

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

Oliver Bothe and Eduardo Zorita

ol.bothe@gmail.com

Received and published: 8 October 2019

Dear referees, dear editor,

We want to thank the referees and the editor for evaluating our manuscript and providing such encouraging comments.

Below we respond to the reviewer comments. We will modify the manuscript to shift the focus towards the uncertainties. We yet are undecided whether we explicitly include a bootstrap uncertainty estimate similar to Neukom et al. (2019).

We are sorry that we took so long for our final response.

C1

On behalf of the authors

Sincerely yours,

Oliver Bothe

C2

Editor

Please for the revisions, you might add in the paper, that compared to PAGES2k, Luterbacher et al. 2016 excluded the Tatra and Albania proxies from their analysis as they lack significant correlations with European summer temperature variability.

Response We make this change and add: Already Luterbacher et al. (2016) noted this and, therefore, did not consider these two proxies in their reconstruction effort.

C3

Referee 1

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

General comments.

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

Response: We thank the reviewer for their positive evaluation.

Our revisions try to improve on the writing.

Regarding the overfitting: See our response below.

C4

Specific comments.

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

Response: We change this to: Combining proxy information and climate model simulations reconciles these sources of information about past climates.

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

Response: We are not convinced that this simply repeats the content of the previous sentence but follow the suggestion of the referee: The analogue or proxy surrogate reconstruction method is a computationally cheap data assimilation approach, which searches in a pool of simulated climate states the best fit to proxy data.

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

Response: We change this accordingly.

P1, L10: Remove “using”?

Response: We change this accordingly.

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

Response: We change this to read: However, the approaches disagree on the

C5

warmest preindustrial decade, which is in the early 15th century for the Euro 2k reconstruction and in the early 18th century for the analogue approach.

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

Response: We try to be consistent in our writing. Therefore, we here change the sentence to: The reconstructions from the analogue method . . .

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

Response: We try to use more commas to ease reading the manuscript.

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

Response: We change this to read: The reconstructions from the analogue method also represent the local variations of the observed proxies. Local uncertainties of the temperature reconstructions tend to be large in areas that are poorly covered by the proxy records.

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

Response: We regard it to be the right expression but change the occurrences according to the referee’s suggestion.

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

Response: We replace this with “searching”.

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

Response: There are further applications. Therefore, we would like to keep the sentence as it is.

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

C6

rewrite.

Response: We rewrite this: The analogue method is a computationally cheap means to contrast information from simulations and reconstructions in the sense of data assimilation. However, it is methodologically less sophisticated than full data assimilation procedures (compare, e.g. Tardif et al., 2019).

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

Response: We rewrite: Here we propose a reconstruction . . .

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

Response: We are sorry for the random line numbers. We use the RMarkdown template and under certain unclear conditions this happens, but can easily be repaired, which we unfortunately did not do.

We rewrite: Here we obtain large-scale fields of summer temperature based on a pool of relevant candidate fields from model simulations and a set of local proxy data records as predictors for the period 1260 to 2003 of the Common Era (CE).

Figure 1: Please add units to axes.

Response: We clarify the Figure.

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

Response: We rewrite: That is, they use recent observations made on properties of paleo-archives. These archives, in turn, recorded the past environmental conditions. . .

P4, L19: “to more than one environmental condition” do you mean that a given proxy

C7

parameter can depend on more than climate or environmental variable? Please clarify.

Response: We rewrite: . . . that the archives recorded signals from more than one climate or environmental variable (e.g. temperature and precipitation . . .

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

Response: We think our choice of words is valid, but we rewrite: Correlations provide a simple measure of the relation between proxy-observations and the climatic environment over a period when reliable (instrumental) observations of the climatic variability exist.

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

Response: We add another table.

Table 1: So, the correlations are between the tree ring data series and the JJA CRU temperature. Please add these details to the caption. How do you deal with seasonality of the proxy data? In Wilson et al. (2016) each proxy site is listed with different seasonal sensitivity to temperature (Table 1) and I believe you are using some of the same data.

Response: 1. We add the details to the caption. 2. As we compare our reconstruction to the Euro 2k reconstruction, we consider the seasonal attribution as used by the publications for this reconstruction. That is, we do not test whether the relation of the proxies is strongest for summer. For the attribution to summer compare Luterbacher et al. (2016) and PAGES 2k Consortium (2013).

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

Response: Yes. We clarify this: The latest start date of any of the remaining eight

C8

proxy indices is the year 1260 CE. Thus, all eight records cover the period 1260 to 2003 CE.

P7, L16: "strong ensemble" or "8 member ensemble"?

Response: We clarify: there exists a multi-model ensemble of simulations for the last 1100 years. A number of additional simulations comply with the PMIP3 protocol but are not included in the effort

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

Response: We remove the sentence.

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

Response: 1. We clarify the rescaling in the caption. 2. Euro 2k is an area mean. 3. We are unclear what the referee is referring to with respect to the CRU data. 4. Since the CRU data are only shown for qualitative comparisons, we rather would not do this. However, we did add such a comparison to the respective panel. 5. We add an Appendix to show additional Figures including one plotting the data by Luterbacher et al. (2016).

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

Response: We thought it was clear enough, but now show the equation.

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

Response: We shortly describe the characteristics of both uncertainty estimates. Both describe realistically part of the uncertainty.

C9

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

Response: We clarify the sentence.

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

Response: We clarify this.

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

Response: We only considered tropical eruptions. We clarify this: We consider a subselection of tropical eruption events in 1286, 1345, 1458, 1601, 1641, 1695, 1809, and 1815. We performed a superposed epoch analysis but do not show the results. We considered fields and area averages. We chose the five calendar years before an eruption year as reference period, which is a common approach (compare, e.g. Sigl et al., 2015).

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

Response: We modify this: The analogues even appear to occasionally capture the relation between the proxies included and those excluded, which obviously might be by chance.

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

Response: We do so.

C10

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

Response: Yes. We clarify the caption.

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

Response: We again clarify the caption by mentioning that these are the correlations.

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

Response: The correlations are between the proxy locations and the medians of 39 analogues. Indeed they are high, but we would hope that the proxies included in our search constrain the search effectively and give good reconstruction results. This holds especially for the median, which is a filter for the data of the reconstruction ensemble members. The aim of data assimilation is to match the observations, i.e. the proxies, closely with the simulated data. Nevertheless, we are unable to exclude the possibility that our data constrains the pool too much and therefore may fail in a prediction exercise.

These correlations indicate only agreement with the proxy records not necessarily with the true temperature. Anyway, locally high correlations do not indicate high skill elsewhere. Indeed, correlations with the observational CRU data are in line with the correlations between the proxies and the CRU data as one may expect from these high correlation coefficients. The comparison to the BEST-data shows, this does not necessarily reflect on how well the reconstruction captures the observed temperature elsewhere.

Regarding the use of proxy system models: An optimal approach would incorporate a

C11

calibrated proxy system model to pre-process the simulation data. Indeed, any reconstruction approach can benefit from pre-processing data with calibrated proxy forward models.

Regarding overfitting, there are no parameters in the analog-setting that are calibrated for a better fit to the predictand. The number of analogs chosen or the distance metric chosen are not optimized for a better fit.

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

Response: We clarify this: The top-left panel for Nuuk highlights that the lack of constraints on the reconstruction can result in potentially artificial spikes in the time-series.

C12

Referee 2

This study is an interesting contribution to the field of climate reconstruction because it adds and compares multiple way of uncertainty estimation to the widely and successfully used analog reconstruction methodology. It fits very well to the scope of Climate of the Past. Hence, I suggest publication after revisions that should make the structure clearer, condense the results including figures with time series and after putting the focus a bit more on the novelty of the uncertainty estimation than on the reconstruction.

Response: We thank the referee for the positive evaluation.

Our revisions try to clarify the structure of the manuscript, put the emphasis on the uncertainty, and clarify the Figures.

However, we indeed think that time-series plots are in the most cases the most appropriate visualisation.

Comments:

Introduction

- Would be worth mentioning the just published global reconstruction by Neukom et al. 2019, which includes an analog approach, too.

Response: Of course. Until the publication of Neukom et al. (2019) we were not aware of their work. In view of the comments of referee 3 we will discuss their approach.

- I would find a list of the content helpful at the end of the introduction, saying that three approaches are tested: 1. best analog only, . . .

Response: We add a short paragraph outlining the manuscript.

C13

- Page 2, line 20: "guestimate" is colloquial language

Response: We regard it an appropriate term, but slightly change the sentence.

Methods

- The entire structure of the study and the used error estimation should be made clearer. Can you add a schematic diagram?

Response: We try to clarify the structure of the methods section and to be more explicit about the error estimation. However, we do not think a schematic diagram is necessary at this point.

- Explain clearly how you come to your three reconstruction experiments, into which the results are separated. I assume the number 39 for the minimal number of analogs in 2003 (page 5, line 20) is the reason for having 39 in section 3.2 but that is not clear to the reader.

Response: We clarify this part.

- Page 4, line 23: "under certain assumption" Which assumptions? Please write more precise.

Response: This refers to the assumptions mentioned in the next paragraphs. We clarify the sentence.

- Page 4, last two paragraphs: I would rather put the equations more prominent in separate lines and not in the middle of the sentence because understanding the error estimation is crucial for this study.

Response: We follow this recommendation.

- Page 5, line 6: modified

C14

Response: We thank the referee for spotting this.

- Page 5, line 15: "dates" You have not mentioned yet that you reconstruct JJA averages at annual resolution

Response: We clarify this now at this location and in section 2.1.1.

- Page 5, line 15ff.: How do you choose the noise SD levels such as 2.57? And why if you write in line 22 that only the 1 SD criterion gives a reasonable number of analogs?

Response: We clarify this now. We choose 2.57SD as it gives a reasonable minimum number for a set of good analogues. We choose 1SD as it gives a reasonable maximum number of analogues for a fixed SD level reconstruction.

Proxies

- Is there a reason to use the gridded CRU data for proxy correlations here and the BEST data later in the paper?

Response: We use the regionally representative series from BEST and we use these for periods before widespread instrumental data is available. We use the CRU data as correlation target as it is commonly used.

- You could explain that the correlation of the excluded location in Slovakia are low because trees are limited to temperatures in another season. Otherwise it seems strange, why they appear in the PAGES data base. However, I am not sure why the Albania chronology with weakly significant negative correlation appears in the PAGES collection. Maybe, it has been removed in the more strictly screened version 2 of the data base? Having this negative correlation in mind, I do not understand why it is used for comparison/verification later in the paper? I would not expect a good match/positive correlation in the analog reconstruction.

C15

Response: Already Luterbacher et al. (2016) removed these two series from their reconstruction effort. We may remove the relevant panels in our revised version but did not yet decide on this.

We cannot recall why the EuroMed 2k network included both chronologies in their initial reconstruction approach. The original publication for the Albanian record (Seim et al., 2012, <https://doi.org/10.3354/cr01076>) identifies a significant negative relation to temperature with, however, only small correlation coefficients. We might presume that initially EuroMed 2k considered this to be enough for this data sparse region.

Model simulations

- Page 7, line 5: Please explain again briefly why the "similar internal variability" of the simulation is important instead of referring to the previous section

Response: We clarify this now.

Results

- Generally, try to shorten the results section and have a clear and consistent structure for the three experiments. I would put more focus on the uncertainty results than the reconstruction itself.

Response: We will try to be more concise in our writing while preserving the relevance of the manuscript. We will try to shift the focus to the uncertainty in our revisions.

- Make clearer, how the three experiments compare and later in the discussion what we can learn from this.

C16

Response: Our revisions try to be clearer about the differences between the three setups.

- Page 9, line 2: *why is the plot relative to Euro-2k and not relative to instrumental data?*

Response: The idea was to compare the reconstruction and its uncertainty against a previous reconstruction based on the same data. If we put everything relative to a 20th century data, the potentially different trends over the 20th century will make such a comparison harder. We consider changing the visualisation and aim to clarify this in our revisions.

- Fig. 3: *It is not surprising that the analogs fit better, where you have spatial proxy clusters than isolated locations. I have not seen this discussed in the paper.*

Response: We are not sure about the point of the reviewer. We are going to add more discussion on the different data availability. However, considering Figure 3, it is not necessarily the case that the analogues fit better in, e.g., the Alps or Scandinavia.

- Page 11, line 26: *It is good that the analog reconstruction generally agrees with previous statistical reconstructions but they are not a reference and it is unclear which ones are closer to reality. Rather just see if they are in your uncertainty range.*

Response: Our revisions aim to be clearer about the evaluation of our reconstruction against the data. However, we would argue in this case that the convergence of both approaches is an important aspect. Indeed, such convergence is, in our view, one strength of the recent work by the PAGES 2k Consortium (2019) and Neukom et al. (2019).

- Page 14, line 30ff: *Have you considered weighting the analogs with respect to their distance?*

Response: We considered weighting the analogs. Indeed, weighting may provide us with a clear posterior distribution. However, weighting the analogues by their distance,

C17

in our understanding, to some extent would counter our approach of using analogues that relate to a certain uncertainty level of the proxies.

- Page 15, line 5: *"visually there is good agreement" Not clear what you are talking about, other series besides Tatra and Albania?*

Response: We clarify in our revisions: Visually there is good agreement between the proxies included in our analogue search and our reconstructed local series. Usually, the range of reconstructed values is relatively narrow for these proxies.

- Page 15, line 29: *Add reference to figure*

Response: We modify and clarify this: Thus, Figure 5 shows that the included proxies anchor the reconstruction to a very narrow range of variability if we choose a fixed number of analogues.

- Page 15, line 35: *How can you have a stronger 20th century warming trend in the reconstruction than in observations and at the same time have trouble to find analogs for exceptionally warm years such as 2003?*

Response: We refer to the warming trend from the early 19th century onwards and thereby the mean warming over time for this period. The lack of analogues is due to the exceptional warm years in the early 21st century. We do not find analogues for these, the specific interrelation among the proxy records, and within a narrow one standard deviation uncertainty range.

- Page 16, line 9: *If you look at the temperature evolution after individual eruptions, this is not a superposed epoch analysis.*

Response: Thank you for highlighting our lack of clarity. First, we indeed did a valid Superposed Epoch Analysis but considered also individual evolutions. Second, as we only mention the individual evolutions here, we skip the reference to the superposed epoch analysis.

C18

- Page 19, line 14 and Fig. 8: Why do you show a mean and not the median in this case? The mean should be influenced by the number of averaged analogs and the numbers are highly different in this case.

Response: The referee is correct. We redid the analysis with the median. We are yet undecided what to show in the revised manuscript because visual differences are negligible and differences in results are small. We will discuss our decision.

- Page 19, line 28ff: Why is the comparison with instrumental data just done for the 1 SD reconstruction?

Response: We considered the fixed number 1SD reconstruction as essential part of the work and therefore did it only for this. We consider including equivalent Figures for the other two approaches in an appendix.

Concluding remarks

- Please avoid 1-sentence paragraphs

Response: We will do so.

Figures

- Generally, please think of a way to reduce the number of figures with time series. Both, the number of really necessary panels in each figure and figures in total. E.g. it is probably not required to see the annual resolution reconstruction for the full period for all three experiments or do multiple smoothing have to be presented?

Response: We will reconsider all Figures.

C19

- I find the uncertainty ranges often impossible to see (e.g. Fig. 2a). I cannot recognize the "envelope", you are talking of. As this a main focus of the paper, please try to find a way to plot uncertainty better visible, e.g. just a smoothed version for the entire period and a subperiod at annual resolution.

Response: We will reconsider all visualisations and try to put maximum emphasis on the uncertainty ranges.

C20

Referee 3

This study proposes a new climate reconstructions for Europe for nearly the full last millennium. The approach is based on the Analog Method, also known in the literature as Proxy Surrogate Reconstruction. One of the main novelties of this manuscript is how the authors extend the methodology to explicitly account for uncertainties. The authors present several reconstructions and compare them to the Euro 2k reconstruction, as well as independent data from the BEST project. Similarities, differences, advantages and caveats are discussed through the manuscript.

General comment

Most classical reconstruction methods produce a single reconstruction which does not explicitly account for uncertainty, although it is acknowledged that it populates this type of data-sets. This is problematic because uncertainty is not only ubiquitous, but it is heterogeneous both in time and space. This is an important limitation that precludes the proper assessment of the limitations of the knowledge we can gather from climate reconstructions. In this sense, I think this study is important and necessary to improve one prominent tool to produce such reconstructions, the Analog Method.

Response: We thank the referee for their evaluation and rating of our manuscript.

The design of the study is sensible, and I have mostly minor comments regarding details I could not fully understand and therefore might deserve clarification. Should not be for the issue I discuss below, I would recommend publication after minor revision.

Response: We thank the referee.

There is however an important aspect that has to be improved in the manuscript under the light of very recent bibliography published even after this discussion was started.

C21

There exists a published extension to the Analog Method that allows to estimate uncertainties. This is part of a recent publication with a more general aim (Neukom et al., 2019). There, authors briefly introduce and apply a methodology which largely differs from the one presented here, but that aims at the same purpose: explicitly assess uncertainties in climate field reconstructions with the Analog Method. I think this work should somehow account for the existence of this already published method. The level of modification applied to the manuscript depends on the authors. At the minimum, the differences between approaches should be discussed (for example, the approach opted by Neukom et al. (2019) does not produce missing values, being in principle an important advantage). At best, the approach adopted by Neukom et al. (2019) could be implemented here as well, and a comparison could be done between both methods. In my opinion, the latter would greatly improve the interest of this manuscript, but it is perhaps a major modification of the work that falls beyond its original scope. I leave it up to the authors and I would not be disappointed if they decide not to tackle this task.

Response: As the referee notes, we became aware of Neukom et al. (2019) after the discussion phase started. In view of their publication, we have to modify various parts of the manuscript. We will thoroughly discuss the differences between our approach and their approach.

At this point it is unlikely that we produce a bootstrap uncertainty estimate following Neukom et al.

The approach of Neukom et al. combines two sources of uncertainty. These are differing pools of candidate fields and differing proxy coverage. The former is to some extent included in our consideration of a set of fields, which agree with the initial uncertainty. The latter is not included in our approaches. Neukom et al. describe the uncertainty if we have less information available than we have. We describe the uncertainty due to the uncertainty in the proxy anchors.

C22

Minor comments

1. Page 2, Line 20: I think the correct citation is Gómez-Navarro et al. (2014)

Response: Gómez-Navarro et al. (2017) discuss the main differences between the analogue search and offline data assimilation approaches.

2. Fig 1: Maybe excluded locations could be shown with grey symbols, as well as the are representative for Central Europe. The location of these proxies is relevant for example to understand Figure 5.

Response: We make these modifications.

3. Page 4, Lines 28-29. I think it is more correct to say that, only when Varres and Varsig are uncorrelated, the total variance is the sum of both (because in that case the covariance term vanishes).

Response: This is what we intended to express. We will clarify our writing.

4. Page 5, Line 6: typo (modified)

Response: We thank the referee for spotting this.

5. Page 5, Lines 17–21: I do not understand where the 2.57 comes from. How it is related to the minimum number of 39 proxies? Please clarify.

Response: We clarify this in the revised manuscript. $2.57SD$ is equivalent to a 99% interval. 39 is the smallest number of analogues found at any date. We therefore later choose this as the size of our fixed number reconstruction ensemble.

6. Page 5, Lines 22-23: why is it the only one? why 2105 is special? why not $1.5SD_{noi}$?

Response: Considering fixed SD_{noi} intervals, the number of valid analogues increases. It may become soon unfeasibly large. We think that the 2105 analogues

C23

for a $1SD_{noi}$ interval are still reasonable. Therefore, we only consider a $1SD_{noi}$ interval for the fixed SD reconstruction.

7. Overall, in the two paragraphs aforementioned, it lies the core of the two reconstructions carried out. I think this is important, and it should be made more explicit that the two approaches represent different method used for real below. Perhaps this can be made more explicit with some structural element, such as an un-ordered list or similar.

Response: We try to clarify the methods section in our revised manuscript.

8. Page 6, Table 1: I assume this is exactly the correlation used to define the SD_{noi} in each proxy location, right? If so, this could be clarified in the main text (especially in section 2.1.2).

Response: Yes it is. We will clarify this in the main text.

9. Page 6, lines 6–8: The criterion to exclude two proxies is not very clear. What is meant by “relevant portion of variance”? In Fig. 5 we learn that the reconstruction in these sites is poor. Would it be better if these sites were part of the network. Surely the answer is yes. I understand that the amount of climate information we get is poorer than in the other locations, but still we could benefit for having some information. At worst, if the proxies were pure noise, it would not be necessarily worse than not having information at all. In other words, I think having poor information is better than having none, and it's not fully obvious to me why proxies should be excluded from the analysis based on relatively low correlation alone.

Response: We clarify this in the revised manuscript.

Already Luterbacher et al. (2016) excluded these proxies because they lack a clear temperature signal.

The referee is correct that, generally, a pure noise record should not be worse than having no information at all. However, this is not the worst case. The worst case would be a record which biases the distance measure in our analogue search towards

C24

a different state.

That is, we follow the common approach to only include proxies with a signal beyond a certain level. We are aware that this has been a controversial decision in the past but we regard it valid in view of past practices.

10. *Page 6 (but relevant for the whole study): why do you restrict the reconstruction to the period 1260 to 2003? The reconstruction could have been applied further back in time. The number of proxies varies in time, but this could be even beneficial for this study, focused on the validation of new methodologies. It would show how the estimates of the uncertainty presented here are sensible to a varying number of proxies. I feel that this choice has unnecessarily limited the scope of the manuscript.*

Response: We thank the referee for their confidence in our approach. The referee is correct that in principle there are no reasons to stop in 1260. However, stopping there, in our opinion, eases the interpretation of results since thereby only an equal number of proxies enters the uncertainty estimation.

We will clarify this in the revised manuscript.

11. *Page 7, Table 2: it could be interesting to write the total number of analogues, i.e. the pool size. It would make more meaningful the number of proxies used to produce ensembles. For example, having 817 analogues (as in Fig. 8) has a clearer meaning when you add that they are 817 out of, let's say, 25000. It shows that you are still selecting a relatively minor number of relatively good analogues.*

Response: We clarify this in the revised manuscript.

12. *Page 7, Lines 6–7: I think having a consistent bias through the pool is not necessarily good, as it seems to be implied by the wording. It ensures that the bias are translated into the reconstruction. This is partly avoided using structurally different models to build the pool. I do not mean that the authors should necessarily rebuild the reconstruction with a larger set of models, but I think that at least they should not imply*

C25

that using a single model is somehow beneficial.

Response: We clarify this in the revised manuscript.

13. *Page 8, Fig. 2: I think a line marking the 0 K anomalies would help to read the series. This pertains mostly panels b and c, where the sign of the anomaly is important, but difficult to appreciate without such a line. This argument applies to Figs 6 and 7 as well.*

Response: Our revisions generally try to make our Figures clearer. We do add a zero line to panel c) but are yet undecided about other panels, because it even more increases the number of elements per panel.

14. *Page 8, Fig. 2: It's not fully clear to me what this figure (as well as Figs 6 and 7) show. Does "summary" mean spatial average?*

Response: "Summary" means summary of the main results of the reconstruction. We modify the captions.

15. *Page 9, Line 10: please change "degree Kelvin" to "Kelvin". Please review it, as there are other locations where I saw this in the manuscript.*

Response: We do so.

16. *Page 9, Lines 9–16. The order of these two paragraphs can be exchanged. It's a bit unusual and therefore confusing to discuss Fig. 2c before Fig. 2b.*

Response: We restructure the section in our revisions.

17. *Page 10, Lines 10–15: I think the fact that the reconstruction underestimate the intra-location variability is a problem of the pool, not the Analog Method itself. Do the authors think that this could be improved if higher resolution models were used to build the pool?*

Response: We did not investigate this feature. We only state it without attributing it to

C26

the method. There are a number of explanations on which we only very shortly touch here. First, the noisy proxy series may overestimate the true intra-location variability. Second, the simulations may be too smooth in space. This, thirdly, might be due to the low resolution and simulations with higher resolutions might help then. Fourth, the chosen distance measure may result in such a feature dependent on the characteristics of the simulation pool, which however should usually not be the case.

We clarify our statement in the revised version.

18. Page 11, Fig. 3: *The list of locations in the caption is misleading (the name and the ID are written all together). It seems a detail, but it puzzled me for a while until I realised that Tor92 and Torneträsk are not two proxies, but the ID and the name of the same one. You could easily remove this by using for instance parenthesis to separate name from ID or vice versa.*

Response: We clarify the caption.

19. Page 11, Line 26: *“The general agreement between the Euro 2k and the analogue. . .” this reads odd at this point, as the reader does not know where to find the information the authors are referring to. It turns out that this comparison is introduced later, in Figure 6 in Page 14.*

Response: We try to clarify in our revisions what is meant at this point.

20. Page 15: Lines 17–23: *The reduced variance could be quantified (how much is notably smaller variance in Line 19?). Further, the lost of variance when more analogues are considered is common in this approach, and generally in any statistical approach, i.e. there is a bias-variance trade off. It could be noted here that this has been comprehensively discussed in the bibliography of the Analog Method.*

Response: We will make these discussions clearer in the revised version.

21. Page 15, Line 32: *do the authors have a theory on what could be the reason for such systematic differences? Are they meaningful, can they be used to discuss*

C27

merits or problems in the reconstructions? Or are they rather low-frequency random fluctuations highly sensitive to method parameters?

Response: In the case of the mid 16th century deviation there are indications that the Euro 2k more validly captures the extremes in this period (Wetter and Pfister, 2011, <https://doi.org/10.5194/cp-7-1307-2011>, 2013, <https://doi.org/10.5194/cp-9-41-2013>), which may again indicate a period where the simulation pool is insufficient. Generally though, we would assume that it is mainly random due to the different sensitivities of the methods.

22. Page 16: Lines 25–26: *The presence of missing values in years with volcanic eruptions is a major caveat of the method, as those are typically the years most interesting in climate studies. Here it would be specially relevant my comment about a comparison with the method presented by Neukom et al. (2019).*

Response: We will discuss this.

23. Page 16, Lines 27–31: *I do not see why it is “unsurprising” this lack of analogues for the recent period. The pool contains this warming as well, so the search should not present more problems for this period than in any other.*

Response: We will clarify this.

Indeed, the pool includes this period but we do not only require a similar mean state but also a similar interrelation between locations, which makes it more likely that the limited size of the pool does not include such a case.

24. Page 19, Lines 18–20: *Maybe I’m miss-evaluating this, but I think that anchoring the reconstruction within a range of 8 K is a poor result. It shows that the 800 analogues are indeed poorly constrained in this region, so we have little idea of how the actual climate was in that period and region. More generally, I have the concern that the spread shown for example in Fig. 7 might provide an optimistic measure of the actual uncertainty. Fig 7e for instance shows the range in the spatial average, which is about 2*

C28

K. But this is after spatial average, where regional differences can cancel out! I wonder how large is the range in each location. This might perhaps be illustrated with a map of (temporally averaged) ranges? Eventually, my guess is that using as many as 800 analogues or more, really far away from "the best" is, as outlined by the authors, too much.

Response: We consider visualising the mean and the range of the temporal temperature ranges.

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