.

Indeed, on his Climate Audit blog, McIntyre has shown how simple (yet not completely unreasonable) substitutions of different proxy series in the Jones and Briffa estimates can yield dramatically different reconstructions [2,3].

In Sections 3.4 and 3.5, we discussed the problem of inconsistency between the different proxy series. On lines 1037-1066, we mentioned the importance of carrying out rigorous “sensitivity studies”. However, this really is such a major problem that it might be worth emphasising it a bit more.

Choosing which types of proxy series to include in a reconstruction is itself quite subjective.

For instance, Briffa, 2000; Esper et al., 2002 and D’Arrigo et al., 2006 all limited their proxy choices to tree ring proxies. On the other hand, Loehle, 2007 specifically avoided tree ring proxies. These choices were often deliberate, and often justifications were given. Yet, clearly, the researchers involved differed in their opinions over the most relevant proxies to use.

But, even within the different types of proxy, there is a huge degree of subjectivity. This can occur both on the part of the researchers constructing the reconstruction, and on the part of the researchers developing the individual proxy series:

- (a) For instance, Figure 4 in the article shows two different versions of the Polar Urals tree ring proxies (the Briffa 1995 version and the Esper 2002 version). Both of these proxies suggest substantially different temperature trends. Moreover, the various versions of the Yamal tree ring proxy series are from essentially the same region (NW Siberia) as the Polar Urals, yet they imply different trends again.

So, if a researcher chooses one proxy out of: Polar Urals (1995), Polar Urals (2002), Yamal (2000), Yamal (2013), or some other further proxy, this could substantially alter the final reconstruction. These decisions are obviously subjective, and different researchers might have different opinions as to which (if any) to include.

- (b) When a researcher is developing individual proxy series, they may succumb to confirmation bias (whether consciously or unconsciously) in deciding whether the series is reliable or not, depending on how closely the series matches the “expected” temperature trends. Proxies which do not match these expectations might be unfairly discarded – the so-called “File-Drawer problem” [4].

In our opinion, these problems deserve much more robust and extensive discussion amongst the paleoclimate community than they have received to date. There has been some brief discussion in the literature, e.g., the comments arising from Osborn & Briffa, 2006 [5], i.e., Bürger, 2007 [6] and Osborn & Briffa, 2007 [7]. However, this should be a top priority.

We think two independent, yet complementary, approaches could be particularly useful in tackling these problems in future reconstructions:

- (a) Providing much more discussion of the justifications for choosing/discarding particular proxy series (perhaps as supplementary information if there are a large number of series involved).
- (b) Carrying out much more extensive and rigorous “sensitivity studies” in which the effects of removing/introducing/substituting each of the relevant proxy series are carefully documented. We think sensitivity experiments along the lines of those carried out by McIntyre on the Climate Audit blog (e.g., [2,3]) should become routine.

In version 0.1, we alluded to these issues, but given their importance, maybe it might be worth adding further emphasis. With that in mind, we propose including an additional paragraph or two into Sections 3.4 and 3.5 further highlighting the significance of these points.

Finally, Craig Loehle also suggests that we should provide more emphasis to the substantial statistical differences between the various reconstructions. By grouping the reconstructions into three separate groups based on the CWP:MWP ratio (Figures 10-12 in the article), we felt that this graphically illustrated that there were considerable inconsistencies between the various reconstructions. This seems to be a new approach, as previous reviews have tended to group all reconstructions into a single graph and/or exclude certain reconstructions.

We also noted the inconsistencies between the different reconstructions over the relative warmth of the 1400s, and over the length and magnitude of the “Little Ice Age”. In addition, in the conclusions (Section 6), we highlighted some reasons to be cautious over even those portions of the reconstructions which appear to agree with each other (i.e., the existence of MWP/LIA/CWP periods).

Still, we agree that it might be useful to explicitly highlight the fact that, when the reconstructions are compared in detail, there are substantial differences between all of the reconstructions – indicating that they are not in as much agreement as is often assumed.

Therefore, we propose adding a paragraph or two to the text of Sections 5 and 6 so that we are comparing *and contrasting* the reconstructions instead of just comparing them.

We would appreciate feedback from Craig Loehle and anyone else on whether these would be useful and sufficient modifications, or if alternative/additional modifications should be made.

References

- [1] Quote Investigator (2011). URL: <http://quoteinvestigator.com/2011/05/13/einstein-simple/>
- [2] S. McIntyre (2006). "Briffa 2000 and Yamal". Climate Audit. URL: <http://climateaudit.org/2006/06/01/briffa-2000-and-yamal/>
- [3] S. McIntyre (2008). "Jones et al 1998: Impact of new versions". Climate Audit. URL: <http://climateaudit.org/2008/08/04/jones-et-al-1998-impact-of-new-versions/>
- [4] J. D. Scargle (2000). "Publication bias: The 'file-drawer' problem in scientific inference". *J. Sci. Exploration* 14, pp 91-106. URL: http://www.scientificexploration.org/journal/jse_14_1_scargle.pdf
- [5] T. J. Osborn and K. R. Briffa. "The spatial extent of 20th century warmth in the context of the past 1200 years". *Science* 311, pp 841-844. DOI: [10.1126/science.1120514](https://doi.org/10.1126/science.1120514)
- [5] G. Bürger. "Comment on 'The spatial extent of 20th century warmth in the context of the past 1200 years'". *Science* 316, p 1844. DOI: [10.1126/science.1140982](https://doi.org/10.1126/science.1140982)
- [7] T. J. Osborn and K. R. Briffa. "Response to comment on 'The spatial extent of 20th century warmth in the context of the past 1200 years'". *Science* 316, p 1844. DOI: [10.1126/science.1141446](https://doi.org/10.1126/science.1141446)

(All online references checked 9th April 2014).