Clim. Past Discuss.,
https://doi.org/10.5194/cp-2019-170-RC2, 2020

Climate of the Past Discussions

Interactive comment on "Technical Note: The analogue method for millennial-scale, spatiotemporal climate reconstructions" *by* Oliver Bothe and Eduardo Zorita

Anonymous Referee #2

Received and published: 3 March 2020

General Comments:

The paper discusses an analogue method of paleoclimate reconstruction. In this method, the researcher starts with a set of paleoclimate records (here, temperature-sensitive records in or near Europe) and searches for similar climate states within a pool of climate simulation outputs. By finding modeled states with match the proxy records, this method can be used to estimate the state of the climate system at locations which do not have local data. This method has been used in previous research, so the main focus of this paper is on the treatment of temporal and magnitude uncertainty of the proxy records.

Printer-friendly version

Discussion paper

In general, the goal of the paper–to better account for uncertainty in a computationally cheap reconstruction method–is worthwhile, so the case study presented in this paper is welcome. However, the method doesn't seem to work very well, which seems to be a major shortcoming. While, in theory, this may be acceptable as a stepping stone to further research, I also have additional concerns about the design and presentation of the research. In particular: 1) descriptions of the paper's methodology are sometimes confusing, and would benefit from further refinement; 2) I have several concerns about the paper's methodology, which seem like they limit the success of finding analogues; a revised methodology may result in a more successful reconstruction and a more interesting paper; and 3) the figures could be improved. These points are expanded upon in the "Specific comments" section below. I feel like these are important points which should be addressed.

Specific comments:

1. In a method-heavy paper, extra care must be taken to ensure that the paper is intuitive. When reading the paper, however, I had a variety of questions about how the method worked and what factors were keeping it from working better. Several of these confusions are listed below:

- The discussion of ellipses, which represent uncertainty in time and magnitude, is somewhat confusing at first, and it took me some time to understand they were used within the methodology.

- The relevance of the 90% vs. 99% vs. 99.99% cutoffs is not clearly explained. It appears that they refer to percentiles of magnitude and time uncertainty, but how are they calculated?

- Some aspects of panels d and e in Figs. 5 and 8 are unclear. As far as I understand, these panels are showing the annual data underlying the selected 101-year means, but I'm not sure what I should take away from them. Can their purpose be better explained, or can they be revised to show the relevant points in a more intuitive manner? In

particular, I don't understand the lines marked as "examples". Also, it may help if the "examples" were solid lines rather than dotted/dashed.

In general, I would encourage the authors to read through the manuscript again with a focus on making explanations clearer and more intuitive.

2. I am concerned about several aspects of the methodology, which seem like they may prevent the method from finding good analogues. My main two concerns are described below, with the second point being the more important of the two. Unless I am misunderstanding something (see point #1 above), I would like to see these concerns discussed or, preferably, directly accounted for within the methodology.

2.1. Uncertainty Ellipse Edge-Effects:

The use of uncertainty ellipses, which have a hard cutoff, may prevent the method from finding good analogues. One example of this may be imagined at the left and right "edges" of the ellipses. At the left and right edges of the ellipse, the vertical extent of an ellipse (representing magnitude uncertainty) becomes very small, eventually reaching 0. If the method is looking for analogues near the edge of one of these ellipses, the range of an "acceptable" analogue would be very narrow, rejecting many potential candidates.

Let's take the scenario in section 2.2.4 as an example. The paper states that there is a hypothetical proxy value at 500 BP, with age uncertainties from 600 to 400 BP. This hypothetical uncertainty ellipse stretches between 600 and 400 BP, with its maximum magnitude uncertainty at 500 BP. If an analogue search is conducted at 500 BP, the method accepts all points within the full uncertainty range of the ellipse. However, if an analogue search is conducted at 401 BP, the uncertainty range of the ellipse (i.e. the height of the ellipse, similar to the ones visualized in Fig. 2b) would be much smaller, therefore rejecting many potential analogues. This seems counter-intuitive to me. Wouldn't it make more sense to broaden the magnitude uncertainty as you get farther from the central age date, since we are less sure that the data point is applicable

Printer-friendly version

Discussion paper

as we get farther from its original dated age?

This issue may only be a problem at the start or end of a proxy record, or near a very long gap, but I expect that it would become more and more of a problem as the method is applied to more proxies, which naturally have different start and end dates. Unless I'm misunderstanding the method, I think that a better handling of these "edge effects" would help the method find more valid matches. Perhaps rectangles could be used instead of ellipses, since I see no reason that magnitude uncertainty should be decreased near the edge of temporal uncertainties. If anything, I would expect a particular point to become less precise toward the edges, not more precise. Since altering the method to address this would likely be too much work, I think that this point should be at least be mentioned in the paper.

2.2. Potential for Outliers to Cause Method Failure:

The paper mentions that the method uses the absolute temperatures calibrated from proxies, rather than anomalies. The authors discuss the problems surrounding the choice of absolute values vs. anomalies, but I'm concerned that biases in the absolute value of a single record (or simply non-climate proxy variations) could cause the method to fail. Consider applying this methodology to a group of proxies where a single proxy has been accidentally calibrated to be too warm by 5 degrees C. An error like this could hypothetically cause every single potential analogue to fail for the entire length of the proxy, as it's possible that no modeled state would show a spike of temperature at that particular location compared to everywhere else in the region. This means that the method would fail even if every other proxy were a perfect recorder of climate.

If a single problematic proxy can cause the whole method to fail, this problem will only become more likely to occur as the method is applied to a larger and larger proxy database. As it is, the method has trouble finding analogues with even a small set of proxies (as little as 7 proxies for the E09 case). This seems like a fundamental problem with the method, limiting its future application. The authors try to widen the group of

Printer-friendly version

Discussion paper

successful analogues by using wider uncertainty bands, including/removing records, and using annual model states rather than 101-year means, but I don't think that any of these solutions fix the underlying problem, which I suspect is the use of a binary match/mismatch dichotomy with the uncertainty ellipses. Using strict match/mismatch criteria probably makes the method overly sensitive to mismatches with single proxies. The use of a skill metric, as used in other work, may help alleviate issues arising from a subset of problematic records. Alternately, perhaps analogues could be accepted even if a certain percentage of the proxies don't match, to account for biases and non-climate noise within the proxy data set.

To the authors' credit, much of the paper does discuss potential problems with the method, and also suggests ways that things could be improved in the future. Indeed, the paper appears to be an exploration of how to account for age/magnitude uncertainties, rather than the presentation of a finished methodology. However, the paper would be much more satisfying to read if some of these issues were implemented directly, hopefully leading to a more complete reconstruction than the one shown in Fig. 8.

If this is not possible, I would at the very least like to see the following: 1) More discussion of the methodological problems mentioned above. 2) A different title, which accurately reflects the fact that the paper's methodology is a work-in-progress rather than a finished method. As-is, the title makes it sound like this paper demonstrates a finished methodology, when it appear to be an exploration of uncertainties which may lead to a better method in the future. Because of this, a better title might be something like: "Considerations of proxy uncertainties within the analogue method of paleoclimate reconstruction".

3. In general, several of the figures could be improved. For example, the black and red colors in Fig. 3 are difficult to distinguish, and the lines in Figs. 5c and 8c are difficult to interpret, since they use similar thicknesses and opacities. Improving the figures may also help make the methodology more intuitive, as I commented about in point #1 above.

Printer-friendly version

Discussion paper

A few other minor questions/concerns: Why only use 101-year means, rather than means which vary site-by-site to better reflect the temporal characteristics of individual proxy records? Also, why does the pseudoproxy experiment only use summer means, as mentioned in line 30 on page 10? And why does the number of sites differ between the pseudoproxy experiment (Fig. 1a) and the real experiment (Fig. 1b)? I had other questions about methodological choices while reading the paper, but the major points discussed throughout the review above seemed like the most important.

A final technical note: some figures (especially Fig. 4) have so many lines that the paper is difficult to print (it gets stuck on a "flattening" step for a long time).

In summary, while the paper focuses on an interesting and useful approach to paleoclimate reconstruction, I think that several things need to be improved before it can be considered for publication. A fundamental problem is that this appears to be a method paper, but the method doesn't work very well. If the method cannot be improved, the concerns above should at least be addressed and the paper should get a new title which better reflects its contents.

Finally, despite all of my comments and concerns, I do think that this is an interesting and potentially useful method, and I hope that further progress is made in the future.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2019-170, 2020.

Printer-friendly version

Discussion paper