Clim. Past Discuss.,
doi:10.5194/cp-2016-98-AC2, 2017

Climate of the Past
Discussions

Interactive comment on "Testing the analog method in reconstructing the global mean annual temperature during the Common Era" *by* Juan José Gómez-Navarro et al.

Juan José Gómez-Navarro et al.

jjgomeznavarro@um.es

Received and published: 13 January 2017

Dear editor of Climate of the Past; I have carefully read the manuscript entitled "Testing the analog method in reconstruct- ing the global mean annual temperature during the Common Era" by Gomez-Navarro et al, submitted for publication in Climate of the Past.

The paper proposes to test the suitability of the Analog Method (AM) for producing climate field reconstructions (CFR) of mean annual temperatures across the globe. The manuscript does not intend to fully compare pros and cons of the AM-CFR method with those of other methods available for producing CFR (eg regression-based methods orREGEM algorithm). Instead, the authors make use of pseudo-proxies experiments (PPE) at various degrees of signal degradation (with climate models being the original

reference signal) to pinpoint and eventually attribute problems relating to the method itself and to the quality / diversity / scarcity of the proxies used to find temperature analogs.

The paper is fairly well written and logical. It builds a strong and convincing argument and clearly shows that the AM-CFR method is well suited to produce spatially complete climate CFR reconstructions at the global scale. The paper builds on a strong literature review where the various concepts associated with CFR reconstruction are explicitly and precisely detailed. The mathematical demonstrations associated with the CFR method are precise, although sometimes not always necessary (correlation eq and RMSE Eq), but still, they reflect the high level of technical know-how of the authors with respect to AM-CFR approaches. I am confident that the paper will become and important reference for the use of CFR for global scale reconstructions

I have a few comments listed below, I would invite the authors to reply to them in order to clarify some aspects of the manuscript, most particularly the discussion of concepts relating to the AM method. The comments are somehow minor, but I expect the authors to address them in order to improve the general quality of the manuscript.

The authors are grateful for the time devoted to the review of the article and the positive view of it

1) Spectral signature of proxies vs PPE. One of the most important finding in this article is without a doubt the fact that the AM method performs better when PPE are used instead of real world proxies. Indeed, real-world proxies seem to be noisier than expected and for that reason the search for analogs tends to be less accurate (than that performed with PPE), resulting in estimations that are less well correlated with observed MAT. However, a major difference between PPE and real-world proxies is that real-world proxies commonly encompass only parts of the full spectrum of variability of the MAT. For example, tree rings, especially when severely standardized, relate to high frequency variations in MAT and while a correlation exists with that climatic field, *it reflects mostly the correlation of high-frequency periodicities. Contrastingly, low frequency proxies (such as pollen) may correlate with MAT, assuming that the correlation is a valid statistic in the presence of auto-correlated series. –and this is not a trivial assumption-. Still, if significant, the correlation would reflect only the low frequency component of the variability in MAT. On top of this, the signature of high-frequency signals probably reflects characteristic of local to regional climate dynamics while lower frequency variations common to proxies and MAT (whether they are internally or externally forced) tend to be visible at larger spatial scales. (and that would make it easier for the AM method).*

On the contrary, PPE resulting from the degradation (addition of white noise) of original climate fields might better preserve the original spectrum of variability characteristic of temperatures series because the added noise tends to propagates evenly throughout the spectrum of variability of original climate fields. In short, I find the discussion (L. 505) around the fact that real-world proxies exhibit lower than expected correlations values with MAT rather incomplete. I invite the authors to fully explore other factors responsible for this drop in correlation and to include these in the discussion. I suggest, at first glance, to first ensure that PPE, real-world proxies and components. I fear that considering only the correlation (between proxies and MAT) when generating PPE might cast shadow on other sources of noise and variability inherent to the proxies and might be one of reasons why the AM performs less well with real-world proxies than with PPE

This is an excellent suggestion, although some members of the dendroclimatological community may disagree on the extent of variance loss at low-frequencies by tree-ring proxies. Relatively modern standardization methods, like RCS, claim a much better variance preservation properties than 'classical' standardization methods. However, the question posed by the reviewer is indeed relevant, since it is not always clear what the spectral properties of proxies, and of their non-climate noise, are. According to previous publications with very a similar methods, but for reconstructing regional pre-

cipitation over Europe, addressed this issue (Gómez-Navarro et al., 2014; cited in the manuscript). The authors found that the difference between white and red noise is in the design of PPE is rather minor. Still we will explore and report the results in the revised version including additional PPE with non-white noise, trying to mimic the behavior of proxy records. Note however that this will always be incomplete, as it is difficult to cover all possible types of proxies available and the wealth of varieties of possible non-climatic noise.

2) Estimation of analogs with low-resolution proxies. The PAGES-FULL dataset contains proxies covering the last 2k around the world. Most of these proxies are highly resolved tree ring series that are clustered in few sites such as the Himalayas, subarctic Canada, western Europe and the Andes. The rest of the proxies found in PAGES–FULL are rather poorly resolved proxies with uncertain dating (pollen and others), during the Common Era. Those proxies contain numerous missing values (actually probably more missing values than observed values during the CE). Early in the manuscript (L. 180) the authors state that they interpolate between observed / dated values to emulate an annual resolution. But I fear that once these series are interpolated, they remain low-resolution proxies in the sense that each year is strongly autocorrelated to the previous or following one and so on, while the high frequency component remains inexistent. When strongly autocorrelated proxies are used to calculate analogs, correlations between observations and predictions are often overestimated (see Guiot et al 2010). This is because the best analog is most often found just before or after the given year for which an analog is required. Consequently, this property does not guarantee that an independent analog could be found for a more distant period. Without a proper assessment of this problem in the paper, I fear that low frequency proxies could contribute to artificially "boost" the correlations between predicted and observed MAT. As a matter of fact, figure 8 shows that the addition of a few low frequency proxies from PAST-SEL to PAST-FULL (N passing from 514 to 641 proxies) actually seems to boost prediction accuracies significantly especially at places where very few proxies are available. What part of this increased prediction accuracy comes from the above-mentioned effects?

We believe that the point raised by the reviewer is not really an issue in this study. It is important to recall that the low frequency in the proxies is in principle unrelated to its counterpart in the pool (actually the pool has no temporal order by definition, so the concept of autocorrelation does not apply). Further, the search of analogs is carried out independently in each time step, disregarding previous time steps. Thus, the fact that some proxies might have artificially high levels of autocorrelation does not boost artificially the correlation between the target and the analog. This is especially true in the case of PPE, where such artificial autocorrelations do not exist and no interpolation is carried out to remove missing values.

3) Not clear how the AM-CFR achieves extrapolation and how reliable it is. At many occasions, the authors claim that the AM method is able to "extrapolate" in order to produce spatially complete MAT fields. I am not sure how this word is used here. To extrapolate means making a prediction "above the limits" of a given calibration dataset. If this is done in the present paper, I request that the authors produce additional analysis to demonstrate how and how well this is done. However, as I understand it (and maybe I am wrong here), the AM cannot extrapolate over the range of observed variability. By definition, an analog is a year (or a pool of years) observed during the Common Era where proxies are similar to those observed in the past, similarity being of course measured by some distance metric. So, in other words, the analog must exist during the Common Era for it to be transferred to the past. Consequently, an analog cannot extrapolate over the bounds of the variability in MAT. On the contrary, the search for analogs is constrained within the bounds of observed variability, resulting in a native incapacity to extrapolate. Therefore, I thus suggest to replace the word extrapolation by the word "prediction" which is more what MA actually does. Following that same idea, I wonder what would be the consequences of an inability of the AM method to extrapolate over the range of observed variability. Since the method aims at producing climate field reconstructions, would it be well suited to reconstruct periods (eg the MWP) that could have been warmer (even locally) than the Common Era? What about the LIA? Would the method be able to reconstruct periods that are much

colder that any of the last decades? So given this discussion, is the method doomed to be extremely conservative? I would suggest to test that, for example, by removing the anomalously warm last decades visible in figure 9, calibrating the AM say during the 1911-1980 period, and predicting MAT for the 1980-201X period. Since this latter period is censored off from the calibration period, is the AM method able to predict it or does it simply underestimates it? This could orient the discussion on the capacity of the AM-CFR method to extrapolate

We believe this issue is rather a misunderstanding, probably caused by not using the term "extrapolation" carefully enough. We do not mean extrapolation in a temporal sense, and certainly the analogs are just pool members (or plain averages of them). Therefore it is clear that the output of the reconstruction is constrained by the variability in the pool, so in this sense the method does not extrapolate anything outside the pool. Still, is worth to note that due to the multiple models merged and the length of each run, the available variability within the pool is larger than the variability observed in the calibration period.

In any case, what we mean when we write "extrapolation" in this context is "spatial extrapolation". It is an expression we use to refer to the general goal aimed by CFR techniques: filling spatial gaps between local reconstructions. Therefore we will re-write the parts where this expression is used to minimize chances of mislead the reader.

4) Evaluation of uncertainty. The uncertainty of reconstructions produced by the AM-CFR method is never shown or discussed. Is this uncertainty larger where proxies are inexistent? How large are the 95% confidence intervals and how dependent are they on the number / density of proxies. Could the authors add uncertainty bands around reconstructions in Figure 9 and / or produce a map of the size of the 95% confidence interval of the reconstructions? That would help measure the robustness and reliability of the AM-CFR method

This is an important caveat of the first version of the manuscript that has also been pointed out by other reviewers. However, this question is actually much deeper than it may seem at first sight, since there is no theory -to our knowledge- to estimate estimation uncertainty in the AM as in other established statistical methods like linear regression.

There may be different sources of uncertainty, and the reviewer seems to be referring only to the uncertainty due to the limited number of proxy records. Other sources of uncertainty will be related to the finite spatial correlation of the temperature field (since the AM method is used to produce a full temperature field based on individual pseudo-proxy records. When using observations as analogue pool, thus source of uncertainty will be smaller than when using model output as analogue pool, since a climate model output will never perfectly represent 'reality'. Other sources of uncertainty are related to the similarity between the target pattern and the selected analogue: when target and analogue are very dissimilar because the analogue pool is too small, the uncertainty should be larger than when the target and the analogue are indeed similar.

To some extent, the sources of uncertainty not related to the proxy network are indirectly estimated with the PPE, comparing the range of reconstructions with the actual 'true' temperature, but of course this is valid only in the setting used here: the GISS model as target and other models as analogue pool. In an extreme case, where the proxy records cover the whole world, the uncertain will be mostly caused by the similarity between the spatial correlation of the 'true' field and the spatial correlation of the fields in the pool of analogues.

However we agree that an uncertainty estimate for our results should be provided. Developing a method that addresses all the above mentioned sources of uncertainty is beyond the scope of this study. Therefore we will produce an estimation of the uncertainty that is at least comparable to those used in linear regression-based CFR methods, following standard procedures that make use of the calibration residuals. We will visualize temporal and spatial changes in these uncertainties, as suggested.

Specific comments Figure 1. I would like these points to show which proxies (tree rings, pollen, ice cores, lake sediments) are located where, perhaps on figure 1b? L70. Able to produce (no d) L104. Cannot result (remove s)

We will consider these comments in the reviewed manuscript.

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2016-98, 2016.