

Paper review

Title: Technical Note: Considerations on using uncertain proxies in the analogue method for spatiotemporal reconstructions of millennial-scale climate

Authors: Oliver Bothe and Eduardo Zorita

The paper describes an analogue method of deriving spatially-complete temperature fields from sparse proxy records by finding modeled climate fields which match the proxy data within uncertainty bounds. The paper first tests this method using a pseudoproxy framework and then does several experiments using real proxy data. The authors also consider alternate experimental designs, such as using different proxy records and different criteria for finding valid matches.

Overall, I think that this is worthwhile research, and it is always good to explore methodologies which can supplement our incomplete perspective on past climate from proxy records. Additionally, the paper is much improved from the first version. The authors do a good job discussing potential pitfalls or improvements to the method, which is useful since future improvements would benefit the methodology. The primary shortcoming of the paper is the fact that the method doesn't work as well as one would hope. Instead of finding insightful analogues for past climate, the method tends to either find too many analogues (resulting in a very broad estimate) or none at all. The authors acknowledge this shortcoming, which is good to see, and the title accurately reflects the paper's "work-in-progress" nature. As it is, I think this is a useful paper, with the understanding that the method presented does not appear to represent a final product but rather a stepping stone toward more successful methodologies in the future. Many of my concerns with the first draft, such as the somewhat confusing description of the methodology and unclear elements of the figures, have been improved. However, I still do not agree with the authors use of uncertainty ellipses. My comments and suggestions are discussed in more detail below.

General comments:

My general comments concern the overall design of the methodology. While the method described in the paper has clear limitations, I think that this is appropriately discussed to a large degree in the paper. The authors have tested, or at least discussed, some ways of improving the method in the future, which is commendable. However, I want to return to two of my comments from my first review:

1. First, the pass/fail dichotomy of the analogue search seems very strict. If one (or more) proxy has a considerable bias, this single record could cause the method to fail. This can be seen in the author's choice to exclude a proxy on p.11, stating that "we find (not shown) that including this record puts very strong constraints on the analogue candidates and can reduce the chance of finding valid analogues." The excluded proxy is "the alkenone unsaturation ratios of Bendle and Rosell-Melé (2007)". This limitation could be mitigated if proxy uncertainty values were varied on a proxy-by-proxy basis (for example, by assigning a large uncertainty to certainty records individually, rather than increasing the uncertainty percentiles for all proxies simultaneously), but it may be difficult to find reasonable uncertainty values for each proxy. The paper decides to use consistent values across proxy records. To the paper's credit, the authors explore alternate experimental designs where analogues are found valid even if they fail to match a certain number/percentage of proxies. This is good. I'm not sure if anything else needs

to be done regarding this point in the paper, but I would like the authors to continue considering this point if they pursue this research more in the future.

2. Secondly, I still feel that the shape of the uncertainty ellipses may lead to difficulties finding valid analogues. As I see it, the main problem with the use of ellipses is that the analogue search becomes stricter, rather than more lenient, for searches farther from the original date of a given datapoint. This presents a problem at the ends of records, and will only get worse as more proxies are included. For example, consider Figure 2b. As a thought experiment, imagine that the data point shown at 7.8 kYa BP is the oldest point in this proxy record. When searching for analogues at 7.8 kYa, any value from ~22-30 degrees C would be deemed acceptable. However, when searching for analogues at ~8.5 kYa (i.e. just inside the edge of the uncertainty ellipse), only a very small range of values around 25 degrees C would be considered acceptable. Why would the analogue search be stricter for a period where the data point is less likely to be valid? Shouldn't it be more lenient, since we don't know if the datapoint is even relevant at that age? Furthermore, if we look for analogues at ~8.6 kYa (i.e. just outside of that uncertainty ellipse), any temperature would be deemed valid for this location (i.e. this data point provides no constraint on the analogues chosen for that date). This shift from somewhat strict (near the original age point) to extremely strict (at the border of the ellipse) to not-a-factor-at-all (outside of the ellipse) doesn't make sense to me. Wouldn't it make more sense for a data point to have less impact on the analogue selection for ages when it is less likely to be relevant? In the authors' response, they ask "*Wouldn't we, in this alternative scenario, then overemphasize the ranges far away from the original dated age?*" Yes, but that seems appropriate. The method would result in larger uncertainty bounds for times when the proxies provide less of a constraint. I'm not sure how this would be implemented mathematically in the methodology (perhaps through the use of rectangular uncertainty bands, as explored in the paper, to the authors' credit), but the current methodology is unsatisfying to me. One idea would be to have proxy uncertainties increase as one gets farther from central age of each proxy data point, and then require that the analogues be within the bounds of all proxy data points. The authors explain their reasoning for their choices in their response to my first review, but I disagree with their rationale. It is possible, however, that I am overlooking something.

I fear that the problems in both of the points above will become worse as more proxy records added. Users of the methodology could deal with this by excluding certain proxies (as mentioned in the example earlier), but this seems time intensive and potentially somewhat arbitrary. Regarding my second point above, it would probably take a lot of time to change the method entirely, but I would like to see a little more discussion (perhaps an extra paragraph) about this point in the paper.

Specific comments

- Page 1, line 19. By "climate state" do you mean the spatial climate field?
- Figure 1 caption. Perhaps mention that the pseudoproxy locations are slightly offset from the actual proxy locations. I know that this is mentioned in the paper's text, but including a brief statement in the Fig. 1 caption should help prevent people from wondering why it appears that the same panel is shown twice.
- Page 6, lines 27-28. The statement "Using anomalies circumvents this issue" makes it sound like you are going to use anomalies, when you don't. Consider rephrasing to make this clearer.
- Page 6, line 32: "Dansgaard-Oeschger (DO)" should be "Dansgaard-Oeschger (DO) events".

- Page 7, line 5. I would change “original temperature units” to “absolute temperature units”, since “original” makes it sound like you would be using the measured proxy units.
- Page 8. Regarding the 90% and 99% uncertainty values (and later 99.9% and 99.99%), are temporal uncertainty values the same for each proxy?
- Page 9, line 17-18. The phrase “the envelope does not necessarily cover all years within the period of interest” doesn’t make much sense to me. Can you rephrase?
- Page 9, lines 32-33: The paper states “...increase the probability to find a valid analogue at a certain date”. But what if widening the time uncertainties made new proxies relevant to some ages? Couldn’t that actually reduce the likelihood of finding an analogue (see page 24, lines 24-26)?
- Page 12, lines 14-15. The paper says “Regarding proxy uncertainty, we try to identify as complete uncertainty estimates as possible from Marcott et al. (2013) and their references. For the sake of simplicity, we decided to assume an uncertainty of $\sigma = 1K$ for all proxies.” If you use a common number for all proxies, why did you look at individual proxy values? Is 1K close to the mean or median value?
- Page 15, lines 1-2: I’m not sure exactly what this sentence means: “The latter modification also avoids that individual data points have an overly strong influence within our envelopes of tolerance.” Can you rephrase?
- Page 15, line 4: I’m not sure what “black lines” refers to in this sentence.
- Page 15, line 5: Remove the word “below”.
- Figure 6, panel c: Perhaps rearrange the labels so that “QUEST FAMOUS target” and “pseudoproxy” are listed next to each other (since the pseudoproxy is generated from the QUEST model output, right?). Also, put the “Analogue median” and “Analogue range” labels next to each other since they’re also related.
- Figure 6, panel d: The “Valid analogue examples” label only has one color, but the examples are shown in two different colors. Could you add the second color next to the label?
- Figure 6. Panels d and e took me a while to understand. If possible, could you try to explain it in a more intuitive way?
- Page 21, line 11: Is there any particular reason that age 8000 was chosen for the examples (or age ~2430 for the examples in figure 9)?
- Page 21, line 14. The paper says “Their local range at no point exceeds 4 degree Celsius”. However, it looks like you are referring to the deviation from the median, rather than the total max-min range. Can you check this number and rephrase if necessary? Please do this for the “20 degrees Celsius” value a few lines later as well.
- Figure 9, panels d and e: In these two panels, the two examples look like time-shifted versions of each other. Why is this? Is it because the same 101-year mean was chosen for each date?
- Figure 11 caption: The caption says “All panels include the median, 90% interval, and full range for the reconstructions”. However, only one range is shown. Is the displayed range the 90% interval or the full range? Please fix this line, both here and in the caption to Fig. 12.
- Page 26, line 5: I would remove the comma after “Please”.
- Page 26, line 6: Consider changing “further” to “also”.
- Figure 12 caption. I would change “are valid if they fail” to “are valid even if they fail”.
- Page 28, line 12: “warmer” than what?
- Page 28, lines 23-25: Consider discussing why the “single best analogue” differs from the default experimental setup (by large amounts for some ages). Is it because the “best” analogue may fail at one or more locations but still produce a better overall fit than other analogues? Since the “single best” methodology sounds similar to previous analogue methods, it would be interesting

to explore why it diverges from your main experiment, and whether your method provides benefits over the “single best” analogue.

- Page 30, line 2. I don't understand the use of “e.g.” in this sentence. Please rephrase.
- Page 30, line 11. What does “seasonal sensitivities” in this sentence refer to? I thought that you already considered seasonality in your proxy-to-model comparisons (as stated on page 12, lines 10-12). Or, if all proxy comparisons are made to annual-mean model data, than this needs to be corrected at several points in the paper (e.g. page 12 and Table 1 “season used”).
- Page 32, line 30. The method succeeds only for some ages.
- Page 34, line 8. What does “central” mean in this sentence?

Despite the quantity of these comments, I thought that the paper provides useful considerations and tests of the analogue method. Although I still disagree with the shape of the uncertainty ellipses, I think this is useful and interesting research.