Dear editor, dear referees,

Hereby we want to thank the editor and the referees for their evaluation of our manuscript and their helpful comments. Below we provide first a general list of changes to the manuscript and then our updated responses to their remarks.

We want to apologize that providing these corrections took us so long.

On behalf of the authors

Yours sincerely

Oliver Bothe

List of changes

- Title
  - Title changed
- Abstract
  - minor changes in text
- Introduction
  - minor changes
- Analog method, assumptions, and data
  - Adapted Figure 1 according to reviewer comments and to new proxy collection
  - Adapted Figure 1 caption
  - reframing in terms of tolerance ranges
  - clarifying our considerations on uncertainty
  - clarifying our method
  - adding ellipse equation
  - adding comments on additional experiments
  - minor changes in caption of Figure 2
  - moved Figure from appendix to this section as new Figure 3
  - different pseudoproxy setup
  - explained changes to pseudoproxy generation
  - updated former Figure 3 (now Figure 4) to show all experiments
  - clarification on QUEST-FAMOUS data
  - minor further changes
- Results
  - Figures changed to improve them
  - Changes due to use of different pseudoproxy setup in text and figures
  - added results for different proxy setups from appendix and relevant Figure
  - added results for different methodological choices
- Discussion
  - Changes to better reflect shortcomings of our approach
- Summary and concluding remarks
  - Changed for new version
- Appendix
  - Largely moved to main text

Editor

Comments to the Author: Dear authors,

Thank you for submitting your reply to the comments of the two reviewers. Both find your manuscript of interest, but suggest major revisions. Please submit a revised version according to the changes you proposed.

Response: We thank the editor for this evaluation and guidance.

Regarding your specific questions about 1) estimating skill from non-independent pseudoproxies: I agree with the reviewer that this is an important point and encourage you to include, as you suggested yourselves, an experiment with independent pseudoproxies based on the QUEST-FAMOUS simulations. Perhaps this would also help to include the information that is now in the appendices in the main text?

Response: We thank the editor for this guidance. We now changed the pseudoproxy test to a setting closer to the real conditions. We do not discuss anymore the other test.

We also move most of Appendix A to the main text and remove other parts of Appendix A.

2) title: I agree with the reviewer that a title that better reflects that this is work in progress would be good. In doing so, climate does not need to be changed to temperature.

Response: We thank the editor for this guidance. We changed the title.

The wording used to describe the ellipses is somewhat confusing, as is also clear from the questions posed by the second reviewer. Your explanation "The ellipses do not represent the uncertainty ranges in the value of the proxies, but rather the confidence with which we claim to know the value of the proxy at that time." explains the problem well and I think that besides an improved description of the method and further discussion, a change in the wording would also help to clarify what these ellipses really mean. Perhaps "confidence ellipse" would be a good alternative?

Response: We aim to be as clear as possible for us in describing the approach.

We first reframed everything in terms of confidence ellipses but then thought it might be even clearer to represent the ellipses as tolerance ranges.

And finally something very minor, in Fig A2 there's an "o" missing in first time chironomids are mentioned.

Response: Thank you for spotting this. We corrected this.

With kind regards, Lukas Jonkers

Referee 1

I think the paper addresses an important topic and provides a useful extension of the Analogue Method, combining the data and dates uncertainties. However, I think the paper lacks clarity in the description of the method (some fundamental steps as the generation of the confidence ellipses are not properly explained) and I have concerns about the pseudoproxy setup.

Response: We thank the referee for their generous evaluation of our manuscript. Below we address these points in more detail. In our revisions, we will particularly take care to clarify our method. Below we also address the concerns about the pseudoproxy setup.

In particular, I think the assessment of the method' skill (of course possible under pseudoproxy conditions) is flawed: the same run used as "truth" is used inside the Analogue Pool leading, therefore, to a potential overestimation of the skill. In addition to that, I can not comprehend why the authors selected a pseudoproxy network design (number of proxies, locations of proxies, period covered, uncertainties, etc.) that do not resemble at all the real-world case they later try to reconstruct. I recommend the authors to re-do the exercise generating a pseudoproxy environment as close to the real-case as possible. Of course, later the here presented pseudoproxy setup could be considered informative as how would the method perform if more proxies were considered, etc. but the generation of a closer to real situation is nonetheless essential and I suggest for it not to be bypassed.

Response: There are two points to address here:

1. We redo our pseudoproxy test with a setup closer to the real-world case. See changes in our section on methods and data.

2. We changed the target of our pseudoproxy setup. We try to be careful not to overestimate the skill of the method.

We now use QUEST-FAMOUS simulation data for the pseudoproxy setup, which, however, does not have interannual resolution.

Here, we want to again point out why we used the setup as criticised by the referee: For one, at the time of our study, the Trace-21ka simulation was the only available simulation providing a continuous deglacial climate trajectory in annual resolution. We assumed that an interannual setup is the most reliable approach. Tests showed that in our chosen setup the method does only find analogues from the Trace-21ka simulation from which we also constructed the pseudoproxies, as we wrote in the manuscript.

Thus, if we exclude Trace-21ka from the candidate pool the method fails completely, and excluding the Trace-21ka simulation from the pseudoproxy construction prevents using a simulation with interannual resolution for the construction of pseudoproxies.

Following the review-comments we now use one of the QUEST-FAMOUS simulations. The FAMOUS-HadCM3 simulations for QUEST use accelerated forcings. That is, the last glacial cycle of approximately 120,000 years of climate forcing was simulated in approximately 12,000 simulation years. Thus, the annual simulation data is only representative of ten years of climate evolution.

General Comments

- The description of the method is not clear enough. How do you define the uncertainty ellipses? I don't see anywhere in the text the methodology followed to find such ellipses. Also, related to that, what's the difference between 90%, 99% or 99.99% uncertainty ellipses? Please, provide a clear methodology to follow to find them. What are the confidence intervals? Please, define.

Response: We rewrote the methods section as well as other parts of the manuscript to clarify these points.

- Why in the pseudoproxy setup you don't mimic the real-world conditions your are trying to reconstruct? I find it confusing that the pseudoproxy and real-world proxy locations and time-spans are not the same. As it is now, because the pseudoproxy and proxy cases differ, the results of the pseudoproxy analysis are not completely transferable to the real-world case. I suggest to generate a pseudoproxy network that is exactly the same as in the real-world case, and show the results in that case.

Response: We redo the pseudoproxy setup with a setup closer to the real-world case.

- The title of the manuscript talks about climate reconstructions. However, the manuscript deals only with surface temperature reconstructions. I suggest to modify the manuscript title to reflect this and to add some discussion on how the method could/could not be applied to reconstruct some other climate variables (particularly, how do you expect the results to change when reconstructing a more challenging variable as precipitation?).

Response: We modified the title.

We shortly note in the discussions that reconstruction success for other variables is less likely than for temperature.

- It would be interesting to compare with the results of having a fixed number of Analogues, for example 1 Analogue.

Response: Indeed this would be an interesting experiment. One may, however, ask whether it is really meaningful considering the large uncertainties of the proxies, as one would only test relative to the best estimate for each proxy.

We shortly discuss such a test in our results section now.

We want to repeat that analyses for shorter time scales (compare recent works from Gómez-Navarro and colleagues in Climate Dynamics and Climate of the Past) have found that fewer analogues lead to higher variability in the reconstructions but also to lower skill. This is plausibly also the case here, as these effects can be explained by statistical sampling.

- How do results change if using a less years for the sliding window-mean? For example 50-years-means? There is a mention to interannual data in the discussion session, but not comparison plots are shown.

Response: We now shortly discuss test-cases with interannual data, 51-year averages and 501-year averages.

- For the pseudoproxy setting: The selected reality is the simulation Trace21k. Most of the Analogues come also for this simulation. It would be fairer for assessing the method's skill if the chosen reality is excluded from the analogues pool. How do the results change if done so? When allowing the same simulation as reality to enter the pool, results might be overly optimistic. For some of the simulations listed there are several runs available, in that case one run could be selected as reality and the other pool of Analogues.

Response: See our response to a previous point: We understand the concern of the referee, and agree that the setup is suboptimal. We reconsidered our writing and hope the new version is sufficiently careful not to overestimate the skill of the method.

We now use a different setup. We try to be as clear as possible about the potential quality of the method.

Here, we want to point out why we used the setup as criticised by the referee: For one, at the time of our study, the Trace-21ka simulation was the only available simulation providing a continuous deglacial climate trajectory in annual resolution. Tests showed that in our chosen setup the method does only find analogues from the Trace-21ka simulation from which we also constructed the pseudoproxies as we write in the manuscript.

Thus, if we exclude Trace-21ka from the candidate pool the method fails completely, and if we don't use Trace-21ka for the pseudoproxy construction, we cannot use a simulation with interannual resolution.

We now use the QUEST FAMOUS simulations to provide pseudoproxies.

Regarding the suggestion of a single model ensemble: we want to emphasize that there is not really a suitable single model ensemble of simulations over the periods of interest available to test the analogue method on the timescales of interest.

Specific Comments

* Abstract:

- In the first paragraph authors talk about the last 21 kyr. However, in the second paragraph the target is reduced to the last 15kyr. Please, rephrase or explain failure in the target.

Response: We will be more clear about the temporal scope of the manuscript. The discrepancy was solely related to the different temporal extent of the pseudoproxy and real-world applications. To be specific, the period of the pseudoproxy setup was limited by the length of the available simulation while the period of the real world setup was limited by the length of the proxy records.

Our changed pseudoproxy setup effectively leads to more comparable periods.

- The authors could emphasize that in the present for the reconstruction method seems to be no better than a long-term mean.

Response: We are not fully clear to what "in the present" refers here. We now note "that the analogue method in the present setting may represent the recent climate worse than simply taking the average over the period of instrumental observations."

- These fields reveal that uncertainty are also large locally. Please, change for . . . uncertainty is also. . .

Response: We change this to "uncertainties are also".

* Introduction:

- Please clarify the definition of nonillion

Response: We do clarify in our revisions that nonillion refers to 10^{30}.

* Section 2:

- Here you sometimes use the word Analog instead of Analogue. Please, unify throughout the manuscript

Response: We are sorry for this oversight. We changed instances of "analog" to "analogue".

- Figure 1: Please, add latitude and longitude. Also introduce the acronyms P01 and E01, as so far they have not been introduced in the text.

Response: We added these. Figure 1 now includes axes for longitude and latitude. The caption now introduces P01 and E01.

- Page 4: 'Our interest is in temperature', please clarify if it is surface, annual mean, etc. What is a temperature calibration?

Response: We clarify this now.

-Page 5: Please explain better the meaning of "at best centennial" Does this mean that there are no proxy records with resolution finer that 100 years?

Response: We clarify this.

- Page 6: why not consider the same period for real and pseudo proxy setups?, how are ellipses of confidence constructed? Please, provide the appropriate ellipse equation for its construction.

Response: Considering our previous pseudoproxy setup we were able to extend the reconstruction period for the pseudoproxy approach back to the Last Glacial Maximum. We regarded this an interesting exercise. Considering the new pseudoproxy setup we also are able to extend it but concentrate our discussions on the shorter period.

We clarified the explanation of how we construct the ellipses and added the equation.

- Page 8: What is a credible interval? Please, define.

Response: We clarified our terminology. We now reframe the uncertainties in terms of tolerance ranges to accept analogues.

- Page 10: The authors say: "randomly chosen pseudo age uncertainties". How are those selected? Is the random process a Gaussian distribution? Which mean and variance? This needs more clarification.

Response: Bothe et al. (2019) include a switch in their script for pseudoproxy calculation. One may use a Gaussian distribution and the parameters of this distribution. A second option is to use an uncertainty dependent on the assumed random smoothing of the pseudoproxy. A third option calculates the uncertainty assuming a constant smoothing of the pseudoproxy record length and a random Gaussian offset. We use this setup but scale it down to reduce the width of the uncertainties.

Figure 3: Isn't it easier to show the plots in the form of line-plots? Specially plot a is difficult to read, as it looks like a huge black block, differences are hard to distinguish.

Response: We think the vertical lines better represent the discrete character of the approach but we changed the visualisation to line-steps.

Figure 4: Please, put all the plots in the same scale

Response: The reviewer's suggestion would make it harder to identify changes in individual series. We now use a common absolute range of the temperature-axes for all panels.

* Section 3:

- Page 14: The authors indicate very little variability in the reconstruction median over certain periods. This probably arises due to too many Analogues are selected in those periods. How could you constrain the Analogue selection?

Response: There is a trade-off between considering the uncertainty of the proxies and constraining the number of analogues. That is, if we want to consider the uncertainty in the way we do, then we allow for weakly constrained analogue ranges. If we allow different levels of proxy uncertainty, we can choose only the best M analogues. We then can limit the number of analogues by another criterion based on their distance to individual proxies or their overall Euclidean distance.

Indeed, a likely explanation for the little variability in central estimates and the generally rather constant character of our reconstructions could be that the space of valid analogues is too unconstrained and too many analogues are considered valid. However, also the single-best analogue approach shows such a behaviour. That is, while the reconstruction is undoubtedly badly constrained, even the best analogues differ little between subsequent dates. Part of this may be due to our choice to consider a rather large temporal range of influence of individual dated records. Our ellipses of tolerance may result in a strong influence of an unlikely value at a specific date. This could potentially be solved by explicitly considering the likelihood of a value at a particular date instead of simply taking a binary criterion. A less complex solution could be obtained by pooling proxy values in temporal windows, weighting them *within these windows, and then performing a reconstruction considering certain ranges of tolerance to accept analogues.*

We add this discussion to the manuscript.

Figure 5: Panel c: please add name like “warmer case”, “colder case” and the respective locations (lon, lat). Panel d: add the subtitle “Regional average” Panel e: add the subtitle “Grid point: (lon, lat)” In panels d and e: I can’t understand what the authors mean by “examples”. Why some of the examples look like dots and some as dashed lines? Are the dots (dashes) associated to the warmer (colder) case shown in panel c? It would be interesting to discuss the moments when the Target is outside the envelope (Figure 5a)

Response: We will try to improve the visualisation of our results, and we will provide a clearer description of the results. We tried to follow the reviewer suggestions on Figure 5. As the pseudoproxy setup changed the discussion of cases outside the envelope would have to change as well. That is, even in the perfect model setup of the preprint, the simulation data and the pseudoproxy differed. Then we could not expect the analogue reconstruction to always include the original target.

Figure 6: Please, add the units directly above the colorbar. Also, indicate the year that is being shown as Example.

Response: We added the units and mention the year in the Figure caption.

Figure 7: Please put all the plots in the same scale.

Response: As for Figure 4, the reviewer’s suggestion would make it hard to identify changes in individual series. Now, we use a common absolute range of the temperature-axes for all panels.

Figure 8: Similar considerations as in Figure 5. Figure 9: similar considerations as in Figure 6.

Response: We clarified all four Figures. We adapted Figure 9 following the suggestions on Figure 6. Note, as we added further Figures, the Figure-numbering changed.

Page 24: In the summary the authors say that the method succeeds in the pseudoproxy setup. I think that sentence might be overestimating the skill of the method, as the authors used one model run (Trace21k) both as truth and as proxy pool. Please, remove the truth from the possible pool of Analogues to be able to properly analyse the method’s skill.

Response: While we would prefer using an interannual input for the pseudoproxies we changed the setup so that the truth is not any longer in the pool of analogues. We are confident that our previous statement was careful enough and think the current manuscript is clearer.

For the real-case the authors say the reconstructions fail. How can you assure failure when you don’t know the truth? I think the sentence should be re-phrased and the only thing that can be known for sure is the failure to find Analogues within the selected pool. I think that it needs to be made clearer that not knowing the truth in the real-case is exactly the reason for making pseudoproxy analysis. Which leads, again, for the importance of the pseudoproxy setup (design of the network, period covered, etc.) to be as similar as possible to the real-case.

Response: We rephrase this to highlight that a failure of the method is equivalent to a failure of finding analogues in the candidate pool.

Referee 2

General Comments:

The paper discusses an analogue method of paleoclimate reconstruction. In this method, the researcher starts with a set of paleoclimate records (here, temperatur-esensitive 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.

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:

Response: We thank the referee for the positive reading of our manuscript. We would particularly thank them for highlighting the manuscript's value as a stepping stone.

1) descriptions of the paper's methodology are sometimes confusing, and would benefit from further refinement;

Response: We hope to have clarified the methodology sufficiently. We did not include an additional figure to illustrate it.

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;

Response: We thank the referee for raising the possibilities to improve the manuscript. We address the comments below.

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.

Response: We tried to improve the visualisations.

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:

Response: We thank the referee for their detailed criticisms.

- 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.

Response: We extend our description of the ellipses and try to be clearer in our terminology throughout the manuscript. We decided against a figure to specifically explain their role.

- 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?

Response: This is part of the calculation of the uncertainty ellipses. We clarify theses aspects in the revised manuscript.

- 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?

Response: We describe the purpose of these panels in more detail.

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.

Response: We tried to clarify the panels and their descriptions.

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.

Response: We thank the referee that they detail their concerns so carefully.

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.

Response: The referee is correct in this description. The ellipse describes a two dimensional interval in which we search. Thus, at this edge, there is little probability of finding a valid analogue considering the age uncertainty and the data uncertainty. An alternative to this approach would be to assume that both uncertainties affect the selection independently and, in turn, taking a rectangle. Even then, we would have edge effects though of a different kind. Here, we use a two dimensional Gaussian to represent the effect of proxy-uncertainty and dating-uncertainty on our tolerance to accept an analogue. Therefore, our current edge effect is not a bug but a feature. We want the data to allow for less analogues in either direction. We will try to clarify this.

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 as we get farther from its original dated age?

Response: Wouldn't we, in this alternative scenario, then overemphasize the ranges far away from the original dated age?

Apparently we were not clear enough in explaining how to interpret the ellipses. The ellipses do not represent the uncertainty ranges in the value of the proxies, but rather the tolerance with which we accept analogues. An analogue that may be numerically close to the target should not be as easily accepted for dates that are far off the median proxy date as they are accepted for dates that are temporally closer to the proxy median age. Essentially the ellipses define a weighting scheme (although with binary weights) according to that tolerance level. If we adopt the scheme suggested by the reviewer, we would select many analogues that appear to match the proxy for dates at the edges of the dating-uncertainty interval, where actually we are very unsure that the proxy is delivering any useful information about the *climate at that point in time. In that situation we do require the analogue to be numerically very close to the proxy, otherwise we would reject it. In contrast, we are laxer for dates that are temporally close to the proxy age.*

We hope that our revisions make these points more clearly.

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.

Response: We try to clarify the description of our method and our assumption on why to use an ellipse and not a rectangle. A rectangular tolerance region would lead to accepting analogues for dates far off the most probable date with the same tolerance as for dates that are temporally closer to the most probable age. This would be fine if we consider the dating as having uniform probability over the dating uncertainty range, which is not plausible. We shortly discuss the effect of a rectangular tolerance range.

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.

Response: We understand the concern of the referee. By considering the uncertainty of the record we would hope to be able to compensate for such an error at least partially. This should be independent of whether it is a systematic bias in the record or whether only a single measurement is erroneous. However, we cannot exclude that such biases lead to a failure of the method.

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 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.

Response: We would again like to argue that including the uncertainty of the records should compensate for this problem. Problems with the reliability of the proxies affect any reconstruction method. One can assume that the method compensates for them or one can accept that unreliable proxies reduce our ability to make reliable estimates about past climates.

We think the failure of finding analogues is rather due to the insufficient pool of analogues and less due to problems with the reliability of the proxies.

We shortly discuss experiments where we allow that it is enough if N-1, N-2, or 75% of all proxy records are matched.

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.

Response: We do not follow up on our suggestion in the initial response to rewrite the manuscript as an exploration of how to handle the uncertainties in the analogue method.

We hope that our rewriting sufficiently addresses the referee's points.

We, still, do want to emphasize that failure of a method may primarily signal that our data (cf. our proxy information or our simulation pool or both) are insufficient to inform us about a problem at hand. We do not claim here that this is the case with our paper, we just want to emphasize that completeness of a reconstruction is not an information about the quality of a method, a paper, or the reconstruction.

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".

Response: We hope that our revisions sufficiently discuss the problems described by the referees and those already mentioned in our submitted manuscript.

We adapt the title.

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.

Response: We reconsider all our visualisations. Particularly, we hope to have addressed the comments with respect to Figures 3, 5, and 8.

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.

Response: The referee is correct. Ideally one should use site-specific means and adapt these for each individual measurement. The information to do so is not necessarily available - as stated in the manuscript. We considered this at one point but did fail to achieve a computationally effective implementation at that point.

We used summer means as we made the assumption that this is a representative season for the proxy locations. We change this now to annual means as we also changed the simulation from which we compute the pseudoproxies.

We also now use the same number of locations in the pseudoproxy experiment as in the main real-world experiment. We decided against using the seasonal representations from the *real-world case.*

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).

Response: We have to prevent this happening. We will reconsider all our visualisations and the output format for these cases. To our knowledge this should not happen anymore.

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.

Response: We thank the referee for their fair evaluation of our manuscript. We, however, want to express our surprise that technical notes should only deal with well working methods. We will try to account for all the reviewer's suggestions.

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.

Response: We want to thank the referee once more for their generous evaluation.

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

Oliver Bothe[1] and Eduardo Zorita[1]

[1]Helmholtz Zentrum Geesthacht, Institute of Coastal Research, 21502 Geesthacht, Germany

Correspondence: Oliver Bothe (ol.bothe@gmail.com)

Abstract. Inferences about climate states and climate variability of the Holocene and the deglaciation rely on sparse paleo-observational proxy data. Combining these ~~sparse~~ proxies with output from climate simulations is a means for increasing the understanding of the climate throughout the last ~~~21 millennia~~tens of thousands of years. The analogue method is one approach to do this. The method takes a number of sparse proxy records and then searches within a pool of more complete information (e.g., model simulations) for analogues according to a similarity criterion. The analogue method is non-linear and allows considering the spatial covariance among proxy records.

Beyond the last two millennia, we have to rely on proxies that are not only sparse in space but also irregular in time and with considerably uncertain dating. This poses additional challenges for the analogue method, which have seldom been addressed previously. The method has to address the uncertainty of the proxy-inferred variables as well as the uncertain dating. It has to cope with the irregular and non-synchronous sampling of different proxies.

Here, we ~~propose~~ describe an implementation of the analogue method including a specific way of ~~dealing with~~ addressing these obstacles. We ~~use uncertainty ellipses~~ include the uncertainty in our proxy estimates by using *ellipses of tolerance* for tuples of individual proxy values and dates~~and, thereby,~~. These ellipses are central to our approach. They describe a region in the plane spanned by proxy dimension and time dimension for which a model analogue is considered to be acceptable. They allow us to consider the dating as well as the data uncertainty. ~~Results~~ They therefore form the basic criterion for selecting valid analogues.

We discuss the benefits and limitations of this approach. The results highlight the potential of the analogue method to reconstruct the climate ~~of the last ~15 millennia~~from the deglaciation up to the late Holocene. However, in the present case, the reconstructions show little variability of their central estimates but large uncertainty ranges. The reconstruction by analogue provides not only a regional average record but also allows assessing the climate state compliant with the used proxy predictors. These fields reveal that ~~uncertainty are also large locally~~ uncertainties are also locally large. Our results emphasize the ambiguity of reconstructions from spatially sparse and temporally uncertain, irregularly sampled proxies.

1 Introduction

It is a pervasive idea in environmental and climate sciences that past states provide us with information about the future (Schmidt et al., 2014a; Kageyama et al., 2018). Therefore, paleoclimatology aims to understand past spatial and temporal climate variability, preferentially using a dynamical understanding of the climate processes. To achieve this, we need spatial and temporal information about past climate states and past climate evolutions. Our understanding of the past, however, relies on spatially and temporally sparse paleo-information. Data assimilation methods and data-science approaches are ways to provide estimates for the gaps in time and space. One simple approach is the analogue method or so called proxy surrogate reconstructions (Gómez-Navarro et al., 2017; Jensen et al., 2018). This method is similar to k- nearest neighbour classification algorithms in machine learning applications. The present manuscript discusses one implementation of the analogue method for reconstructing surface temperature over timescales including the Holocene and the last deglaciation.

If we want to use the analogue method beyond approximately the last two millennia, we have to tackle additional challenges, which usually can be evaded for the Common Era. For example, our proxy records are not only spatially sparse but they also have a coarse temporal resolution on these timescales. Furthermore, the sampling generally is irregular for each individual proxy. Indeed, sample dates differ between ~~predictors~~ proxies on these timescales, and ~~sample ages, i.e. dates , are~~ these dates are also uncertain. Recently, Jensen et al. (2018) use the approach to reconstruct the climate at the Marine Isotope Stage 3 (MIS3, 24,000 to 59,000 years before present~~(~~; 24–59kyr BP) ~~)~~ addressing such challenges. Including part of a deglacial period, as we do here, ~~complicates applications further~~ further complicates applications as we consider a climate trajectory with strong trends.

The basic idea of the analogue method is simple. An analogy tries to explain an item based on the item's resemblance or equivalence to something else. In the analogue method, one uses a set of sparse ~~predictors~~ proxies, i.e. predictors, and searches for analogues for them in a pool of ~~more complete candidates~~candidates that are spatially more complete. In paleoclimatology, the predictors can be local proxy records and the candidate analogues can be fields from climate model simulations. One assesses the similarity of the simulation output and the proxy records at the proxy locations to find valid analogues. The reconstructed field is then the complete field given by the analogue.

It is important to note that comparable approaches suffer from a trade-off between accuracy and reliability of reconstructions as shown by Annan and Hargreaves (2012) for a particle filter method. This depends on quality and quantity of the available proxy records. This drawback also affects the analogue method as shown by Franke et al. (2010) and Gómez-Navarro et al. (2015), who find that the skill accumulates at the predictor locations~~of the proxy records~~. Similarly, Talento et al. (2019) highlight that the analogue method may perform badly in regions with little proxy coverage.

Most paleoclimate applications of the analogue method focussed on the Common Era of the last 2,000 years (e.g., Franke et al., 2010; Trouet et al., 2009; Gómez-Navarro et al., 2015, 2017; Talento et al., 2019; Neukom et al., 2019). In this context, Graham et al. (2007) call the results of the ~~analog~~ analogue method a "proxy surrogate reconstruction". Gómez-Navarro et al. (2017) provide a comparison of the analogue approach to more complex common data assimilation-techniques. Applications often only consider the single best analogue, which may not necessarily be appropriate especially for predictors ~~under~~ affected by uncertainty. Paleo-applications of the analogue method generally try to upscale the local proxy information but the analogue method was also applied for downscaling of large-scale information (e.g., Zorita and von Storch, 1999).

Here, we describe another approach to obtain reconstructions by analogue over millennial timescales with based on spatially and temporally sparse and uncertain proxies. It differs in some aspects from the approach so far applied to shorter and more recent periods. Our approach tries to explicitly consider not only age uncertainties (compare with Jensen et al., 2018) but also the uncertainties of the proxy values or, similarly, of the temperature reconstructions inferred from these proxies. We make specific assumptions on the uncertainty of the data and the dates of the proxy predictors. We further account for the temporal irregularity of the sampling of different predictors. Our As explained in more detail later, our approach considers an analogue candidate as valid if it complies with our assumptions on the uncertainty of the proxy predictors. We apply the method over time periods encompassing parts of the last deglaciation until the late 20th century of the Common Era (CE). That is, we try to apply the analogue method over a period when the climate cannot validly be described as stationary at local, regional, and global spatial scales.

Beyond the mentioned challenges for analogue reconstructions on millennial timescales, the method strongly relies on is also constrained by the pool of available analogue fields. van den Dool (1994) considers how likely it is to observe two atmospheric flows over the northern Hemisphere that resemble each other within the observational uncertaintyat the time of that study. The study finds that a pool would have to include a nonillion, i.e., 10^{30} of potential analogues to achieve this. Obviously, we aim for less accuracy in paleoclimatology due to larger uncertainties. However, there are still only few climate simulations for relevant timescales, and these simulations also only cover cover only parts of the time periods of interest. Furthermore, these simulations stem from different climate models whose reliability on these timescales may not have been shown yet (Weitzel et al., 2018; Kageyama et al., 2018).

The next section first summarizes again the main characteristics of analogue searches for paleo-reconstructions. Afterwards, we present our way of dealing with uncertain tuples of data and date, that is with uncertain paleo-observations with uncertain datingdescribing ranges of tolerance for which we choose analogues. We also describe how we consider the fact that different proxies are sampled at different times. The section also presents our selection of a simulation pool. We present results over time periods encompassing parts of the last deglaciation until the late 20th century of the Common Era for a pseudoproxy setup (compare Smerdon, 2012) over the last ∼22 millennia and a realistic setup over the last ∼15 millennia for the European-North-Atlantic sector. Our appendix shows We also shortly describe results for alternative proxy setups. Finally, we discuss our assumptions and results. Thereby, we We aim to emphasize the opportunities of the analogue method while also highlighting its challenges.

2 Analog Analogue method, assumptions, and data

2.1 General Method

In an analogue search one tries to complement incomplete information from one dataset by data from other more complete datasets. One ranks the more complete data by its similarity to the incomplete available information in the first data set.

In paleoclimatology this usually means that one uses a set of spatially sparse proxy records and wants to find fields from simulations or reanalyses that are most analogous to the proxy records at their locations. The pool of candidate fields depends on the available simulations and reanalyses.

If, for example, one uses proxies for temperature, such a ranking may simply provide the ~~model output~~ simulated temperature field that has the smallest Euclidean distance to the sparse proxy information at their locations. Alternatively, one can consider not just one but a small number of good ~~fits~~ analogues with small distances (Franke et al., 2010; Gómez-Navarro et al., 2015, 2017; Talento et al., 2019). However, it is also possible to define a range of tolerable deviations from the proxy predictor values and consider all ~~candidates~~ analogue-candidates that are within this range as valid analogues (compare Bothe and Zorita, 2019). Matulla et al. (2008) discuss the effect of the choice of ~~distance~~ similarity measures for a ~~downscaling exercise~~different application.

An important aspect of a paleoclimate reconstructions is the uncertainty of the reconstructed data. To our knowledge, only Jensen et al. (2018) and Neukom et al. (2019) consider the uncertainty of the final reconstruction among earlier paleo-applications of the analogue method, and only Jensen et al. use proxies with prominent age uncertainties in their work on MIS3. They perform multiple reconstructions to obtain reconstruction uncertainties by shifting the dates of their proxies within the stated age uncertainties. Uncertainty information is particularly relevant for applications like the one of Jensen et al. where one has to deal with predictors that are sparse, irregular, and ~~uncertain in time.~~ uncertainly dated.

2.2 Present application of the analogue method

We use spatially and temporally sparse ~~as well as data and time uncertain proxies~~proxies, affected by uncertainties in their values and their dating for analogue searches on millennial timescales. Next, we detail our simplifying assumptions about what the data represents, its uncertainties, and the dating uncertainties. We also describe how we choose the dates for which we perform the climate reconstruction.

2.2.1 Variable of interest

Our interest is in temperature. Specifically, we concentrate on means of seasonal or annual temperature at the surface. We consider proxies ~~that have a temperature calibration and~~ for which the literature previously reports a sensitivity to temperature in form of a calibration relation. We search for analogues within fields of simulated surface temperature. To do the comparison, we consider the model variable “surface temperature” over the European-North-Atlantic domain shown in Figure 1. The reconstruction also uses these fields.

Theoretically, the variable or variables to be reconstructed can be different from the variable or multiple variables represented by the paleo-observational predictors. Indeed, we here assume that it is possible to reconstruct annual temperatures from proxy records with diverse seasonal attributions.

Using temperature in a multi-proxy comparison requires a number of assumptions. First, we assume that the proxy recorders indeed were temperature-sensitive. More importantly here, we assume that all the different recorders, aquatic or otherwise, rep-

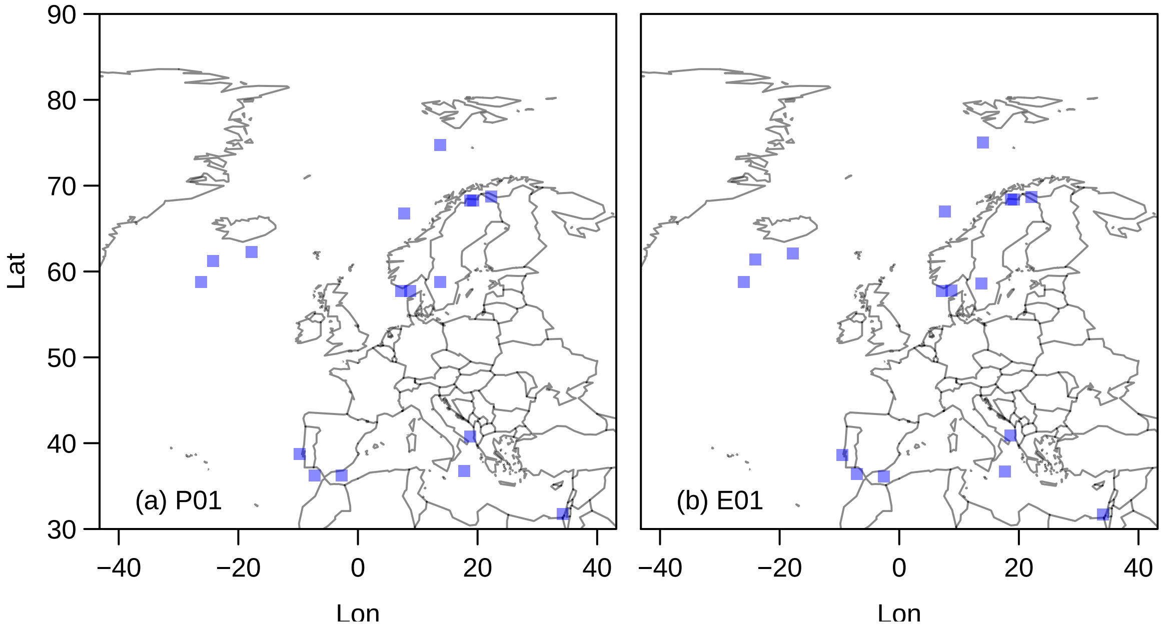

Figure 1. Map of the reconstruction domain and the proxy predictors: ~~(a)~~ **(a)** for the pseudoproxy setup, ~~(b)~~ experiment P01, **(b)** for the main proxy setup, experiment E01.

resent temperature at the surface. This is an assumption of convenience in view of potential habitat biases of the proxy records ~~(Telford et al., 2013; Tierney and Tingley, 2015; Jonkers and Kučera, 2017, 2018; Rebotim et al., 2017; Tierney and Tingley, 2018; Dolma~~

2.2.2 Data handling and use of model simulation pool

Section 2.3.1 gives details on our selected proxies. In short, we choose 17 proxies at locations in the European-North Atlantic sector (Figure 1b) from the compilation of Marcott et al. (2013). These are from a variety of different proxy systems. We take these as published by either Marcott et al. (2013) or the original publications. Therefore, calibrations and uncertainty estimates have diverse origins. Considering the proxy ages, we ~~use these~~ adopt those as published.

Optimally one would choose consistent proxy parameters, a consistent recalibration, and a consistent calibration target. Consistency among parameters and calibration ensures a relation among the proxy predictors, which, one can assume, increases the chance that the proxy records lead to a selection of physically meaningful analogues. In this case, the proxies can effectively anchor the analogue ~~procedure~~selection. We here assume that all chosen proxy-types reliably represent the target of interest and a multi-proxy approach is viable.

The analogue method allows searching for analogues at dates when there is information. One can pool the predictor dates into consistent intervals of, for example, 500 years, and search for analogues for these 500-year pools. One can follow the example of Jensen et al. (2018) and interpolate the proxy records to consistent time steps using the age models for the individual records. We choose here a different approach; we identify all the years for which at least one of the chosen proxy records includes a dated value. We perform analogue searches for these dates according to our considerations on uncertainty, which we describe in the following subsection 2.2.3.

Each data point of a proxy series potentially represents a time-interval of a specific length and the comparison should consider this temporal resolution. That is, if one data point represents a 50-year accumulation and ~~one~~ another data point represents a 500-year accumulation, the procedure ideally accounts for these differences. We decide to use typical resolutions instead of individual resolution estimates to simplify the procedure ~~. Considering the accumulation estimates for the proxies from Marcott et al. (2013)~~ and allow a computationally more efficient analogue search for data and time uncertain proxies. Indeed, it is not necessarily the case that a proxy-record publication includes the information to estimate the pointwise temporal resolution. Considering information provided by Marcott et al. (2013) on their proxies, we conclude ~~that the average resolution of the proxies is~~ for our chosen subset of these (compare Table 1) that the proxies have at best centennial average resolutions. While there are proxies with higher and lower resolutions, a reasonable estimate for the overall average resolution is centennial.

Therefore, we decide to compare the proxy estimates to 101-year averages of the model simulation output. That is, we compare them to 101-year mean values, which we obtain by using a 101-year moving mean on the simulation output time series that is closest to the proxy location. ~~Additionally, in~~

In one test case, we do not preprocess the simulation output but use the annually resolved values of the output for the comparison. For this specific test, we also include the simulation data from the FAMOUS-HadCM3 simulations for the QUEST-project ~~(compare Smith and Gregory, 2012)~~(compare Smith and Gregory, 2012, and section 2.3.2) to which we refer as QUEST FAMOUS simulation. The latter simulation used accelerated forcings and the output data is only ~~available as 10-year snapshots~~representative for decadal forcing conditions. The data is available in monthly resolution for the full simulation period for air temperature in 1.5 meter height, and as snaphots every ten simulation years for surface temperature. We do two more tests with differing resolutions that use 51- and 501-year averages respectively.

We test ~~our~~ the whole approach by using pseudoproxies. We ~~calculate~~ construct the pseudoproxies following the procedure of Bothe et al. (2019a, more specifically their ensemble approach). This approach takes simulated grid-point data and transforms it in multiple steps into a pseudoproxy record. The steps follow the framework of a proxy system model including a sensor model, an archive model, and an observation model (see Evans et al., 2013). Bothe et al. (2019a) first add a noise estimate for environmental non-temperature influences at the sensor stage. This stage also includes adding a bias term due to changing insolation. Next, the archive stage primarily represents a smoothing of this record, which is meant to reflect effects of, e.g., bioturbation. The measurement stage adds another noise term. After sampling this record at a specific number of dates, the procedure, finally, also adds an error term for the proxy data reflecting effective dating uncertainties. ~~Because we separate dating and data uncertainty, we modify~~ We set this term to zero in the script of Bothe et al. (2019a) because we do not aim to transfer dating uncertainty to ~~set the effective dating uncertainty error to zero~~the data uncertainty. Pseudoproxy locations are simulation data grid-points close to the real proxy locations. Figure 1a shows the ~~28 pseudoproxy locations~~ 17 pseudoproxy locations and allows to identify their slight offset to the real proxy locations (Figure 1b) due to the discrete character of the simulation data. The pseudoproxy generation smooths each record to mimic the temporal filtering effects of the real environmental archive. The smoothing length is randomly chosen but temporally uniform for each record. The search for analogues again uses 101-year mean estimates from the simulation pool (compare the ~~previous paragraph~~ paragraph above).

Simulations ~~or simulation projects~~ potentially differ in their modern day climate mean (compare, e.g., Zanchettin et al., 2014). Using anomalies circumvents this issue. One can consider simulation output as anomalies to the climatology over the 20th century or over the full simulation period or over the longest period common to all simulations. For example, Jensen et al. (2018) ~~take anomalies from any proxy or simulation record relative to~~ construct the anomaly record for a data series by subtracting the temporal mean calculated over the full period of ~~this record~~ the record of interest. Their period of interest backs this decision. The proxy records of Jensen et al. (2018) suggest an overall rather stable climate in the North Atlantic during Marine Isotope Stage 3 although a number of Dansgaard-Oeschger (DO) occurred during this period. We presume that using anomalies allows ~~for~~ to include a wider range of simulations and analogue candidates for each date.

In the present case, the period of interest includes mainly the last 15kyr~~for the real proxies and even the last 22kyr for the pseudoproxy test~~. Thus, it spans part of ~~or even the full~~ the deglaciation from the Last Glacial Maximum to the Holocene optimum. Our selection of simulations can only piecewise cover ~~the~~ that period of interest, which complicates the construction of a surface temperature candidate pool. Indeed, the most recent dates differ among the proxy records, and, thus, there is ~~not a~~ no simple procedure to provide anomalies relative to a consistent modern climate. Additionally, using anomalies may introduce climatic inconsistencies if we are interested in climate variables other than temperature. For these reasons, we decide that we cannot reasonably use anomalies. Instead, we try to find analogues for the ~~proxies~~ local proxy reconstructions in their original temperature units without subtracting any climatology.

2.2.3 Proxy uncertainty

We are interested in millennial timescales from the last ~~glacial maximum through the~~ deglaciation until the recent past. On these timescales, uncertainty affects our proxy predictors ~~twofold~~in two ways. First, we have to consider the age or dating uncertainty. Second, the measured proxy data and the temperatures inferred from them are affected by various sources of uncertainty (compare, e.g., Dolman and Laepple, 2018; Reschke et al., 2018; Jensen et al., 2018, and their references).

To our knowledge, previous applications of the analogue method usually did not consider proxies with considerable age uncertainties except for Jensen et al. (2018). Jensen et al. consider the age uncertainty by shifting the date of each proxy by $\pm$ 500 years. Thereby, they obtain an ensemble of 2^{14} reconstructions from which they calculate confidence intervals for their final reconstruction. They do not separately consider the uncertainty of the proxy/reconstruction value. For details, see Jensen et al. (2018).

We choose a different approach (Figure 2). We interpret each data point in a proxy series together with its dating as a data point in the two ~~dimensions~~ dimensional space spanned by temperature and time. Each proxy data point is located on this two dimensional temperature-time plane and each point is surrounded by uncertainty ranges along both dimensions. We can utilize the uncertainty ranges in the two dimensions as our *area of tolerance*, in which the analogue candidates should be located to be considered as good analogues. We can do so at different levels of proxy-time uncertainty, e.g., 90% or 99%, comparable to common expressions of uncertainty intervals. These choices of uncertainty yield increasingly larger areas of tolerance.

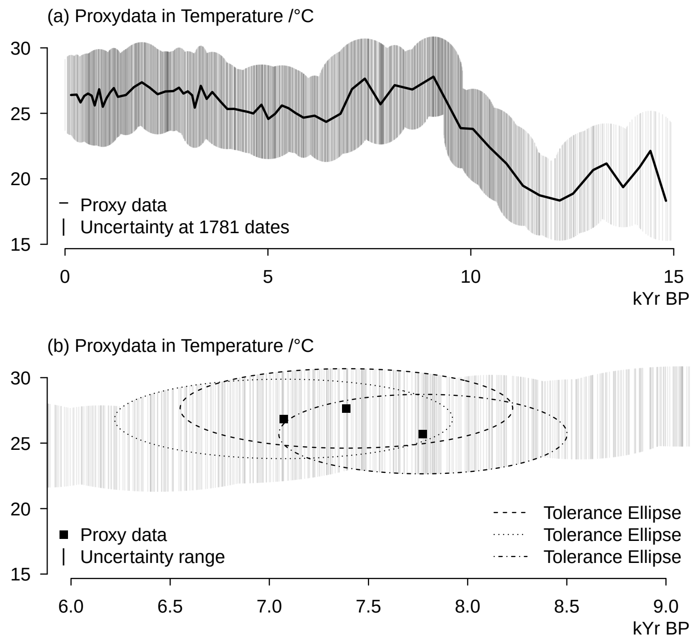

Figure 2. ~~Visualisation of our approach to~~ Considerations on uncertainty and constructing tolerance envelopes: ~~(a)~~ **(a)** Example proxy data (line) and assumed data uncertainty at all dates when we reconstruct values. Number of dates are all dates when any of the proxies included has a dated data point. ~~(b)~~ **(b)** Proxy data at three example dates and ~~uncertainty~~ tolerance ellipses for these dates using data uncertainty and age/date uncertainty.

To define these areas of tolerance we still have to define their shape. Our interest is in finding analogues that agree with the proxy data but also account for these uncertainties. Then, we could take the uncertainty estimates of temperature and time ~~.~~ ~~The uncertainties of the data point allow to~~ to construct a two-dimensional uniform estimate in form of a rectangle of tolerance. Analogue candidates would be valid analogues if they fall locally within these rectangles. If they fall outside of the rectangle they would not be considered valid analogues. Although the uncertainties in temperature and time are commonly taken to be Gaussian, the rectangular approach is the best one if we consider the uncertainties of date and temperature isolated from each other. Then, our tolerance for the temperature data has the same structure at the border of our temporal tolerance range as it has at the central estimate for the date. However, in our application, we do not see both tolerance ranges in isolation. We assume that our tolerance range is a two-dimensional construct in time and temperature. Then, our tolerance construct takes the shape of a two-dimensional Gaussian. This implies that our tolerance areas are ellipses. According to the tolerance ellipses, therefore, we accept fewer analogues for dates far away from the median proxy age estimate. For these dates, analogue candidates need to be numerically very close to the proxy. In contrast, we accept more analogues close to the central age estimate of the proxy, and tolerate that they may more strongly differ from the numerical central estimate of the proxy.

Indeed, two-dimensional Gaussian tolerance areas can be visualised as ellipses. As we have estimates of the uncertainties of the data point, we can construct ellipses of ~~confidence~~ tolerance around each data point. We use the R (R Core Team, 2019) package ellipse (Murdoch and Chow, 2018) whose default ellipse-function follows Murdoch and Chow (1996) by implementing the ellipse equation as

$$(x,y) = (\alpha \cdot \sigma_x \cdot cos(\theta + d/2) + \mu_x, \alpha \cdot \sigma_y \cdot cos(\theta - d/2) + \mu_y), \tag{1}$$

where x is our time-dimension and y is our temperature dimension. Furthermore, α is the tolerance level of interest transformed to a t-test statistic as implemented by Murdoch and Chow (2018), σ_i are the one standard deviation levels of the uncertainties in x- and y-direction, μ_i are the best estimates of the values in x- and ~~these~~ y-direction, i.e. date and data, and $\theta \in [0, 2\pi]$. $cos(d) = \rho$ is the correlation between temperature and time uncertainties. However, we do not consider potential non-zero covariances between dating uncertainty and proxy uncertainty. For simplicity, we also do not take account of the likely correlations between subsequent tuples of data and date.

A two-dimensional tolerance ellipse represents tolerance levels for two-dimensional normal distributed data. However, as in the simple case of a tolerance rectangle, our interest is only in the ellipse as a binary decision criterion to consider the data included in the ellipse and to neglect the data outside of the ellipse. That is, we use the ellipse as an area of tolerance to identify valid analogues from our analogue candidate pool. The ellipses will provide the maximal acceptable distance for a simulated data to be considered an analogue (Figure 2b).

~~We have to assume a certain confidence value for the ellipse construction. The envelopes as shown in~~ That is, the ellipses are not meant to represent the uncertainty ranges in the value of the proxies. They are rather meant to define a limit beyond which an analogue candidate is not considered any more. Essentially the ellipses define a weighting scheme (although with binary weights) based on the proxy and dating uncertainties.

The ellipses are defined from points in the proxy-time space (see Figure 2b~~are pointwise uncertainties~~). We construct ellipses for those data points for which a published record provides ages. ~~The uncertainty envelope for a date and~~ Our tolerance range for a specific date as well as the tolerance envelope for the full proxy record follows from the superposition of the ~~uncertainties~~ tolerance ellipses from successive data points (see panels of Figure 2 and later Figures ~~4 and 7~~5 and 8). This envelope generally provides for each date ~~an upper and a lower limit of potential values or, potentially, results in lack of defined values~~ upper and lower limits of values that the analogue candidates need to fall between. However, the envelope may also result in the impossibility to define analogues for certain locations or even for all locations for certain dates.

That is, ~~while a two-dimensional uncertainty ellipse represents uncertainty levels for two-dimensional normal distributed data, our interest is only in the ellipse as a decision criterion to consider the data included in the ellipse and to neglect the part of the distribution outside of the ellipse. Effectively, we treat the data as if they are uniform distributed inside of the ellipse for one two-dimensional data point or inside the envelope of multiple overlapping data points~~the superposition of ellipses constructs a tolerance envelope (Figure 2a), which we use to identify valid analogues from our candidate pool. The ellipses around the data points ~~, therefore,~~ mark the limit of their ~~2-dimensional~~ pointwise two-dimensional area of influence in our search for spatially resolved analogues~~:~~. Their superposition is essential for identifying those simulated data ~~that fall within the ellipse are~~ to be equally considered as analogues~~and those that fall outside the ellipse are not considered at all. For simplicity,~~

~~we do not take account of the likely correlations between subsequent tuples of data and date. We also do not consider potential non-zero covariances between dating uncertainty and proxy uncertainty~~. If the tolerance ranges for multiple data points in a record overlap for a given year, we simply take their maximal ranges. Simulated data that fall outside the tolerance ranges are rejected.

Because we provide ~~such pointwise estimates and envelopes for the climate of the~~ reconstructions only for those years for which one of the chosen proxy records includes a dated value, and because our tolerance estimates are essentially pointwise, the envelope does not necessarily cover all years within the period of interest. Furthermore, because we use the envelopes as decision criterion, it can happen that the method fails to find any valid analogues for given years. ~~The~~

Our pointwise estimates are compliant with the initial uncertainty of the proxies and our final reconstruction uncertainties are an expression of this initial confidence in the local data. This is in contrast to Jensen et al. (2018), who provide an ensemble of reconstructions. Their uncertainty estimate measures the uncertainty of the initial reconstruction relative to shifted ages. That is, the two different applications of the analogue method consider different things in their uncertainty estimates. The reconstruction uncertainty in our approach originates from the selected analogues. ~~In our approach, if uncertainties for multiple data points in a record overlap for a given year, we simply take their maximal ranges.~~

2.2.4 Analogue search

The ellipses of ~~uncertainty~~ tolerance allow in theory to produce reconstructions for each year included in the dating uncertainty. That is, if a proxy series has a value dated to the year 500 BP ~~and~~ with a dating uncertainty of $\sigma = 50$yr, and if we decide to consider dates within $\pm 2\sigma$ then we can ~~consider the proxy record for the~~ search for analogues from 600 to 400 BP. However, we decide to only reconstruct values at those dates at which at least one proxy is dated. That is, if only this hypothetical proxy has a dated value between 600 and 400 BP and it only has this one dated value, we perform the reconstruction only for the year 500 BP. Our assumption is that this maximises the link between the reconstruction and the underlying proxies. Thus, if we increase the width of the ~~uncertainty~~ tolerance envelope, we usually do not obtain reconstructed values at more dates but only increase the probability to find a valid analogue at a certain date.

~~The~~ In other applications of the analogue method, the choice of a valid analogue usually relies on a distance metric. This is commonly the Euclidean distance (compare Franke et al., 2010; Gómez-Navarro et al., 2017; Talento et al., 2019), although Jensen et al. (2018) use an unweighted root mean square error (RMSE) as distance metric between their proxies and the analogue candidates from their simulation pool. Based on such a distance, one can select the best fit, a small number of good fits, e.g., the ten analogues with the smallest distance, or a composite or interpolation of a small number of good fits.

Here, we deviate from this and decide neither on a fixed number of analogues nor on a defined metric. Candidates in our pool are valid analogues if they are within ~~a certain uncertainty interval around the proxy predictors~~ the tolerance range (compare section 2.2.3). That is, as described above, we have an envelope of ~~credible~~ tolerance values for certain years and each proxy record. ~~A~~ For our standard approach, a candidate is a valid analogue for a date if ~~all proxies defined at this date include the candidate values in their credible interval. Credible intervals~~ it falls within the ellipse of tolerance for all proxies. We also mention tests where an analogue is valid if it is outside the ellipses at one location, at two locations or at 25% of the locations.

We consider only a small set of potential ellipses. These are 90% and 99.9% for the pseudoproxy approach, and either 99% or 99.99% for the various proxy setups.

We additionally show one instance of a reconstruction using just one best analogue. For this test, we choose the analogue with the smallest Euclidean distance to our proxy values. As we deal with proxy records that are irregularly spaced in time, we have to find a way to select dates for which to do a single-best analogue reconstruction and get the proxy values for these dates. To do so, we consider the proxy values valid at all dates within their a 90% dating uncertainty, then identify the range of these values, and take the mid-point of the range as the proxy value for this date.

In short, our reconstruction is based on the following workflow. We have a set of sparse proxy predictors and a pool of simulated fields. As our proxies are not only sparse in space and uncertain in their values but also irregular and uncertain in time, we have to decide, (a) when to compare them, (b) in which resolution to compare them, and (c) how to consider the uncertainties in time and value. Therefore, we decide to (i) compare the proxies and simulated data for all dates when one proxy is dated, to (ii) compare the proxies to 101-moving means of the simulated data, and (iii) to take the proxy data values as valid within an ellipse of tolerance around the dated value in time and temperature space. Then analogue candidates are valid analogues if they are within these tolerance ranges around all proxy records included in the search.

2.3 Data

2.3.1 Proxies

We concentrate on a European-North-Atlantic domain (Figure 1). There, we choose 17 locations with proxy-inferred temperature records from the collection of Marcott et al. (2013, see also Tables 1 and 1). Nine of these series use alkenone $U_{37}^{K'}$ but the set also includes temperatures derived from foraminifera Mg/Ca (2 records), pollen (2), chironomids (2), TEX86 (1), and foraminiferal assemblages (1) (compare Table 1 and ~~Appendix Figure ??~~Figure 3). For the various proxy types see, e.g., Rosell-Melé et al. (2001, and their references) or Tierney and Tingley (2018, and their references) for $U_{37}^{K'}$, Anand et al. (2003, and their references) or Tierney et al. (2019, and their references) for foraminiferal Mg/Ca, Kim et al. (2008, and their references) or Tierney and Tingley (2015, and their references) for TEX86, Seppä and Birks (2001) and Seppä et al. (2005) for the specific pollen records, Larocque and Hall (2004) for the specific chironomid records, and Sarnthein et al. (2003a) for the specific record using foraminiferal assemblages.

Figure 1b shows the proxy locations. These are not all records within the domain from Marcott et al. (2013). We ~~exclude~~ do not consider additional seasonal attributions for the foraminifera assemblage data of Sarnthein et al. (2003a, compare also Marcott et al., 2013; Sarnthein et al., 2003b). We further excluded the alkenone unsaturation ratios of Bendle and Rosell-Melé (2007, see also Marcott et al., 2013) as well after initial tests due to concerns about the potential influence of sea-ice in simulations. Indeed, 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. We exclude two more records because they are co-located with other proxies. That is, we do not use the stacked radiolaria assemblage records of Dolven et al. (2002, see also Marcott et al., 2013) because the upper part of the record is from the same upper core as the $U_{37}^{K'}$ data of Calvo et al. (2002, see also Marcott et al., 2013).

Table 1. Information about the considered proxy records: IDs, geographical location, seasonal attribution according to Marcott et al. (2013), proxy type, seasonal attribution used here, and analogue search setups that use the record. All proxy data are from the supplement of Marcott et al. (2013). Table 1 provides references to original publications and data sets. Proxy setups refer to those analogue search tests where this proxy is included (see Appendix ??compare also Figure 3).

Proxy ID	Lat	Lon	Season in Marcott et al. (2013)	Proxy type	Season used	Proxy setups
MD95-2043	36.1	-2.6	Annual	$U_{37}^{K'}$	Annual	1-3, 4-6
M39-008	36.4	-7.1	Annual	$U_{37}^{K'}$	Annual	1-2, 4-7, 9
MD95-2011	67	7.6	Summer	$U_{37}^{K'}$	Summer	1-4
ODP984	61.4	-24.1	Winter	Mg/Ca (N. pachyderma d.)	Winter	1, 7-9
GeoB 7702-3	31.7	34.1	Summer	TEX86	Summer	1, 5-9
IOW225517	57.7	7.1	Spring to Winter	$U_{37}^{K'}$	Annual	1-4, 6
IOW225514	57.8	8.7	Spring to Winter	$U_{37}^{K'}$	Annual	1-4
M25/4-KL11	36.7	17.7	Spring to Winter	$U_{37}^{K'}$	Annual	1-7
AD91-17	40.9	18.6	Annual (seasonal bias likely)	$U_{37}^{K'}$	Annual	1-6
Lake 850	68.4	19.2	Summer	Chironomid transfer function	Summer	1, 7-8
Lake Nujulla	68.4	18.7	Summer	Chironomid transfer function	Summer	1, 7-8
MD95-2015	42 58.8	-26	Annual	$U_{37}^{K'}$	Annual	1-4
D13882	38.6	-9.5	Summer	$U_{37}^{K'}$	Summer	1-6,8
GIK23258-2	75	14	Summer	Foram transfer function	Summer	1, 4-9
Flarken Lake	58.6	13.7	Annual	Pollen MAT	Annual	1, 7-9
Tsuolbmajavri Lake	68.7	22.1	Summer	Pollen MAT	Summer	1, 5-9
RAPID-12-1K	62.1	-17.8	Late Spring to early Summer	Mg/Ca (G. bulloides)	Summer	1, 6-9
GeoB 5901-2	36.4	-7.1	Annual	$U_{37}^{K'}$	Annual	3

Similarly, we ad hoc decide to use the $U_{37}^{K'}$ data of Cacho et al. (2001, see also Marcott et al., 2013) instead of that of Kim et al. (2004a, see also Marcott et al., 2013), which are basically co-located. We use the data of Kim et al. (2004a) in one alternative proxy setup(see Table 1 , Appendix ??, and Appendix Figure ??). Table 1 and Figure 3 provide details on our different proxy setups. All in all, we consider nine different setups of proxy networks, which we name E01 to E09. In a pseudoproxy-setup, we use a network of locations equivalent to E01 and therefore name this pseudoproxy-setup P01.

We consider the seasonal attributions of individual proxy records in our search for analogues. We generally take the attributions , the calibrations, and the uncertainties and the calibrations for the records as published by Marcott et al. (2013) but also check the references provided by them. Seasonal attributions are diverse for the various proxy records. The majority is either for summer season (7) or annual (8) according to Marcott et al. (2013). We compare the proxies to the simulation output season

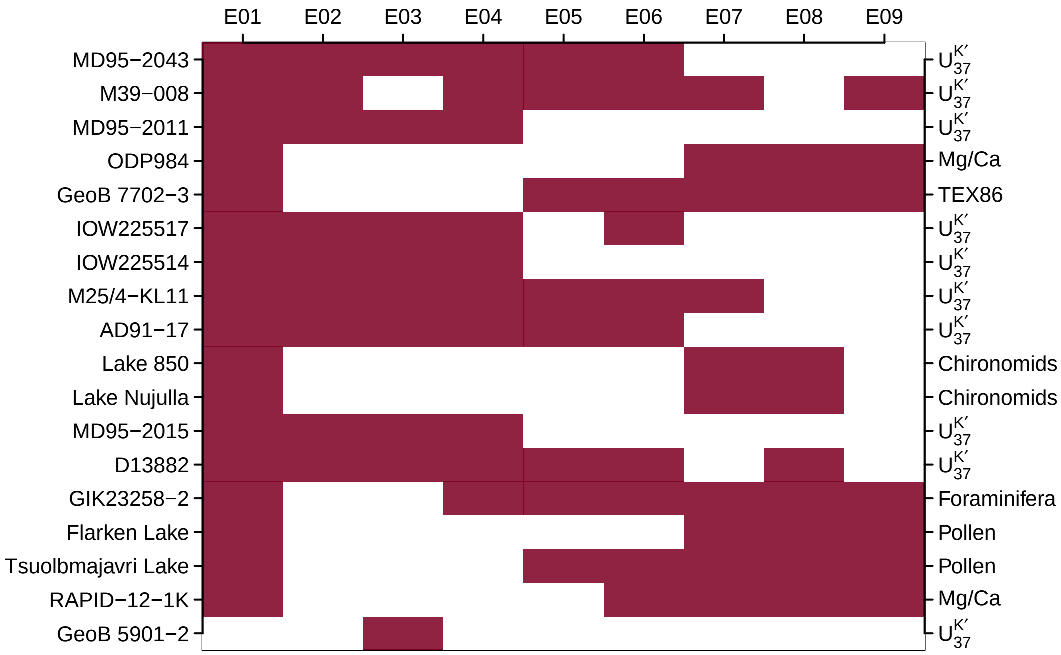

Figure 3. Information about the different proxy setups: Matrix of proxy records against proxy setup (E01-E09). For more information see Table 1. White-out means that the relevant proxy is not included in the respective proxy setup.

that is close to the seasonal attribution as given by Marcott et al. (2013) or the original publication. For simplicity's sake, we only consider the modern meteorological seasons DJF (December to February), MAM (March to May), JJA (June to August), and SON (September to November) as well as the calendar annual simulation means (compare Table 1). We do ignore possible calendar effects (e.g., Bartlein and Shafer, 2018; Kageyama et al., 2018).

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 ~~a uniform~~ an uncertainty of $\sigma = 1K$ for all proxies. This reduces the uncertainty for some records and potentially increases the uncertainty for others. We regard this to be a reasonable simplification for records ~~where~~ for which we are unable to infer full uncertainties for the temperature reconstructions either from Marcott et al. (2013) or from the original publications.

We performed reconstruction exercises for various proxy setups. ~~In the main manuscript we~~ We concentrate on the full set of proxies mentioned above (see Figure 1b and first 17 lines of Table 1). Figure ~~3~~4b visualizes how many of these 17 proxies are available for the dates for which we aim to reconstruct temperature. The ~~panel~~ Figure shows this for two different assumptions on uncertainty (red and grey lines, see section 2.2.3).

~~Appendix ?? includes information on the additional attempts at reconstruction by analogue (Appendix Figures ?? to ??, compare also Table 1). Appendix 3.2.1 provides the results~~We perform additional reconstruction exercises. Figure 3 and Table 1 give a first impression of the setups. We shortly ~~discuss~~ describe the results for these alternative setups in our results section below. Most notably among these alternative tests are setups that use only $U_{37}^{K'}$ proxies (Appendix Figure ??c,d, see also Appendix Figure ??Figure 3). The difference between the two $U_{37}^{K'}$ setups is that E03 (Figure ??) uses record GeoB 5901-2 instead of record M39-008 (compare Table 1 and Figure ??). Figure ?? supplements the information in Figure 3for all reconstruction attempts3).

2.3.2 PseudoproxiesPseudoproxies

We test our analogue method using pseudoproxies. We calculate the pseudoproxies following the procedure of Bothe et al. (2019a, more specifically their ensemble approach) but omit their effective dating uncertainty error term. They provide pseudoproxies based on simulated annual mean temperature and for a global selection of grid points from the TraCE-21ka simulation (He, 2011; Liu et al., 2009). Here, we calculate the pseudoproxies for summer season (June, July, August; JJA) and annual average data and for the chosen European-North-Atlantic domain only. The seasonal choice is due to the large number of proxies attributed to summer in our real proxy selection. The approach provides also approach also provides randomly chosen pseudo age uncertainties. The Following Bothe et al. (2019a) and their repository (Bothe et al., 2019b), these base on assumptions on the smoothing of the pseudoproxies and a Gaussian term.

Here, the pseudoproxy computation uses the TraCE-21ka simulation (He, 2011; Liu et al., 2009). As Bothe et al. (2019a)QUEST FAMOUS simulation data (Smith and Gregory, 2012). Specifically, we use the simulation ALL-5G (see tables 2 and 5). For details on this and the other QUEST-FAMOUS simulations, please see Smith and Gregory (2012). The FAMOUS-HadCM3 simulations for QUEST use accelerated forcings (compare Smith and Gregory, 2012). That is, the last glacial cycle of approximately 120,000 years of climate forcing was simulated in approximately 12,000 simulation years. Thus, the annual simulation data is only representative of ten years of climate evolution. Data is available in simulation monthly resolution for air temperature in 1.5 meter height but only as snaphots every ten simulation years for surface temperature. We use the simulation year annual means of the air temperature data for the construction of the pseudoproxies. The FAMOUS-HadCM3 simulations use a very low resolution atmospheric model with a 5 degree latitude by 7.5 degree longitude grid. Therefore, we use the cdo application from the Max Planck Institute for Meteorology (https://code.mpimet.mpg.de/projects/cdo/, last access: 18 August 2020) to remap the data to a 0.5 by 0.5 degree grids and use this for the pseudoproxy calculations. In this remapped data we follow Bothe et al. (2019a) and use grid point data close to proxy locations identified in the work by ?, ?, or Marcott et al. (2013). ? and ? focus on the deglaciation while Marcott et al. (2013) consider the Holocene. We use this larger collection for the pseudoproxy setup with the aim to test the method for an as large as possible collection of locations that have been previously consideredused in the realistic setup.

We modify the pseudoproxy script of Bothe et al. (2019a) to account for the reduced temporal resolution of the available QUEST FAMOUS data. This primarily means considering the default parameter settings that are given in time units. This also includes ad-hoc scaling the randomly chosen dating uncertainty to approximate the distribution of the observed dating uncertainties. The latter modification also avoids that individual data points have an overly strong influence within our envelopes of tolerance.

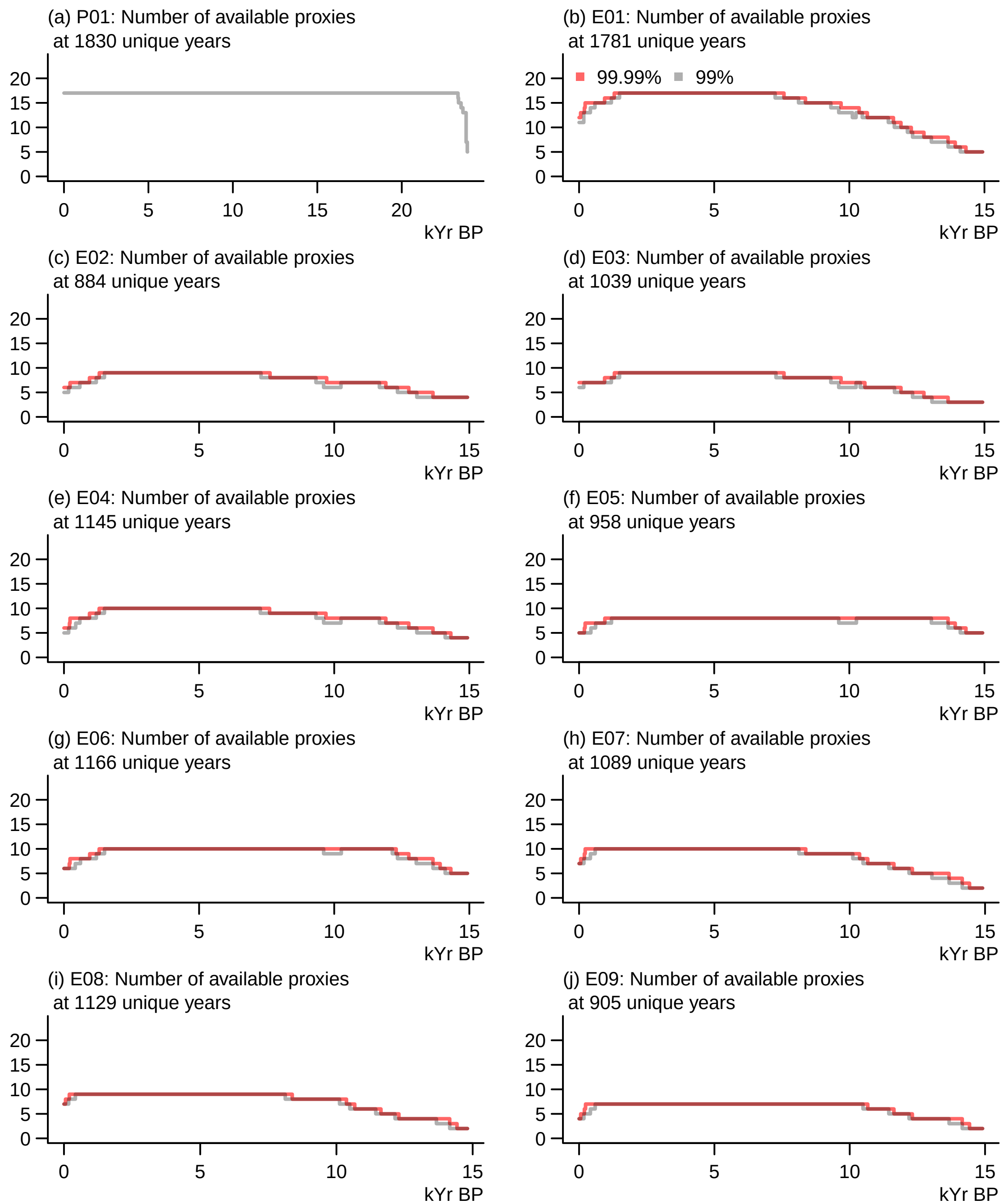

Figure 4. Information about the number of available proxies for the dates to be reconstructed: ~~(a)~~ **(a)** the pseudoproxy setup, ~~(b)~~ **(b)** to **(j)** the various proxy ~~setup~~setups according to Figure 3 and Table 1. In ~~(b)~~ **(b)** to **(j)** we show results for two different assumptions on the uncertainties, a 99% envelope and a 99.99% envelope (compare section 2.2.3).

Table 2. Information about the pool of simulation data: Model name, the project for which the simulations were performed, the simulated periods from this model output, the number of total years. All simulation data are remapped to 0.5 by 0.5 degree grids. References and data locations are provided in Appendix Table 2. The Appendix also lists all individual simulations used in tables 3 to 5. Note FAMOUS-HadCM3 uses accelerated forcings~~and there are only snapshots available~~. We, thus, chose to exclude this simulation for most cases.

Model	Project	Periods	Total years
CNRM-CM5	PMIP3	LGM, MidHolocene	400
COSMOS-ASO	PMIP3	LGM	600
CSIRO-Mk3L-1-2	PMIP3	LGM	500
GISS-E2-R	PMIP3	LGM, MidHolocene, past1000	9309
HadCM3	PMIP3	past1000	1001
HadGEM2-CC	PMIP3	MidHolocene	35
HadGEM2-ES	PMIP3	MidHolocene	102
IPSL-CM5A-LR	PMIP3	LGM, MidHolocene, past1000	1701
MIROC-ESM	PMIP3	LGM, MidHolocene, past1000	1200
MPI-ESM-P	PMIP3	LGM, MidHolocene, past1000	1400
CESM1	Last Millennium ensemble	past1000, pre-industrial control, industrial	33156
CCSM3	TraCE-21ka	LGM to present	22040
MPI-ESM-Cosmos	MILLENNIUM COSMOS	past1000, pre-industrial control, industrial, projection	5909
FAMOUS-HadCM3	Quaternary QUEST	Last Glacial Cycle	6014

Figure 1a shows the ~~28~~ 17 pseudoproxy locations. ~~Figure 3a~~ These are close to the realistic proxy locations. Figure 4 visualizes the number of pseudoproxy locations with data against the dates at which we try to reconstruct values (black lines). The pseudoproxy records are shown in Figure ~~4~~ 5 below. The figure also visualizes our assumptions on the uncertainty of the pseudoproxies (compare section 2.2.3) in terms of the tolerance envelope.

2.3.3 Simulations

Table 2 provides a general overview of the various simulations in our pool of candidates. Supplementary Tables 2 to 5 give additional information. We only consider previously published simulations. These stem from a variety of projects and were performed with a variety of models. The projects are TraCE-21ka "Simulation of Transient Climate Evolution over the last 21,000 years" (Liu et al., 2009), the Paleoclimate Modelling Intercomparison Project Phase III (PMIP3, Braconnot et al., 2011, 2012), the CESM Last Millennium Ensemble Project (Otto-Bliesner et al., 2015), the Max Planck Institute Community Simulations of the last Millennium (Jungclaus et al., 2010), and Quaternary QUEST (e.g. Smith and Gregory, 2012).

We use simulations for various different time periods to increase the candidate pool. We assume that simulation climatologies can differ over a relatively wide range (e.g., Zanchettin et al., 2014). Simulations from the TraCE-21ka and the QUEST projects are transient over periods covering the last approximately 22kyr and the last glacial cycle respectively. Otherwise, the simulations are transient over the last millennium, or time-slices for the Mid-Holocene and the Last Glacial Maximum. Additionally we also include pre-industrial control simulations. Such a multi-model and multi-time-period candidate pool effectively follows suggestions of Steiger et al. (2014). We note that considering simulations for the last millennium as candidate for the Last Glacial Maximum can introduce climatological inconsistencies if the method identifies these fields as analogues.

The FAMOUS-HadCM3 simulations for QUEST use accelerated forcings (compare Smith and Gregory, 2012) and the output for surface temperature is only available in steps of ten simulation years. Therefore, we use this simulation only for a test case and exclude it for the main discussions.

We remap all simulation output to a 0.5 by 0.5 degree grid for the construction of pseudoproxies and for the search for analogues. The motivation is that thereby fewer proxies are close to the same grid point. However, resulting differences ~~are likely small~~ between grid points are likely small. We use the original resolution for the final regional average reconstructions and the evaluation of field data. Local grid point evaluations are done against the remapped files.

3 Results

3.1 Pseudoproxy application

The pseudoproxy application allows highlighting the possibilities of our implementation of the analogue method. It further already provides a glimpse at potential problems.

Our implementation of the analogue method searches for analogues within the full pool of simulation fields but excluding the FAMOUS-HadCM3 output from the QUEST project. Pseudoproxies are ~~for the virtual climate in the TraCE-21ka simulation~~ derived from this latter simulation to which we refer as the QUEST FAMOUS simulation data. We compare the pseudoproxy predictors to 101-year moving averages of the simulation output. We ~~use~~ concentrate on 90% ~~uncertainty~~ tolerance ellipses in the pseudoproxy application of the analogue search but also include results for 99.9% tolerance ellipses. Valid analogues are those simulation fields that are within ~~these resultant uncertainty~~ the resultant tolerance envelopes for all pseudoproxy locations available for a date.

Temperatures are reconstructed for the full domain of the European-North-Atlantic sector including the Arctic as shown in Figure 1. Figure 1a shows the ~~proxy~~ pseudoproxy locations. The time period of interest ranges from the last glacial maximum until the late 20th century of the Common Era~~, i.e. it is the full period of the pseudoproxies from the Trace-21ka simulation. Figure 3~~. Figure 4a highlights that most pseudoproxies are defined at all dates. That is, the chosen sample dates of the pseudoproxies are ~~very~~ close to each other and, thereby, the generated dating uncertainties result in relatively large overlaps. Figure ~~4~~ 5 presents the pseudoproxies including their ~~uncertainty~~ tolerance envelopes.

In this setting, the analogue search tries to identify analogues for ~~4,947~~ 1830 dates. Our implementation finds between ~~6 and 14,343 analogues at 4,935 dates~~ 1 and 7,919 analogues at 531 dates (Figure 6b); it fails to find analogues for ~~12~~ dates

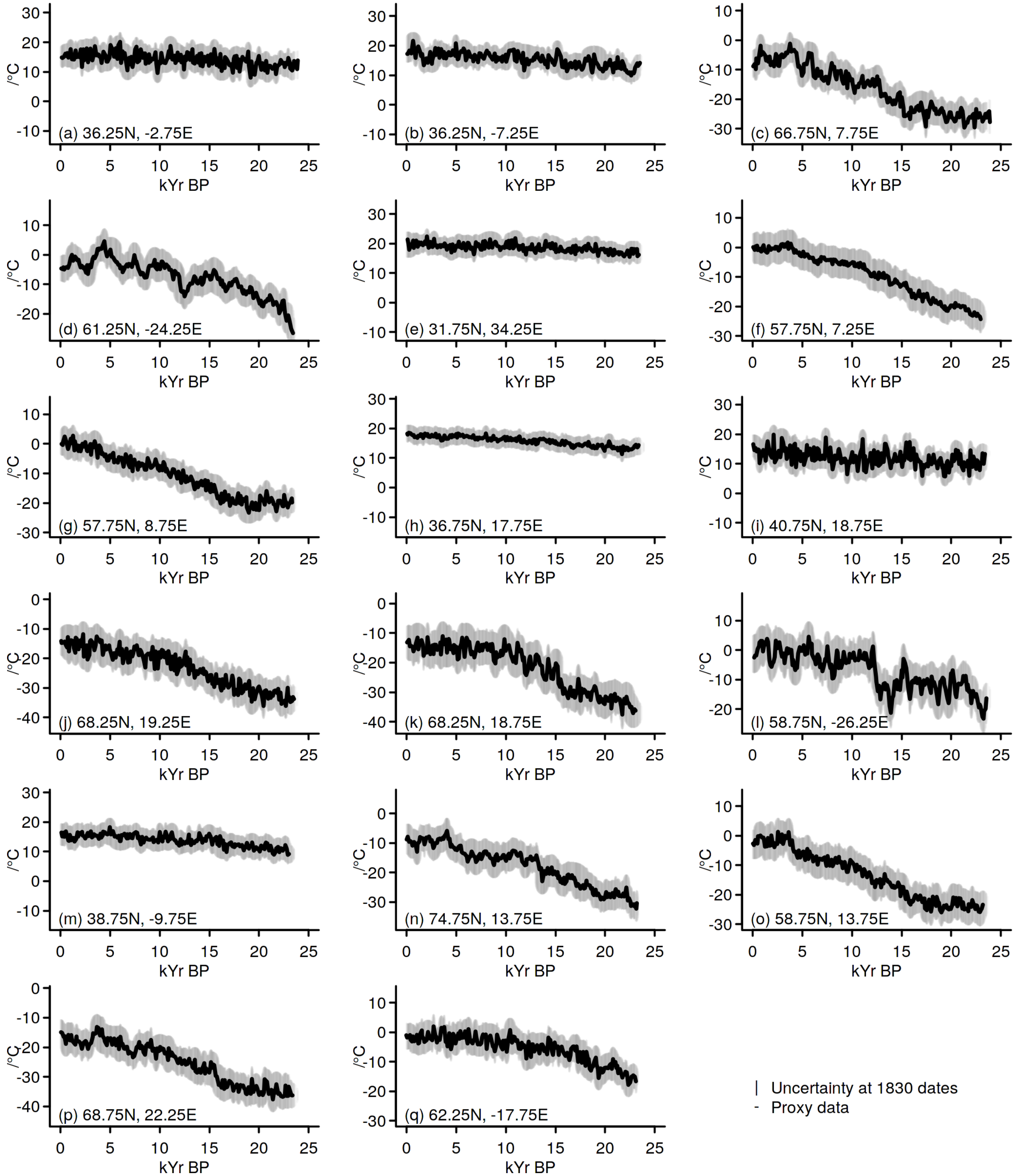

Figure 5. Pseudoproxy data and assumed uncertainties for the ~~28~~ 17 locations in our pseudoproxy application.

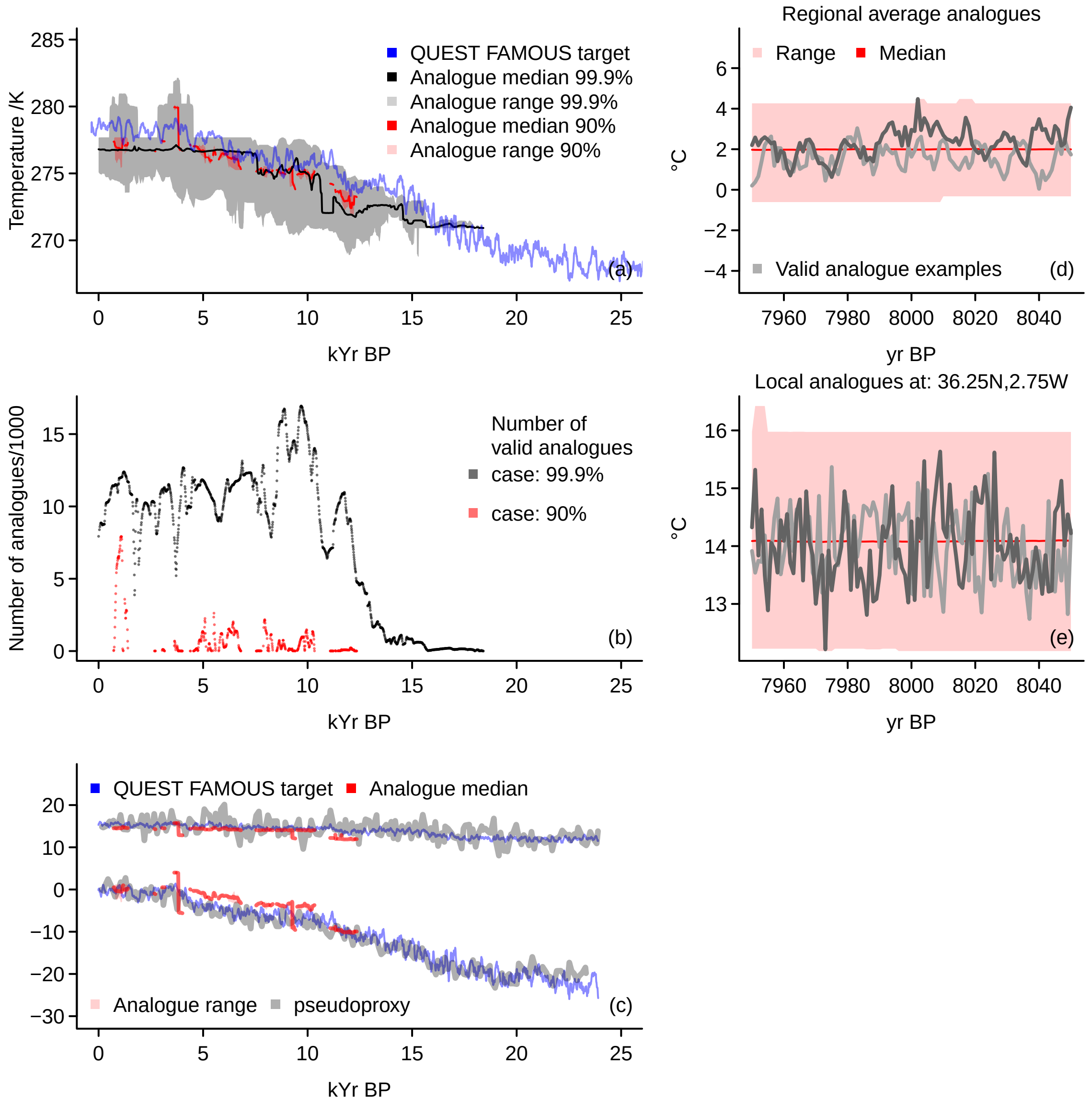

Figure 6. ~~Visualising reconstruction~~ Reconstruction results for the pseudoproxy application of the analogue method: ~~(a)~~ **(a)** regional averages for 101-year moving averages ~~,~~ for two different tolerance envelope levels (~~b~~90%, reds, 99.9%, greys), **(b)** number of analogues for both setups, ~~(c)~~ **(c)** local results for 101-year moving averages for 90% tolerance envelopes only, ~~(d,e)~~ **(d,e)** expansion of regional ~~(d)~~ **(d)** and local ~~(e)~~ **(e)** 101-year moving averages in regional and local 101 year series ranges for 90% tolerance envelopes only. ~~(a)~~ **(a)** shows the 101-year moving average regional target in the TraCE-21ka simulation (blue), the median of all analogues (~~black)~~90%, ~~the range of all analogues (light~~ red, 99.9%, black), and the ~~90%~~ range of ~~the~~ all analogues for the respective tolerance ranges (~~red~~colored shading). ~~(b)~~ **(b)** shows the number of analogues found for each of the dates considered. ~~(c)~~ **(c)** adds for two locations (warmer case: 36.25N, 2.75W; colder case: 57.75N, 8.75E) the local pseudoproxy data (grey), the local target (blue), the range of all local analogue values (light red), and the local median of the analogues (red). (d,e) show the median (red), the range (light red), and two valid analogue examples of the extension of 101-year mean analogues into ~~101 year~~ 101-year long time-series for the ~~(d)~~ **(d)** regional averages and ~~(e)~~ **(e)** one location. Due to the coarse resolution for the QUEST FAMOUS data, panels **(c)** and **(e)** use the remapped data of the simulation.

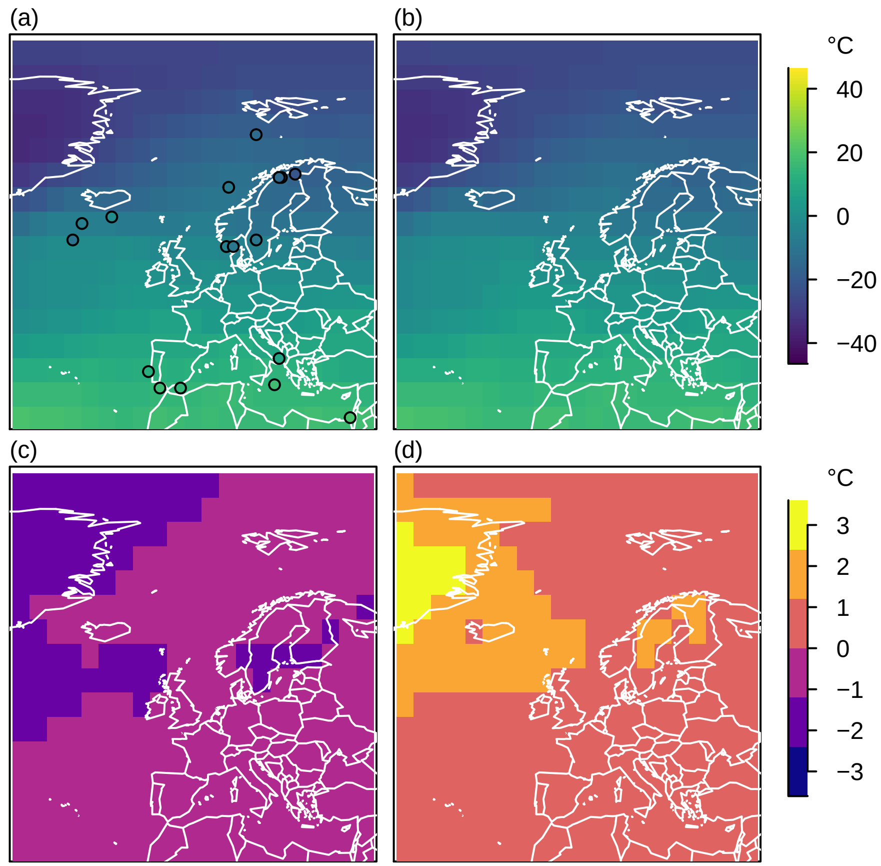

Figure 7. ~~Visualising~~ Temperature field ~~information~~ reconstructions in °C for the pseudoproxy approach: Example of a 101-year mean annual temperature analogue reconstruction for the European-North-Atlantic sector in °C ~~. (a)~~ centered in the year 8,000 before 1950. **(a)** One example analogue, ~~(b)~~ **(b)** local median of all analogues, ~~(c)~~ **(c)** local minimum of all analogues, and ~~(d)~~ **(d)** local maximum of all analogues. Panels ~~(c)~~ **(c)** and ~~(d)~~ **(d)** show differences to the median in °C.

during the deglaciation and the glacial maximum. Analogues stem ~~only~~ mainly from the Trace-21ka simulation~~from which we also constructed the pseudoproxies for this test. This is not surprising because the local climatologies of~~ . Occasionally, output from the ~~TraCE-21ka simulation are slightly distinct from the other simulations (not shown). Using anomalies might compensate for this but please see our reasoning why we do not use anomalies in section 2.2.2~~PMIP3 past1000 simulation with IPSL-CM5A-LR is also classified as valid analogues.

Results change if we consider a wider tolerance envelope. For an 99.9% tolerance envelope instead of a 90% one, we are able to find between 1 and 16,944 valid analogues at 1,438 of 1,830 dates (Figure 6b). Analogues stem from four additional simulations compared to the smaller tolerance envelope. These are the PMIP3 midHolocene and lgm setups of IPSL-CM5A-LR, the COSMOS-ASO lgm setup, and the GISS-E2-R past1000 ensemble member r1i1p122.

Figures ~~5 and 6~~ 6 and 7 provide information on which type of ~~results we can obtain from analogue reconstructions. They plot results for the reconstruction from pseudoproxies~~reconstructions we obtain from our analogue method. Panel (a) of Figure ~~5~~ 6 is for the area mean reconstructions ~~and it~~ for two different tolerance envelopes. It shows the resulting reconstruction ~~median in black , the full range of potential analogues in light red , and a~~ medians in black and red for 99.9% and 90% ~~interval in darker red~~tolerance assumptions respectively. The blue line in the panel is the 101-year moving average regional temperature from the simulation, i.e. the reconstruction target.

~~It is encouraging that the target rarely falls outside of the envelope~~ Shading in the panel shows the full range of potential analogues. ~~However~~

The results are encouraging but problems are obvious. We are able to find valid analogues for both tolerance ranges. Analogues are regularly relatively close to the target for the narrow tolerance range. However, their number is often small and there are periods without any valid analogues. The range does seldom include the target. Further, the reconstruction with a narrow tolerance assumption does not provide valid analogues earlier than approximately year 13,~~we note that~~ 500 BP.

On the other hand, the range of potential analogues is ~~very wide~~only weakly constrained for the wider tolerance range. For example, the analogue search ~~regards up to approximately two thirds of the simulated time period~~ may regard more than 17,000 records of the TraCE-21ka simulation as valid analogues ~~during the TraCE-21ka-equivalent of the Bølling-Allerød~~ around the year 10,000 BP (compare Figure ~~5~~6b). ~~Additionally, the proxies~~This wide range often includes the target. However, the target is mostly above the median estimate. The reconstruction gives a rather constant estimate from a small number of analogues for the period earlier than 16 thousand years before present.

The pseudoproxies, together with their uncertainties, are a weak constraint during most of the ~~Holocene. The reconstruction envelope gives a constant upper limit for the most recent ~9kyr and approximately the oldest 5kyr, and the lower limit is approximately constant between ~15kyr and ~18kyr BP~~period of interest if we assume a wider tolerance but they fail to capture the target if we assume a stronger knowledge about their value. In addition, the reconstruction envelopes and medians show rather little variability and often give nearly constant values over long periods. That is, the set of valid analogues has a notable overlap for these periods. The lacking variability among analogues together with the ~~very~~ potentially wide range of analogues ~~results in very little~~ is reflected in the small variability in the reconstruction median~~over certain periods~~.

Besides the regional average, the results also allow to extract the local representations. Figure ~~5~~6c shows two examples ~~. We will~~ for the narrow 90% tolerance assumption. These are for the pseudoproxies at 36.25N,2.75W and 57.75N, 8.75E. We refer to those as the warmer ~~and colder cases~~southern and colder northern locations respectively. The panel plots again the target simulation output in blue, the full analogue range in light red, and the analogue median in red. We also add the pseudoproxy in grey.

~~In the warmer case~~At both locations, the range is very small for the narrow tolerance range. At the southern location, the reconstruction median ~~captures well the average characteristics of the pseudoproxy, which already follows the underlying simulation targetrather closely. Again~~is generally below the target and the range is hardly identifiable and does not include the target. This is comparable to the northern location, where, however, the median is generally above the target. Even for the wide tolerance range the target is more often outside than within the full analogue range at the southern location while at the northern location the range includes the target regularly (not shown). Thus, the range of potential analogue cases ~~and its median show little variability. The narrow reconstructed range does not necessarily cover the moving averages of the target data. More importantly, it also fails to cover the full variabilityof the pseudoproxy~~is still relatively narrow at the southern location but can be already quite wide at the northern location. Also locally, analogue range and median show little variability. In the ~~colder~~ northern case, the analogue ~~median fails~~ medians fail for both tolerance assumptions to capture the average characteristics of the pseudoproxy except for approximately the most recent ~~5kyr. The pseudoproxy also deviates slightly more clearly from the target data than for the warmer example. Indeed, the analogue median differs less from the target data than from the pseudoproxy. The range of the analogues mostly includes the moving averages of the target data but not the pseudoproxy for the earliest portion and for colder climates. The pseudoproxy reconstruction~~ 3kyr.

The pseudo-reconstruction results suggest that the approach can provide local information in addition to the regional average. Relatively wide tolerance appears to be necessary to capture the local characteristics at the two chosen locations. This is more successful for some periods but success always varies regionally.

Since we search analogues among temporal moving window averages, the analogue search provides one more result of interest. Any analogue state represents a temporal average and we can expand it to provide information about the time-variations underlying the analogue average climate state. That is, we obtain climate evolutions that comply with our proxy-constraints. This, for example, allows to get an impression of how temperature changed on sub-centennial, e.g. interannual timescales, or to obtain an estimate of the interannual variability. Panels (d) and (e) of Figure ~~5 provide these information~~6 provide these informations under a narrow tolerance assumption. The panels show the range and the median of 101-year series. They also add two examples of 101-year time-series. Panel (d) is for the regional average, and panel (e) for ~~a grid point~~ the grid point at 36.25N, 2.75W. Both show 101-year expansions around the average centred in the year 8,000 BP.

~~There is a very wide~~ Although we consider a narrow tolerance range, which results in very narrow ranges around the mean analogue state, the expanded range of potential analogues ~~. However, the examples~~ is still notably wide. The two examples of valid analogues highlight how much two climates may differ over the period although ~~they result in a valid analogue for the proxies under uncertainty.~~ both are valid analogues considering the proxy uncertainty. Wider tolerance ranges give larger ranges of reconstructions and result in larger differences between the 101-year time-series.

Finally, ~~the~~ our reconstruction approach allows considering the spatial fields ~~for~~ of valid analogues. Figure ~~6~~ 7 adds an example for 101-year mean ~~summer~~ annual temperature. It shows one valid analogue field in panel (a) and the local median, minimum, and maximum values of all analogues in panels (b) to (d) respectively. The chosen date is the year ~~8006 BP~~ 8,000 BP for the narrow tolerance range. Panel (a) also adds the values for the pseudoproxies that enter the analogue search. The example analogue and the pseudoproxies agree to some extent but ~~there is notable disagreement especially at more northern latitudes~~disagreement is notable south of Iceland. There are more than ~~14~~1,000 analogues for this year. Their local range ~~may exceed 18 degree Celsius~~ at no point exceeds 4 degree Celsius for the narrow tolerance setup. Local positive deviations from the median may differ ~~the most southwest of Svalbard where, indeed, one proxy constrains our search~~ most strongly over Greenland and in Scandinavia. In the latter region, proxies should constrain our search for analogues. Local negative deviations may become largest ~~southeast of Greenland, over southern Finland, and in the Barents Sea. These are all regions where we do~~

~~not have direct proxy information but which are relatively close to proxy records in our collection~~over comparable domains. We do not show the equivalent Figure for the wide tolerance assumption but note that in this case the local range of results may exceed 20 degree Celsius and that largest positive excursions occur southwest of Svalbard, where a proxy constrains our search. The largest negative excursions are located at the eastern border of our domain in the Barents Sea, where our search is effectively unconstrained.

The pseudoproxy application of our implementation of an analogue search shows the viability of such approaches for reconstructing past climates from spatially sparse proxies with temporally sparse, irregular, and uncertain ages. The pseudoproxy tests also ~~highlights that the large uncertainties~~ show that the results depend on our assumptions on how tolerant we are with respect to our confidence in the proxy-input. For example, assuming wide tolerance ranges can lead to a wide range of potential analogues~~. The proxies~~ , while narrow tolerance ranges may lead to failures of capturing the target. The method may even fail to find valid analogues. Thus, the pseudoproxies are only weak constraints on the potential climate.

3.2 Application to real proxies

Already the pseudoproxy test highlights the potential ~~and~~ but also the associated problems in using the analogue method for the type of proxies we are interested in, together with a limited pool of candidate fields. The analogue reconstruction ~~captures~~ is able to capture the target data but the search ~~often provides~~ may provide either a very wide ~~uncertainty range~~ or a too narrow uncertainty range relative to the target. Wide ranges occur mostly due to the large number of valid analogues ~~. Conversely~~while narrow ranges signal that there are only few analogues fitting the proxy data under the made assumptions on the fidelity of the proxies. Finally, the method ~~occasionally may~~ may overall fail to provide valid analogues.

Our focus here is on a multi-archive and multi-proxy reconstruction using 17 proxies (compare section 2.3.1) for the European-North Atlantic sector for approximately the last 15kyr. Preliminary tests showed that using a 90% ~~intervals~~ tolerance level leads to ranges that are too narrow to find any suitable analogues (not shown). We only show the results for using 99% and 99.99% ~~envelopes~~levels in the estimation of our tolerance envelopes around proxy records. For the meaning of these levels see the descriptions for equation 1.

Figure ~~7~~ 8 shows the proxies and their ~~uncertainties~~ constructed tolerance envelopes for the locations in Figure 1b. The panels highlight that the real proxy values are less equally distributed through time, are generally smoother, and differ more in their lengths compared to the pseudoproxy setup. Figure ~~3~~ 4b already showed how the number of available proxies increases from 11 to 17 but then again decreases until only 5 proxies are available for the earliest dates. ~~The appendix compares~~ Below we compare the full 17-proxy setup to different sets of proxies. Table 1 and ~~Appendix Figures ?? to ??~~ Figure 3 to 4 give details for the different sets.

In the case of the main set of 17 proxies, our implementation tries to find analogues for 1,781 dates. There are between 1 ~~to~~ and 900 analogues for 141 dates for ~~a~~ 99% tolerance envelopes (see Figure ~~8~~9b). Analogues come from two different simulations. It is obvious that the method often fails ~~.~~ to provide a valid analogue.

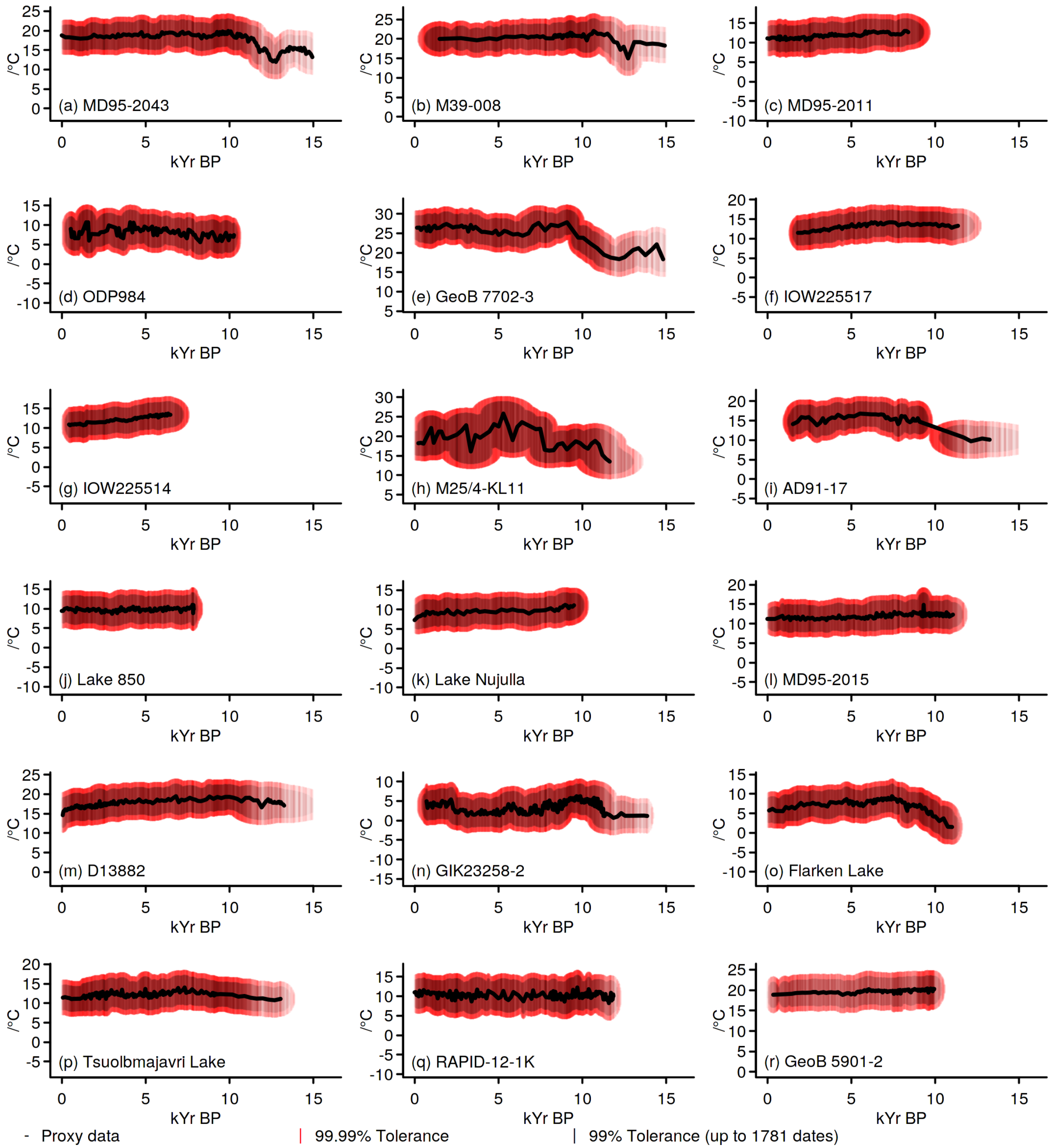

Figure 8. Proxy data and assumed uncertainties for ~~the 17 main~~ all proxy record locations in our analogue search under two different ~~uncertainty~~ tolerance envelopes.

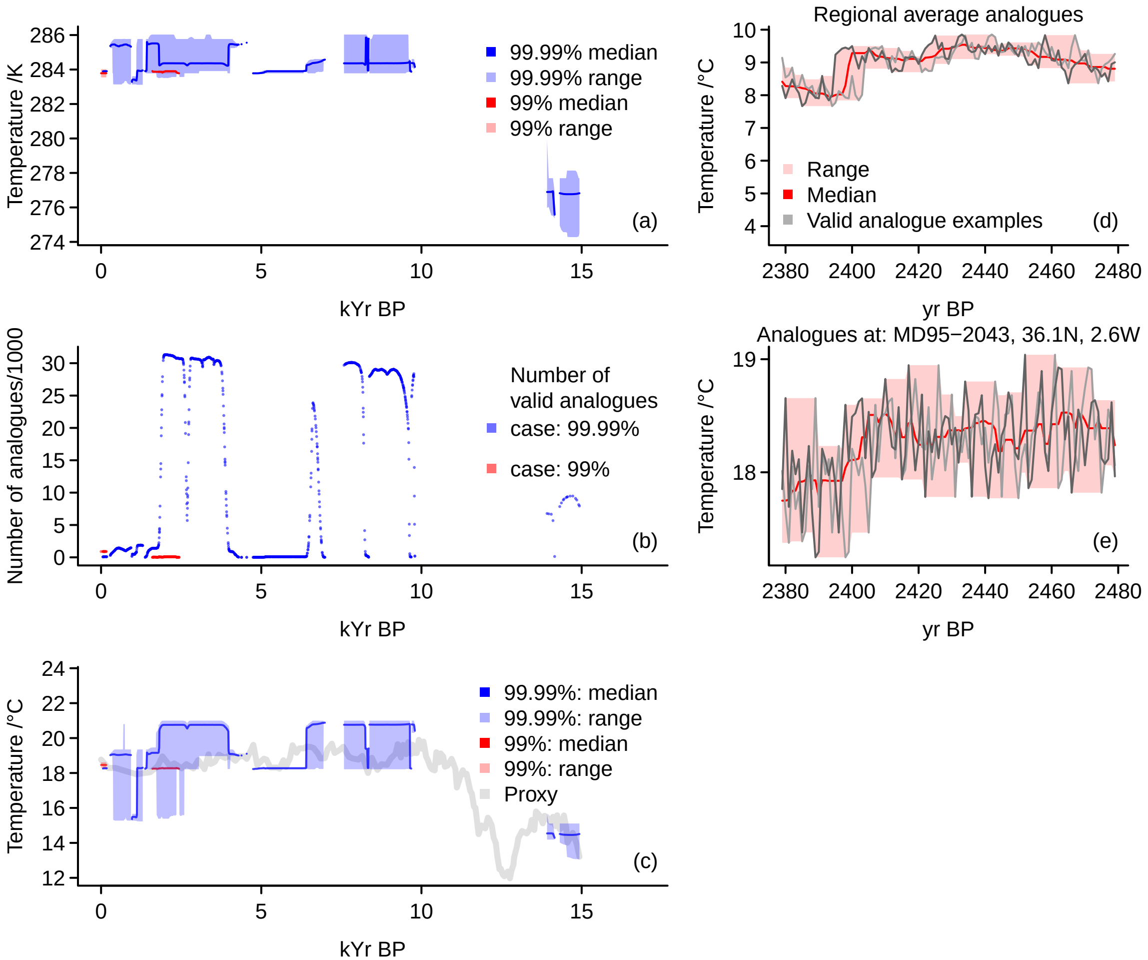

Figure 9. ~~Visualising reconstruction~~ Reconstruction results for the analogue method under two ~~assumed uncertainty envelopes~~different tolerance assumptions: ~~(a)~~ **(a)** regional averages for 101-year moving averages, ~~(b)~~ **(b)** number of analogues, ~~(c)~~ **(c)** local results for 101-year moving averages, ~~(d,e)~~ **(d,e)** expansion of regional ~~(d)~~ **(d)** and local (~~e~~36.1N,2.6W, MD95-2043) **(e)** 101-year moving averages in ~~101 year~~ 101-year series ranges. ~~(a)~~ **(a)** shows median and range of all analogues for an assumed 99% ~~uncertainty~~ tolerance envelope (red) and a 99.99% envelope (blue) for the proxy data. ~~(b)~~ **(b)** shows the number of analogues found for each of the dates considered; red: 99% envelope, blue: 99.99% envelope. ~~(c)~~ **(c)** adds for the location of the record MD95-2043 the proxy data (grey), and the range and median of all local analogue values for a 99% envelope (red) and a 99.99% envelope (blue). ~~(d,e)~~ **(d,e)** show the median (red), the range (light red), and two valid analogue examples of the extension of 101-year mean analogues into 101 year long time-series for ~~(d)~~ **(d)** the regional averages and ~~(e)~~ **(e)** the location of MD95-2043. ~~(d)~~ **(d)** and ~~(e)~~ **(e)** only show results for the 99% ~~uncertainty~~ tolerance envelope.

For the 99.99% envelope, these basic results change. The method identifies 1 to 31,304 analogues at 1,288 dates (see Figure 89b). These come from 42 different simulations. There are no valid analogues between ~10kyr BP and ~14kyr BP. Otherwise, there are extended periods with very many analogues and other periods with few analogues.

For the narrower uncertainty envelopetolerance assumption, the method finds valid analogues only for the recent past millennia (Figure 89). Even then, it is only successful for few periods (Figure 89b). In this case, the range of the area average reconstruction (Figure 89a) and at the local proxy location (Figure 89c) is very narrow. There is very little regional or local temporal variability in the analogues. However, the reconstruction may reflect well the average state of the local proxy series (Figure 89c). As for the pseudoproxy test, we can expand the analogues, i.e., the 101-year moving means, to reveal the potentially show the underlying time-variations (Figure 89d and e). These also reflect the again provide an impression of interannual variability that is compliant with our proxy constraints on the centennial average. These panels emphasize the very narrow range of potential analogues for the regional average but also for the local series.

For the wider uncertainty tolerance envelope, the method identifies valid analogues for more dates (Figure 89) and, generally, there are more valid analogues for these dates (Figure 89b). However, there are more proxies available for some dates (compare Figure 34) and this increases the number of constraints on the analogue candidates for these dates. Thus, there are dates when the range of the regional average reconstruction for a 99.99% uncertainty tolerance envelope does not necessarily include the 99% envelope data.

The range of the reconstruction may be wide regionally or locally wide for the 99.99% envelope, but this does not ensure that it locally includes the proxy values (Figure 89c). There is little temporal variability in the reconstructed data. This is mainly because of the large number of analogues and the relatively low temporal variation in the set of valid analogues (Figure 89b). Further, the reconstruction is rather constant.

Figure 9 10 plots examples of a field and of the local minima, median, and maximum maxima of potential analogues for the two envelopes. It uses different dates for the different examples (2429 BP and 14105 BP)different tolerance envelopes. The upper row uses the 99% envelope reconstruction for the year 2,429 BP and the lower row uses the 99.9% envelope reconstruction for the year 14,105 BP. For both dates, all valid analogues are from only one simulation each. The examples in panels (a) and (e) of Fig. 9 also include 10 also include as dots the proxy values available for the respective datesas dots. These highlight that, for the late Holocene date, the found analogues capture the proxies rather well though with exceptions over Scandinavia. However, the date at ~14kyr BP strongly disagrees for the only valid with the one proxy at high northern latitudes. The range of analogues is very narrow for the late Holocene example from the narrow uncertainty applicationtolerance case. Differences become largest over Greenland and along the sea-ice edge. For the deglacial example from the wider uncertainty applicationand the wider tolerance case, differences become largest east and west of Iceland.

3.2.1 Results for different proxy setups

Table 1 introduces a number of additional proxy setups . The sets (E02 to E09). These use different sub-selections of proxies from our initial selection. Appendix ?? Further, most of them test sparser sets of locations around central Europe (compare

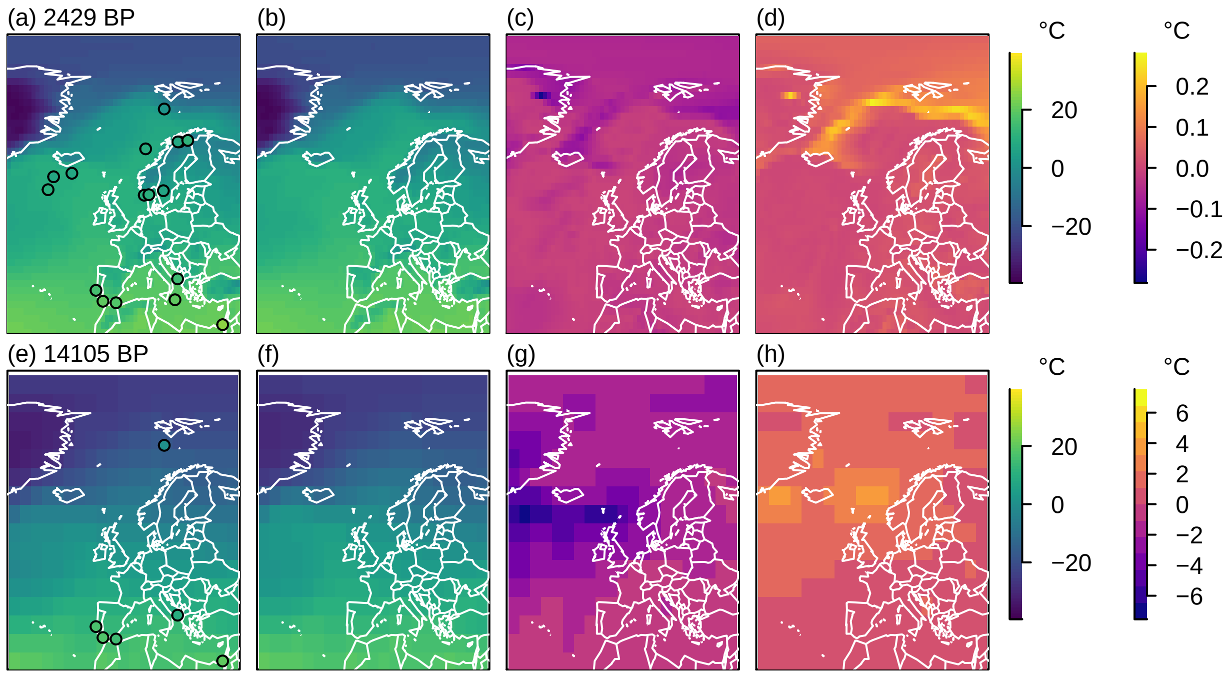

Figure 10. ~~Visualising field~~ Field information for the analogue search: Two examples of 101-year mean annual temperature analogue ~~reconstruction~~ reconstructions for the European-North-Atlantic sector in °C. ~~(a,e)~~ **(a,e)** One example analogue, ~~(b,f)~~ **(b,f)** local median of all analogues, ~~(c,g)~~ **(c,g)** difference of local minimum to local median of all analogues, and ~~(d,h)~~ **(d,h)** difference of local maximum to local median of all analogues. ~~(a)~~ **(a)** to ~~(d)~~ **(d)** are for the 99% ~~uncertainty~~ tolerance envelope and the year 2,~~(e)~~ 429 BP, **(e)** to ~~(h)~~ **(h)** for the 99.99% ~~uncertainty~~ tolerance envelope and the year 14,105 BP.

Table 1). Figure 3 provides additional information ~~and presents details of the results in its subsection 3.2.1~~about which records are included in the different setups and their proxy types. Here, we shortly present the results.

Experiment E01 is our main setup. It was described in the previous section. It uses the 17 chosen proxy locations, which we also use for the pseudoproxy setup. Setups E02 and E03 are based only on alkenone $U_{37}^{K'}$ records and E03 replaces M39-008 by GeoB 5901-2, as both are co-located. E04 to E09 include different numbers of other proxy types instead of $U_{37}^{K'}$. Figure 4 shows the availability of proxies for the different setups. Figure 8 presents the proxy data and assumed uncertainties including record GeoB 5901-2. For more information see Table 1 and Figure 3.

Figure 11 shows the reconstruction results for the proxy setups E01 to E09. All panels plot the reconstructions using the 99% and the 99.99% tolerance envelopes. Panel (a) adds for our main setup a reconstruction where we consider interannual data for the simulations and include the QUEST FAMOUS-HadCM3 simulations. Please, see section 2.3.3 for details on these simulations. The panel further includes the results of testing an analogue approach where only the single best analogue is considered at each date.

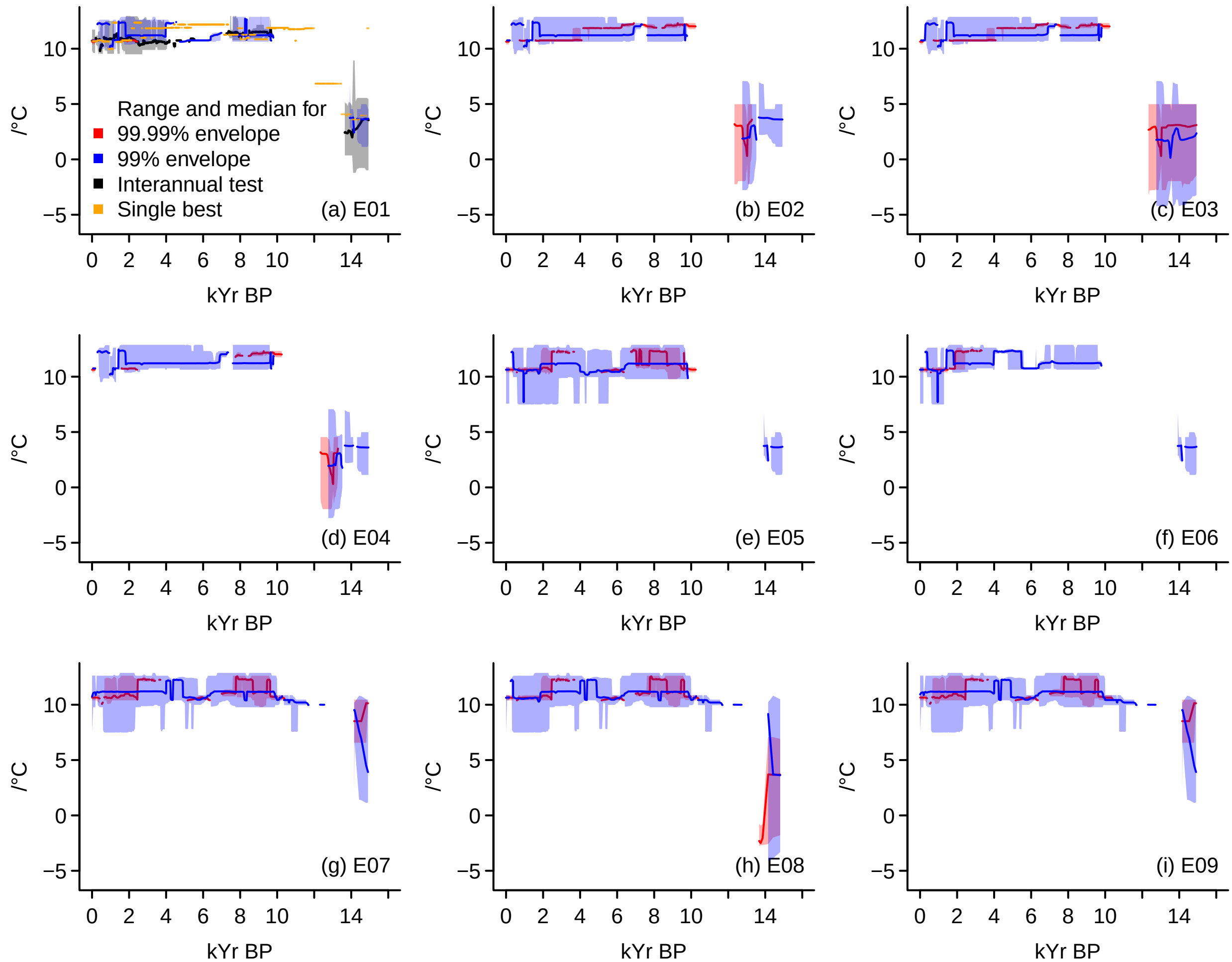

Figure 11. Visualising the reconstructions for the various proxy setups: **(a)** E01, **(b)** E02, **(c)** E03, **(d)** E04, **(e)** E05, **(f)** E06, **(g)** E07, **(h)** E08, **(i)** E09. All panels include the median, 90% interval, and full range for the reconstructions under a 99% tolerance envelope (red) and a 99.99% envelope (blue). Panel **(a)** additionally includes a setup in black where we do not consider 101-year moving averages of simulation data but all simulation output as provided including the FAMOUS-HadCM3 simulations for QUEST. Orange in panel **(a)** is a test considering only the single best analogues for each date.

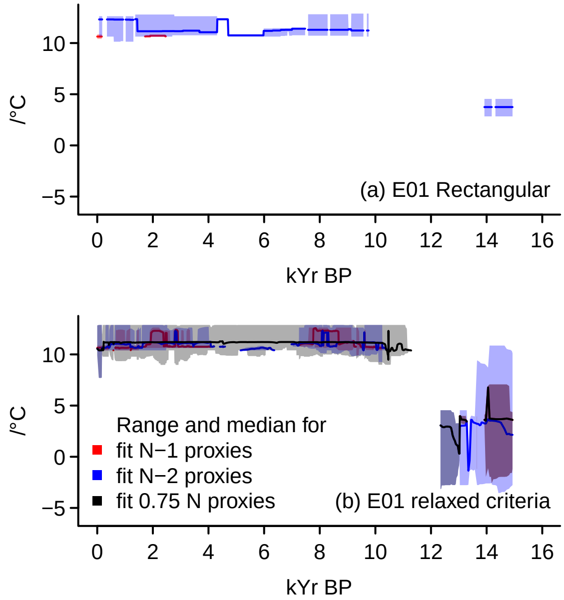

Figure 12. Visualising alternative reconstructions: **(a)** E01 with rectangular tolerance range, **(b)** E01 with relaxed criteria for analogue selection. Panel **(a)** includes the median, 90% interval, and full range for the reconstructions under a 99% tolerance envelope (red) and a 99.99% envelope (blue). Panel **(b)** shows range and median for setups where the analogue candidates are valid if they fail at one (red) or two (blue) locations or at 25% of all locations (black).

The ~~different sets confirm~~ panels of Figure 11 highlight that loosening the ~~uncertainty constraint~~ tolerance constraint and thereby widening the tolerance envelope leads to valid analogues for notably more dates as well as a wider range of valid analogues. We also obtain analogues at more dates if we keep the ~~uncertainty~~ tolerance envelope at the lower level but do not preprocess the simulation output to 101-year moving means (Figure 11a black lines). This inclusion of interannual data increases the number of analogues throughout the reconstruction period. This variation of the experiment also uses more simulation data by including the QUEST FAMOUS data, but this only affects the reconstruction success in the 15th millennium BP in this setup. We performed further tests with different averaging periods of 51 and 501 years, respectively, while keeping the narrow tolerance envelope (not shown). Increasing the averaging period to 501 years reduces the number of valid analogues and the number of dates with any valid analogues. Reducing the averaging period to 51 years allows to find a few valid analogues in the 15th and 16th millennium BP. In this setting, the approach also finds more valid analogues in recent millennia.

Generally, the reconstruction success appears to be better for proxy setups that only include $U_{37}^{K'}$ records (Figure 11b,c). Such consistent sets of proxies provide a more continuous reconstruction for both local ~~uncertainty levels~~tolerance assumptions. Nevertheless, we fail to obtain valid analogues, i.e. reconstructed values at the end of the deglaciation. While results are quite similar over much of the period between both reconstruction attempts E02 and E03, the second setup allows a wider and potentially colder range in the period before ~12kyr BP.

Further panels of Figure 11 add different setups. Panel (d) complements the $U_{37}^{K'}$ proxies by one foraminiferal assemblage record. Panels (e) and (f) also test different setups dominated by $U_{37}^{K'}$ but including other proxies. Panels (g) to (i) use different small setups of proxies around the European area.

Multi-archive setups with fewer proxies give generally wider ranges of possible analogues. Otherwise all setups tend to be in a comparable range regarding their median and their range considering the last 10 millennia. Differences between all setups are largest in the 14th millennium BP due to a larger range for some reconstructions.

Both multi-proxy setups in panels (e) and (f) fail to provide analogues before the deglaciation for the narrower tolerance assumption. The setups in panels (g) and (i) are notably warmer in the 14th millennium BP compared to results in panel (h) but also compared to other setups. This holds for both tolerance envelopes. A common difference is the inclusion of M39-008 while excluding the $U_{37}^{K'}$ record D13882 (compare Table 1 and Figure 3). The latter record is thought to represent summer temperatures off the west coast of Portugal while the former is meant to represent annual temperatures in the Gulf of Cadiz. We note that panels (e) and (f) also are warmer in the 14th millennium BP for the wider tolerance range. These also include M39-008 and exclude D13882. Please note, the supplement of Marcott et al. (2013) refers to D13882 as D13822.

Generally, we find that the reconstructions from different setups differ in their ability to reconstruct climate for certain periods. Indeed, different setups may provide notably different climates, particularly for the early part of the time period of interest. Certain proxies appear to shift the results for the earlier part of our reconstruction between a warmer and a colder deglacial estimate. It is beyond the scope of this paper to disentangle the reasons for this. All setups provide rather constant reconstruction ranges.

As noted, Figure 11a adds a single-best analogue reconstruction, where the reconstructed value for a given date is the analogue candidate with the smallest Euclidean distance to the proxy values for that date. During the past approximately four millennia as well as during the period from eight to 10 thousand years BP, the single-best estimate is included in the ranges of our other reconstruction efforts. Indeed it is close to the test with interannual data throughout the Common Era of the last 2,000 years. In the period between four and eight thousand years BP, the result from the single best analogue setup differs notably from the other efforts. However, the approach allows to obtain estimates when the other approaches fail between 10 to 14 thousand years BP. Comparably to our other reconstruction attempts, the single-best analogue reconstruction shows only little variability. Noteworthy are the reconstructed values in the 15th millennium BP where the single-best analogue represents a Holocene level warm climate and not a deglacial climate.

We consider two more modifications of our approach. Figure 12 shows, first, results using a rectangular tolerance region, and, secondly, reconstructions for tests where an analogue candidate is valid although it falls outside the tolerance region at one, two, or 25% of the locations. The rectangular setup has minimal influence on the reconstruction but gives more homogeneous ranges of valid analogues and succeeds at slightly more dates in finding valid analogues. Relaxing the tolerance criteria results in very wide ranges in the early part of the study period. Due to few available proxies in that period, the criterion to fit 75% of locations is stricter than the criterion that allows to fail at two locations. The resulting reconstructions still have little variability. They also either give a wide and nearly constant range of potential values or a very narrow range.

4 Discussions

Our implementation of an analogue search method for reconstructing surface temperature over multimillennial timescales relies on a number of decisions, which are uncommon compared to other paleo-reconstruction efforts on multimillennial timescales. Central to our assumptions is that taking account of ~~uncertainty~~ the uncertainty in our underlying data is indispensable in analogue approaches for paleoclimatology and~~particularly~~ , particularly, if one uses spatially and temporally sparse as well as data and age uncertain proxies. There is one prime motivation behind our specific handling of uncertainty in terms of tolerance ranges and our selection of reconstruction dates: The analogue search for a chosen date should use as much information about this date as possible, including the uncertainty of other data points whose age ~~uncertainty~~ uncertainties include the currently given date of interest.

This leads to the use of ~~uncertainty~~ tolerance ellipses. Assumptions here are that, firstly, data and date are inseparable; secondly, that this assumption also holds for the tuple and its two-dimensional uncertainty; and, thirdly, that a reconstruction exercise has to consider both parts of the uncertainty to sufficiently estimate the range of reconstructed values. ~~Our~~ Admittedly, our procedure is a simplified approach to incorporating these assumptions. More correctly, one would calculate the multivariate ~~mixture distributionand identify the relevant uncertainty range from it~~joint distribution, and use a measure of likelihood to select the analogues. As a ~~sidenote, the full multi-predictor~~ side note, the highly dimensional space for all ~~dates also is a multivariate mixture~~ proxies also follows a multivariate distribution, which one could then employ in more sophisticated data science approaches.

~~Our handling of uncertainty causes that we cannot easily implement a distance measure like the Euclidean. A more formal definition of similarity should take into account the multivariate and correlated nature of uncertainty: in time and across-proxies.~~

~~We are confident~~ We trust that considering both parts of the uncertainty enables better and more reliable reconstruction estimates. We concede that this procedure may exaggerate the range of potential climates and thereby may reduce the precision of the reconstruction (compare also Annan and Hargreaves, 2012). We postulate that this, however, is only partly due to the assumptions on uncertainty, which may transfer uncertainty to too many records. We think it is also because the simulation pool ~~does not reflect the climate relations among proxy record locations as given by the proxy records~~is not fully consistent with all the proxies simultaneously. It is beyond the scope of the present study to investigate whether this, in turn, is because of unreliable simulations, lacking overlap between reconstructed and simulated climates, or lacking reliability of the proxy records, that is their errors.

With respect to the lacking precision of the reconstructions, Annan and Hargreaves (2012) already identified a similar issue in their particle filter data assimilation approach. Annan and Hargreaves note that in a setup where one has only few and highly uncertain proxy predictors the reconstruction tends to lack accuracy. We think that for the analogue method one could remedy this by weighing the valid analogues by a distance measure relative to the pattern of proxy predictors or by their agreement with each individual predictor. We note that the analogue method in the present setting may represent the recent climate worse than simply taking the average over the period of instrumental observations.

Our handling of uncertainty in terms of tolerance results in difficulties in implementing a distance measure like the Euclidean. A more formal definition of similarity should take into account the multivariate and correlated nature of uncertainty: in time and across proxies.

Related to our handling of uncertainty is our approach of reconstructing data for those years when at least one proxy predictor is dated. This also may contribute to the wide range of the reconstructions by neglecting information in between these dates. Alternatively, one could pool the proxy dates into constant intervals of, for example, 100 years. ~~Such an approach makes as strong assumptions as~~ The underlying assumptions here are as strong as in our procedure. We note that Jensen et al. (2018) use the published age models to interpolate their proxy records to consistent time steps. They compare their proxies to 10-year averages of the simulation pool. Incorporating, presumably Bayesian, age models maximises the available prior information used from, e.g., the original measurements. Nevertheless, we decide against interpolation procedures, even based on Bayesian age models, assuming that this may result in ~~overconfidence in the subsequent reconstruction~~overconfident reconstructions. For example, interpolation could suggest more certainty in reconstructed values where and when we have little ~~to~~ or no proxy information (see, e.g. Figure ~~7~~8i between approximately 9kyr and 11kyr BP).

Additional assumptions relate to characteristics of the considered proxy predictors. This includes our decision to generally compare the proxy predictors to centennial averages of the simulation output. Thereby, we do not allow for the fact that the proxy sensor might record extreme-like events. Similarly, we also do not consider the differing resolutions for each date and each location. Further, we compare the proxy predictors and the simulation pool in terms of temperatures instead of using surrogate proxies in proxy units from the simulation pool. Finally, the use of temperature for the surface and for an attributed and calibrated season does not account for the sensor specific habitats and seasonal sensitivities or their changes (compare Jonkers and Kučera, 2017; Kretschmer et al., 2018). Our comparison, thus, is based on the assumption that the proxy inferred climate property and the proxy record ~~reasonably well relate~~ relate reasonably well to the parameter of interest (annual surface temperature) and that, in turn, comparisons to the equivalent simulated output are valid. ~~Therein~~ In doing that we rely on the previously published information about the considered proxy record. Similarly, our expansion of the temporal average reconstructions into 101-year time-series relies on the quality of the proxy data and on appropriate assumptions on the temporal representativeness of the data. The possibility for such a temporal downscaling is a unique feature of analogue search reconstructions from temporal averages and of comparable data assimilation techniques.

Possible improvements of the method would respect more explicitly the irregular resolution of the proxy records and the different resolutions between the records. Similarly, applications benefit if we can discriminate whether a proxy sensor records mean climatic conditions or extreme-like events. Including the proxy specific habitat and growth season also leads to a more appropriate comparison as does employing proxy forward models to make the comparison in proxy units.

~~Published proxy records do not necessarily provide all the information to assess, e.g., the resolution. The available published information is also generally not sufficient to identify whether a record represents events or mean states. Regarding uncertainty, producers of proxy records do not always report calibration uncertainties. Even if these are known, assumed uncertainties may not capture all potential error sources.~~

Better understanding of the proxy systems and availability of the full simulation output data would allow for proxy series specific analogue searches analogue searches that are more specific for each proxy series. It further would enable the use of locally calibrated process-based forward integrations by proxy system models. The advent of proxy system forward models in principle allows producing proxy parameter representations in the virtual environment of the simulations (Schmidt, 1999; Tolwinski-Ward et al., 2011; Thompson et al., 2011; Evans et al., 2013; Dee et al., 2015, 2016, 2017, 2018; Jones and De (Schmidt, 1999; Tolwinski-Ward et al., 2011; Thompson et al., 2011; Evans et al., 2013; Dee et al., 2015, 2016, 2017, 2018; Jones and De but there are still gaps in the understanding on how the sensor recording of the biological, physical, chemical, or geological process reacts to the environment. Additionally, records may lack necessary information. While such applications are quickly developing (see Dee et al., 2016; Jones and Dee, 2018; Dolman and Laepple, 2018; Konecky et al., 2019), data assimilation of this kind of information is still not operational even for the Common Era with its potentially high resolution and potentially high quality proxies (Hakim et al., 2016; Tardif et al., 2019; Emile-Geay et al., 2017).

It is generally advisable to use consistent proxy parameters, a consistent recalibration, and a consistent calibration target. This should increase the probability of the proxy predictors constraining the pool of potential analogues (compare the results in Appendix section 3.2.1). Often such consistency is an implicit or explicit assumption (compare, e.g., Reschke et al., 2018). On the other hand, the analogue approach, in theory, should allow using different parameters and calibrations if the comparison is to the same target. Indeed, ideally, it should also compensate even a comparison of different parameters. This, however, depends on how much proxy records indeed constrain the ultimate target property for the reconstruction.

Our reconstruction is only for the approximate domain of the proxy predictors. However, it may be possible that a set of proxy predictors from, for example, Europe also provides information on larger scale climate variables. Further, we deal only with temperature reconstructions. However, climate is more than simply temperature. Indeed, if there is evidence that the proxy predictors are relevant constraints on other climate fields beyond, in this example, temperature, the pool of analogues can provide information on other climate variables.

Regarding the However, reconstructing other variables for hydrology or climate dynamics depends on a sufficient number of proxy records that reliably represent these. That is, there are two conditions on the proxy records, they have to represent the variable and there have to be enough of them. In addition, we have to be confident that the simulation pool reliably represents the climate variable and its spatial distribution. Considering the number of available reliable proxies for, e.g., precipitation and the quality of simulations' representation of it, we would expect that reconstruction success using the analogue method may be worse for these other variables than for temperature (compare also Gómez-Navarro et al., 2015).

Regarding the temporal resolution, a test of our method suggests that, for a given uncertainty intervalassumed tolerance level, the analogue search is more successful in finding valid analogues if we consider higher resolution data and less successful if we reduce the resolution of the data. That is, the method performs slightly better using 51-year averaged simulation data than using 101-year averaged data, and it performs even better using interannual data. While such an interannual analogue search may misinterpret what the proxy data represents, it may be a more truthful comparison considering the potential level of environmental noise in the proxy data relative to the targeted temperature signal.

Similarly, we find more valid analogues if we use less stringent criteria in our search for valid analogues. A single-best analogue reconstruction also gives a more continuous reconstruction.

However, all approaches have in common that reconstruction medians as well as reconstruction ranges are relatively constant over time. The reconstructions show little variability and are lacking clear differences in climate between the late and the early Holocene.

A likely reason for the little variability in central estimates and the generally rather constant character of our reconstructions could be that the space of valid analogues is too unconstrained and the method labels too many candidates as valid analogues. However, also the single-best approach shows such a behaviour. That is, while the reconstruction is undoubtedly only weakly constrained, even the best analogues differ little between subsequent dates. Part of this may be due to our choice to consider a rather large temporal range of influence of individual dated records. Our ellipses of tolerance may result in a strong influence of an unlikely value at a specific date. This could potentially be solved by considering explicitly the likelihood of a value at a date instead of simply taking a binary criterion. A less complex solution could be obtained by pooling proxy values in temporal windows, weighting them within these windows, and then performing a reconstruction considering certain ranges of tolerance.

Our aim here is to use the local proxy uncertainty to select analogues. There is a trade-off between considering the uncertainty of the proxies and constraining the number of analogues. That is, if we want to consider the uncertainty in the way we do, then we allow for weakly constrained analogue ranges. If we allow different levels of proxy uncertainty, we can choose only the best M analogues. We, in turn, can limit the number of analogues or weigh them by certain criteria, e.g., based on their distance to individual proxies or their overall Euclidean distance.

Beyond these methodological aspects the size and character of the pool of analogue candidates influences the quality of the results. ~~As apparent from this paper,~~ Indeed, the lacking sensitivity to differences in climate and the lacking variability in our results may be a sign of an insufficient pool size or an insufficient overlap between simulated climate and the environmental conditions described by the proxy records.

Our results suggest that a pool including mid-Holocene, Last Glacial Maximum, and transient deglacial simulations does not ensure finding valid analogues for the time-period of the deglaciation and the Holocene. An insufficient large pool of candidate analogues requires ~~wider~~ more tolerant assumptions on uncertainty to obtain valid analogues. Thereby, the analogues remain unconstrained. A small pool also allows for non-uniqueness of analogues. Additionally climatological inconsistencies become more likely if the range of simulated periods in the model pool is wide.

We do not use anomalies. If there was a large ensemble of simulations over ~~this period~~ our period of interest, the use of anomalies would be advisable. Similarly, if all proxy records had common modern age data, there might be a valid anomaly building process. However, we include simulations for time-slices with notable different climatologies, and proxy records begin at various modern dates. One solution could be a sliding climatology for the proxies, which is added again for the final reconstruction. We note that using anomalies also might result in climatic inconsistencies. Furthermore, if we want to apply proxy forward models based on the calibration between measured property and temperature we do not use anomalies either because calibration relations frequently need temperature on either the Celsius or Kelvin scales.

This section outlined a number of potential improvements of the approach. Some of these would increase the number of necessary computations. While the increase in costs is not prohibitive, we decided against including such procedures here. However, it appears particularly worthwhile to try to implement a workflow that combines feasible data science methods, some version of simple data assimilation, and a proxy system model framework like PRYSM (Dee et al., 2015, 2016, 2018; Jones and Dee, 2018) in future attempts of spatiotemporally resolved reconstructions if the interest is in a dynamical understanding of the climate variability over multimillennial timescales.

5 Summary and concluding remarks

The analogue method is a computationally cheap data assimilation approach. Here, we discuss a specific application for time uncertain, sparse, and irregularly sampled proxies. We focus on the North Atlantic sector and the time period from ~~about ~~~~~approximately 15kyr BP to the late 20th century.

The approach succeeds in providing ~~a reconstruction~~ reconstructions in a pseudoproxy setup. ~~It is less successful~~ Already this setup highlights two potential problems. The method may either fail to find valid analogues or provide a wide range of potential analogues, which do not necessarily include a target climate. These problems relate to assumptions on the uncertainty in the proxy input data.

The approach performs comparable for realistic proxy setups. ~~Then, reconstructions fail particularly~~ However, then, the analogue search often fails to find valid analogues as none of our candidate fields comply with our criteria for a valid analogue. That is the method fails to provide a climate reconstruction because of a lack of valid analogues. In the present case, this particularly occurs over the late deglaciation and early Holocene.

~~Reconstructions~~ Furthermore, our reconstruction by analogue are generally rather imprecise for the used time uncertain, sparse, and irregularly sampled proxies and a limited pool of simulation data. The range of potential analogue values ~~is~~ can become very wide for a given date. Regional average reconstruction medians show little variation over time.

The analogue method is non-linear and considers the spatial covariances between the proxy records. ~~Thereby, resulting fields provide~~ While it lacks precision in our setup, it nevertheless provides us with spatial field estimates of past climate states that are consistent with the regional inter-relations as presented by the proxy predictors.

Data availability. We provide lists of valid analogues per date and experiment at https://osf.io/pj9eg. This allows identifying valid climate states for dates. We also provide files for area mean analogue ranges and medians.

The proxy data we use is available from the supplement of Marcott et al. (2013) at https://doi.org/10.1126/science.1228026 (see also https://science.sciencemag.org/content/suppl/2013/03/07/339.6124.1198.DC1, both links last accessed December 30, 2019). Primary data citations are Cacho et al. (2006), Grimalt and Calvo (2006), Came et al. (2007b), Castañeda et al. (2010b), Emeis et al. (2003b), Emeis et al. (2000a), Giunta and Emeis (2006), Larocque and Hall (2006), Grimalt and Marchal (2006), Rodrigues et al. (2010), Sarnthein et al. (2003b), Sundqvist et al. (2014a, see also Digerfeldt, 2010, Digerfeldt, 2009, and Sundqvist et al., 2014b), Sundqvist et al. (2014a, see also Voeltzel, 2010a, Voeltzel, 2010b, and Sundqvist et al., 2014b), Thornalley et al. (2009b), and Kim et al. (2004b).

Simulation data is available from a number of sources. Data from simulations for PMIP3 can be obtained from the Earth System Grid Federation, e.g., at the node https://esgf-data.dkrz.de/projects/esgf-dkrz/ (last accessed December 30, 2019). Last Millennium ensemble data and TraCE-21ka output are available at https://www.earthsystemgrid.org/ (last accessed December 30, 2019). Millennium COSMOS simulation data is best accessed via https://cera-www.dkrz.de/WDCC/ui/cerasearch/ (last accessed December 30, 2019). Quaternary QUEST data may be obtained via https://catalogue.ceda.ac.uk/uuid/a43dcfaccfae4824ab9ab2b572703e72 (last accessed December 30, 2019).

Map of the reconstruction domain and the proxy predictors: (a) for the pseudoproxy setup; (b) to (j) for the different proxy setups. For more details see Table 1 and Figure ??.

Figure ?? shows the proxy distribution for the various proxy setups within the reconstruction domain. Figure ?? gives more information about which records are included in the different setups and their proxy types. Panel (a) of Figure ?? repeats the pseudoproxy setup from Figure 1, which has a higher density than any of the real proxy setups in panels (b) to (j). E01 uses all available locations. Setups E02 chosen proxies and E03 in panels (c) to (d) are pure alkenone $U_{37}^{K'}$ setups and E03 replaces M39-008 by GeoB 5901-2 as both are co-located. Most of these setups test sparser sets of locations around central Europe. E04 to E09 include different numbers of other proxy types instead of $U_{37}^{K'}$.

Information about the different proxy setups: Matrix of proxy records against proxy setup (E01-E09). For more information see Table 1. White-out means that the relevant proxy is not included in the respective proxy setup.

Figure ?? complements Figure 3 for the various setups. Panels show the number of available proxies and the number of dates for which analogues are searched.

Information about the number of available proxies for the dates to be reconstructed: (a) the pseudoproxy setup, (b) to (j) the various proxy setups according to Figure ?? and Table 1. In (b) to (j) we show results for two different assumptions on the uncertainties, a 99% envelope and a 99.99% envelope (compare section 2.2.3).

Proxy data and assumed uncertainties for GeoB 5901-2.

Figure ?? adds the proxy data and assumed uncertainties for the record GeoB 5901-2 and thereby supplements Figure 7. For more information see Table 1.

Visualising the reconstructions for the various proxy setups: (a) E01, (b) E02, (c) E03, (d) E04, (e) E05, (f) E06, (g) E07, (h) E08, (i) E09. All panels include the median, 90% interval, and full range for the reconstructions under a 99% uncertainty envelope (red) and a 99.99% envelope (blue). Panel (a) additionally includes a setup where we do not consider 101-year moving averages of simulation data but all simulation output as provided including the FAMOUS-HadCM3 simulations for QUEST. the simulation pool]Information Additional information on the additional proxy setups

Map of the reconstruction domain and the proxy predictors: (a) for the pseudoproxy setup; (b) to (j) for the different proxy setups. For more details see Table 1 and Figure ??.

Figure ?? shows the proxy distribution for the various proxy setups within the reconstruction domain. Figure ?? gives more information about which records are included in the different setups and their proxy types. Panel (a) of Figure ?? repeats the pseudoproxy setup from Figure 1, which has a higher density than any of the real proxy setups in panels (b) to (j). E01 uses all available locations. Setups E02 chosen proxies and E03 in panels (c) to (d) are pure alkenone $U_{37}^{K'}$ setups and E03 replaces M39-008 by GeoB 5901-2 as both are co-located. Most of these setups test sparser sets of locations around central Europe. E04 to E09 include different numbers of other proxy types instead of $U_{37}^{K'}$.

Information about the different proxy setups: Matrix of proxy records against proxy setup (E01-E09). For more information see Table 1. White-out means that the relevant proxy is not included in the respective proxy setup.

Figure ?? complements Figure 3 for the various setups. Panels show the number of available proxies and the number of dates for which analogues are searched.

Information about the number of available proxies for the dates to be reconstructed: (a) the pseudoproxy setup, (b) to (j) the various proxy setups according to Figure ?? and Table 1. In (b) to (j) we show results for two different assumptions on the uncertainties, a 99% envelope and a 99.99% envelope (compare section 2.2.3).

Proxy data and assumed uncertainties for GeoB 5901-2.

Figure ?? adds the proxy data and assumed uncertainties for the record GeoB 5901-2 and thereby supplements Figure 7. For more information see Table 1.

Visualising the reconstructions for the various proxy setups: (a) E01, (b) E02, (c) E03, (d) E04, (e) E05, (f) E06, (g) E07, (h) E08, (i) E09. All panels include the median, 90% interval, and full range for the reconstructions under a 99% uncertainty envelope (red) and a 99.99% envelope (blue). Panel (a) additionally includes a setup where we do not consider 101-year moving averages of simulation data but all simulation output as provided including the FAMOUS-HadCM3 simulations for QUEST. the simulation pool

0.1 Results References for the different chosen proxy setsrecords

Table 1 introduces already a number of additional proxy setups. The sets use different sub-selections of proxies from our initial selection. Figures ?? to ?? provide additional information on the various setups.

Figure ?? shows the reconstruction results for the proxy setups E01 to E09. All panels plot the reconstructions using the 99% and the 99.99% uncertainty envelopes. Panel (a) adds for our main setup also a reconstruction where we consider interannual data for the simulations and include the FAMOUS-HadCM3 simulations. Please see section 2.3.3 for details on these simulations.

Figure ??a shows three versions of reconstructions with the full set of proxies. If one widens the uncertainty envelope the method provides analogues at notably more dates as shown by the blue lines and shading compared to the red lines and shading in Figures ??a and 8a. Similarly, we are able to obtain analogues at more dates, if we keep the uncertainty envelope at the lower level but relax the condition on 101-year means and thereby allow for inclusion of the QUEST-FAMOUS simulations and comparison to interannual fields (see black colored results in Figure ??a). Including the QUEST-FAMOUS simulations increases the number of analogues in the 15th millennium BP, whereas including interannual data increases the number throughout (not shown).

It appears that the internal coherence among the considered proxies matters as, generally, the reconstruction success appears to be better for proxy setups that only include $U_{37}^{K'}$ records, i.e. proxy setups E02 and E03 (see Figure ??). We show these in panels (b) and (c). The setup in panel (c) differs from the one in panel (b) only in so far as we replace the record M39-008 with GeoB 5901-2. Both proxies are co-located.

These consistent sets of proxies provide a more continuous reconstruction for both local uncertainty levels. Nevertheless, we fail to obtain reconstructions at the end of the deglaciation. While results are quite similar over much of the period between both, the second setup allows a wider and potentially colder range for the period before ~12kyr BP.

Further panels of Figure ?? add different setups. Panel (d) complements the $U_{37}^{K'}$ proxies by one foraminiferal assemblage record. Panels (e) and (f) also test different setups dominated by $U_{37}^{K'}$ but including other proxies. Panels (g) to (i) use different small setups of proxies around the European area.

Both setups in panels (e) and (f) fail to provide analogues before the deglaciation for the narrower uncertainty envelope. The setups in panels (g) and (i) are notably warmer in the 14th millennium BP compared to results in panel (h) but also compared to other setups. This holds for both uncertainty envelopes. A common difference is the inclusion of M39-008 while excluding the $U_{37}^{K'}$ record D13882 (compare Table 1 and Figure ??). The latter record is thought to represent summer temperatures off the west coast of Portugal while the former is meant to represent annual temperatures in the Gulf of Cadiz. We note that panels (e) and (f) also are warmer in the 14th millennium BP for the wider uncertainty envelope. These also include M39-008 and exclude D13882. Please note, the supplement of Marcott et al. (2013) refers to D13882 as D138221 provides references to the original publications for the individual proxy records.

The multi-archive setups with fewer proxies give generally wider ranges of possible analogues. Otherwise, all setups tend to be in a comparable range regarding their median and their range considering the last 10 millennia. Differences between all setups are largest in the 14th millennium BP due to a larger range for some reconstructions. However, the few reconstructions with valid analogues in the 12th and 13th millennium BP give similarly wide ranges then.

Generally, reconstruction ranges are rather constant. We, further , find that the reconstructions from different setups differ in their ability to reconstruct climate for certain periods. Indeed, different setups may provide notably different climates particularly for the early part of the time period of interest. Certain proxies appear to shift the results for the earlier part of our reconstruction between a warmer and a colder deglacial estimate. It is beyond the scope of this paper to disentangle the reasons for thistable further adds references to the datasets directly and thereby the repositories where the records are available.

1 Additional information on the chosen proxies and the simulation pool

0.1 References for the chosen proxy records

Table 1 provides references to the original publications for the individual proxy records

0.1 Additional information on the simulation pool

Table 2 provides references for the various models from which we include simulations in the candidate pool. The table further adds references gives links to the repositories where the records are availableinterested researchers can obtain the simulation data.

0.2 Additional information on the simulation pool

Table 2 provides references for the various models from which we include simulations in the candidate pool. The table further gives links to the repositories where interested researchers can obtain the simulation data.

Table 1. Additional information for the used proxy records: Proxy ID, main reference, and reference for the data sets. For additional information see Table 1.

Proxy ID	Original Publication	Data References
MD95-2043	Cacho et al. (2001)	Cacho et al. (2006)
M39-008	Cacho et al. (2001)	Cacho et al. (2006)
MD95-2011	Calvo et al. (2002)	Grimalt and Calvo (2006)
ODP 984	Came et al. (2007a)	Came et al. (2007b)
GeoB 7702-3	Castañeda et al. (2010a)	Castañeda et al. (2010b)
IOW225517	Emeis et al. (2003a)	Emeis et al. (2003b)
IOW225514	Emeis et al. (2003a)	Emeis et al. (2003b)
M25/4-KL11	Emeis et al. (2000b)	Emeis et al. (2000a)
AD91-17	Giunta et al. (2001)	Giunta and Emeis (2006)
Lake 850	Larocque and Hall (2004)	Larocque and Hall (2006)
Lake Nujulla	Larocque and Hall (2004)	Larocque and Hall (2006)
MD95-2015	Marchal et al. (2002)	Grimalt and Marchal (2006)
D13882	Rodrigues et al. (2009)	Rodrigues et al. (2010)
GIK23258-2	Sarnthein et al. (2003a)	Sarnthein et al. (2003b)
Flarken Lake	Seppä et al. (2005)	Sundqvist et al. (2014a)
Tsuolbmajavri Lake	Seppä and Birks (2001)	Sundqvist et al. (2014a)
RAPID-12-1K	Thornalley et al. (2009a)	Thornalley et al. (2009b)
GeoB 5901-2	Kim et al. (2004a)	Kim et al. (2004b)

Tables 3 to 5 complement tables 2 and 2. They give the central simulation IDs. This allows finding the simulations more easily in the repositories.

Table 2. Additional information about the pool of simulation data: Model name, main reference, and link to the provider of the data. For additional information see Table 2.

Model	References	Link
CNRM-CM5	Voldoire et al. (2013)	https://esgf-data.dkrz.de/
COSMOS-ASO	Budich et al. (2010)	https://esgf-data.dkrz.de/
CSIRO-Mk3L-1-2	Phipps et al. (2011)	https://esgf-data.dkrz.de/
GISS-E2-R	Schmidt et al. (2014b)	https://esgf-data.dkrz.de/
HadCM3	Collins et al. (2001)	https://esgf-data.dkrz.de/
HadGEM2-CC	Jones et al. (2011)	https://esgf-data.dkrz.de/
HadGEM2-ES	Jones et al. (2011)	https://esgf-data.dkrz.de/
IPSL-CM5A-LR	Dufresne et al. (2013)	https://esgf-data.dkrz.de/
MIROC-ESM	Sueyoshi et al. (2013)	https://esgf-data.dkrz.de/
MPI-ESM-P	Giorgetta et al. (2013)	https://esgf-data.dkrz.de/
CESM1	Otto-Bliesner et al. (2015)	https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CESM_CAM5_LME.html
CCSM3	Liu et al. (2009)	https://www.earthsystemgrid.org/project/trace.html
MPI-ESM-Cosmos	Jungclaus et al. (2010)	https://cera-www.dkrz.de/WDCC/ui/cerasearch/project?acronym=MILLENNIUM_COSMOS
FAMOUS-HadCM3	Smith and Gregory (2012)	https://catalogue.ceda.ac.uk/uuid/a43dcfaccfae4824ab9ab2b572703e72

Table 3. Information on indidividual simulations: model, simulation, and period.

Model	Simulation ID	Period
CNRM-CM5	lgm_r1i1p1	Last Glacial Maximum
CNRM-CM5	midHolocene_r1i1p1	Mid Holocene
COSMOS-ASO	lgm_r1i1p1	Last Glacial Maximum
CSIRO-Mk3L-1-2	midHolocene_r1i1p1	Mid Holocene
GISS-E2-R	lgm_r1i1p150	Last Glacial Maximum
GISS-E2-R	lgm_r1i1p151	Last Glacial Maximum
GISS-E2-R	midHolocene_r1i1p1	Mid Holocene
GISS-E2-R	past1000_r1i1p121	Last Millennium
GISS-E2-R	past1000_r1i1p122	Last Millennium
GISS-E2-R	past1000_r1i1p1221	Last Millennium
GISS-E2-R	past1000_r1i1p123	Last Millennium
GISS-E2-R	past1000_r1i1p124	Last Millennium
GISS-E2-R	past1000_r1i1p125	Last Millennium
GISS-E2-R	past1000_r1i1p126	Last Millennium
GISS-E2-R	past1000_r1i1p127	Last Millennium
GISS-E2-R	past1000_r1i1p128	Last Millennium
HadCM3	past1000_r1i1p1	Last Millennium
HadGEM2-CC	midHolocene_r1i1p1	Mid Holocene
HadGEM2-ES	midHolocene_r1i1p1	Mid Holocene
IPSL-CM5A-LR	lgm_r1i1p1	Last Glacial Maximum
IPSL-CM5A-LR	midHolocene_r1i1p1	Mid Holocene
IPSL-CM5A-LR	past1000_r1i1p1	Last Millennium
MIROC-ESM	lgm_r1i1p1	Last Glacial Maximum
MIROC-ESM	midHolocene_r1i1p1	Mid Holocene
MIROC-ESM	past1000_r1i1p1	Last Millennium
MPI-ESM-P	lgm_r1i1p1	Last Glacial Maximum
MPI-ESM-P	lgm_r1i1p2	Last Glacial Maximum
MPI-ESM-P	midHolocene_r1i1p1	Mid Holocene
MPI-ESM-P	midHolocene_r1i1p2	Mid Holocene
MPI-ESM-P	past1000_r1i1p1	Last Millennium

Table 4. List of simulations continued

Model	Simulation ID	Period
CESM1	0850cntl.001.cam.h0	pre-industrial control
CESM1	001.cam.h0	Last Millennium
CESM1	002.cam.h0	Last Millennium
CESM1	003.cam.h0	Last Millennium
CESM1	004.cam.h0	Last Millennium
CESM1	005.cam.h0	Last Millennium
CESM1	006.cam.h0	Last Millennium
CESM1	007.cam.h0	Last Millennium
CESM1	008.cam.h0	Last Millennium
CESM1	009.cam.h0	Last Millennium
CESM1	010.cam.h0	Last Millennium
CESM1	011.cam.h0	Last Millennium
CESM1	012.cam.h0	Last Millennium
CESM1	013.cam.h0	Last Millennium
CESM1	850forcing.003.cam.h0	Last Millennium
CESM1	GHG.001.cam.h0	Last Millennium
CESM1	GHG.002.cam.h0	Last Millennium
CESM1	GHG.003.cam.h0	Last Millennium
CESM1	LULC_HurttPongratz.001.cam.h0	Last Millennium
CESM1	LULC_HurttPongratz.002.cam.h0	Last Millennium
CESM1	LULC_HurttPongratz.003.cam.h0	Last Millennium
CESM1	ORBITAL.001.cam.h0	Last Millennium
CESM1	ORBITAL.002.cam.h0	Last Millennium
CESM1	ORBITAL.003.cam.h0	Last Millennium
CESM1	OZONE_AER.001.cam.h0	1850CE-2005CE
CESM1	SSI_VSK_L.001.cam.h0	Last Millennium
CESM1	SSI_VSK_L.003.cam.h0	Last Millennium
CESM1	SSI_VSK_L.004.cam.h0	Last Millennium
CESM1	SSI_VSK_L.005.cam.h0	Last Millennium
CESM1	VOLC_GRA.001.cam.h0	Last Millennium
CESM1	VOLC_GRA.002.cam.h0	Last Millennium
CESM1	VOLC_GRA.003.cam.h0	Last Millennium
CESM1	VOLC_GRA.004.cam.h0	Last Millennium
CESM1	VOLC_GRA.005.cam.h0	Last Millennium

Table 5. List of simulations continued.

Model	Simulation ID	Period
CCSM3	trace	LGM to present
MPI-ESM-Cosmos	mil0001	pre-industrial control
MPI-ESM-Cosmos	mil0006	Last Millennium up 2005CE
MPI-ESM-Cosmos	mil0021	Last Millennium to 2100CE
MPI-ESM-Cosmos	mil0025	Last Millennium to 2100CE
MPI-ESM-Cosmos	mil0026	Last Millennium to 2100CE
FAMOUS-HadCM3 (accelerated)	ALL-5G	Last Glacial Cycle
FAMOUS-HadCM3 (accelerated)	GHG	Last Glacial Cycle
FAMOUS-HadCM3 (accelerated)	ORB	Last Glacial Cycle
FAMOUS-HadCM3 (accelerated)	ALL-ZH Last Glacial Cycle	
FAMOUS-HadCM3 (accelerated)	ICE	Last Glacial Cycle

Author contributions. OB designed and conducted the study and was the main author. Both authors discussed the methods, the results, and their implications.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Funding for this research is by the German Federal Ministry of Education and Research (BMBF) within the Research for Sustainability initiative (FONA; www.fona.de) through the first and second phases of the PalMod project (FKZ: 01LP1509A, FKZ: 01LP1926B). Discussions with Marlene Klockmann and Sebastian Wagner helped to improve the manuscript. We thank two reviewers and the editor for their valuable comments.

References

Anand, P., Elderfield, H., and Conte, M. H.: Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series, Paleoceanography, 18, n/a–n/a, https://doi.org/10.1029/2002PA000846, 2003.

Annan, J. D. and Hargreaves, J. C.: Identification of climatic state with limited proxy data, Climate of the Past, 8, 1141–1151, https://doi.org/10.5194/cp-8-1141-2012, 2012.

Bartlein, P. J. and Shafer, S. L.: Paleo calendar-effect adjustments in time-slice and transient climate-model simulations (PaleoCalAdjust v1.0): impact and strategies for data analysis, Geoscientific Model Development Discussions, pp. 1–36, https://doi.org/10.5194/gmd-2018-283, 2018.

Bendle, J. and Rosell-Melé, A.: High-resolution alkenone sea surface temperature variability on the North Icelandic Shelf: implications for Nordic Seas palaeoclimatic development during the Holocene, The Holocene, 17, 9–24, https://doi.org/10.1177/0959683607073269, 2007.

Bothe, O. and Zorita, E.: Proxy surrogate reconstructions for Europe and the estimation of their uncertainties, Climate of the Past Discussions, pp. 1–27, https://doi.org/10.5194/cp-2019-81, 2019.

Bothe, O., Wagner, S., and Zorita, E.: Simple noise estimates and pseudoproxies for the last 21 000 years, Earth System Science Data, 11, 1129–1152, https://doi.org/10.5194/essd-11-1129-2019, ~~2019.~~ 2019a.

Bothe, O., Wagner, S., and Zorita, E.: Simple noise estimates and pseudoproxies for the last 21k years, https://doi.org/10.17605/OSF.IO/ZBEHX, https://osf.io/zbehx/, last access: 20 August 2020, 2019b.

Braconnot, P., Harrison, S. P., Otto-Bliesner, B., Abe-Ouchi, A., Jungclaus, J., and Peterschmitt, J. Y.: The Paleoclimate Modeling Intercomparison Project contribution to CMIP5, CLIVAR Exchanges, 56, 15–19, 2011.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nature Climate Change, 2, 417–424, https://doi.org/10.1038/nclimate1456, http://dx.doi.org/10.1038/nclimate1456, 2012.

Budich, R., Giorgetta, M., Jungclaus, J. H., and Reick, C. H.: The MPI-M Millennium Earth System Model : An Assembling Guide for the COSMOS Configuration, Tech. rep., https://pure.mpg.de/rest/items/item{_}2193290/component/file{_}2193291/content, 2010.

Cacho, I., Grimalt, J. O., Canals, M., Sbaffi, L., Shackleton, N. J., Schönfeld, J., and Zahn, R.: Variability of the western Mediterranean Sea surface temperature during the last 25,000 years and its connection with the Northern Hemisphere climatic changes, Paleoceanography, 16, 40–52, https://doi.org/10.1029/2000PA000502, 2001.

Cacho, I., Grimalt, J. O., Canals, M., Sbaffi, L., Shackleton, N. J., Schönfeld, J., and Zahn, R.: Western Mediterranean d18O and Uk37 Data and SST Reconstructions., iGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2006-106. https://www.ncdc.noaa.gov/paleo-search/study/6374, last access: 13 January 2020, 2006.

Calvo, E., Grimalt, J., and Jansen, E.: High resolution U 37K sea surface temperature reconstruction in the Norwegian Sea during the Holocene, Quaternary Science Reviews, 21, 1385–1394, https://doi.org/10.1016/S0277-3791(01)00096-8, 2002.

Came, R. E., Oppo, D. W., and McManus, J. F.: Amplitude and timing of temperature and salinity variability in the subpolar North Atlantic over the past 10 k.y., Geology, 35, 315, https://doi.org/10.1130/G23455A.1, 2007a.

Came, R. E., Oppo, D. W., and McManus, J. F.: Subpolar North Atlantic (ODP984) Holocene Planktic d18O and Mg/Ca Data, https://www.ncdc.noaa.gov/paleo-search/study/5593, last access: 13 January 2020, 2007b.

Castañeda, I. S., Schefuß, E., Pätzold, J., Sinninghe Damsté, J. S., Weldeab, S., and Schouten, S.: Millennial-scale sea surface temperature changes in the eastern Mediterranean (Nile River Delta region) over the last 27,000 years, Paleoceanography, 25, https://doi.org/10.1029/2009PA001740, 2010a.

Castañeda, I. S., Schefuß, E., Pätzold, J., Sinninghe Damsté, J. S., Weldeab, S., and Schouten, S.: Isoprenoidal GDGT and alkenone-based proxies of sediment core GeoB7702-3, https://doi.org/10.1594/PANGAEA.736909, 2010b.

~~Clark, P. U., Shakun, J. D., Baker, P. A., Bartlein, P. J., Brewer, S., Brook, E., Carlson, A. E., Cheng, H., Kaufman, D. S., Liu, Z., Marchitto, T. M., Mix, A. C., Morrill, C., Otto-Bliesner, B. L., Pahnke, K., Russell, J. M., Whitlock, C., Adkins, J. F., Blois, J. L., Clark, J., Colman, S. M., Curry, W. B., Flower, B. P., He, F., Johnson, T. C., Lynch-Stieglitz, J., Markgraf, V., McManus, J., Mitrovica, J. X., Moreno, P. I., and Williams, J. W.: Global climate evolution during the last deglaciation., Proceedings of the National Academy of Sciences of the United States of America, 109, E1134–42, , 2012.~~

Collins, M., Tett, S. F., and Cooper, C.: The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments, Climate Dynamics, 17, 61–81, https://doi.org/10.1007/s003820000094, http://link.springer.com/10.1007/s003820000094, 2001.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220–1247, https://doi.org/10.1002/2015MS000447, 2015.

Dee, S., Parsons, L., Loope, G., Overpeck, J., Ault, T., and Emile-Geay, J.: Improved spectral comparisons of paleoclimate models and observations via proxy system modeling: Implications for multi-decadal variability, Earth and Planetary Science Letters, 476, 34–46, https://doi.org/10.1016/J.EPSL.2017.07.036, https://www.sciencedirect.com/science/article/pii/S0012821X1730420X, 2017.

Dee, S. G., Steiger, N. J., Emile-Geay, J., and Hakim, G. J.: On the utility of proxy system models for estimating climate states over the common era, Journal of Advances in Modeling Earth Systems, https://doi.org/10.1002/2016MS000677, 2016.

Dee, S. G., Russell, J. M., Morrill, C., Chen, Z., and Neary, A.: PRYSM v2.0 : A Proxy System Model for Lacustrine Archives, Paleoceanography and Paleoclimatology, https://doi.org/10.1029/2018PA003413, 2018.

Digerfeldt, G.: Pollen profile FLARKTOT, Lake Flarken, Sweden, https://doi.org/10.1594/PANGAEA.711884, https://doi.org/10.1594/PANGAEA.711884, 2009.

Digerfeldt, G.: Age determination of sediment core FLARKTOT, Lake Flarken, Sweden, https://doi.org/10.1594/PANGAEA.740343, https://doi.org/10.1594/PANGAEA.740343, 2010.

Dolman, A. M. and Laepple, T.: Sedproxy: A forward model for sediment-archived climate proxies, Climate of the Past, 14, 1851–1868, https://doi.org/10.5194/cp-14-1851-2018, 2018.

Dolven, J. K., Cortese, G., and Bjørklund, K. R.: A high-resolution radiolarian-derived paleotemperature record for the Late Pleistocene-Holocene in the Norwegian Sea, Paleoceanography, 17, 24–1–24–13, https://doi.org/10.1029/2002pa000780, 2002.

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the

IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Emeis, K.-C., Struck, U., Schulz, H.-M., Rosenberg, M., Bernasconi, S. M., Erlenkeuser, H., Sakamoto, T., and Martinez-Ruiz, F. C.: Sea surface temperature reconstruction for Mediterranean Sea samples, supplement to: Emeis, Kay-Christian; Struck, Ulrich; Schulz, Hans-Martin; Rosenberg, Mark; Bernasconi, Stefano M; Erlenkeuser, Helmut; Sakamoto, Tatsuhiko; Martinez-Ruiz, Francisca C (2000): Temperature and salinity variations of Mediterranean Sea surface water over the last 16, 000 years from records of planktonic stable oxygen isotopes and alkenone unsaturation ratios. Palaeogeography, Palaeoclimatology, Palaeoecology, 158(3-4), 259-280, https://doi.org/10.1594/PANGAEA.735959, 2000a.

Emeis, K.-C., Struck, U., Schulz, H.-M., Rosenberg, R., Bernasconi, S., Erlenkeuser, H., Sakamoto, T., and Martinez-Ruiz, F.: Temperature and salinity variations of Mediterranean Sea surface waters over the last 16,000 years from records of planktonic stable oxygen isotopes and alkenone unsaturation ratios, Palaeogeography, Palaeoclimatology, Palaeoecology, 158, 259–280, https://doi.org/10.1016/S0031-0182(00)00053-5, http://www.sciencedirect.com/science/article/pii/S0031018200000535, 2000b.

Emeis, K.-C., Struck, U., Blanz, T., Kohly, A., and Voß, M.: Salinity changes in the central Baltic Sea (NW Europe) over the last 10 000 years, The Holocene, 13, 411–421, https://doi.org/10.1191/0959683603hl634rp, http://hol.sagepub.com/content/13/3/411.abstract, 2003a.

Emeis, K.-C., Struck, U., Blanz, T., Kohly, A., and Voss, M.: Sea-surface temperature reconstruction of sediment cores from the Skagerrak, supplement to: Emeis, Kay-Christian; Struck, Ulrich; Blanz, Thomas; Kohly, Alexander; Voss, Maren (2003): Salinity changes in the central Baltic Sea (NW Europe) over the last 10000 years. The Holocene, 13(3), 411-421, https://doi.org/10.1594/PANGAEA.738458, 2003b.

Emile-Geay, J., McKay, N. P., Kaufman, D. S., Von Gunten, L., Wang, J., Anchukaitis, K. J., Abram, N. J., Addison, J. A., Curran, M. A., Evans, M. N., Henley, B. J., Hao, Z., Martrat, B., McGregor, H. V., Neukom, R., Pederson, G. T., Stenni, B., Thirumalai, K., Werner, J. P., Xu, C., Divine, D. V., Dixon, B. C., Gergis, J., Mundo, I. A., Nakatsuka, T., Phipps, S. J., Routson, C. C., Steig, E. J., Tierney, J. E., Tyler, J. J., Allen, K. J., Bertler, N. A., Björklund, J., Chase, B. M., Chen, M. T., Cook, E., De Jong, R., DeLong, K. L., Dixon, D. A., Ekaykin, A. A., Ersek, V., Filipsson, H. L., Francus, P., Freund, M. B., Frezzotti, M., Gaire, N. P., Gajewski, K., Ge, Q., Goosse, H., Gornostaeva, A., Grosjean, M., Horiuchi, K., Hormes, A., Husum, K., Isaksson, E., Kandasamy, S., Kawamura, K., Kilbourne, K. H., Koç, N., Leduc, G., Linderholm, H. W., Lorrey, A. M., Mikhalenko, V., Mortyn, P. G., Motoyama, H., Moy, A. D., Mulvaney, R., Munz, P. M., Nash, D. J., Oerter, H., Opel, T., Orsi, A. J., Ovchinnikov, D. V., Porter, T. J., Roop, H. A., Saenger, C., Sano, M., Sauchyn, D., Saunders, K. M., Seidenkrantz, M. S., Severi, M., Shao, X., Sicre, M. A., Sigl, M., Sinclair, K., St George, S., St Jacques, J. M., Thamban, M., Thapa, U. K., Thomas, E. R., Turney, C., Uemura, R., Viau, A. E., Vladimirova, D. O., Wahl, E. R., White, J. W., Yu, Z., and Zinke, J.: A global multiproxy database for temperature reconstructions of the Common Era, Scientific Data, 4, 170 088, https://doi.org/10.1038/sdata.2017.88, 2017.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Science Reviews, 76, 16–28, https://doi.org/10.1016/j.quascirev.2013.05.024, 2013.

Franke, J., González-Rouco, J. F., Frank, D., and Graham, N. E.: 200 years of European temperature variability: insights from and tests of the proxy surrogate reconstruction analog method, Climate Dynamics, 37, 133–150, https://doi.org/10.1007/s00382-010-0802-6, 2010.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, https://doi.org/10.1002/jame.20038, 2013.

Giunta, S. and Emeis, K.-C.: Age and alkenone-derived Holocene sea-surface temperature records of sediment core AD91-17, https://doi.org/10.1594/PANGAEA.438366, 2006.

Giunta, S., Emeis, K. C., and Negri, A.: Sea surface temperatures reconstruction of the last 16,000 years in the eastern Mediterranean Sea, https://doi.org/10.13130/2039-4942/5447, 2001.

Gómez-Navarro, J., Werner, J., Wagner, S., Zorita, E., and Luterbacher, J.: Precipitation in the Past Millennium in Europe—Extension to Roman Times, in: Integrated Analysis of Interglacial Climate Dynamics (INTERDYNAMIC), edited by Schulz, M. and Paul, A., SpringerBriefs in Earth System Sciences, pp. 133–139, Springer International Publishing, https://doi.org/10.1007/978-3-319-00693-2_22, http://dx.doi.org/10.1007/978-3-319-00693-2_22, 2015.

Gómez-Navarro, J. J., Werner, J., Wagner, S., Luterbacher, J., and Zorita, E.: Establishing the skill of climate field reconstruction techniques for precipitation with pseudoproxy experiments, Climate Dynamics, 45, 1395–1413, https://doi.org/10.1007/s00382-014-2388-x, 2015.

Gómez-Navarro, J. J., Zorita, E., Raible, C. C., and Neukom, R.: Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era, Climate of the Past, 13, 629–648, https://doi.org/10.5194/cp-13-629-2017, 2017.

Graham, N., Hughes, M., Ammann, C., Cobb, K., Hoerling, M., Kennett, D., Kennett, J., Rein, B., Stott, L., Wigand, P., and Xu, T.: Tropical Pacific – mid-latitude teleconnections in medieval times, Climatic Change, 83, 241–285, https://doi.org/10.1007/s10584-007-9239-2, 2007.

Grimalt, J. O. and Calvo, E.: Age and alkenone-derived Holocene sea-surface temperature records of sediment core MD95-2011, https://doi.org/10.1594/PANGAEA.438810, 2006.

Grimalt, J. O. and Marchal, O.: Age and alkenone-derived Holocene sea-surface temperature records of sediment core MD95-2015, https://doi.org/10.1594/PANGAEA.438814, 2006.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research, 121, 6745–6764, https://doi.org/10.1002/2016JD024751, 2016.

He, F.: Simulating Transient Climate Evolution of the Last Deglaciation with CCSM3, Ph.D. thesis, University of Wisconsin-Madison, http://www.cgd.ucar.edu/ccr/TraCE/doc/He_PhD_dissertation_UW_2011.pdf, 2011.

Jensen, M. F., Nummelin, A., Nielsen, S. B., Sadatzki, H., Sessford, E., Risebrobakken, B., Andersson, C., Voelker, A., Roberts, W. H. G., Pedro, J., and Born, A.: A spatiotemporal reconstruction of sea-surface temperatures in the North Atlantic during Dansgaard–Oeschger events 5–8, Climate of the Past, 14, 901–922, https://doi.org/10.5194/cp-14-901-2018, 2018.

Jones, C. D., Hughes, J. K., Bellouin, N., Hardiman, S. C., Jones, G. S., Knight, J., Liddicoat, S., O'Connor, F. M., Andres, R. J., Bell, C., Boo, K. O., Bozzo, A., Butchart, N., Cadule, P., Corbin, K. D., Doutriaux-Boucher, M., Friedlingstein, P., Gornall, J., Gray, L., Halloran, P. R., Hurtt, G., Ingram, W. J., Lamarque, J. F., Law, R. M., Meinshausen, M., Osprey, S., Palin, E. J., Chini, L. P., Raddatz, T., Sanderson, M. G., Sellar, A. A., Schurer, A., Valdes, P., Wood, N., Woodward, S., Yoshioka, M., and Zerroukat, M.: The HadGEM2-ES implementation of CMIP5 centennial simulations, Geoscientific Model Development, 4, 543–570, https://doi.org/10.5194/gmd-4-543-2011, 2011.

Jones, M. D. and Dee, S. G.: Global-scale proxy system modelling of oxygen isotopes in lacustrine carbonates: New insights from isotope-enabled-model proxy-data comparison, Quaternary Science Reviews, https://doi.org/10.1016/J.QUASCIREV.2018.09.009, 2018.

Jonkers, L. and Kučera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, Climate of the Past, 13, 573–586, https://doi.org/10.5194/cp-13-573-2017, 2017.

Jonkers, L. and Kučera, M.: Sensitivity to species selection indicates the effect of nuisance variables on marine microfossil transfer functions, Climate of the Past Discussions, pp. 1–19, https://doi.org/10.5194/cp-2018-107, 2018.

Jungclaus, J. H., Lorenz, S. J., Timmreck, C., Reick, C. H., Brovkin, V., Six, K., Segschneider, J., Giorgetta, M. A., Crowley, T. J., Pongratz, J., Krivova, N. A., Vieira, L. E., Solanki, S. K., Klocke, D., Botzet, M., Esch, M., Gayler, V., Haak, H., Raddatz, T. J., Roeckner, E., Schnur, R., Widmann, H., Claussen, M., Stevens, B., and Marotzke, J.: Climate and carbon-cycle variability over the last millennium, Climate of the Past, 6, 723–737, https://doi.org/10.5194/cp-6-723-2010, 2010.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L. L., Fischer, H., Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Lunt, D. J., Mahowald, N. M., Richard Peltier, W., Phipps, S. J., Roche, D. M., Schmidt, G. A., Tarasov, L., Valdes, P. J., Zhang, Q., and Zhou, T.: The PMIP4 contribution to CMIP6 - Part 1: Overview and over-arching analysis plan, Geoscientific Model Development, 11, 1033–1057, https://doi.org/10.5194/gmd-11-1033-2018, 2018.

Kim, J.-H., Rimbu, N., Lorenz, S. J., Lohmann, G., Nam, S.-I., Schouten, S., Rühlemann, C., and Schneider, R. R.: North Pacific and North Atlantic sea-surface temperature variability during the Holocene, Quaternary Science Reviews, 23, 2141–2154, https://doi.org/10.1016/j.quascirev.2004.08.010, 2004a.

Kim, J.-H., Rimbu, N., Lorenz, S. J., Lohmann, G., Nam, S.-I., Schouten, S., Rühlemann, C., and Schneider, R. R.: Age and alkenone-derived Holocene sea-surface temperature records of sediment core GeoB5901-2, https://doi.org/10.1594/PANGAEA.438384, https://doi.pangaea.de/10.1594/PANGAEA.438384, 2004b.

Kim, J.-H., Schouten, S., Hopmans, E. C., Donner, B., and Sinninghe Damsté, J. S.: Global sediment core-top calibration of the TEX86 paleothermometer in the ocean, Geochimica et Cosmochimica Acta, 72, 1154–1173, https://doi.org/10.1016/j.gca.2007.12.010, 2008.

Konecky, B., Dee, S. G., and Noone, D.: WaxPSM: A forward model of leaf wax hydrogen isotope ratios to bridge proxy and model estimates of past climate, Journal of Geophysical Research: Biogeosciences, p. 2018JG004708, https://doi.org/10.1029/2018JG004708, 2019.

Kretschmer, K., Kucera, M., and Schulz, M.: Modeling the distribution and seasonality of Neogloboquadrina pachyderma in the North Atlantic Ocean during Heinrich Stadial 1, Paleoceanography, 31, 986–1010, https://doi.org/10.1002/2015PA002819, 2016.

Kretschmer, K., Jonkers, L., Kucera, M., and Schulz, M.: Modeling seasonal and vertical habitats of planktonic foraminifera on a global scale, Biogeosciences, 15, 4405–4429, https://doi.org/10.5194/bg-15-4405-2018, 2018.

Larocque, I. and Hall, R. I.: Holocene temperature estimates and chironomid community composition in the Abisko Valley, northern Sweden, Quaternary Science Reviews, 23, 2453–2465, https://doi.org/10.1016/j.quascirev.2004.04.006, 2004.

Larocque, I. and Hall, R. I.: Abisko Valley, Northern Sweden, Chironomid-inferred Temperature Reconstruction, https://www.ncdc.noaa.gov/paleo-search/study/6349, last access: 13 January 2020, 2006.

Liu, Z., Otto-Bliesner, B. L., He, F., Brady, E. C., Tomas, R., Clark, P. U., Carlson, A. E., Lynch-Stieglitz, J., Curry, W., Brook, E., Erickson, D., Jacob, R., Kutzbach, J., and Cheng, J.: Transient simulation of last deglaciation with a new mechanism for Bolling-Allerod warming., Science (New York, N.Y.), 325, 310–4, https://doi.org/10.1126/science.1171041, 2009.

Malevich, S. B., Vetter, L., and Tierney, J. E.: Global Core Top Calibration of δ18O in Planktic Foraminifera to Sea Surface Temperature, Paleoceanography and Paleoclimatology, 34, 1292–1315, https://doi.org/10.1029/2019PA003576, 2019.

Marchal, O., Cacho, I., Stocker, T. F., Grimalt, J. O., Calvo, E., Martrat, B., Shackleton, N., Vautravers, M., Cortijo, E., van Kreveld, S., Andersson, C., Koç, N., Chapman, M., Sbaffi, L., Duplessy, J.-C., Sarnthein, M., Turon, J.-L., Duprat, J., and Jansen, E.: Apparent long-term cooling of the sea surface in the northeast Atlantic and Mediterranean during the Holocene, Quaternary Science Reviews, 21, 455–483, https://doi.org/10.1016/S0277-3791(01)00105-6, 2002.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of regional and global temperature for the past 11,300 years., Science (New York, N.Y.), 339, 1198–201, https://doi.org/10.1126/science.1228026, 2013.

Matulla, C., Zhang, X., Wang, X. L., Wang, J., Zorita, E., Wagner, S., and Storch, H.: Influence of similarity measures on the performance of the analog method for downscaling daily precipitation, Climate Dynamics, 30, 133–144, https://doi.org/10.1007/s00382-007-0277-2, 2008.

Murdoch, D. and Chow, E. D.: ellipse: Functions for Drawing Ellipses and Ellipse-Like Confidence Regions, https://CRAN.R-project.org/package=ellipse, r package version 0.4.1, 2018.

Murdoch, D. J. and Chow, E. D.: A Graphical Display of Large Correlation Matrices, American Statistician, 50, 178–180, https://doi.org/10.1080/00031305.1996.10474371, 1996.

Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P.: No evidence for globally coherent warm and cold periods over the preindustrial Common Era, Nature, 571, 550–554, https://doi.org/10.1038/s41586-019-1401-2, 2019.

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model, Bulletin of the American Meteorological Society, 97, 735–754, https://doi.org/10.1175/bams-d-14-00233.1, 2015.

Phipps, S. J., Rotstayn, L. D., Gordon, H. B., Roberts, J. L., Hirst, A. C., and Budd, W. F.: The CSIRO Mk3L climate system model version 1.0 - Part 1: Description and evaluation, Geoscientific Model Development, 4, 483–509, https://doi.org/10.5194/gmd-4-483-2011, 2011.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, last access: 27 November 2019, 2019.

Rebotim, A., Voelker, A. H., Jonkers, L., Waniek, J. J., Meggers, H., Schiebel, R., Fraile, I., Schulz, M., and Kucera, M.: Factors controlling the depth habitat of planktonic foraminifera in the subtropical eastern North Atlantic, Biogeosciences, 14, 827–859, https://doi.org/10.5194/bg-14-827-2017, 2017.

Reschke, M., Rehfeld, K., and Laepple, T.: Empirical estimate of the signal content of Holocene temperature proxy records, Climate of the Past Discussions, pp. 1–27, https://doi.org/10.5194/cp-2018-154, 2018.

Rodrigues, T., Grimalt, J. O., Abrantes, F. G., Flores, J. A., and Lebreiro, S. M.: Holocene interdependences of changes in sea surface temperature, productivity, and fluvial inputs in the Iberian continental shelf (Tagus mud patch), Geochemistry, Geophysics, Geosystems, 10, n/a–n/a, https://doi.org/10.1029/2008GC002367, 2009.

Rodrigues, T., Grimalt, J. O., Abrantes, F. F., Naughton, F., and Flores, J.-A.: Sea surface temperature reconstruction from sediment core D13882, https://doi.org/10.1594/PANGAEA.761811, 2010.

Rosell-Melé, A., Bard, E., Emeis, K.-C., Grimalt, J. O., Müller, P., Schneider, R., Bouloubassi, I., Epstein, B., Fahl, K., Fluegge, A., Freeman, K., Goñi, M., Güntner, U., Hartz, D., Hellebust, S., Herbert, T., Ikehara, M., Ishiwatari, R., Kawamura, K., Kenig, F., de Leeuw, J., Lehman, S., Mejanelle, L., Ohkouchi, N., Pancost, R. D., Pelejero, C., Prahl, F., Quinn, J., Rontani, J.-F., Rostek, F., Rullkötter, J., Sachs, J., Blanz, T., Sawada, K., Schulz-Bull, D., Sikes, E., Sonzogni, C., Ternois, Y., Versteegh, G., Volkman, J. K., and Wakeham, S.: Precision of the current methods to measure the alkenone proxy U 37 K' and absolute alkenone abundance in sediments: Results of an interlaboratory comparison study, Geochemistry, Geophysics, Geosystems, 2, n/a–n/a, https://doi.org/10.1029/2000GC000141, 2001.

Sarnthein, M., Van Kreveld, S., Erlenkeuser, H., Grootes, P. M., Kucera, M., Pflauman, U., and Schulz, M.: Centennial-to-millennial-scale periodicities of Holocene climate and sediment injections off the western Barents shelf, 75°N, Boreas, 32, 447–461, https://doi.org/10.1111/j.1502-3885.2003.tb01227.x, 2003a.

Sarnthein, M., Van Kreveld, S., Erlenkeuser, H., Grootes, P. M., Kucera, M., Pflauman, U., and Schulz, M.: Sea surface temperatures of sediment core GIK23258-2, PANGAEA - Data Publisher for Earth & Environmental Science, https://doi.org/10.1594/PANGAEA.114683, in supplement to: Sarnthein, M et al. (2003): Centennial-to-millennial-scale periodicities of Holocene climate and sediment injections off western Barents shelf, 75°N. Boreas, 32(3), 447-461, https://doi.org/10.1111/j.1502-3885.2003.tb01227.x, 2003b.

Schmidt, G. A.: Forward modeling of carbonate proxy data from planktonic foraminifera using oxygen isotope tracers in a global ocean model, Paleoceanography, 14, 482–497, https://doi.org/10.1029/1999PA900025, 1999.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., Tremblay, L. B., and Yiou, P.: Using palaeo-climate comparisons to constrain future projections in CMIP5, Climate of the Past, 10, 221–250, https://doi.org/10.5194/cp-10-221-2014, 2014a.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y.-H., Cheng, Y., Clune, T. L., Del Genio, A., de Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, Journal of Advances in Modeling Earth Systems, 6, 141–184, https://doi.org/10.1002/2013MS000265, 2014b.

Seppä, H. and Birks, H.: July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: pollen-based climate reconstructions, The Holocene, 11, 527–539, https://doi.org/10.1191/095968301680223486, 2001.

Seppä, H., Hammarlund, D., and Antonsson, K.: Low-frequency and high-frequency changes in temperature and effective humidity during the Holocene in south-central Sweden: implications for atmospheric and oceanic forcings of climate, Climate Dynamics, 25, 285–297, https://doi.org/10.1007/s00382-005-0024-5, 2005.

~~Shakun, J. D., Clark, P. U., He, F., Marcott, S. A., Mix, A. C., Liu, Z., Otto-Bliesner, B., Schmittner, A., and Bard, E.: Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation., Nature, 484, 49–54, , , 2012.~~

Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, WIREs Clim Change, 3, 63–77, https://doi.org/10.1002/wcc.149, http://dx.doi.org/10.1002/wcc.149, 2012.

Smith, R. S. and Gregory, J.: The last glacial cycle: transient simulations with an AOGCM, Climate Dynamics, 38, 1545–1559, https://doi.org/10.1007/s00382-011-1283-y, 2012.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction, Journal of Climate, 27, 426–441, https://doi.org/10.1175/JCLI-D-12-00693.1, 2014.

Sueyoshi, T., Ohgaito, R., Yamamoto, A., Chikamoto, M. O., Hajima, T., Okajima, H., Yoshimori, M., Abe, M., O'ishi, R., Saito, F., Watanabe, S., Kawamiya, M., and Abe-Ouchi, A.: Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM, Geoscientific Model Development, 6, 819–836, https://doi.org/10.5194/gmd-6-819-2013, https://www.geosci-model-dev.net/6/819/2013/, 2013.

Sundqvist, H. S., Kaufman, D. S., McKay, N. P., Balascio, N. L., Briner, J. P., Cwynar, L. C., Sejrup, H. P., Seppä, H., Subetto, D. A., Andrews, J. T., Axford, Y., Bakke, J., Birks, H. J. B., Brooks, S. J., de Vernal, A., Jennings, A. E., Ljungqvist, F. C., Rühland, K. M., Saenger, C., Smol, J. P., and Viau, A. E.: Arctic Holocene Proxy Climate Database, https://www.ncdc.noaa.gov/paleo-search/study/15444, last access: 13 January 2020, 2014a.

Sundqvist, H. S., Kaufman, D. S., McKay, N. P., Balascio, N. L., Briner, J. P., Cwynar, L. C., Sejrup, H. P., Seppä, H., Subetto, D. A., Andrews, J. T., Axford, Y., Bakke, J., Birks, H. J. B., Brooks, S. J., de Vernal, A., Jennings, A. E., Ljungqvist, F. C., Rühland, K. M., Saenger, C., Smol, J. P., and Viau, A. E.: Arctic Holocene proxy climate database – new approaches to assessing geochronological accuracy and encoding climate variables, Climate of the Past, 10, 1605–1631, https://doi.org/10.5194/cp-10-1605-2014, 2014b.

Talento, S., Schneider, L., Werner, J., and Luterbacher, J.: Millennium-length precipitation reconstruction over south-eastern Asia: a pseudo-proxy approach, Earth System Dynamics, 10, 347–364, https://doi.org/10.5194/esd-10-347-2019, 2019.

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, Climate of the Past, 15, 1251–1273, https://doi.org/10.5194/cp-15-1251-2019, 2019.

Telford, R. J., Li, C., and Kucera, M.: Mismatch between the depth habitat of planktonic foraminifera and the calibration depth of SST transfer functions may bias reconstructions, Climate of the Past, 9, 859–870, https://doi.org/10.5194/cp-9-859-2013, https://www.clim-past.net/9/859/2013/, 2013.

Thompson, D. M., Ault, T. R., Evans, M. N., Cole, J. E., and Emile-Geay, J.: Comparison of observed and simulated tropical climate trends using a forward model of coral δ18O, Geophysical Research Letters, 38, L14 706+, https://doi.org/10.1029/2011gl048224, 2011.

Thornalley, D. J. R., Elderfield, H., and McCave, I. N.: Holocene oscillations in temperature and salinity of the surface subpolar North Atlantic., Nature, 457, 711–4, https://doi.org/10.1038/nature07717, 2009a.

Thornalley, D. J. R., Elderfield, H., and McCave, I. N.: Subpolar North Atlantic Holocene Temperature and Salinity Reconstructions., iGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2009-097. https://www.ncdc.noaa.gov/paleo-search/study/8623, last access: 13 January 2020, 2009b.

Tierney, J. E. and Tingley, M. P.: A TEX 86 surface sediment database and extended Bayesian calibration, Scientific Data, 2, 150 029, https://doi.org/10.1038/sdata.2015.29, 2015.

Tierney, J. E. and Tingley, M. P.: BAYSPLINE: A New Calibration for the Alkenone Paleothermometer, Paleoceanography and Paleoclimatology, 33, 281–301, https://doi.org/10.1002/2017PA003201, http://doi.wiley.com/10.1002/2017PA003201, 2018.

Tierney, J. E., Malevich, S. B., Gray, W., Vetter, L., and Thirumalai, K.: Bayesian calibration of the Mg/Ca paleothermometer in planktic foraminifera, https://doi.org/10.31223/osf.io/y3xdg, eartharxiv.org/y3xdg, 2019.

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannual variation in tree-ring width, Climate Dynamics, 36, 2419–2439, https://doi.org/10.1007/s00382-010-0945-5, 2011.

Trouet, V., Esper, J., Graham, N. E., Baker, A., Scourse, J. D., and Frank, D. C.: Persistent Positive North Atlantic Oscillation Mode Dominated the Medieval Climate Anomaly, Science, 324, 78–80, https://doi.org/10.1126/science.1166349, 2009.

van den Dool, H. M.: Searching for analogues, how long must we wait?, Tellus A, 46, 314–324, https://doi.org/10.1034/j.1600-0870.1994.t01-2-00006.x, 1994.

Voeltzel, D.: Age determination of sediment core TSUOLJVR, Lake Tsuolbmajavri, Finland, https://doi.org/10.1594/PANGAEA.740821, https://doi.org/10.1594/PANGAEA.740821, 2010a.

Voeltzel, D.: Pollen profile TSUOLJVR, Lake Tsuolbmajavri, Finland, https://doi.org/10.1594/PANGAEA.739916, https://doi.org/10.1594/PANGAEA.739916, 2010b.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M. P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L., and Chauvin, F.: The CNRM-CM5.1 global climate model: Description and basic evaluation, Climate Dynamics, 40, 2091–2121, https://doi.org/10.1007/s00382-011-1259-y, http://link.springer.com/10.1007/s00382-011-1259-y, 2013.

Weitzel, N., Wagner, S., Sjolte, J., Klockmann, M., Bothe, O., Andres, H., Tarasov, L., Rehfeld, K., Zorita, E., Widmann, M., Sommer, P., Schädler, G., Ludwig, P., Kapp, F., Jonkers, L., García-Pintado, J., Fuhrmann, F., Dolman, A., Dallmeyer, A., and Brücher, T.: Diving into the Past: A Paleo Data–Model Comparison Workshop on the Late Glacial and Holocene, Bulletin of the American Meteorological Society, 100, ES1–ES4, https://doi.org/10.1175/bams-d-18-0169.1, 2018.

Zanchettin, D., Bothe, O., Müller, W., Bader, J., and Jungclaus, J.: Different flavors of the Atlantic Multidecadal Variability, Climate Dynamics, 42, 381–399, https://doi.org/10.1007/s00382-013-1669-0, 2014.

Zorita, E. and von Storch, H.: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods, J. Climate, 12, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012%3C2474:tamaas%3E2.0.co;2, 1999.