

Interactive comment on “Proxy surrogate reconstructions for Europe and the estimation of their uncertainties” by Oliver Bothe and Eduardo Zorita

Anonymous Referee #1

Received and published: 1 August 2019

Bothe and Zorita present a study where they investigate different ways of obtaining an uncertainty estimate for climate reconstructions using the analogue method, also known as the proxy surrogate reconstruction method. They authors describe the downside of single member reconstructions, and produce both single member and multi ensemble member reconstructions, which are compared to other reconstructions and observations. Then they go on to describe how an uncertainty can be assigned based on i) the fit of the analogue ii) assumptions on the standard deviation of the noise or iii) the ensemble spread when using a fixed number of ensemble members. Finally, the authors conclude on the pros and cons of the different approaches.

C1

General comments.

This study is overall well executed, thorough and timely. However, the writing is somewhat uneven, especially in the introduction, which I have commented on in detail below, but please go through the entire manuscript as I might have run out of steam. I have few major comments about the methodology itself, but I am wondering if the method is overfitting the model data to the proxies (see specific comments below). If the writing is brushed up as well as taking my other comments into account, I think this work could be suitable for publication.

Specific comments.

P1, L1: Please rewrite this sentence. It is the combination itself that reconciles the two sources, so if this is possible “allows” is redundant. Also, if one is reconciled with the other then they are both reconciled, making “both” redundant.

P1, L3: “. . . to benefit from the advantages of both data sources” this is in a way a repetition from previous sentence. Why not say something about the technique? E.g. “The analogue or proxy surrogate reconstruction method is a computationally cheap data assimilation approach which samples a model ensemble based on the best match to proxy data”.

P1, L9: Replace “had been” with “was”?

P1, L10: Remove “using”?

P1, L12-L14: “The approaches do not agree. . . “ this sentence is not easy to read. Perhaps rewrite “However, the two approaches do not agree on the warmest preindustrial decades, which for the Euro 2k reconstruction is during the early 15th century, and for the analogue approach is during the early 18th century”.

P1, L15: “The surrogate reconstructions...” I suggest that you early in the manuscript choose to call the reconstructions either surrogate or analogue, even if you have said it means the same - just to make it easier to read.

C2

P1, L15: Insert comma before “but”. Please use more commas to help the reader.

P1, L15-L16: Actually, I don’t understand the sentence. You lose me around “even under uncertainty”. Please rewrite.

P1, L20: Is “paleo-observations” the right word? Why not simply “proxy data”?

P2, L1: Replace “the search for” with “finding”?

P2, L7: Why “not only”, maybe cut this?

P2, L14-L15: “The analogue method... “ This sentence is hard to follow. Please rewrite.

P3, L1: Either write “Here we propose ...” or “In this study we provide...”.

P3, L?? (something strange happens with the line numbers): “Here, we obtain... “ (skip comma after “Here”). Please be clear on what is model and what is proxy. I suppose “pool of relevant candidate fields” is model out put and “local indices” is proxy data?

Figure 1: Please add units to axes.

P4, L15-L16: “That is, they use recent observations, which measured archives... “. I don’t follow this sentence. Please rewrite.

P4, L19: “to more than one environmental condition” do you mean that a given proxy parameter can depend on more than climate or environmental variable? Please clarify.

P4, L21-22: Is “environmental condition” the correct word, or is “climate state” more accurate.

P4, L26: I suggest you make a sheet or table with mathematical abbreviations that you use in the paper.

Table 1: So, the correlations are between the tree ring data series and the JJA CRU temperature. Please add these details to the caption. How do you deal with seasonality

C3

of the proxy data? In Wilson et al. (2016) each proxy site is listed with different seasonal sensitivity to temperature (Table 1) and I believe you are using some of the same data.

P6, L13: “The last of the remaining eight proxy indices starts in 1260” meaning that all remaining records cover 1260 to 2003 CE?

P7, L16: “strong ensemble” or “8 member ensemble”?

P7, 2nd paragraph, L8: “Since the current manuscript is not least a proof of concept... “ this formulation sounds off.

Figure2: (a) rescaled temperature? Please specify which scaling is used in the caption, so you don’t have to look for it. Euro 2k is an area mean? It’s hard to see that the difference to the CRU temp. Can you show this in (c)? Luterbacher et al. (2016) is discussed a lot in relation to Figure 2, it would be helpful if you show this data as well.

P9, L17: “calculated as the square root ...” why not write out the equation?

P9, c. L25: So which uncertainty is realistic? And why?

P9, L29-L30: “The coldest century was until 1648 CE in the best-analogue reconstruction but until 1678 CE in the Euro 2k record” please write the interval of the coldest century. This formulation is unclear.

P10, L1-L? (again random line numbers): When discuss interval please write them out instead of just giving the end year. It’s much easier to read. Just write “the warmest century was 1353-1452 CE”.

P10: About the volcanic analysis. Did you look at high latitude eruptions, e.g. Laki? How did you do the super imposed epoch analysis? Maps of field anomalies, or time series? How did you define the reference period before the eruptions?

P10, 2nd paragraph, L7-L8: “Interestingly, the analogues even appear to occasionally capture the relation between the proxies included and those excluded” couldn’t this be completely random? Then it’s not very interesting.

C4

P10, 2nd paragraph, L10: Replace “1947. Then” with “1947, where”.

Figure 4: What are the numbers next to the site name e.g. “(a) Tor92 0.91”? Is 0.91 the correlation?

Figure 5: Again, what are the numbers next to the site names?

P15, L4: Correlations “between 0.84 and 0.98” for proxies and reconstructed temperature. These correlations are a good bit higher on average than the data in Table 1. Are you overfitting the data, or how can you explain this? Wouldn't you need forward proxy modeling of tree growth to give a more realistic link between model and proxy data (e.g. Tardif et al. 2019)?

P21, L8: Is it really “strange variability” since the reconstruction is unconstrained in Greenland?

References.

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, *Clim. Past*, 15, 1251-1273, <https://doi.org/10.5194/cp-15-1251-2019>, 2019.

Wilson, R., Anchukaitis, K., Briffa, K.R., Büntgen, U., Cook, E., D'Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Helama, S., Klesse, S., Krusic, P.J., Linderholm, H.W., Myglan, V., Osborn, T.J., Rydval, M., Schneider, L., Schurer, A., Wiles, G., Zhang, P., Zorita, E., 2016. Last millennium northern hemisphere summer temperatures from tree rings: Part I: the long term context. *Quat. Sci. Rev.* 134, 1e18. <http://dx.doi.org/10.1016/j.quascirev.2015.12.005>.

Interactive comment on *Clim. Past Discuss.*, <https://doi.org/10.5194/cp-2019-81>, 2019.