

Dear David Thornalley,

along with this document we submit the revised version of our manuscript “Empirical estimate of the signal content of Holocene temperature proxy records”.

According to the major comments of the reviewer we made the following major changes:

- adding a method section 3.2, discussing the spatial correlation structure in the models vs. the spatial correlation structure in reanalysis data,
- extending the discussion section 5.1 by adding potential model shortcomings that may lead to an overestimation of the spatial coherency in the models, and
- adding a new discussion section 5.5 on implications and future steps forward.

More details on these changes and other ones according to the minor questions/comments are described in the response to the reviewer comments attached below. As a few changes are slightly different to the originally suggested ones, we add the new changes to the original response as set in green normal font. The marked-up version of the revised manuscript is attached to the end of this document.

Thank you again for considering our manuscript.

Kind regards,
Maria Reschke

Dear Oliver Bothe,

Thank you very much for taking the time to review our discussion paper and for your constructive and detailed comments. Below we respond to your comments. Referee comments are set in blue italic font and author comments in black normal font.

Response to Major Comments:

Your results and how much we can infer from them strongly hinges on the assumptions you make. You state those clearly and discuss them already. You do this in your discussion section. From my point of view it is necessary that you extend on these discussions already when you present your assumptions.

I understand that others may disagree with discussions taking place in the methods section. Indeed it may be that in a follow up review I say, I was wrong, because the manuscript reads better in this version. However considering this version of the manuscript, the lack of a discussion of your assumptions' appropriateness in the method section clouds the reading.

1. Section 3.1: Let me extend on my short summary. First, I think your assumptions are reasonable and well stated and do not invalidate your approach. However, they also can invite strong criticisms. You counter these mostly later in the manuscript, but I think you have to show early that you are thinking about this and why the assumptions are appropriate.

A number of questions you should probably deal with early on are, for example: Aren't models thought to be more homogenous than observations? Isn't it unlikely that the proxies really recorded the same 4D-signal? Less important is possibly, whether you can really capture the uncertainty about the signal in the simple estimates you take.

We agree that it may be helpful to add more information on the assumptions earlier in the method section. Our study does not assume the same 4D-signal, but only assumes that the correlation structure in the models is realistic. We argue that simulating a reasonable correlation structure might be easier than simulating the right phase and amplitudes of the climate variability.

As we agree that our results hinge on the correlation structure, we suggest to address this point in the revised version early on by 1.) adding a new section 3.2, discussing the spatial correlation structure in the models vs. the spatial correlation structure in reanalysis data, and 2.) extending the discussion section to include a list of potential model shortcomings that may lead to an overestimation of the spatial coherency in the models (see also our response to your specific comments).

We suggest to leave the remaining part of the discussion (assumption of additive noise) in the discussion section.

Concerning 1.)

To check the realism of the correlation structure in the model simulations, we now further analysed the correlation structure of the surface temperature field in the 20C3M reanalysis product (Compo et al., 2006) (Fig.R1). Interestingly, analysing the full time-period of 1871 to 2011 results in a much higher decorrelation length than estimated for the Holocene, likely caused by the coherent anthropogenic forcing. Removing the last decades to minimise the

human influence, e.g., analysing 1871-1950 results in a correlation structure resembling the spatial correlation of MPI6k.

As we expect that the climate does not get more localised on longer time scales, but if anything, more spatially coherent (e.g., Jones et al., 1997; Kim and North 1991) this suggests that the decorrelation lengths used in this study might not be unrealistically large.

Thus, instead of relying on climate model simulations one could even obtain similar results based on the reanalysis correlation structure and assuming that the correlation structure is similar on longer time scales than on the time scales sampled by the instrumental data. To make this point, we suggest adding the reanalysis correlation structure estimated from the proxy positions in the manuscript Figure 3 (Fig.R2).

One could still argue that fine-scale structures (e.g., at the coast or at shelves) not resolved by the models (as well as by the reanalysis) might lead to localised variations as we already discuss in Section 5.1, but we do not see a clear evidence for this on inter-annual and longer time scales from analysing high-resolution model simulations (e.g., the AWI-FESOM simulation in an eddy-permitting resolution). However, as this latter work is still preliminary we would not include it and just discuss this possibility.

Concerning 2.)

There are several shortcomings in present climate model simulations such as the two simulations used here that may lead to an overestimation of the coherency in the models. Possibilities include that models underestimate internal climate variability that is generally more localised than externally forced climate variability (e.g., Laepple and Huybers, PNAS 2014). One possibility (Laepple and Huybers, GRL 2014) is that the model effective horizontal diffusivity may be too large which would reduce internal variability and lead to larger correlation structures. Further, the low, non-eddy permitting resolution of the model simulations used here might suppress small scale features and the role of persistent coastal currents.

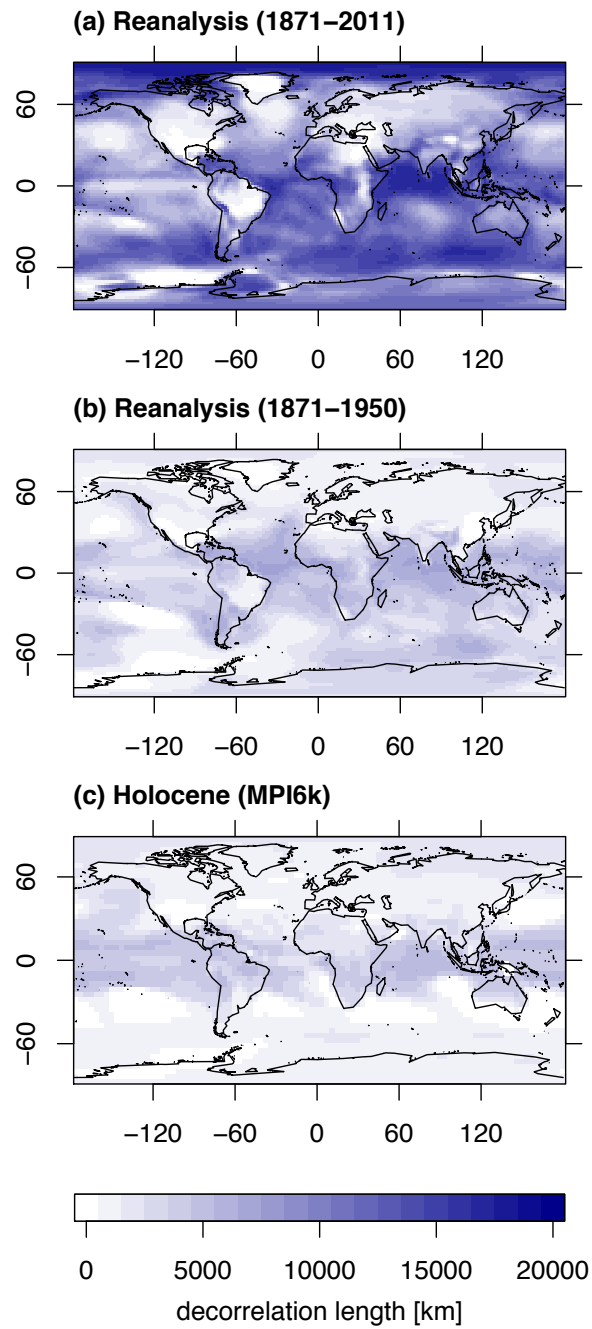


Fig.R1: Decorrelation length of reanalysis data and the 6ky simulation of MPI6k. The decorrelation length is similar for the Holocene and reanalysis data from 1871 to 1950 indicating that the Holocene spatial correlations are realistic.

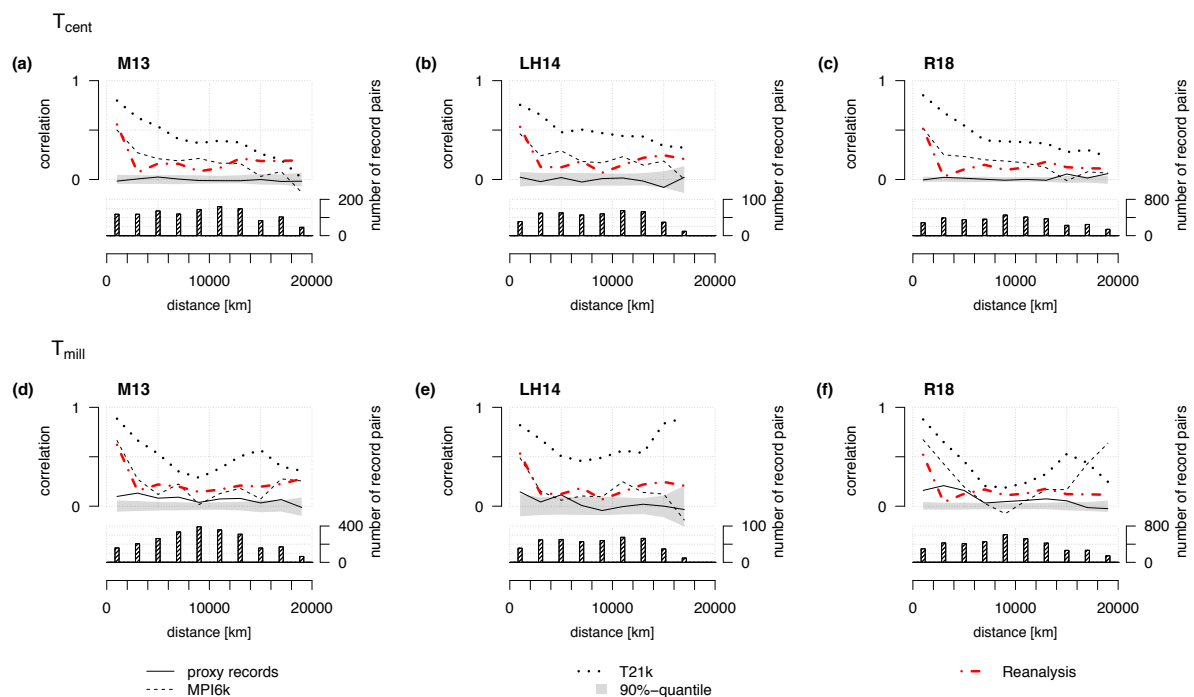


Fig.R2: Spatial correlation of reanalysis data for the time window from 1871 to 1950 (red lines). As the correlation over distance plots for the reanalysis data are very similar to the ones of the MPI6k this indicates that the spatial correlations of the model data are realistic.

2. Reference to Reschke et al. (particularly page 6, line 2): Please provide some information in the methods section or at least in an appendix on the method (not least since Reschke et al. is not openly accessible).

We will add more information in the method section on the reference Reschke et al. to render the manuscript more independent. See our response to the detailed comments below. We further added an unformatted version of the manuscript in the publicly AWI publication database and Researchgate website to allow an easier access to this reference.

Response to Minor Comments:

General:

Maybe mention your focus on the last 6kyr already in the title or at least in the abstract.
We will add this in the abstract.

Abstract:

Suggestion: Maybe rethink the abstract to clarify how you proceed (e.g., page 1, line 17 and following). That is, state which simulations you use before mentioning how they influence the results, or don't mention the specific models at all but just highlight the contrasting results.

We will mention the specific models in the abstract before explaining the results.

Suggestion: Mention the use of the correlation structure before discussing its influence.
We will introduce the correlation structure in the first half of the abstract.

Introduction:

Page 2, line 9: If I understand your point correctly, this is not only about non-climatic influences but also about climatic influences different from the specific signal we are interested in.

Exactly, we consider everything except the specific signal of interest as noise (thus non-climatic signals or climate signals not related to the specific signal of interest). We will rewrite lines 4-10 to make this clearer.

Data:

Page 3, line 12: You possibly should introduce the abbreviation for Methylation of Branched Tetraether - or simply skip mentioning them.

We will follow your suggestion and skip the naming of the 'other' proxies.

Page 3, line 24: Please discuss, why annual temperature is an appropriate choice.

We will add the explanation of our annual mean temperature choice. The reason of our choice is to be consistent to the proxy dataset that we assume to record annual mean temperatures following the standard interpretation of these datasets. The annual mean interpretation is often chosen because of the lack of accurate information about the proxy and location-specific seasonality.

Page 4, line 3: Are all the mentioned forcing factors really continuously transient in the simulation?

Not all forcing factors are continuously transient for the entire T21k simulation. With the disappearance of the Eurasian (~8ky BP) and the Laurentide Ice Sheet (~6ky BP) the transient continental ice sheet forcing ended at around 6ky. As the retreat of the ice sheets ended, there is also no meltwater forcing in the northern hemisphere since 6ky BP. The meltwater fluxes for the southern hemisphere ended at 5ky BP (He, 2011). Thus, only the orbital forcing and the greenhouse gas concentrations are remaining. We will include this in the revised manuscript.

Method:

Page 5, line 8: Can you please provide slightly more information for what this reference is here? I do not directly see how it relates to the sentence.

This reference explains the method used for filtering but as we describe the method in the next paragraph, we will remove the reference here.

Page 5, line 13: What do you mean by resampled in this context? Further, could you give more details on your block averaging (e.g., block length).

Our aim is to derive a time series from the annual model time series that resembles the proxy time series in having the same number and ages of the proxy observations. For this, we apply block averaging. To get the observation for the observation time t_i we average all observations between half the difference to the previous observation time ($t_i - \Delta t_i/2$) and half the difference to the next observation time ($t_i + \Delta t_{i+1}/2$). We chose the approach of averaging the annual time series instead of interpolating, as for marine records samples often

include adjacent depths or the sample distance is smaller than the typical mixing depth in the sediment (Berger and Heath, 1968).

We will rephrase the paragraph and include the details of the block averaging in the revised version.

Suggestion: Page 5, line 16, "For each : :": I am not sure whether this description gives the reader enough information to redo your analyses. But I am neither sure that it does not. Maybe rethink this.

We will rewrite this sentence to: 'For each proxy compilation (M13, LH14, R18), we estimated the time scale dependent (T_{cent} , T_{mill}) correlations between all possible proxy record pairs. We further estimated the time scale dependent correlations between all model time series pairs.' We further plan to make the R scripts available in an online repository.

Page 5, line 20: Please give some more details here on what your reference work describes in this context.

We will extend this description to make it more independent.

'For this step, the irregularly sampled time series were linearly interpolated onto a regular grid ($\Delta t = 10y$) and subjected to a Gaussian filter with cut-off frequency $1/400y$ (T_{cent}) or $1/1000y$ (T_{mill}). This approach has been shown to deliver good results in tests using surrogate data with the sampling properties of Holocene marine sediment cores for the estimation of time scale dependent correlations (Reschke et al., 2019).'

We slightly changed the description: 'For this step, the irregularly sampled time series were linearly interpolated onto a regular grid ($\Delta t = 10y$) and subjected to a Gaussian filter with cut-off frequency $1/400y$ (T_{cent}) and linear detrending or alternatively to a Gaussian filter with cut-off frequency $1/1000y$ (T_{mill}) omitting the detrending step. This approach has been shown to deliver good results for the estimation of time scale dependent correlations in tests using surrogate data with the sampling properties of Holocene marine sediment cores (Reschke et al., 2019).'

Results:

Suggestion: Page 6, line 25: Would you be willing to discuss how realistic you regard the simulated correlation coefficients, and, possibly, how this may affect the relevance of your SNRs, e.g., if we assume the simulated correlation coefficients are not realistic.

In the discussion section 5.1 'Spatial correlation structure of model simulations' we already discuss the possibility of overestimating model correlations and their role for the SNRs and we add an additional discussion on potential shortcomings of the climate model simulations. We would like to keep this structure and not to discuss this in the result section. However, as described in our first answer, we suggest to add a new section 3.2 in the method part, discussion the spatial correlation structure in the models vs. the spatial correlation structure in reanalysis data. This demonstrates that model correlations are not unrealistic and that the main conclusions of the manuscript could be obtained without the use of climate models, just relying on the (mainly inter-annual) instrumental correlation structure and assuming that on longer time scales the correlation should not decrease.

Page 7, line 20: Maybe you should be more explicit in writing about the results for T21k (see also short technical comment below).

We will rewrite the results of T21k-based estimates to be more explicit and precise.

Discussions:

Suggestion: I think it would be interesting to be more specific in what your results imply for interpreting the proxy records and derived larger scale reconstructions. And what it would mean, if your results are either too pessimistic or even too optimistic.

We agree that such a discussion would be useful and will add a new discussion section on 'implications and future steps forward'. First, care should be taken in maximising the SNR when creating Holocene climate records. This includes an optimal measurement design (e.g., the choice of the sample size) for example supported by proxy forward modelling. Second, Holocene studies relying on a small number of records might be associated with a large uncertainty (except if these records have a higher quality than the average). Third, Holocene stacks relying on a large number of records, such as the stack in Marcott et al. (2013), will be robust if the errors are independent across sites but it will be very difficult to extract spatio-temporal patterns from these datasets.

If our results are too pessimistic (e.g., the true climate is more regional than simulated by the used model simulations), this would imply that individual proxy records in the Holocene can be safely interpreted as regionally representative climate signal as it is currently done in the literature. On the other hand, if our SNR estimates are too optimistic, the value of singular proxy reconstructions without additional expert knowledge would be very limited and stacks such as used by the tree ring community might be needed.

Suggestion: Page 8, last paragraph: Are there potentially other reasons that may result in higher correlations, e.g., how the models are built. Did your department's earlier work hint to any further explanations, or did the PAGES project CVAS come up with some additional explanations?

We agree that it would be useful for the reader to include a discussion on possible model shortcomings that could lead to an overestimation of spatial correlations (= underestimation of spatial degrees of freedom).

Possibilities include that models underestimate internal climate variability that is generally more localised than externally forced climate variability. One suggestion (Laepfle and Huybers, GRL 2014) was that the model effective horizontal diffusivity may be too large which would reduce internal variability and lead to larger correlation structures. Further, the low, non-eddy resolving resolution of the models might suppress small scale features and the role of persistent coastal currents. We will add a discussion of these points.

Page 9, line 7: Do I miss it, or do you omit to specify "N".

We actually missed to specify N which is the number of sites ranging from 3 to 50. We will add this.

Suggestion: Page 9, line 18: Maybe make the points of this paragraph already stronger when you present the results.

We mention this already in page 7, line 23. 'An analysis of the proxy-specific SNRs yielded higher uncertainties due to the relatively small number of record pairs (see Fig. S6-S15 for the complete set of results)' but will make this point clearer and link it to the sensitivity study discussion.

Suggestion: Page 10, line 1: Can you discuss, how assuming a more appropriate

seasonal and depth choice would influence your results?

Currently, we interpret all records from proxy types as annual mean surface temperature. As different proxies are recording different parts of the climate component, we expect that the correlation among time series from different proxies is lower than for time series of the same proxy which recorded the temperature from a more similar climate component. This is already discussed in lines 3-8.

However, a different season or depth might also have a different correlation structure in the model which will influence our results. Calculating the correlation structure of summer and winter in both models suggests that this can increase or decrease the correlation and seems to be model dependent. Thus, the net-effect on the SNRs is not clear.

Finally, even for one proxy type and proxy carrier (e.g., foraminifera), the recorded season and depth is location-specific and this will reduce the correlation compared to the correlation of the climate sampled at any globally fixed season or depth. However, this reduction in the SNR (that is defined at the moment for annual mean temperatures, but could be changed to any globally fixed season or depth) is real.

We will add a discussion of the latter two points.

Page 10, line 1: Isn't the work of Jonkers and Kucera and further of their colleagues relevant here?

This is right. We will add appropriate reference (e.g., Jonkers and Kucera, 2017).

Page 10, line 1: I may be wrong, but I think, the work of Jessica Tierney and colleagues on TEX86 calibrations is relevant here.

Our point is that we have seasonal and depth-specific differences in the recorded climate component. In case of TEX86 it was suggested that this proxy records sub-surface temperatures. We will additionally add Tierney and Tingley (2015) proposing a calibration for the upper 200m.

Conclusion:

Suggestion: Page 11, line 18: I think you could be more explicit about the relevance of your work.

We will add a new section in the discussion on 'Implications and future steps forward' as described above and in the response to reviewer 2. We will further extend/modify this conclusion statement to:

Nevertheless, our SNR estimates are still relevant for synthesis and model comparison efforts (e.g., Marcott et al., 2013), that usually interpret all proxy records together. While in the ideal case, most errors will be averaged out in global stacks based on a large number of records, the interpretation of spatio-temporal patterns will be very uncertain.

We changed this statement slightly: 'Nevertheless, our SNR estimates are still relevant for synthesis and model comparison efforts (e.g., Marcott et al., 2013), that usually interpret all proxy records together. While in the ideal case global stacks based on a large number of records will average out most of the error contributions, the interpretation of spatio-temporal patterns will remain uncertain.'

Response to Technical Comments:

General: I don't mind seeing "years" written out instead of abbreviated. At least in the

abstract I think it would be better to write “400 years” instead of “400y”.

We agree and will change this in the abstract.

Page 1, line 18: Maybe skip “rather”

We agree.

Page 1, line 20: If I understand the sentence correctly, the second “SNRs” plus its article is superfluous.

We agree.

Page 1, line 24: I don’t think the first sentence of the paragraph is necessary.

At the moment, one of the shortcomings of our study is that we are not able to make robust statements about specific proxy types as the amount of proxy-specific records is too low as shown in our sensitivity study. In this sense, we would prefer to keep the sentence in the abstract.

Page 4, line 1: If I understand the sentence correctly, it is incomplete.

Thanks for spotting this. It is missing an ‘and’ which we will correct.

Page 4, line 7ff: Does this sentence and the next refer to both models or do you mean that you use for TraCE all three mentioned variables? Please clarify.

This refers to both models. We will clarify this.

Page 7, line 19 “differ ..”: Please clarify: do you mean they differ between the two simulations?

Yes. We will rewrite it to: ‘For all three proxy compilations (M13, LH14, R18) the SNRs obtained for mixed proxy types depend on the choice of the model simulation.’

Page 7, line 20 “consistent with”: Please clarify: Do you mean they are consistent with the model or they are consistent in the analyses using this model.

The latter, we will add that the SNRs estimated for the three datasets M13, LH14 and R18 are more similar and therefore more consistent if the analysis uses the T21k simulation. Based on MPI6k the SNR estimates are more different so that the results are less consistent than the estimates based on T21k.

Once again, thank you for your comments,
Maria Reschke

References:

- Berger, W. H., and Heath, G. R.: Vertical mixing in pelagic sediments, *J. mar. Res.*, 26, 134-143, 1968.
- He, F.: Simulating transient climate evolution of the last deglaciation with CCSM3, PhD thesis, University of Wisconsin-Madison, 2011.
- Jones, P. D., Osborn, T. J., and Briffa, K. R.: Estimating Sampling Errors in Large-Scale Temperature Averages, *J. Climate*, 10, 2548-2568, 1997.

- Jonkers, L., and Kucera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, *Clim. Past*, 13, 573-586, doi:10.5194/cp-13-573-2017, 2017.
- Kim, K.-Y., and North, G. R.: Surface Temperature Fluctuations in a Stochastic Climate Model, *J. Geophys. Res.*, 96(D10), 18573-18580, doi:10.1029/91JD01959, 1991.
- Laepple, T., and Huybers, P.: Global and Regional Variability in Marine Surface Temperatures, *Geophys. Res. Lett.*, 41(7), 2528-2534, doi:10.1002/2014GL059345, 2014.
- Laepple, T., and Huybers, P.: Ocean Surface Temperature Variability: Large Model-Data Differences at Decadal and Longer Periods, *P. Natl. Acad. Sci. USA*, 111(47), 16682-16687, doi:10.1073/pnas.1412077111, 2014.
- 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*, 339(6124), 1198-1201, doi:10.1126/science.1228026, 2013.
- Reschke, M., Kunz, T., and Laepple, T.: Comparing methods for analysing time scale dependent correlations in irregularly sampled time series data, *Comp. Geosci.*, 123, 65-72, doi:10.1016/j.cageo.2018.11.009, 2019.
- Tierney, J. E., and Tingley, M.: A TEX86 surface sediment database and extended Bayesian calibration, *Scientific data*, 2, 150029, doi:10.1038/sdata.2015.29, 2015.

Dear Reviewer,

Thank you very much for reviewing our discussion paper and your constructive comments. Below we respond to your comments set in blue italic font. The author comments are set in black normal font.

Substantial Concerns:

1. The assumption of model simulation as reality:

o Although the authors are upfront regarding their underlying assumptions and essentially state that they are taking the model output at face value, I strongly recommend exploring more ways in which the model simulations might be oversimplifying their results. For example, in their Discussion section, the authors discuss the role of the “spatial correlation structure of model simulations” and how these correlations might be overestimated. The authors should add a discussion here about how biases in the simulation of climate variability itself in these transient models can lead to biases in correlation distances.

o In other words, if a (hypothesized) transient simulation from 6 ka to present showed the same, coherent changes across the entire Northern Hemisphere, the calculated SNE, as the authors propose, would be exceedingly low - however, we know that such a transient simulation is an unlikely representation of reality. Thus, I feel that the manuscript would greatly benefit if the authors included text on how typical (and atypical) shortcomings of MPI6k and T21k are influencing their results.

We agree with the reviewer that including a better discussion of the correlation structure, possible shortcomings in the model simulations and their effect of our results would be useful. This was also asked by Reviewer 1.

We suggest to address this point in the revised version by 1.) adding a new section 3.2, discussing the spatial correlation structure in the models vs. the spatial correlation structure in reanalysis data, and 2.) extending the discussion section to include a list of potential model shortcomings that may lead to an overestimation of the spatial coherency in the models.

Concerning 1.)

To check the realism of the correlation structure in the model simulations, we further analysed the correlation structure of the surface temperature field in the 20C3M reanalysis product (Compo et al., 2006) (Fig.R1). Interestingly, analysing the full time-period of 1871-2011 results in a much higher decorrelation length than estimated for the Holocene, likely caused by the coherent anthropogenic forcing. Removing the last decades to minimise the human influence, e.g., analysing 1871-1950 results in a correlation structure resembling the spatial correlation of MPI6k.

As we expect that the climate does not get more localised on longer time scales, but if anything, more spatially coherent (e.g., Jones et al., 1997; Kim and North, 1991) this suggests that the decorrelation lengths used in this study might not be unrealistically large.

Thus, instead of relying on climate model simulations one could even obtain similar results based on the reanalysis correlation structure and assuming that the correlation structure is similar on longer time scales than on the time scales sampled by the instrumental data. To make this point, we suggest adding the reanalysis correlation structure estimated from the proxy positions in the manuscript Figure 3 (Fig.R2).

One could still argue that fine-scale structures (e.g., at the coast or at shelves) not resolved by the models (as well as by the reanalysis) might lead to localised variations as we already discuss in Section 5.1, but we do not see a clear evidence for this on inter-annual and longer time scales from analysing high-resolution model simulations (e.g., the AWI-FESOM simulation in an eddy-permitting resolution). However, as this latter work is still preliminary we would not include it and just discuss this possibility.

Concerning 2.)

There are several shortcomings in present climate model simulations such as the two simulations used here that may lead to an overestimation of the coherency in the models. Possibilities include that models underestimate internal climate variability that is generally more localised than externally forced climate variability (e.g., Laepple and Huybers, PNAS 2014). One possibility (Laepple and Huybers, GRL 2014) is that the model effective horizontal diffusivity may be too large which would reduce internal variability and lead to larger correlation structures. Further, the low, non-eddy permitting resolution of the model simulations used here might suppress small scale features and the role of persistent coastal currents.

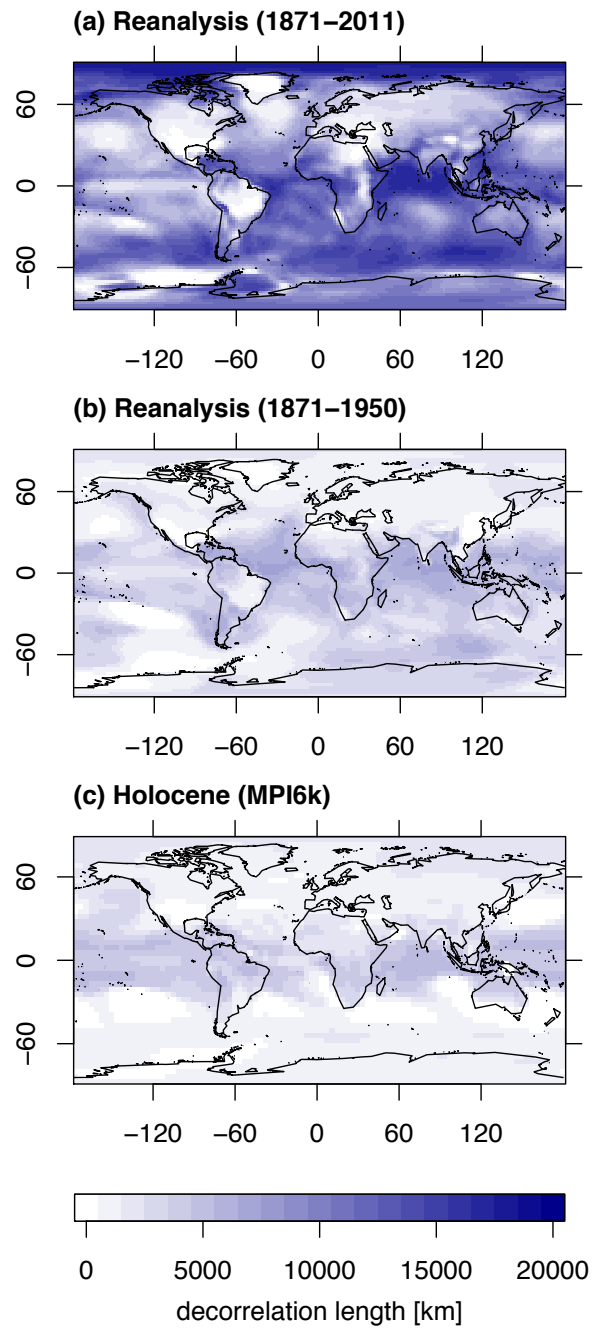


Fig.R1: Decorrelation length of reanalysis data and the 6ky simulation of MPI6k. The decorrelation length is similar for the Holocene and reanalysis data from 1871 to 1950 indicating that the Holocene spatial correlations are realistic.

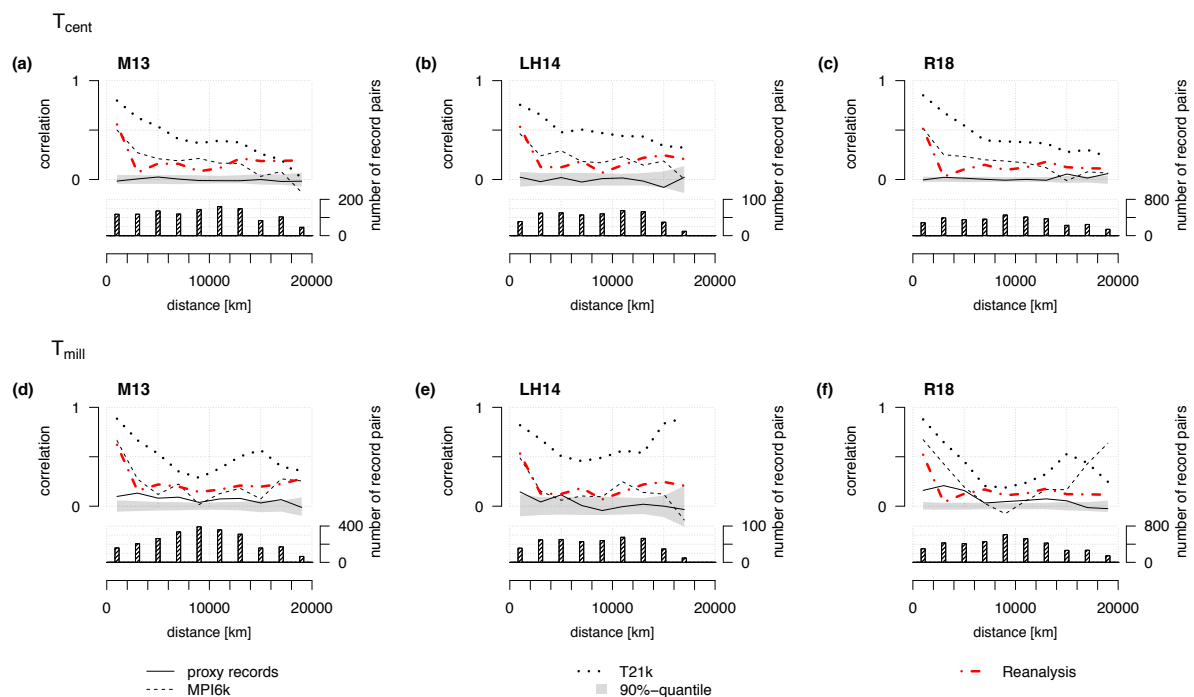


Fig.R2: Spatial correlation of reanalysis data for the time window from 1871 to 1950 (red lines). As the correlation over distance plots for the reanalysis data are very similar to the ones of the MPI6k this indicates that the spatial correlations of the model data are realistic.

2. Clarity on the separation of “multi-proxy syntheses” versus individual paleoclimate datasets and suggestions for improvement:

In their abstract, the authors state that “The estimated low signal content of Holocene temperature records should caution against over-interpretation of these kinds of datasets until further studies are able to facilitate a better characterisation of the signal content in paleoclimate records.” Here (and later on in their manuscript) the authors need to be very clear about what “these kind of datasets” mean. If they are implying that a broad-brush collation of datasets such as R18 or M13 is over-interpreted, I might agree with them that their analysis tends to demonstrate this aspect. However, this is untrue for a myriad of individual paleoclimate datasets (many of which are subsamples of aforementioned synthesis data sets) that are carefully vetted with high sensitivity to temperature and/or other variables such as precipitation, vegetation, salinity, productivity, etc. and more so, to seasonality - both aspects put together which are not addressed in this paper at all. I strongly recommend rewriting the above statement in the abstract as well as the final statement in the introduction (“more reliable interpretations of proxy records”; amongst other places) as it unnecessarily detracts from what the authors are proposing.

We agree with the reviewer and will be clearer in the abstract and conclusions to separate between multi-proxy datasets and individual paleoclimate datasets. Specifically, we will precise that ‘these kind of data’ are large multi-proxy and multi-site data compilations.

Such statements are also arguably misleading (e.g., modern monitoring and culturing will lead to far better interpretations of proxy datasets compared to estimates of SNR with a climate model) especially considering the point above that their analyses hinge on taking model output at face value.

We would argue that both methods, case studies such as modern monitoring, culturing, sediment traps, etc. as well as global statistical approaches such as used in this study, will lead to complementary information about the signal contained in the proxy records.

Case studies will be much more precise on the aspects they are analysing, but might omit other effects which are also present in the down-core record. Global statistical approaches include all effects influencing the down-core record, but suffer from the need to make (strong) assumptions.

Ideally, both methods converge to the same results giving credibility that the proxy system and its limitations are completely understood.

o The authors' work is a significant advance concerning model-data comparison. In its current version, suggestions on how model simulations or proxy development or the comparison of the two might be improved for better comparative metrics are lacking. I feel that some discussion on how their analyses might be developed further could be helpful.

We agree that this discussion would be useful and will add a new discussion section on implications and future steps forward.

Recent progress in computing power has enabled climate models to perform high resolution, often eddy permitting model simulations (e.g., HighResMIP project) and long simulations (>1000 year) are getting in reach. This is an important step to resolve the spatial scales and regions (mainly shelf areas and coasts) sampled by the proxies. Confronting these results with (replicated) sediment records, ideally accounting for seasonal/depth habitat using heuristic (Jonkers and Kucera, 2017) or complex ecological models (PLAFOM) would allow to better constrain the centennial spatial structures and climate variability as well as to refine the estimates of the proxy signal content shown in this study.

While our assumption of ignoring variations in the seasonal and depth habitat of the proxy recorders and the potential shortcoming in the current model correlation structure might have led to pessimistic SNR estimates, our results still underline the challenge of resolving the small Holocene temperature variations with current marine proxy records. Further improving our understanding of the proxy systems using modern monitoring, culturing and sediment traps and implementing this knowledge into ecological models (Jonkers and Kucera, 2017; PLAFOM) and proxy system models (Dolman and Laepple, 2018) is needed. Forward modelling the proxy records allows to better estimate the signal content and to optimise the sampling (e.g., replication of cores) and measurement process (e.g., sample size, number of foraminiferal tests). Although labour intensive, replicate records would allow to separate local climate variability from non-climate variability and thus provide an important step forward in understanding the proxy and climate variability.

Minor questions and comments:

- Perhaps I missed it, but why are there no counterpart plots to the T-cent in Fig. 1d-e shown in the main text for T-mill?

We agree that the naming of the figures was misleading. The aim of these figures is to provide a visual impression of the decorrelation lengths based on MPI6k and T21k. To only show one set of maps, we combined here both time scales (time scales larger than 400y, no detrending). The effect of the two time scales can be seen in Figure 3. We will change the nomenclature and describe this more clearly.

- Why does the correlation in T-mill with T21k (Fig. 3e) as well as with Uk'37 and Mg/Ca (Fig. 4b) show an uptick after 15000 KM distance?

This uptick is likely the result of the orbital forcing that is partly symmetric (effect of obliquity) and antisymmetric (precession) between the hemispheres. For the LH14 dataset (manuscript Fig.3e), the positive correlations at distances >15000km are between the tropics and the northern or southern hemisphere temperate zone as well as between sites of the northern and southern hemisphere temperate zone. There is only one time series pair with negative correlation.

- What are spatially important regions for proxy record development? Considering that the authors' work is specifically geared towards correlation distances, do their analyses pinpoint which regions are particularly data-deficient (e.g., Indian Ocean, South Atlantic, etc.) and would assist in their comparative metric?

There are some regions with a low number of sites such as the Southern Oceans. However, for this kind of study, more important than reducing the lack of single site data in these regions would be enhancing the number of replicate cores (= cores from nearby deployments that were subject to the same climate signal). This would allow to improve estimations of the signal content of proxy records and to test our understanding of proxy formation processes. This is shortly mentioned in 5.1, but we suggest to add this in the new section on implications and future steps forward.

- Is there any particular reason that the authors have not performed a similar analysis with the combined multiproxy datasets of R18, LH14, and M13?

The datasets were collected with a different focus (M13: reconstruction global temperature; LH14 and R18: temperature variability analysis) and currently use self-consistent, but different calibration and age-modelling approaches. Thus, we use them to test the sensitivity of the results on the choice of the dataset, but combining all datasets would necessitate recalibrations which is beyond our study.

- Again, I would suggest adding up front in the discussion that their analysis explicitly discounts the seasonality of proxies.

We will add that we neglected in our study the proxy-specific recording and especially seasonality. We will further discuss the effect of ignoring seasonality in more detail as suggested by Reviewer 1.

- Section 5.1: Is there a reference for anthropogenic forcing strongly increasing correlation decay length? Why necessarily, should this be the case? I feel there ought to be a statement explaining this here.

The correlation decay length observed in instrumental data (ignoring the last decades) and unforced models is largely consistent with a diffusive energy balance model (Kim and North, 1991) with increasing correlation lengths related to longer time scales (= more time to diffuse). In contrast, forced variability has a correlation length dominated by the spatial pattern of the forcing. For example, for a global forcing such as increasing greenhouse gases this leads to a globally coherent signal overlaying the internal climate variability.

This has been noted by Jones et al. (1997) and to some extent by Sutton et al. (2015), but to our knowledge there is no separate publication on this. However, it is clearly visible when analysing the decorrelation length of the surface temperature field in the reanalysis data (Compo et al., 2006) (Fig.R1). Focussing on the entire reanalysis time period results in a mean

decorrelation length of ~9150km. Contrary, analysing the time window from 1871-1950 the mean decorrelation length is ~3020km (Fig.R1), a finding consistent to the role of anthropogenic forcing.

We will add a short explanation and reference to Jones et al. (1997) in the revised manuscript.

- Although the Reschke et al. in review citation is provided, is there any reason for the 1/400y cut-off for the centennial time scale as opposed to something else?

Given a set of time series and their sampling resolution, the optimal cut-off frequency is the highest frequency that can be still resolved by the sampling without introducing a strong bias in the metric of interest, here the correlation.

Simulating surrogate records with the same sampling properties as the true records, Reschke et al. (2019) found that 1/400y is the optimal cut-off for a reasonably large subset of the data used in this study. Due to the Nyquist theorem, one needs at least 2 observations per period, and for typical non-equidistant paleo-data, four times the mean sampling frequency seems to be a rule of thumb appearing from several studies (Laepfle and Huybers, PNAS 2014; Reschke et al., 2019) although this will depend on the sampling properties and thus testing this individually using Monte Carlo experiments is the safest option.

Once again, thank you for your comments,
Maria Reschke

References:

- Compo, G. P., Whitaker, J. S., and Sardeshmukh, P. D.: Feasibility of a 100-Year Reanalysis Using Only Surface Pressure Data, *Bull. Amer. Met. Soc.*, 87, 175-190, 2006.
- Dolman, A. M., and Laepfle, T.: Sedproxy: a forward model for sediment-archived climate proxies, *Clim. Past*, 14, 1851-1868, doi:10.5194/cp-14-1851-2018, 2018.
- Jones, P. D., Osborn, T. J., and Briffa, K. R.: Estimating Sampling Errors in Large-Scale Temperature Averages, *J. Climate*, 10, 2548-2568, 1997.
- Jonkers, L., and Kucera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, *Clim. Past*, 13, 573-586, doi:10.5194/cp-13-573-2017, 2017.
- Kim, K.-Y., and North, G. R.: Surface Temperature Fluctuations in a Stochastic Climate Model, *J. Geophys. Res.*, 96(D10), 18573-18580, doi:10.1029/91JD01959, 1991.
- Laepfle, T., and Huybers, P.: Global and Regional Variability in Marine Surface Temperatures, *Geophys. Res. Lett.*, 41(7), 2528-2534, doi:10.1002/2014GL059345, 2014.
- Laepfle, T., and Huybers, P.: Ocean Surface Temperature Variability: Large Model-Data Differences at Decadal and Longer Periods, *P. Natl. Acad. Sci. USA*, 111(47), 16682-16687, doi:10.1073/pnas.1412077111, 2014.
- Reschke, M., Kunz, T., and Laepfle, T.: Comparing methods for analysing time scale dependent correlations in irregularly sampled time series data, *Comp. Geosci.*, 123, 65-72, doi:10.1016/j.cageo.2018.11.009, 2019.
- Sutton, R., Suckling, E., and Hawkins, E.: What Does Global Mean Temperature Tell Us about Local Climate?, *Phil. Trans. R. Soc. A*, 373(2054), 20140426, doi:10.1098/rsta.2014.0426, 2015.

Empirical estimate of the signal content of Holocene temperature proxy records

Maria Reschke^{1,2}, Kira Rehfeld^{1,3}, Thomas Laepple¹

¹Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Telegrafenberg A45, 14473 Potsdam, Germany

5 ²Institute of Earth and Environmental Sciences, University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany

³Institut für Umweltphysik, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany

Correspondence: M. Reschke (mreschke@awi.de)

10 **Abstract.** Proxy records from climate archives provide evidence about past climate changes, but the recorded signal is affected by non-climate related effects as well as time uncertainty. As proxy based climate reconstructions are frequently used to test climate models and to quantitatively infer past climate, we need to improve our understanding of the proxy records' signal content as well as the uncertainties involved.

In this study, we empirically estimate signal-to-noise ratios (SNRs) of temperature proxy records used in global compilations of the mid to late Holocene (last 6000 years). This is achieved through a comparison of the correlation of proxy time series from close-by sites of three compilations and model time series extracted at the proxy sites from two transient climate model simulations, a Holocene simulation of the ECHAM5/MPI-OM model and the Holocene part of the TraCE-21ka simulation.

In all comparisons, we found the mean correlations of the proxy time series on centennial to millennial time scales to be low ($R < 0.2$), even for nearby sites, which resulted in low SNR estimates. The estimated SNRs depend on the assumed time uncertainty of the proxy records, the time scale analysed, and the model simulation used. Using the spatial correlation structure of the ECHAM5/MPI-OM simulation, the estimated SNRs on centennial time scales ranged from 0.05 - assuming no time uncertainty - to 0.5, for a time uncertainty of 400 years. On millennial time scales, the estimated SNRs were generally higher. Use of the TraCE-21ka correlation structure resulted generally in lower SNR estimates than for ECHAM5/MPI-OM.

25 As the number of available high-resolution proxy records continues to grow, a more detailed analysis of the signal content of specific proxy types should become feasible in the near future. The estimated low signal content of Holocene temperature compilations should caution against over-interpretation of these multi-proxy and multi-site syntheses until further studies are able to facilitate a better characterisation of the signal content in paleoclimate records.

Gelöscht: data

Gelöscht: two

Gelöscht: Holocene

Gelöscht: climate model simulations

Gelöscht: rather

Gelöscht: the SNRs

Gelöscht: model simulation

Gelöscht: records

Gelöscht: kinds of datasets

1 Introduction

Improving our understanding of the climate systems and its variability requires knowledge about the climate of the pre-instrumental period. Proxy records from different climate archives are available for determining past climate conditions (e.g., Bartlein et al., 2011; Hugué et al., 2006; Johnsen et al., 2001; Li et al., 2006; Luckman et al., 1997). However, as any observational estimate, paleoclimate proxies are affected by uncertainties (e.g., Breitenbach et al., 2012; Lohmann et al., 2013).

The signal that can be retrieved from paleoclimate archives record depends on various temporal (seasonal recording, dating), geological (mixing, transport, sorting), biological (life-time of organisms, habitat depth, bioturbation), and chemical (preservation and dissolution) processes (e.g., Bard, 2001; Berger and Heath, 1968; Goreau, 1980; Leduc et al., 2010; Lohmann et al., 2013; Mollenhauer et al., 2003; Ohkouchi et al., 2002; Rehfeld et al., 2016; Rosell-Melé and Prah, 2013; Schneider et al., 2010; Telford et al., 2004; van Sebille et al., 2015).

Therefore, the proxy variations do not only contain the climate signal of interest (e.g., annual mean temperature), but also other climatic influences as well as non-climate variability. This poses a challenge to the interpretation of proxy signals, especially in systematic model-data comparisons and quantitative data synthesis efforts. Different approaches have been proposed in an effort to alleviate this problem and improve analyses:

- (i) obtain a better statistical or mechanistic understanding of how and what a proxy actually records (e.g., Fisher et al., 1985; Grauel et al., 2013; Ho and Laepple, 2016; Münch et al., 2016, 2017; Richey et al., 2011; Rosén et al., 2003; Thirumalai et al., 2013);
- (ii) modelling of the proxy signal (e.g., Dee et al., 2011, 2015; Dolman and Laepple, 2018; Evans et al., 2013; Roche et al., 2018); and
- (iii) through a detailed, expertise-driven analyses of single sites (e.g., Stebich et al., 2015).

In this study, we use a comparison of proxy records and model simulations to improve the characterisation of proxy uncertainties through empirical estimates of the signal-to-noise ratio (SNR) in temperature-related proxies. At present, studies on SNRs in proxies are rare and mainly focussed on the instrumental period (e.g., Mann et al., 2007, 2008; Smerdon, 2012, Münch and Laepple, 2018). In contrast, the present study focusses on the pre-instrumental Holocene period which has received considerable attention in the community (e.g., Bakker et al., 2017; Gajewski, 2015; Mangerud and Svendsen, 2018; Marcott et al., 2013; Mischel et al., 2017; Moossen et al., 2015; Sejrup et al., 2016; Thibodeau et al., 2018; Wanner et al., 2015; Zhang et al., 2017). In particular, we focus on estimating SNRs in temperature-sensitive proxy records to improve analyses of Holocene temperature evolution and variability. A better understanding of Holocene proxy time series SNRs will lead to improved and more reliable interpretation of proxy records in multi-proxy and multi-site data compilations and should raise awareness for the need of careful and critical evaluations of paleoclimate reconstructions.

Gelöscht: P

Gelöscht: proxies

Gelöscht: very specific climate information that depends on

Gelöscht: In addition,

Gelöscht: the information of interest can be obscured by co-recorded non-climatic components which represents one possible cause for misinterpretation of

Gelöscht: s

2 Data

This study builds on existing compilations of recalibrated high-resolution Holocene temperature-sensitive proxy records to facilitate intercomparison of multiple time series. The analysis is based on three proxy datasets and two model simulations to test the robustness of our results and the sensitivity to the choice of a particular climate model.

Gelöscht: s

2.1 Proxy records

We focus on globally distributed multi-archive and multi-proxy compilations of the Holocene temperature evolution from a wide variety of locations (Fig. 1a-c, Tab. 1; Tab. S1-S3), namely

- (1) M13: the compilation of Marcott et al. (2013) that was used to reconstruct the global and regional temperature evolution of the past 11.3ky;
- (2) LH14: Uk37 and Mg/Ca proxy data compiled in the extended dataset of Laepple and Huybers (2014a) that was used to reconstruct regional temperature variability and builds on the compilation of Leduc et al. (2010);
- (3) R18: the Holocene part of the compilation of Rehfeld et al. (2018) that was used to compare Glacial and Holocene temperature variability.

The datasets mostly originated from marine sediment cores and the proxy types include Uk37, planktonic foraminifera Mg/Ca, TEX₈₆, terrestrial bio-indicators (fossil pollen modern analogue technique, fossil chironomid transfer function), ice-core stable isotopes ($\delta^{18}\text{O}$, $\delta^2\text{H}$) and several others. As the early Holocene was influenced by deglaciation following the Last Glacial Maximum (e.g., Kaplan and Wolfe, 2006), we restricted the time series to the last 6ky (6ky BP to present day, where BP denotes years before 1950). We only analysed time series containing climate information on at least centennial to millennial time scales (i.e., a mean inter-observation time step of $\Delta t < 500\text{y}$). Due to the limited number of available high-resolution time series, the datasets overlap (Tab. 1) to some degree and are thus not independent.

Gelöscht: (microfossil assemblages, MBT, tree ring width)

Gelöscht: all

2.2 Climate model simulations

We analysed surface air temperature data from simulations of two coupled atmosphere-ocean general circulation models: A 6ky transient Holocene simulation from ECHAM5/MPI-OM (henceforth abbreviated as MPI6k) (Fischer and Jungclauss, 2011) and the TraCE-21ka (T21k) (Liu et al., 2009) simulation from the CCSM3 model, both of which have been used frequently in recent studies (e.g., Gregoire et al., 2016; Heinemann et al., 2009; Koldunov et al., 2010; Lu et al., 2018; Matei et al., 2012; May, 2008; Müller and Roeckner, 2008; Pausata and Löfverström, 2015; Werner et al., 2016). For the present study, annual means of temperatures of both model simulations were extracted at the nearest grid-box related to the proxy record locations of M13, LH14 and R18. Our choice of annual means is consistent with the standard interpretation of these multi-proxy datasets to represent annual mean temperatures (Marcott et al., 2013). This interpretation is a pragmatic choice motivated by the lack of accurate information about the proxy and location-specific seasonality across all records forming such a multi-proxy dataset.

Gelöscht: 4

MPI6k (Fischer and Jungclauss, 2011) is a 6ky transient run using ECHAM5/MPI-OM (Jungclauss et al., 2006) which consists of the atmosphere component ECHAM5 (Roeckner et al., 2003), the ocean component MPI-OM (Marsland et al., 2003), and the land surface model JSBACH (Raddatz et al., 2007) with dynamic vegetation module (Brovkin et al., 2009). The model outputs atmospheric variables on a regular longitude/latitude model grid with 96 by 48 horizontal grid-boxes (T31 resolution corresponding to 3.75° in latitude/longitude). The simulation is forced only orbitally with greenhouse gas concentrations set to pre-industrial values. We extracted annual mean surface temperatures at an elevation of 2m from this model ([model variable temp2](#)).

The TraCE-21ka dataset (Liu et al., 2009) [is](#) originated from a simulation of the transient climate between 22ky BP and 1990 CE [and](#) based on a fully-coupled CCSM3 simulation with an atmospheric resolution of T31_gx3 (96 by 48 horizontal grid corresponding to 3.75° in latitude/longitude). Transient forcing factors [in the time-period analysed here \(last 6ky BP\) are changes in the orbitally driven insolation, greenhouse gas concentrations and the meltwater fluxes for the southern hemisphere in the period earlier than 5ky BP](#). Our analysis is independent of the absolute changes and only relies on the simulated spatial correlation structure. For the time scales analysed and the proxy positions of our compilations, this correlation structure is not sensitive to the particular choice of temperature variable (sea surface temperature versus surface temperature or near surface air temperature) [in either model](#).

Gelöscht: is

Gelöscht: included

Gelöscht: , continental ice sheets and coastlines, and meltwater fluxes across the North Atlantic and Gulf of Mexico

Gelöscht: For the 6ky BP period, the changes in orbital insolation and changes in the greenhouse gas concentrations represent the most important forcings while changing continental ice sheets and meltwater fluxes are only of minor importance.

Gelöscht: of the variability of annual surface temperatures

3 Method

3.1 Approach and assumptions

SNRs can be estimated by comparing proxy records that experienced the same or very similar climate signals, e.g., different proxies from the same site or the same proxy from different sites in close spatial proximity. If a pair of records contains the same signal, an independent local noise component, and no time uncertainty, the SNR is given as $R/(1 - R)$ where R is the correlation between both time series (Fisher et al., 1985). Ideally, SNRs would be estimated from local replicates. This is often difficult, or impossible, due to the limited availability of replicated datasets. To increase the number of records and thus improve the robustness of estimates, we extended this approach to also include records from locations that are further apart. This increased spatial separation between sites requires knowledge of the signal covariance (as the climate signal will have been slightly different at each location) and we rely on climate models to provide this information.

The underlying assumptions are thus: (1) when relying on model data, we must assume correctness of the model-based correlation structure; (2) when using different proxies, we must assume that all proxies recorded the same temporal (and spatial) variability of the climate signal (more specifically: annual mean surface temperature); (3) we must assume that differences in the spatial correlation structure between models and proxy observations are due solely to a site-independent additive noise and time uncertainty. [With assumption \(2\) we discount the seasonality of proxies in this study, but discuss the effects of this strong assumption in section 5.3.](#)

Based on these assumptions, we can estimate the SNR by matching the spatial correlation of proxy records and model time series while accounting for time uncertainty and additive noise which can both lead to a deterioration in the spatial correlation. E.g., low correlations among time series can be caused by both: a low time uncertainty in combination with a high noise level and a high time uncertainty in combination with a high SNR (low noise level). Due to this relationship, we quantify SNR estimates as a function of time uncertainty.

Gelöscht: eir

Sites that are very far apart only share a weak climate signal which does not represent any constraint on the SNR as both the climate and proxy correlations will be close to zero. For our SNR estimate, we therefore only included proxy pairs with spatial separations of up to 5000 km, which we found to be a typical decorrelation distance on centennial time scales in the model simulations as we later show.

As climate variability is a function of time scale, we expect that both the spatial correlation structure and SNR will also be time scale dependent. However, the limited number of records and samples in each record prevents a more thorough time scale dependent estimate which could be carried out using a spectral approach for instance (Münch and Laepple, 2018). In order to balance accounting for time scale and estimate robustness, we distinguish between a centennial time scale T_{cent} (with a cut-off frequency of $1/400y$ and by removing the linear trend of the time series) and a centennial to millennial time scale T_{mill} (using a cut-off frequency of $1/1000y$ and including the trend). To estimate T_{cent} , we only used records with a mean sampling interval of less than 200y while all records were included for estimating T_{mill} .

Gelöscht: (cf., Reschke et al., in review)

3.2 Spatial correlation structure of model vs. reanalysis data

As our study depends on the model-based correlation structure, we first analyse this correlation structure at the grid cell level by fitting an exponential $R = e^{-x/l_d}$ to the decay of correlations R as a function of site separation x for time scales larger than 400y with included trend (Fig. 1d, e). We further compare the simulated spatial correlation structure with the spatial correlation structure estimated from reanalysis data using the same method. For this aim, we analyse the annual mean surface temperature field of the 20C3M reanalysis (Compo et al., 2006) (Fig. 1f).

Analysing the entire reanalysis period from 1871 to 2011 results in a high estimate of the mean correlation decay length l_d (~9150 km) that is considerably larger than the correlation decay length found in the MPI6k Holocene model simulation (~2240 km) when analysing the same time scale (unfiltered annual data) as for the reanalysis data. Reducing the human influence (i.e., anthropogenic forcing) by analysing 1871 to 1950 reduces the correlation decay length and results in a similar estimate (~3020 km) than the annual estimate from the MPI6k Holocene simulation (Fig. S1). This result indicates that the model correlation decay lengths used in this study are not unrealistically large and the larger centennial (Fig. 1) than interannual (Fig. S1) decay lengths are consistent with the expectation that temperature fields on longer time scales are more spatially coherent (e.g., Jones et al., 1997; Kim and North, 1991). The general similarity between the model correlations and the correlations in the reanalysis data also holds when we only compare the correlation between the proxy sites (Fig. 3), suggesting that similar conclusions could be also drawn when using the reanalysis correlation structure instead of the model simulations.

3.3 Processing steps

Gelöscht: 2

3.3.1 Estimation of the spatial correlation structure

Gelöscht: 2

From the MPI6k and T21k model time series we extracted annual mean temperatures at those grid cells that contain the location of the proxy record site. As our aim is deriving a time series from the annual model time series that resembles the proxy time series in having the same number and ages of the proxy observations, we apply block averaging. To get a data point for the observation time t_i we average all observations between half the difference to the previous observation time ($t_i - \Delta t_i/2$) and half the difference to the next observation time ($t_i + \Delta t_{i+1}/2$). We chose to use averages rather than interpolation because sediment and ice samples, in particular, often include adjacent depths or have a sample distance that is smaller than the typical mixing depth in the sediment (Berger and Heath, 1968) or diffusion length in ice-cores. For each proxy record pairs (M13, LH14, R18), we estimated the time scale dependent (T_{cent} , T_{mill}) correlations between all possible proxy record pairs. We further estimated the time scale dependent correlations between all model time series pairs. For this step, the irregularly sampled time series were linearly interpolated onto a regular grid ($\Delta t = 10y$) and subjected to a Gaussian filter with cut-off frequency $1/400y$ (T_{cent}) and linear detrending or alternatively to a Gaussian filter with cut-off frequency $1/1000y$ (T_{mill}) omitting the detrending step. This approach has been shown to deliver good results for the estimation of time scale dependent correlations in tests using surrogate data with the sampling properties of Holocene marine sediment cores (Reschke et al., 2019).

Gelöscht: Through block averaging, data were then resampled at the same ages as in the corresponding proxy records.

Gelöscht: dataset

Gelöscht: and

Gelöscht: more

Gelöscht: for an irregularly sampled dataset with regard to

Gelöscht: in review

The spatial separation between two sites was used to place the pair into 2000 km-sized bins (thus containing separations of 0-2000km, 2000-4000km, etc.) and averaging the correlations from proxy/model site pairs contained within the same bin. An overview of the processing steps is given in Fig. 2.

We performed a significance test of the spatial correlation structure based on spatially uncorrelated surrogate time series with a temporal power-law scaling of $\beta = 1$, which is a typical value for Holocene sediment records (Laepple and Huybers, 2014a). In a Monte Carlo procedure with 1000 repetitions, we generated annual surrogate records that were analysed using the same procedure as the true proxy observations, using the 90% quantile of the binned correlations of the surrogate time series as confidence intervals.

3.3.2 Estimation of the SNRs

Gelöscht: 2

The SNR estimate was obtained from a Monte Carlo simulation with 1000 repetitions. Through block averaging, we resampled the annual model data at the same resolution as the corresponding proxy records. We then added time uncertainties (between 0 and 400y) and noise levels ($0.01 < SNR < 100$), before estimating the mean correlation using the interpolation method of Reschke et al. (2019). We estimated the SNRs as a function of the time uncertainty by minimising the absolute difference of the mean correlations of proxy records and modified model simulations.

Gelöscht: in review

We generated the modified model data by separately distorting the time axis and adding noise to the observations of the resampled model time series. As a simple heuristic to simulate time uncertainty, we defined four time control points at the

ages of 1y, 2ky, 4ky, and 6ky and randomly shifted these points by adding a random value from a normal distribution (mean $\mu = 0$, standard deviation $\sigma =$ time uncertainty [y]) except for the value of 1y. The new time axis was then created by linearly interpolating between the time control points. Noisy observations were generated by adding normally distributed noise, ε (with $\varepsilon \sim N$, mean $\mu = 0$ and variance σ^2 as $\frac{\sigma_{\text{model, resampled}}^2}{\text{SNR}}$), to the resampled model time series. Fig. 2 gives an overview of the processing steps.

4 Results

4.1 Spatial correlation structure and correlation decay length

The correlation analysis using all proxy types and locations yielded, unsurprisingly, a general decrease in correlation for larger spatial separations between proxy sites (Fig. 3). Both model simulations exhibit statistically significant spatial correlations at both analysed time scales (T_{cent} and T_{mill}) and for most inter-site separation distances. Throughout all datasets and separation distances, T21k yielded higher correlations than MPI6k, which is consistent with the generally higher correlation decay lengths l_d for T21k, estimated at grid cell level (Fig. 1d, e).

While for T_{cent} the correlation of both model simulations decreases with increasing site separation (Fig. 3a-c), the T_{mill} estimate (Fig. 3d-f) shows a more complex pattern that includes a partial increase in correlation for separation distances larger than 8000 km. This is likely related to variations in orbital forcing affecting the temperature trend that is partly symmetric (effect of obliquity) and antisymmetric (precession) between both hemispheres. Especially for MPI6k, the correlation is weak for separation distances from 4000 to 6000 km.

The spatial correlations obtained from the proxy records differ systematically from those obtained from model simulation data. The mean correlation for close proxy site pairs (separation < 5000 km) was 0.004 to 0.014 for T_{cent} and 0.101 to 0.186 for T_{mill} and thus lower than for model data (MPI6k: 0.303 to 0.338 for T_{cent} , 0.202 to 0.461 for T_{mill} ; T21k: 0.634 to 0.719 for T_{cent} , 0.674 to 0.710 for T_{mill}). For T_{cent} , none of the proxy based correlations is statistically significant and no clear pattern emerges with regard to separation distance. All three datasets yielded a statistically significant correlation at T_{mill} for smaller separation distances, although visibly decreasing for longer separation distances (e.g., 6000-8000 km; cf. Fig. 3d, f).

Comparisons of temperature estimates from different proxy types face the additional challenge that the actual recorded variable (e.g., summer atmospheric temperature vs. mixed layer winter temperature) may depend on the proxy type. We therefore also analysed the proxy-specific results (Fig. 4, Tab. 2). By performing separate analyses for each proxy type (instead of analysing all proxies together) we obtained in all three datasets a higher mean correlation on the T_{mill} time scale for sites within a 5000 km range. For T_{mill} , the proxy-specific mean correlations across all datasets and proxies are between 0.149 and 0.357 compared to 0.101 to 0.186 when correlating sites across proxy types. For T_{cent} , most correlations are indistinguishable from zero and we observed no consistent increase when analysing proxy-specific correlations (Tab. 2). Unfortunately, restricting the analysis to a single proxy type greatly reduces the number of available proxy pairs at any given

Gelöscht: by fitting an exponential, $R = e^{-x/l_d}$, to the decay of correlations R as a function of site separation x

Gelöscht: are

Gelöscht: d with

distance and thus leads to less robust correlation estimates and rather large confidence intervals. We therefore only provide results for the most data-abundant proxy types Uk37 and Mg/Ca and one dataset (LH14) as example in the main manuscript (Fig. 4). The remaining data are shown in the Supplement (Fig. S2-S5). For LH14, both Mg/Ca and Uk37 show a decrease in correlation with increasing separation distance for both time scales. The correlations in this proxy-specific analysis are stronger than the analysis across proxy types (Fig. 3). They are, however, only statistically significant for Uk37 on T_{mill} with separation distances smaller than 5000 km and for a single distance bin (2000-4000km) for Mg/Ca.

Gelöscht: 1
Gelöscht: 4

4.2 SNR estimates

The estimated SNRs of proxy records are a function of time uncertainty because correlations deteriorate due to both, time uncertainty and noise. In general, we found that low (high) SNRs were related to low (high) time uncertainties (Fig. 5). In most cases, the estimated signal content for Holocene temperature-sensitive proxy records was quite low (<0.5).

By using the spatial correlation structure of MPI6k and assuming a time uncertainty (1 sd) of 220y (mean uncertainty in M13) we obtain an estimated SNR of between 0.05 and 0.2 for the T_{cent} time scale and 0.2 for the M13 and R18 datasets on the T_{mill} time scale. The LH14 dataset yielded a SNR of 10 at the T_{mill} time scale.

For all three proxy compilations (M13, LH14, R18) the SNRs obtained for mixed proxy types depend on the choice of the model simulation. Using the T21k simulation generally leads to lower SNR estimates (T_{cent} : $SNR_{T21k, T_{cent}} < 0.05$; T_{mill} : $0.05 < SNR_{T21k, T_{mill}} < 0.2$) than using MPI6k as the correlation of spatially close (separation <5000 km) time series pairs is generally higher in T21k. Interestingly, the SNRs estimated using T21k are more similar between the three proxy compilations and thus more consistent than using MPI6k (Fig. S6).

Gelöscht: The
Gelöscht: and
Gelöscht: differ for the
Gelöscht: data
Gelöscht: Generally, the SNRs were more consistent with T21k compared to MPI6k, although lower due to higher correlations of spatially close (separation <5000 km) pairs (Fig. S5). In addition, the SNRs are lower for
Gelöscht: (
Gelöscht:) compared to
Gelöscht: (
Gelöscht: 6
Gelöscht: 5

An analysis of proxy-specific SNRs yielded higher uncertainties due to the relatively small number of record pairs and potentially caused statistically non-robust estimates for some proxy types (see Fig. S7-S16 for the complete set of results and section 5.2 for a sensitivity test of SNR estimates on the number of record pairs). The dependence of SNR estimates on time uncertainty is very sensitive to how the proxies are compiled and the type of model simulation. However, the overview of all proxy-specific SNR estimates (Fig. 6) suggests some proxy-specific tendencies. On T_{cent} ice-cores show the highest SNR. Mg/Ca shows a high SNR for the LH14 dataset but a low SNR in the two other compilations. Uk37 and terrestrial bio-indicators have the lowest SNR estimate on this time scale. In contrast analysing the T_{mill} time scale that also includes trends in the dataset leads to different results; Uk37 shows the highest SNRs whereas the other proxy types only show a small increase compared to the T_{cent} analysis.

5 Discussion

High-resolution temperature-sensitive proxy records for the Holocene are sparse, irregularly distributed, and from different proxy types. Thus, estimating the SNR in such datasets requires some simplifying assumptions. We assumed that: (1) the spatial correlation of the climate model simulations was realistic, (2) all proxy types were recording the same climate

Gelöscht: rely on

variable, and (3) any non-climatic components of the proxy signal can be fully accounted for through a combination of time uncertainty and additive noise. As we analysed large multi-proxy and multi-site datasets, in our study we neglected the proxy-specific effects such as seasonality in the recording.

The SNRs we estimated, based on these assumptions, generally suggest a low signal content of Holocene temperature records on centennial time scales (T_{cent}). We found a higher signal content on millennial time scales (T_{mill}), but the results were rather sensitive to the choice of the proxy compilation and model simulation. We now discuss the role of the different assumptions on the results.

5.1 Spatial correlation structure of model simulations

Our SNR estimates critically depend on the model-based temperature correlation structure as lower spatial temperature correlations in the models would lead to higher SNR estimates for the proxies and vice versa. In most regions, the model simulation MPI6k shows correlation decay lengths of 1295 to 6030 km (mean decay length: 3995 km) and the correlation decay length of T21k is generally in the range of 2130 and 8705 km (mean decay length: 5920 km) for time scales larger than 400y with included trend (Fig. 1d, e). This is higher than previous estimates of correlation decay lengths from instrumental datasets in the range of 1000 to 3000 km (e.g., Hansen and Lebedeff, 1987; Jones et al., 1997; Madden et al., 1993). However, such a difference is plausible as an increase with time scale is to be expected. For example, Jones et al. (1997) found lower correlation decay lengths related to annual (2100 km) than to decadal (3800 km) time scales. Indeed, when calculating the correlation decay length for MPI6k on unfiltered annual data, it is consistent to the decay length from instrumental data (Jones et al., 1997) as well as from reanalysis data (Fig. S1).

Nevertheless, spatial correlation could be overestimated in the model simulations for two reasons. Firstly, the spatial correlation of instrumental datasets always includes the anthropogenic forcing which strongly increases the correlation decay length (see Fig. S1 and Jones et al., 1997). This effect is not or only weakly present in the 6ky time-period of our analysis. Instrumental records from the industrial period and pre-industrial model simulations might thus be in agreement for the wrong reasons. Secondly, the grid cell size of the models was on the order of several hundred kilometres whereas the records might be representative of a smaller spatial area. Hence, it is possible that proxy based correlations are lower compared to those obtained from the model due to the former being influenced by subgrid-scale temperature variations. Thirdly, there are several shortcomings in present climate model simulations potentially causing an overestimation of the coherency in the two simulations used in this study. One possibility is that models underestimate internal climate variability that is generally more localised than externally forced climate variability (Laepplé and Huybers, 2014a). One mechanism could be a too large effective horizontal diffusivity in the models that would reduce internal variability (Laepplé and Huybers, 2014b) and cause larger spatial correlation structures. Further, small scale features and the role of persistent coastal currents might be suppressed by the relatively low, non-eddy permitting resolution of the models used in this study.

We also found that T21k yielded higher spatial correlations compared to MPI6k (Fig. 1d, e) which in turn resulted in lower SNR estimates if relying on this particular model simulation (Fig. S6). This difference might be related to the presence of

Gelöscht: the T_{cent}

Gelöscht: but

Gelöscht: 5

transient greenhouse gas forcing in T21k (Timm and Timmermann, 2007), although the changes in forcing were small during the analysed time period.

Thus, there remains the possibility that the true temperature variations are more localised than suggested by the model simulations. In this case our estimates of the proxy signal content would be pessimistic. Ultimately, more replicate proxy records are needed to distinguish between these hypotheses.

5.2 Finite number of proxy records

Despite the strong overlap among records, we found our estimates of the spatial correlation structure and SNRs to be sensitive to the choice of proxy compilation (Tab. 2), which suggests that the number of records N may have been limiting the robustness of our estimates. To test this, we performed a sensitivity analysis using different numbers N of surrogate time series ranging from 3 to 50. N 6ky annual surrogate time series were generated from the sum of a common pseudo climate time series modelled as a random process that follows a power-law ($\beta = 1$) scaling and a separate non-climate component that is simulated as uncorrelated white noise. The noise amplitude is chosen to yield $SNR = 0.15$. Irregular sampling times were used to mimic the observed sampling times of the M13 records. Surrogate inter-observation time steps were drawn from a gamma distribution (shape = rate = 2.25), rescaled with a mean inter-observation time step of 108.56y (cf., Reschke et al., 2019). The final pseudo proxy time series were then obtained by block averaging the annual time series to the irregular sampling times. The SNR of the surrogate time series were then calculated following the same method as the proxy records in the main study and repeated for different sites using a Monte Carlo-based procedure with 2000 repetitions.

We found that the uncertainty of SNR estimates that are based on a small number of records can be high (Fig. 7). For a low number of only 15 records (105 correlation pairs), for instance, the uncertainty range of SNRs (90% quantiles of 0.08 to 0.26) is higher than the true SNR value of 0.15. Although we used more than 15 sites per compilation in our analysis (Fig. 7), there were often fewer than 15 time series per proxy type (Tab. 1) which might explain the strong scatter in the proxy-type-specific SNR estimates.

To improve the robustness of SNR estimates, it is unavoidable to significantly increase the number of records that are collected not too far apart from one another (distances < 5000 km). Additionally, a better global coverage of site locations would likely lead to more robust results. Since we sampled the models at the locations of the proxy sites our results should be independent of the spatial sampling distribution if the models were perfect. In reality, however, spatial differences and shifts in the simulated correlation structure are likely and can be overcome by sampling from a wide variety of sites from all over the globe.

5.3 Proxy-specific recording of climate variables

All proxy types used in this study have been reported in the literature as temperature-sensitive and are usually calibrated to the mean annual surface air or surface water temperature. However, this is a gross oversimplification as the true climate variable influencing the recorded signal is proxy-specific and generally more complex. For example, signals reconstructed

Gelöscht: hypothesis

Gelöscht: in review

Gelöscht: s

from marine organism-based proxies such as Mg/Ca, Uk37, and TEX₈₆ are affected by the seasonal and depth-specific preferred habitat of the organism (Ho and Laepple, 2016; [Jonkers and Kucera, 2017](#); Leduc et al., 2010; Lohmann et al., 2013; [Tierney and Tingley, 2015](#)). As we currently interpret all records from different proxy types as annual mean surface temperatures, this might influence our results in various ways. Analysing different proxy types with different recording preferences likely leads to an underestimation of the spatial temperature correlations. Indeed, in our study we found the spatial correlations related to records of the same proxy type for T_{mill} to be higher compared to those for all types (Tab. 2). To gain a better understanding of proxies and their effect on the analyses, we suggested to use proxy-specific SNR estimates instead. However, this is currently hampered by the low number of records in close proximity to one another (Tab. 1; Tab. S4). For many proxy types, this leads to statistically non-significant correlations and unreliable SNR estimates (Fig. S7, S16). Additionally, even for one proxy type and proxy carrier (e.g., foraminifera) we expect a site-specific season and depth habitat. Such differences would reduce the correlation compared to the correlation of the climate component sampled at any globally fixed season or depth and would thus bias the SNR estimates low. Assuming annual mean sea surface temperatures instead of one specific season or depth also influences the correlation structure derived from the models. Calculating the correlation structure of summer and winter in both models (not shown) suggests an increase or decrease of the correlation depending on the choice of the model so that the net-effect on the SNRs is not clear. Finally, analysing the spatial correlation among records of the same proxy type can also lead to overly optimistic results as the correlation among records of the same proxy type could also stem from spatially correlated proxy-specific non-climatic components. A case in point would be the dissolution of foraminiferal shells (Lea, 2003) which could generate spatially correlated noise as the preferential dissolution of carbonate depends on the water depth (Brown and Elderfield, 1996; Dekens et al., 2002), the carbonate ion concentration, and the salinity of the surrounding seawater (Huguet et al., 2006; Lea, 2003; Spero et al., 1997).

Gelöscht: T
Gelöscht: he consequence is a temporal and spatial averaging of the climate signal. ... [1]

Gelöscht: 6
Gelöscht: 5

Gelöscht: .
Gelöscht: On the other hand

5.4 Time uncertainty and non-climatic components of the proxy signal

Our SNR estimates depend on the assumed time uncertainty of the records. While we assumed a mean time uncertainty of 220y (as provided in the M13 dataset), the true time uncertainty for marine records might be considerably higher due to spatially varying reservoir effects (Ascough et al., 2005). This would imply that our SNR estimates are conservative, especially on centennial time scales. On the other hand, using one mean uncertainty value will clearly be too pessimistic for ice-core data that is only subject to much smaller dating uncertainties. Using more sophisticated models to account for time uncertainty (e.g., Blaauw, 2010; Blockley et al., 2007; Telford et al., 2004) and the proxy- and site-specific information on the chronologies would allow to obtain more precise SNR estimates.

We modelled the transfer function between the temperature time series and the calibrated proxy records as a combination of time uncertainty and additive temporally uncorrelated noise. Our approach thus neglects other distortions of the signal and non-additive parts of noise. Multiplicative noise can arise from aliasing due to subsampling that leads to errors that are

proportional to short-term climate variability (Laepple and Huybers, 2013). Variable sedimentation rates, bioturbation and/or bioturbation depths varying over time have a low-pass filtering effect that is similar to irregular sampling. Proxy archive accumulation processes undergo temporal changes due to changes in bioturbation depths and advection (Mollenhauer et al., 2003) or spatial changes in ocean currents (van Sebille et al., 2015) that could introduce additional non-additive noise in the obtained proxy records. Finally, even for a single proxy type, the data quality (i.e., the signal content) is site specific and will depend on the sampling and measurement protocol. For example, the SNR estimates using the LH14 dataset that is mainly based on very high-resolution records (mean sample distance <100y), are higher than estimates based on the two larger proxy compilations. Thirumalai et al. (2018) showed that foraminiferal records based on a large number (70-100) of foraminiferal tests per sample were consistent between cores collected in close proximity to one another leading to much higher correlations compared to our study. As we rely on datasets of opportunity that consist of proxy records measured by various labs over a period of two decades, it seems conceivable that a small number of records could be of a relatively lower quality which would reduce our mean correlation and thus the SNR estimate. New studies, especially when based on a careful design (Thirumalai et al., 2018), could help alleviate this situation.

5.5 Implications and future steps forward

Our results underline the challenge of resolving the small Holocene climatic variations in current climate archives, but also challenge the strong spatial coherency of centennial to millennial temperature variations simulated in current climate models. On the proxy side a continuation of the work on understanding the proxy systems is warranted. Examples are the use of modern monitoring systems, sediment traps and culturing studies. Implementing these findings into ecological models of various complexity (e.g., Jonkers and Kucera, 2017; Kretschmer et al., 2018) and proxy system models (e.g., Dolman and Laepple, 2018) is needed to generalise the knowledge and make it usable in global studies. Forward modelling of proxy records will allow estimates of the signal content complementing the empirical estimates provided here. Finally, a better proxy understanding implemented in proxy system models will also allow to optimise the sampling (e.g., sampling and replication strategy) and measurement process (e.g., number of foraminiferal tests). Finally, although labour intensive, a more frequent generation and analysis of replicate records would allow to separate local and climate and non-climate variability and thus provide a key step in understanding the proxy and climate variability as well as the proxy formation process.

Progress in climate modelling is needed to resolve the spatial scales and regions such as shelf areas and coasts sampled by the proxies. Due to the increase in computing power, climate models will be able to perform long (>1000y) and high resolution, often eddy permitting model simulations (e.g., Haarsma et al., 2016). Confronting these simulations with (replicated) sediment records, ideally accounting for the seasonal and depth habitat of the proxy carriers would allow to better constrain the spatial structures of climate variability as well as refining the estimates of the proxy signal content. If our SNR estimates are realistic, Holocene studies relying on a small number of records might be associated with large uncertainties, except if the quality of the analysed records is considerably higher than the average of the records analysed

here. Holocene stacks relying on a large number of records such as used in Marcott et al. (2013) would be robust if the errors are independent across sites. However, extracting spatio-temporal patterns from such datasets will be difficult. If our results are actually too pessimistic, e.g., as the true climate is more regional than simulated by the used model simulation, this would support the current interpretation of individual Holocene proxy records as a regionally representative climate signal. Otherwise, in case of too optimistic SNR estimates, the value of singular Holocene proxy reconstructions without additional expert knowledge would be limited and regional stacks might be needed to extract regional Holocene signals in analogue to the strategy used by the tree ring community.

6 Conclusion

In this study, we estimated SNRs of Holocene temperature-sensitive proxy records by comparing proxy- and model-based spatial correlations. We found that spatial correlations between proxy records were significantly lower than those computed for temperature time series extracted from climate models. Simply put, the proxy records varied more independently from site to site, whereas the model simulations suggested spatially coherent temperature variations. This in turn led to low SNR estimates in multi-proxy-type analyses if we assume that the correlation structure that we obtained from the model simulations is reasonable.

The low SNRs of Holocene proxy records are likely the result of processes occurring during the formation, preservation and measurement of the proxy signal. For the Holocene, even small uncertainties in the process chain between the climate signal and the climate reconstruction play an important role compared to the small temperature variations. In addition, as evidenced by the difference when comparing results between proxy types and within one proxy type, the proxy-specific recording of different temporal and spatial parts of the temperature (for example summer vs. winter) also affects the SNR of multi-proxy datasets. Nevertheless, our SNR estimates are still relevant for synthesis and model comparison efforts (e.g., Marcott et al., 2013), that usually interpret all proxy records together. While in the ideal case global stacks based on a large number of records will average out most of the error contributions, the interpretation of spatio-temporal patterns will remain uncertain.

The precision of the SNR estimates is strongly dependent on the number of available proxy records. Due to the small number of spatially close records of the same proxy type, the uncertainty in our proxy-type-specific SNR estimates was very high.

Our SNR estimates implicitly depend on the expected time uncertainty, as well as on the model choice. However, for both tested models the multi-proxy-type estimates on centennial time scales (T_{cent}) were smaller ($SNR_{MPI6k, T_{cent}} < 0.5$; $SNR_{T21k, T_{cent}} < 0.05$) than on longer time scales T_{mill} ($SNR_{MPI6k, T_{mill}} \approx 0.2$; $0.05 < SNR_{T21k, T_{mill}} < 0.2$).

Our results of the low signal content of multi-proxy and multi-site datasets, especially on centennial time scales, suggests that caution and a critical evaluation are in order when analysing and interpreting such large datasets. Furthermore, optimising the sampling and measurement procedure is likely needed to faithfully reconstruct small climate variations over the Holocene. As the number of high-resolution proxy records continues to grow, a more detailed analysis of the signal

Gelöscht: these

content of specific proxy types and a model-independent estimate of the spatial correlation structure of climate variations will get feasible and will enable and improve prospects for the interpretation and reconstruction of past climate changes.

Code availability. Software to reproduce the main analyses presented in this paper is available as an R code under <https://link>.

Gelöscht: R scripts to estimate the spatial correlation structure and the signal-to-noise ratios are available on request from the authors and will be made available in an online repository after the manuscript is accepted

5 *Data availability.* The datasets used in this study are available at www.sciencemag.org/cgi/content/full/339/6124/1198/DC1 (M13), at <https://www.nature.com/articles/nature25454#supplementary-information> (R18) and at the PANGAEA database (<https://www.pangaea.de>) under <https://doi.org/10.1594/PANGAEA.XXXXXX> (LH14).

Gelöscht: All

Gelöscht: from the authors on request and will be made available in PANGAEA after the manuscript is accepted

Author contributions. M.R. and T.L. designed the research. M.R. performed the analysis and wrote the first draft of the manuscript. [M.R., K.R. and T.L.](#) contributed to the interpretation and to the preparation of the final manuscript.

Gelöscht: All authors

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

Acknowledgements. We would like to acknowledge Igor Kröner for fruitful discussions and comments on the manuscript. [The work profited from discussions at the CVAS working group of the Past Global Changes \(PAGES\) programme.](#) This study was supported by the Initiative and Networking Fund of the Helmholtz Association Grant VG-NH900 as well as the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 716092). It further contributes to the German BMBF project PALMOD. K.R. acknowledges funding by the German Research Foundation grants DFG RE3994-1/1 and DFG RE3994-2/1.

References

- Ascough, P., Cook, G., and Dugmore, A.: Methodological approaches to determining the marine radiocarbon reservoir effect, *Prog. Phys. Geog.*, 29(4), 532-547, 2005.
- 20 Bakker, P., Clark, P. U., Golleger, N. R., Schmittner, A., and Weber, M. E.: Centennial-scale Holocene climate variations amplified by Antarctic Ice Sheet discharge, *Nature*, 541, 72-76, doi:10.1038/nature20582, 2017.
- Bard, E.: Paleoceanographic implications of the difference in deep-sea sediment mixing between large and fine particles, *Paleoceanography*, 16(3), 235-239, 2001.
- Bartlein, P. J., Harrison, S. P., Brewer, S., Connor, S., Davis, B. A. S., Gajewski, K., Guiot, J., Harrison-Prentice, T. I.,
25 Henderson, A., Peyron, O., Prentice, I. C., Scholze, M., Seppä, H., Shuman, B., Sugita, S., Thompson, R. S., Viau, A.

- E., Williams, J., and Wu, H.: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, *Clim. Dyn.*, 37, 775-802, doi:10.1007/s00382-010-0904-1, 2011.
- Berger, W. H., and Heath, G. R.: Vertical mixing in pelagic sediments, *J. mar. Res.*, 26, 134-143, 1968.
- Blaauw, M.: Methods and code for 'classical' age-modelling of radiocarbon sequences, *Quat. Geochronol.*, 5, 512-518, 2010.
- Blockley, S. P. E., Blaauw, M., Bronk Ramsey, C., and van der Plicht, J.: Building and testing age models for radiocarbon dates in Lateglacial and Early Holocene sediments, *Quaternary Sci. Rev.*, 26, 1915-1926, 2007.
- Breitenbach, S. F. M., Rehfeld, K., Goswami, B., Baldini, J. U. L., Ridley, H. E., Kennett, D. J., Prufer, K. M., Aquino, V. V., Asmerom, Y., Polyak, V. J., Cheng, H., Kurths, J., and Marwan, N.: CONstructing Proxy Records from Age models (COPRA), *Clim. Past*, 8, 1765-1779, doi:10.5194/cp-8-1765-2012, 2012.
- Brovkin, V., Raddatz, T., Reick, C. H., Claussen, M., and Gayler, V.: Global biogeophysical interactions between forest and climate, *Geophys. Res. Lett.*, 36, L07405, doi:10.1029/2009GL037543, 2009.
- Brown, S. J., and Elderfield, H.: Variations in Mg/Ca and Sr/Ca ratios of planktonic foraminifera caused by postdepositional dissolution: Evidence of shallow Mg-dependent dissolution, *Paleoceanography*, 11(5), 543-551, 1996.
- 15 [Compo, G. P., Whitaker, J. S., and Sardeshmukh, P. D.: Feasibility of a 100-year reanalysis using only surface pressure data. *B. Am. Meteorol. Soc.*, 87\(2\), 175-190, doi:10.1175/BAMS-87-2-175, 2006.](#)
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137, 553-597, 2011.
- 20
- Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D. M.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, *J. Adv. Model. Earth Syst.*, 7, 1220-1247, doi:10.1002/2015MS000447, 2015.
- 25
- Dekens, P. S., Lea, D. W., Pak, D. K., and Spero, H. J.: Core top calibration of Mg/Ca in tropical foraminifera: Refining paleotemperature estimation, *Geochem. Geophys. Geosyst.*, 3(4), 1-29, doi:10.1029/2001GC000200, 2002.
- Dolman, A. M., and Laepple, T.: Sedproxy: a forward model for sediment archived climate proxies, *Clim. Past*, 14, 1851-1868, doi:10.5194/cp-14-1851-2018, 2018.
- 30
- Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Sci. Rev.*, 76, 16-28, 2013.
- Fischer, N., and Jungclauss, J. H.: Evolution of the seasonal temperature cycle in a transient Holocene simulation: orbital forcing and sea-ice, *Clim. Past*, 7, 1139-1148, doi:10.5194/cp-7-1139-2011, 2011.

Gelöscht: Discuss.

Gelöscht: -31

Gelöscht: 2018-13

Gelöscht: in review,

- Fisher, D. A., Reeh, N., and Clausen, H. B.: Stratigraphic noise in time series derived from ice cores, *Ann. Glaciol.*, 7, 76-83, 1985.
- Gajewski, K.: Impact of Holocene climate variability on Arctic vegetation, *Global Planet. Change*, 133, 272-287, 2015.
- Goreau, T. J.: Frequency sensitivity of the deep-sea climatic record, *Nature*, 287, 620-622, 1980.
- 5 Grauel, A.-L., Leider, A., Goudeau, M.-L. S., Müller, I. A., Bernasconi, S. M., Hinrichs, K.-U., de Lange, G. J., Zonneveld, K. A. F., and Versteegh, G. J. M.: What do SST proxies really tell us? A high-resolution multiproxy (UK'37, TEXH86 and foraminifera $\delta^{18}O$) study in the Gulf of Taranto, central Mediterranean Sea, *Quaternary Sci. Rev.*, 73, 115-131, doi:10.1016/j.quascirev.2013.05.007, 2013.
- 10 Gregoire, L. J., Otto-Bliesner, B., Valdes, P. J., and Ivanovic, R.: Abrupt Bølling warming and ice saddle collapse contributions to the Meltwater Pulse 1a rapid sea level rise, *Geophys. Res. Lett.*, 43, 9130-9137, doi:10.1002/2016GL070356, 2016.
- [Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fuckar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project \(HighResMIP v1.0\) for CMIP6, *Geosci. Model Dev.*, 9, 4185-4208, doi:10.5194/gmd-9-4185-2016, 2016.](#)
- 15 Hansen, J., and Lebedeff, S.: Global Trends of Measured Surface Air Temperature, *J. Geophys. Res.*, 92(D11), 13345-13372, 1987.
- Heinemann, M., Jungclaus, J. H., and Marotzke, J.: Warm Paleocene/Eocene climate as simulated in ECHAM5/MPI-OM, *Clim. Past*, 5, 785-802, 2009.
- 20 Ho, S. L., and Laepple, T.: Flat meridional temperature gradient in the early Eocene in the subsurface rather than surface ocean, *Nat. Geosci.*, 9(8), 606-610, doi:10.1038/NGEO2763, 2016.
- Huguet, C., Kim, J.-H., Damsté, J. S. S., and Schouten, S.: Reconstruction of sea surface temperature variations in the Arabian Sea over the last 23 kyr using organic proxies (TEX86 and UK'37), *Paleoceanography*, 21, PA3003, doi:10.1029/2005PA001215, 2006.
- 25 Johnsen, S. J., Dahl-Jensen, D., Gundestrup, N., Steffensen, J. P., Clausen, H. B., Miller, H., Masson-Delmotte, V., Sveinbjörnsdóttir, A. E., and White, J.: Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP, *J. Quaternary Sci.*, 16(4), 299-307, doi:10.1002/jqs.622, 2001.
- 30 Jones, P. D., Osborn, T. J., and Briffa, K. R.: Estimating Sampling Errors in Large-Scale Temperature Averages, *J. Climate*, 10, 2548-2568, 1997.
- [Jonkers, L., and Kucera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, *Clim. Past*, 13, 573-586, doi:10.5194/cp-13-573-2017, 2017.](#)

- Jungclauss, J. H., Keenlyside, N., Botzet, M., Haak, H., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U., and Roeckner, E.: Ocean Circulation and Tropical Variability in the Coupled Model ECHAM5/MPI-OM, *J. Climate*, 19, 3952-3972, 2006.
- Kaplan, M. R., and Wolfe, A. P.: Spatial and temporal variability of Holocene temperature in the North Atlantic region, *Quaternary Res.*, 65, 223-231, doi:10.1016/j.yqres.2005.08.020, 2006.
- 5 [Kim, K.-Y., and North, G. R.: Surface Temperature Fluctuations in a Stochastic Climate Model, *J. Geophys. Res.*, 96\(D10\), 18573-18580, doi:10.1029/91JD01959, 1991.](#)
- Koldunov, N. V., Stammer, D., and Marotzke, J.: Present-Day Arctic Sea Ice Variability in the Coupled ECHAM5/MPI-OM Model, *J. Climate*, 23, 2520-2543, doi:10.1175/2009JCLI3065.1, 2010.
- 10 [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, doi:10.5194/bg-15-4405-2018, 2018.](#)
- Laepple, T., and Huybers, P.: Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability, *Earth Planet. Sc. Lett.*, 375, 418-429, doi:10.1016/j.epsl.2013.06.006, 2013.
- Laepple, T., and Huybers, P.: Ocean surface temperature variability: Large model–data differences at decadal and longer periods, *P. Natl. Acad. Sci. USA*, 111(47), 16682-16687, doi:10.1073/pnas.1412077111, 2014a.
- 15 [Laepple, T., and Huybers, P.: Global and regional variability in marine surface temperatures, *Geophys. Res. Lett.*, 41, 2528-2534, doi:10.1002/2014GL059345, 2014b.](#)
- Lea, D. W.: Elemental and Isotopic Proxies of Past Ocean Temperatures, In: *Treatise on Geochemistry*, Vol. 6, Editor: Elderfield, H., Executive Editors: Holland, H. D., and Turekian, K. K., Elsevier, pp. 365-390, doi:10.1016/B0-08-043751-6/06114-4, 2003.
- 20 Leduc, G., Schneider, R., Kim, J. H., and Lohmann, G.: Holocene and Eemian sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry, *Quaternary Sci. Rev.*, 29, 989-1004, doi:10.1016/j.quascirev.2010.01.004, 2010.
- Li, J., Gou, X., Cook, E. R., and Chen, F.: Tree-ring based drought reconstruction for the central Tien Shan area in northwest China, *Geophys. Res. Lett.*, 33, L07715, doi:10.1029/2006GL025803, 2006.
- 25 Liu, Z., Otto-Bliessner, 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 Bølling-Allerød Warming, *Science*, 325(5938), 310-314, doi:10.1126/science.1171041, 2009.
- Lohmann, G., Pfeiffer, M., Laepple, T., Leduc, G., and Kim, J.-H.: A model-data comparison of the Holocene global sea surface temperature evolution, *Clim. Past*, 9, 1807-1839, doi:10.5194/cp-9-1807-2013, 2013.
- 30 Lu, F., Ma, C., Zhu, C., Lu, H., Zhang, X., Huang, K., Guo, T., Li, K., Li, L., Li, B., and Zhang, W.: Variability of East Asian summer monsoon precipitation during the Holocene and possible forcing mechanisms, *Clim. Dynam.*, 1-21, doi:10.1007/s00382-018-4175-6, 2018.

- Luckman, B. H., Briffa, K. R., Jones, P. D., and Schweingruber, F. H.: Tree-ring based reconstruction of summer temperatures at the Columbia Icefield, Alberta, Canada, AD 1073-1983, *The Holocene*, 7(4), 375-389, 1997.
- Madden, R. A., Shea, D. J., Branstator, G. W., Tribbia, J. J., and Weber, R. O.: The Effects of Imperfect Spatial and Temporal Sampling on Estimates of the Global Mean Temperature: Experiments with Model Data, *J. Climate*, 6, 1057-1066, 1993.
- 5 Mangerud, J., and Svendsen, J. I.: The Holocene Thermal Maximum around Svalbard, Arctic North Atlantic; molluscs show early and exceptional warmth, *The Holocene*, 28(1), 65-83, doi:10.1177/0959683617715701, 2018.
- Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: Robustness of proxy-based climate field reconstruction methods, *J. Geophys. Res.*, 112, D12109, doi:10.1029/2006JD008272, 2007.
- 10 Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *P. Natl. Acad. Sci. USA*, 105(36), 13252-13257, doi:10.1073/pnas.0805721105, 2008.
- 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*, 339(6124), 1198-1201, doi:10.1126/science.1228026, 2013.
- 15 Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, *Ocean Model.*, 5(2), 91-127, doi:10.1016/S1463-5003(02)00015-X, 2003.
- Matei, D., Pohlmann, H., Jungclaus, J., Müller, W., Haak, H., and Marotzke, J.: Two Tales of Initializing Decadal Climate Prediction Experiments with the ECHAM5/MPI-OM Model, *J. Climate*, 25, 8502-8523, doi:10.1175/JCLI-D-11-00633.1, 2012.
- 20 May, W.: Climatic changes associated with a global "2°C-stabilization" scenario simulated by the ECHAM5/MPI-OM coupled climate model, *Clim. Dynam.*, 31, 283-313, doi:10.1007/s00382-007-0352-8, 2008.
- Mischel, S. A., Scholz, D., Spötl, C., Jochum, K. P., Schröder-Ritzrau, A., and Fiedler, S.: Holocene climate variability in Central Germany and a potential link to the polar North Atlantic: A replicated record from three coeval speleothems, *The Holocene*, 27(4), 509-525, doi:10.1177/0959683616670246, 2017.
- 25 Mollenhauer, G., Eglinton, T. I., Ohkouchi, N., Schneider, R. R., Müller, P. J., Grootes, P. M., and Rullkötter, J.: Asynchronous alkenone and foraminifera records from the Benguela Upwelling System, *Geochim. Cosmochim. Ac.*, 67(12), 2157-2171, doi:10.1016/S0016-7037(00)00168-6, 2003.
- Moossen, H., Bendle, J., Seki, O., Quillmann, U., and Kawamura, K.: North Atlantic Holocene climate evolution recorded by high-resolution terrestrial and marine biomarker records, *Quaternary Sci. Rev.*, 129, 111-127, doi:10.1016/j.quascirev.2015.10.013, 2015.
- 30 Müller, W. A., and Roeckner, E.: ENSO teleconnections in projections of future climate in ECHAM5/MPI-OM, *Clim. Dynam.*, 31, 533-549, doi:10.1007/s00382-007-0357-3, 2008.
- Münch, T., and Laepple T.: What climate signal is contained in decadal- to centennial-scale isotope variations from Antarctic ice cores?, *Clim. Past*, 14, 2053-2070, doi:10.5194/cp-14-2053-2018, 2018.

Gelöscht:

Gelöscht: Discuss.

Gelöscht: 1-25

Gelöscht: 2018-112, in review

- Münch, T., Kipfstuhl, S., Freitag, J., Meyer, H., and Laepple, T.: Regional climate signal vs. local noise: a two-dimensional view of water isotopes in Antarctic firn at Kohnen Station, Dronning Maud Land, *Clim. Past*, 12, 1565-1581, doi:10.5194/cp-12-1565-2016, 2016.
- Münch, T., Kipfstuhl, S., Freitag, J., Meyer, H., and Laepple, T.: Constraints on post-depositional isotope modifications in East Antarctic firn from analysing temporal changes of isotope profiles, *The Cryosphere*, 11, 2175-2188, doi:10.5194/tc-11-2175-2017, 2017.
- Ohkouchi, N., Eglinton, T. I., Keigwin, L. D., and Hayes, J. M.: Spatial and Temporal Offsets Between Proxy Records in a Sediment Drift, *Science*, 298(5596), 1224-1227, doi:10.1126/science.1075287, 2002.
- Pausata, F. S. R., and Löfflerström, M.: On the enigmatic similarity in Greenland $\delta^{18}\text{O}$ between the Oldest and Younger Dryas, *Geophys. Res. Lett.*, 42, 10470-10477, doi:10.1002/2015GL066042, 2015.
- Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.-G., Wetzol, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, *Clim. Dynam.*, 29, 565-574, doi:10.1007/s00382-007-0247-8, 2007.
- Rehfeld, K., Trachsel, M., Telford, R. J., and Laepple, T.: Assessing performance and seasonal bias of pollen-based climate reconstructions in a perfect model world, *Clim. Past*, 12, 2255-2270, doi:10.5194/cp-12-2255-2016, 2016.
- Rehfeld, K., Münch, T., Ho, S. L., and Laepple, T.: Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene, *Nature*, 554, 356-359, doi:10.1038/nature25454, 2018.
- Reschke, M., Kunz, T., and Laepple, T.: Comparing methods for analysing time scale dependent correlations in irregularly sampled time series data, *Comp. Geosci.*, [123, 65-72, doi:10.1016/j.cageo.2018.11.009, 2019](https://doi.org/10.1016/j.cageo.2018.11.009).
- Richey, J. N., Hollander, D. J., Flower, B. P., and Eglinton, T. I.: Merging late Holocene molecular organic and foraminiferal-based geochemical records of sea surface temperature in the Gulf of Mexico, *Paleoceanography*, 26, PA1209, doi:10.1029/2010PA002000, 2011.
- Roche, D. M., Waelbroeck, C., Metcalfe, B., and Caley, T.: FAME (v1.0): a simple module to simulate the effect of planktonic foraminifer species-specific habitat on their oxygen isotopic content, *Geosci. Model Dev.*, 11, 3587-3603, doi:10.5194/gmd-11-3587-2018, 2018.
- Roeckner, E., Bäuml, G., and Bonaventura, L.: The atmospheric general circulation model, 2003.
- Rosell-Melé, A., and Prahl, F. G.: Seasonality of UK' 37 temperature estimates as inferred from sediment trap data, *Quaternary Sci. Rev.*, 72, 128-136, doi:10.1016/j.quascirev.2013.04.017, 2013.
- Rosén, P., Segerström, U., Eriksson, L., and Renberg, I.: Do Diatom, Chironomid, and Pollen Records Consistently Infer Holocene July Air Temperature? A Comparison Using Sediment Cores from Four Alpine Lakes in Northern Sweden, *Arct. Antarct. Alp. Res.*, 35(3), 279-290, 2003.
- Schneider, B., Leduc, G., and Park, W.: Disentangling seasonal signals in Holocene climate trends by satellite-model-proxy integration, *Paleoceanography*, 25, PA4217, doi:10.1029/2009PA001893, 2010.

Gelöscht: in review

- Sejrup, H. P., Seppä, H., McKay, N. P., Kaufman, D. S., Geirsdóttir, Á., de Vernal, A., Renssen, H., Husum, K., Jennings, A., and Andrews, J. T.: North Atlantic-Fennoscandian Holocene climate trends and mechanisms, *Quaternary Sci. Rev.*, 147, 365-378, doi:10.1016/j.quascirev.2016.06.005, 2016.
- 5 Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, *WIREs Clim. Change*, 3, 63-77, doi:10.1002/wcc.149, 2012.
- Spero, H. J., Bijma, J., Lea, D. W., and Bemis, B. E.: Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes, *Nature*, 390, 497-500, 1997.
- 10 Stebich, M., Rehfeld, K., Schlütz, F., Tarasov, P. E., Liu, J., and Mingram, J.: Holocene vegetation and climate dynamics of NE China based on the pollen record from Sihailongwan Maar Lake, *Quaternary Sci. Rev.*, 124, 275-289, doi:10.1016/j.quascirev.2015.07.021, 2015.
- Telford, R. J., Heegaard, E., and Birks, H. J. B.: All age–depth models are wrong: but how badly?, *Quaternary Sci. Rev.*, 23, 1-5, doi:10.1016/j.quascirev.2003.11.003, 2004.
- Thibodeau, B., Bauch, H. A., and Knies, J.: Impact of Arctic shelf summer stratification on Holocene climate variability, *Quaternary Sci. Rev.*, 191, 229-237, doi:10.1016/j.quascirev.2018.05.017, 2018.
- 15 Thirumalai, K., Partin, J. W., Jackson, C. S., and Quinn, T. M.: Statistical constraints on El Niño Southern Oscillation reconstructions using individual foraminifera: A sensitivity analysis, *Paleoceanography*, 28, 401–412, doi:10.1002/palo.20037, 2013.
- Thirumalai, K., Quinn, T. M., Okumura, Y., Richey, J. N., Partin, J. W., Poore, R. Z., and Moreno-Chamarro, E.: Pronounced centennial-scale Atlantic Ocean climate variability correlated with Western Hemisphere hydroclimate, *Nat. Commun.*, 9, 392, doi:10.1038/s41467-018-02846-4, 2018.
- 20 [Tierney, J. E., and Tingley, M. P.: A TEX86 surface sediment database and extended Bayesian calibration, *Sci. Data*, 2, 150029, doi:10.1038/sdata.2015.29, 2015.](#)
- Timm, O., and Timmermann, A.: Simulation of the Last 21 000 Years Using Accelerated Transient Boundary Conditions, *J. Climate*, 20, 4377-4401, doi:10.1175/JCLI4237.1, 2007.
- 25 van Sebille, E., Scussolini, P., Durgadoo, J. V., Peeters, F. J. C., Biastoch, A., Weijer, W., Turney, C., Paris, C. B., and Zahn, R.: Ocean currents generate large footprints in marine palaeoclimate proxies, *Nat. Commun.*, 6, 6521, doi:10.1038/ncomms7521, 2015.
- Wanner, H., Mercolli, L., Grosjean, M., and Ritz, S. P.: Holocene climate variability and change; a data-based review, *J. Geol. Soc. London*, 172, 254-263, doi:10.1144/jgs2013-101, 2015.
- 30 Werner, M., Haese, B., Xu, X., Zhang, X., Butzin, M., and Lohmann, G.: Glacial–interglacial changes in H218O, HDO and deuterium excess – results from the fully coupled ECHAM5/MPI-OM Earth system model, *Geosci. Model Dev.*, 9, 647-670, doi:10.5194/gmd-9-647-2016, 2016.
- Zhang, Y., Renssen, H., Seppä, H., and Valdes, P. J.: Holocene temperature evolution in the Northern Hemisphere high latitudes – Model-data comparisons, *Quaternary Sci. Rev.*, 173, 101-113, doi:10.1016/j.quascirev.2017.07.018, 2017.

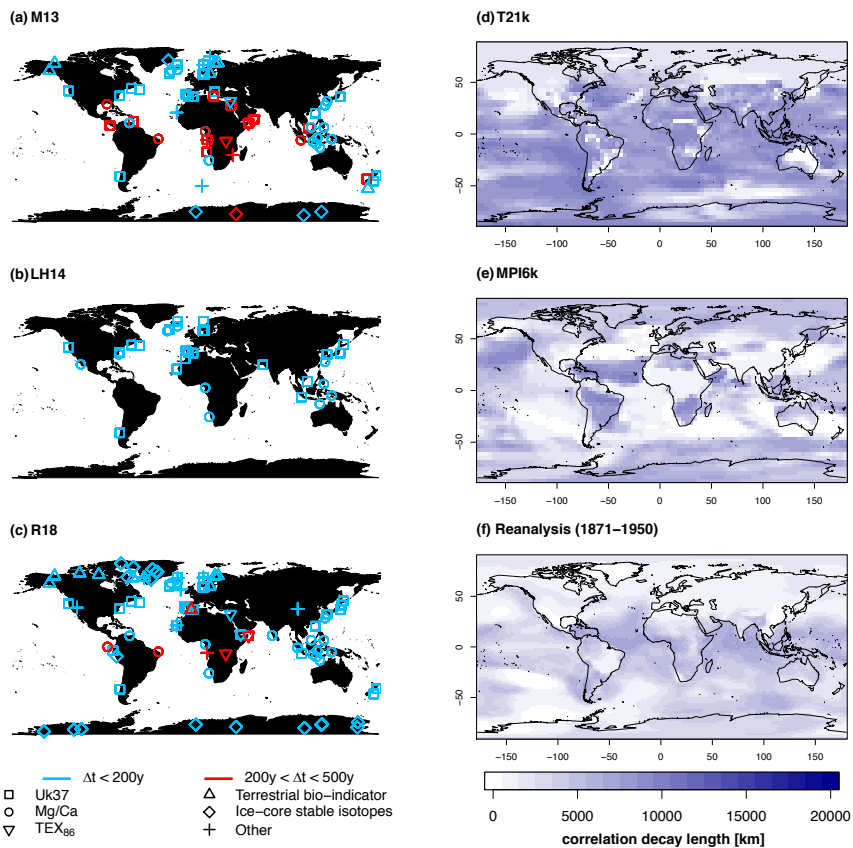


Figure 1: Overview of proxy and model datasets. Site locations of the proxy compilations (a) M13: Marcott et al. (2013), (b) LH14: Laepple and Huybers (2014a), and (c) R18: Rehfeld et al. (2018) used in this study. Proxy types are indicated by symbols and the mean inter-observation time step by colours. Correlation decay length of (d) the T21k and (e) the MPI6k simulations estimated on time scales larger than 400y with included trend, and (f) reanalysis data from 1871 to 1950 estimated from annual data. The spatial correlation decay length is generally higher for T21k than for MPI6k. For a comparison of the model and reanalysis correlation structure on the same time scale, see Fig. S1.

Gelöscht: ... [2]

Gelöscht: centennial
 Gelöscht: (T_{cent} ; not detrended)

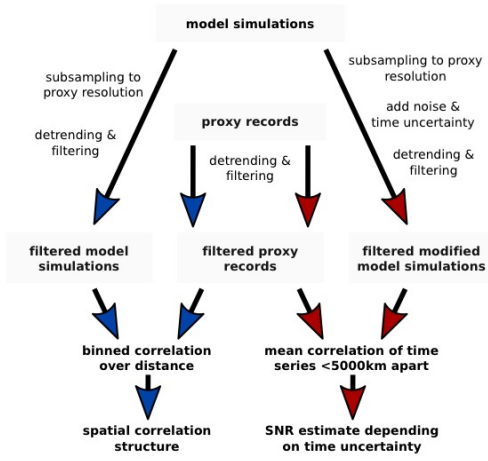
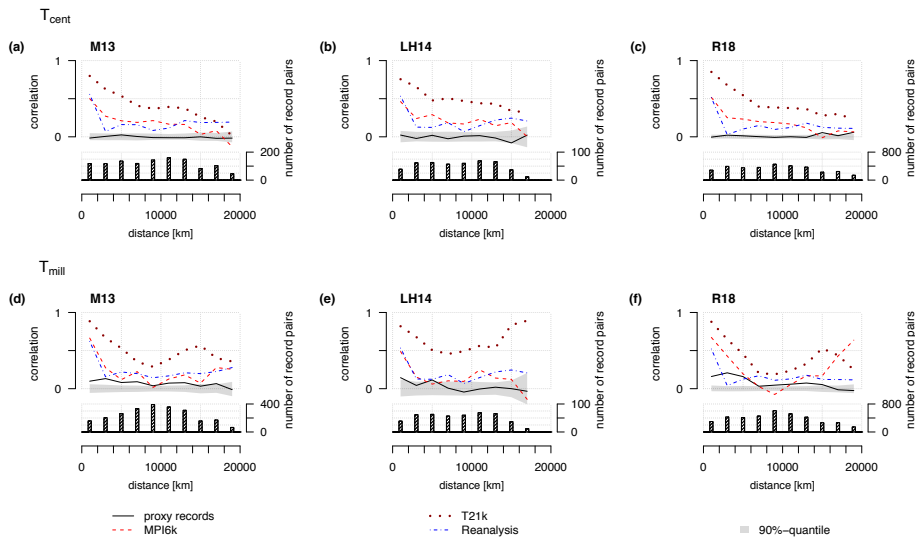
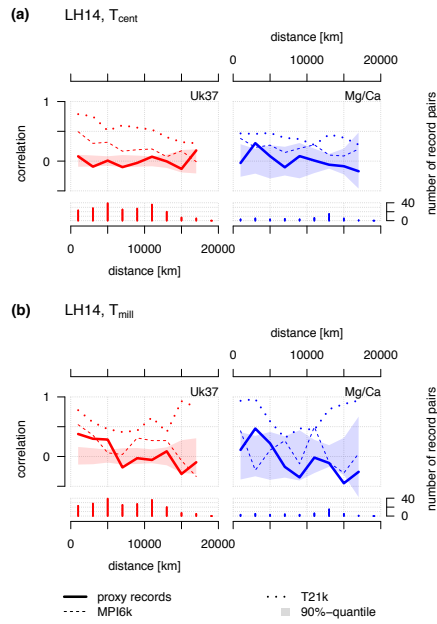


Figure 2: Processing steps for the proxy and model time series. Blue paths illustrate the analysis of the spatial correlation structure. Red paths represent the estimation of SNRs of proxy records as a function of time uncertainty.



Gelöscht:

Figure 3: Spatial correlation structure of Holocene temperature proxy records and simulated surface temperatures based on three multi-proxy datasets and related to (a-c) centennial T_{cent} and (d-f) centennial to millennial time scales T_{mill} . In each panel, the upper part shows the mean correlation of the model simulation (for 2000 km sized bins as a function of the separation distance between record pairs) and reanalysis data (1871-1950) evaluated at the proxy locations (dotted/dashed line) and the proxy dataset (continuous line). The grey polygon represents the 90%-quantile of mean correlations of uncorrelated surrogate time series with a power-law scaling of $\beta = 1$. The lower parts of the panels show the number of record pairs used in each estimate. The spatial correlation structure of the model time series is generally higher than of proxy records which are only statistically significant on T_{mill} at neighbored sites. The highest correlations belong to sites with separation distances less than 4000-6000 km.



5 **Figure 4: Proxy-type-specific (Uk37, Mg/Ca) spatial correlation structure related to (a) centennial T_{cent} and (b) centennial to millennial time scales T_{mill} based on the LH14 dataset.** The upper parts of the panels show mean correlations of 2000 km sized bins as a function of the separation distance between record pairs in the proxy dataset (continuous line) and model simulations evaluated at proxy locations (dotted/dashed line). Polygons represent the 90%-quantiles of mean correlations of uncorrelated surrogate time series with a power-law scaling of $\beta = 1$. The lower parts of the panels show the number of record pairs used for each estimate. The spatial correlation structure of proxy records is non-significant for individual proxy types, except for close (separation <6000 km) sites of Uk37 temperature records at T_{mill} .

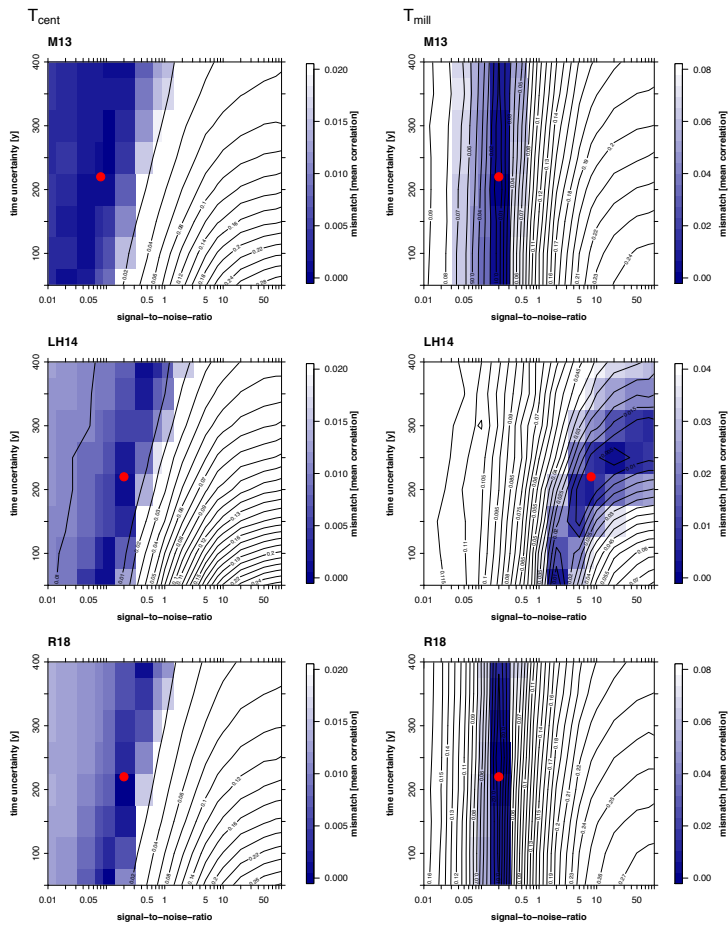


Figure 5: SNR_{MPI6k} estimates of proxy records as a function of time uncertainty related to centennial T_{cent} and millennial time scales T_{mill} . Colour coating and contour lines in each panel show the mismatch between mean correlations of close-by (separation <5000 km) proxy records and time series extracted from the MPI6k simulation at proxy locations as a function of time uncertainty (vertical axis) and SNR (horizontal axis). Areas with the lowest mismatch are represented by the darkest colours and mark suitable combinations of SNR_{MPI6k} estimates and time uncertainties. The red dots illustrate SNR estimates for a time uncertainty of 220y.

5

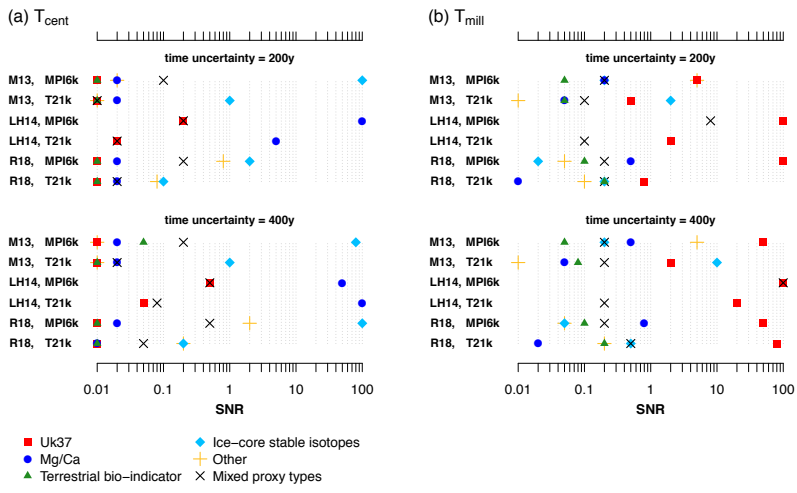
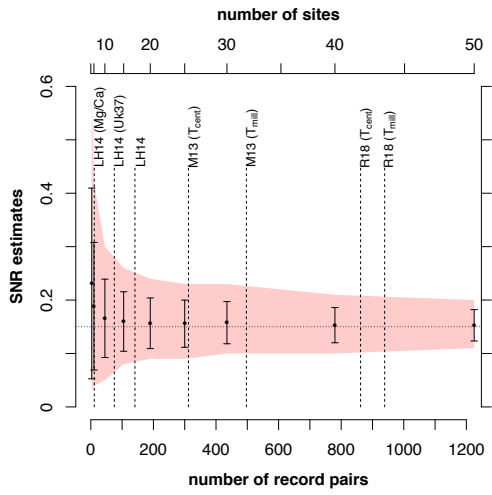


Figure 6: Overview of proxy-specific SNR estimates on (a) centennial T_{cent} and (b) millennial time scales T_{mill} . The symbols represent the SNRs estimated from the different proxy compilations using the simulations of MPI6k and T21k. Upper panels show the results for an assumed time uncertainty of 200y and lower panels for 400y. The SNRs are proxy-type-specific different, but generally higher on the T_{mill} than on the T_{cent} time scale.

5



5 **Figure 7: Sensitivity of the SNR estimates on the number of sites/record pairs based on surrogate time series.** The time series were generated with a predefined SNR = 0.15 (horizontal line). SNR estimates with standard deviations based on 2000 repetitions are shown as dots with error bars. The uncertainties in the SNR are illustrated as polygons showing the 90%-quantiles of the estimates. The uncertainty of SNR estimates is high when only considering a small number of sites. Vertical lines show the numbers of selected sites/record pairs contained in each data compilation. This indicates that for single proxy type analysis the uncertainties in the SNR estimates are high.

Table 1: Numbers of records and their overlap in the proxy compilations used in this study. The total number of time series is separated by proxy type for each proxy compilation (upper part). T_{mill} refers to the number of time series with a mean inter-observation time step of $\Delta t < 500y$ and T_{cent} counts time series with $\Delta t < 200y$. The overlap is shown for each pair and for all compilations (lower part).

	Sum of records	Uk37	Mg/Ca	TEX₈₆	Terrestrial bio-indicator	Ice-core stable isotopes	other
M13 – T_{mill}	70	28	19	4	8	5	6
M13 – T_{cent}	49	18	14	1	8	4	4
LH14 – T_{mill}	31	21	10	-	-	-	-
LH14 – T_{cent}	31	21	10	-	-	-	-
R18 – T_{mill}	88	27	19	4	11	18	9
R18 – T_{cent}	81	26	17	2	10	18	8
M13 \cap LH14	20	13	7	-	-	-	-
M13 \cap R18	45	16	14	3	6	4	2
LH14 \cap R18	22	16	6	-	-	-	-
M13 \cap LH14 \cap R18	19	13	6	-	-	-	-

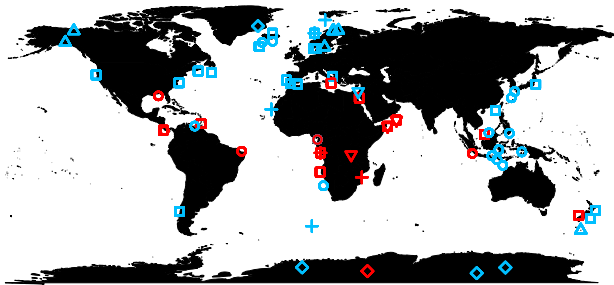
Table 2: Mean correlations of proxy time series with separation distances <5000 km for different proxy types. For each dataset, the mean correlation was estimated for millennial time scales T_{mill} and proxy time series with a mean inter-observation time step of $\Delta t < 500y$ and related to centennial time scales T_{cent} for proxy time series with $\Delta t < 200y$. Mixed proxy types contain all combinations of time series pairs independent of the proxy type. The mean of single proxy types summarises the proxy-type-specific mean correlations weighted by the number of record pairs of each proxy type. Correlations in brackets are not statistically significant ($p = 0.1$).

	Mixed proxy types	Mean of single proxy types	Uk37	Mg/Ca	TEX₈₆	Terrestrial bio-indicator	Ice-core stable isotopes	other
M13 – T_{mill}	0.101	0.149	0.211	(0.068)	(-0.105)	(0.095)	0.414	(-0.14)
M13 – T_{cent}	0.004	-0.009	(-0.003)	(-0.006)	-	(-0.09)	(0.244)	(-0.388)
LH14 – T_{mill}	0.12	0.357	0.365	0.304	-	-	-	-
LH14 – T_{cent}	0.012	0.031	(0.013)	0.151	-	-	-	-
R18 – T_{mill}	0.184	0.208	0.347	(0.034)	(-0.188)	0.17	0.23	(0.107)
R18 – T_{cent}	0.014	0.009	(-0.017)	(-0.024)	(0.148)	(-0.057)	0.1	(0.05)

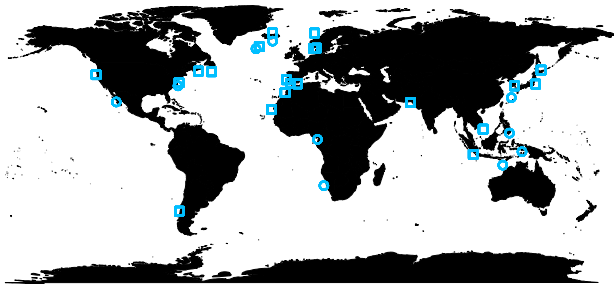
he consequence is a temporal and spatial averaging of the climate signal.

Thus, a

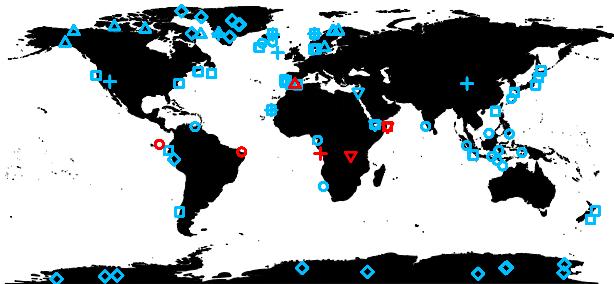
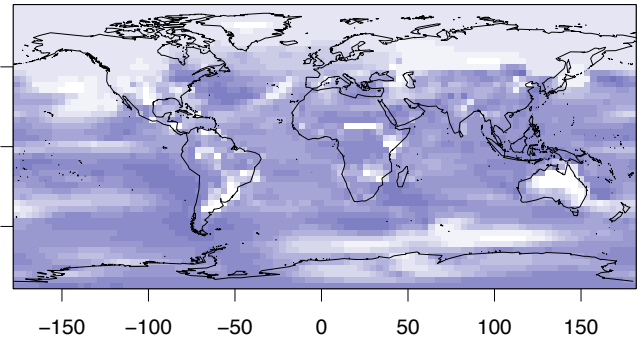
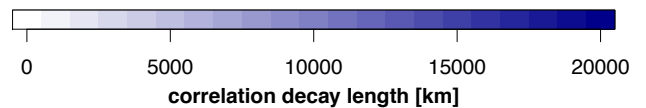
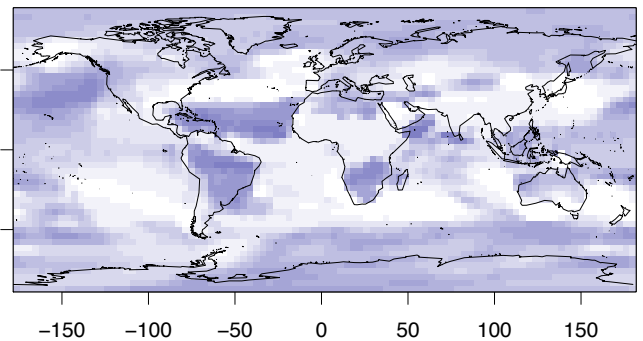
(a) M13



(b) LH14



(c) R18

(d) T21k (T_{cent})(e) MPI6k (T_{cent})

- | | |
|---|---|
| <ul style="list-style-type: none"> □ Uk37 ○ Mg/Ca ▽ TEX₈₆ | <ul style="list-style-type: none"> △ Terrestrial bio-indicator ◇ Ice-core stable isotopes + Other |
|---|---|