

Interactive comment on “Spatio-temporal variability of Arctic summer temperatures over the past two millennia: an overview of the last major climate anomalies” by Johannes P. Werner et al.

Johannes P. Werner et al.

johannes.werner@geo.uib.no

Received and published: 21 June 2017

General Comment

(paraphrased) No scientific question explicitly posed, article lacking focus.

Reply:

We will remove the Arctic Amplification chapter, which will result already in a more coherent manuscript, making connections between the different chapters much clearer.

C1

[Printer-friendly version](#)

[Discussion paper](#)



After revising the reconstruction we will also try and make stronger statements about e.g. the spatial consistency of the warming and cooling episodes.

Other comments

- Abstract, lines 1-2: I think you need to be a little more clear about what is actually unique here. I understand that it is BOTH spatially resolved AND millennial in length, though there are several reconstructions that are one or the other.
- R: we will strengthen that statement in the final version of the paper and stress that it is both the spatial character and temporal aspect (as guessed correctly)
- Paragraph including lines 53-60: This paragraph comes across as a kind of special attack on the glacier advances work in a tone that I'm not sure the authors intended. This summer temperature reconstruction (with skill primarily over Europe) is really different than a glacial reconstruction given the memory of glaciers, the different seasons and climate factors a glacier is responding to, etc. So I don't think a clear declaration against that work is necessarily warranted.
- R: We apologize for the perceived tone. We did not mean to criticize glacial reconstructions from moraines in general and cosmogenic dating in particular. The sentence will be removed as it is also based on a misinterpretation of the results by Young et al. (pers. comm. from Young set this straight).
- define LOC
- R: it stands for "local regression", though mostly when the method is referred to in other articles the abbreviation LOC seems to be used. Still, we will change this.

- Lines 124-125: Need a more specific criticism here or not discuss the issue at all. What constitutes a “strange” correlation? And on what firm basis can you reject the use of the BE product?
- R: the issue is that the correlation between grid cells as a function of the distance (both chordal and orthodromic) is very long (10 000 km) and contains oscillatory parts (not as obvious in the attached figure as in other evaluations). Without analysing any details of the regriding method used in the BEST data, this looks too much like an artefact of an expansion in spherical functions. Truncating after a certain order can lead to spurious oscillations on the sphere. See the attached figure.
- Lines 147, 444: Question mark issues.
- R: missing reference in the bibTeX file. Will be fixed
- Lines 199-202: How are the response parameters being determined? Do your reconstructions happen to take account of the specific choice of parameter values?
- R: The parameters are estimated using the described Gibbs sampler. The reconstructions are conditional on the estimated joint distribution of all parameters (proxy response and climate field). That is, it explicitly takes the uncertainty in the parameters into account.
- Lines 211-222: I think it’s important to note that only Europe has spatially coherent skill, otherwise it’s fairly patchy skill (at least in my reading of Fig. A1).
- R: The skill shown in Figure A1 depends on the length (and quality) of the instrumental data in the specific grid cell. The data coverage over Europe is highest (space and time), other regions (absence of colour in Fig. A1) are less well covered (thus the patchy appearance). Any estimates relying on short time series

[Printer-friendly version](#)[Discussion paper](#)

are thus to be interpreted carefully. This is what we mean by “Thus, these results not only reflect a possibly weak reconstruction but more likely the lack of actual instrumental data to construct any meaningful comparison statistics over the validation period.” (l. 550-551)

- Figure A1: It wasn't clear to me what is meant by “CRPS CE” and “CRPS RE”? As a related issue, CRPS is challenging to interpret because it doesn't have a reference. Perhaps use the skill score version of CRPS that takes account of a reference (e.g., your prior)?

R: We did indeed skim over this issue. We will modify the last two paragraphs accordingly: “Additionally the skill of the reconstruction beyond forecasting the calibration or validation period mean is evaluated. In palaeoclimate reconstructions this is often assessed by the Coefficient of Efficiency and the Reduction of Error statistics (Cook et al., 1994). However, these are not proper scoring rules (Gneiting and Raftery, 2007) and should thus not be used analysing the results of a probabilistic reconstruction method.

To generate a similar statistic, ensembles of surrogates for each location with instrumental data are constructed using the mean and standard deviation over the validation interval from the reconstructions. For these, the $CRPS_{pot}$ is calculated, comparing the surrogates against the instrumental target. This value is then subtracted from the $CRPS_{pot}$ over the calibration (validation) interval, resulting in $CRPS_{pot-CE}$ ($CRPS_{pot-RE}$). As with the CE (RE) a value above zero shows a skilful reconstruction, i.e. a reconstruction that performs better than the climatology over the calibration (validation) interval.

About half of the grid cells with instrumental data have a $CRPS_{pot-CE}$ and $CRPS_{pot-RE}$ that is above zero – and these grid cells are actually also those that have the longest instrumental time series (inside and outside the calibration interval). Thus, these results not only reflect a possibly weak reconstruction

[Printer-friendly version](#)[Discussion paper](#)

but more likely the lack of actual instrumental data to construct any meaningful comparison statistics over the validation period.”

- Fig 2b: Why is there so much precision right up to the end of the calibration interval, but a complete loss of annual precision from 1980 to the present? Are important proxies dropping out here?
- R: The precision appears indeed relatively low, and Fig.2b greatly emphasises this over Fig.2c, especially since we have the "calibration" interval so prominently in there. One issue is that "calibration" is not the entirely correct term, as the reconstruction over this period is in fact mostly determined by the instrumental data (though not entirely, not the relatively high instrumental uncertainty and the spatially sparse coverage). This issue can be addressed by doing what is called a predictive run (see Luterbacher et al. 2016, or Tingley and Huybers 2013), which would in turn give another means of evaluating the reconstruction quality. However, as can be seen from 2d, this is also very likely a proxy availability effect.
- Figure 5: no color on the color labels
- R: These were present in the initially uploaded pdf, we will look into the technical issues behind that. Most likely once the original artwork is uploaded and the LaTeX process takes place at Copernicus this issue will disappear.

Typos / grammatical issues:

- Lines 458-459 “which gets sparser going back in time”
- Line 5:14 “used for these chronologies”

Thanks for catching these!

Figure 1: Correlation between the gridded instrumental series of July SAT of the BE dataset as a function of the orthodromic distance between the grid nodes. Nodes located within 45-90N were used in the analysis.

Interactive comment on Clim. Past Discuss., <https://doi.org/10.5194/cp-2017-29>, 2017.

CPD

Interactive
comment

Printer-friendly version

Discussion paper



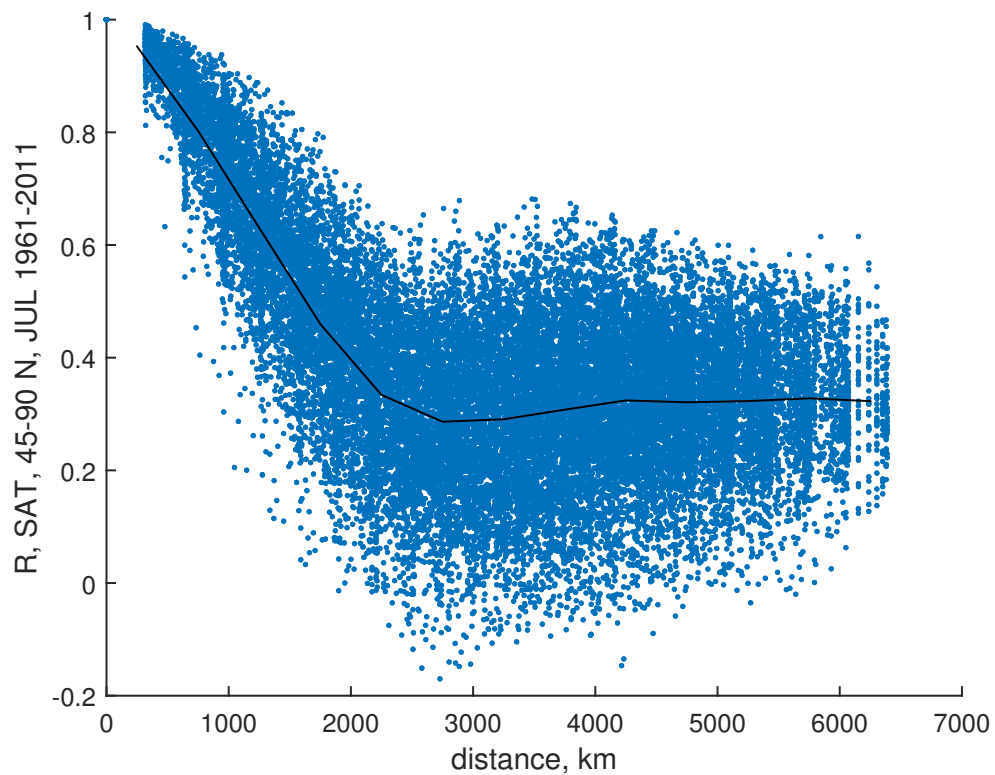
[Interactive comment](#)

Fig. 1.

[Printer-friendly version](#)[Discussion paper](#)