

Review of

‘Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium – Part 3: Practical considerations, relaxed assumptions, and using tree ring data to address the amplitude of solar forcing.

by A. Moberg, R. Sundberg, H. Grudd and A. Hind

### **Recommendation: minor revisions**

This manuscript extends the work on a statistically sound framework to compare climate simulations with proxy data published by Sundberg et al. (2012) (SUN12). In my view this earlier paper constitutes a substantial and much needed step forward in paleoclimatology, and further studies that address practical questions when applying the SUN12 framework and that use it to answer open questions in climate science are in principle well justified. The current manuscript falls into this category. It provides two methodological developments, namely the possibility for autocorrelation in the simulated series, and the direct comparison of two forced simulations, as well as an application to a set of GCM simulations and 15 tree ring records for 1000 AD – 1849 AD to investigate whether the comparison supports weak or strong solar forcing. Although the results of the application are inconclusive, potentially because not enough proxy data have been used, there is enough new content to warrant publication. The writing style is in general good but I have some points that I ask the authors to change or clarify.

### **Specific comments:**

1.)

The extended framework allows for autocorrelation in the simulated series. However the proxy data have autocorrelation too. Why is this not included in the framework? If it is not needed for reasons that have to do with the construction of the test statistics this should be explained.

2.)

p2692, end of abstract: an additional reason for the inconclusive results could be the small ensemble size for the simulations; in an average over a larger ensemble the signal to noise ratio would get higher. A short comment on this should be added.

3.)

p2631, line 4: clarify whether these are reconstructions for local or larger-scale temperatures (I presume it is the former).

4.)

p2632, last paragraph: The calibration needed for the unbiased ranking is in other contexts, e.g. in statistical downscaling, known as ‘inflated regression’. If the goal is to estimate  $z$  from  $\tau$  this is fundamentally wrong, as pointed out for instance by von Storch (J. Clim 1999). In order to avoid confusion it should be pointed out that there is no conflict between the different statements on inflated regression, as the context is different.

5.)

p2639, second paragraph: It is not clear whether the instrumental error mentioned here is the error in the actual measurements, which is what the name implies, or the error in the gridded data, which I presume it is, because these are used as ‘instrumental records’ in the analysis. Please add a comment to avoid confusion.

6.)

P2640, line 26: ‘separate the climate signal from the raw data’ is not precisely phrased. Standardization tries to remove non-climate-related low-frequency variability, but there is still a lot of non-climate-related high-frequency variability left in the standardized tree ring records, so it is not correct to say that the climate signal is separated (or extracted,, which I think is what the authors meant (interpreting ‘separated’ in the usual way would mean that the end result is the climate signal and the raw data)).

7.)

p2642, last paragraph, p2643 first paragraph: I’m not convinced by the arguments given for selecting the size of the region for calibration. Individual tree-ring records respond to the local climate, and therefore, as pointed out by the authors, they are less closely linked to large-scale climate (the correlation gets lower). I would expect that this effect compensates the fact that large-scale temperatures have a higher ratio of forced to unforced variability in an analysis that tries to decide which of two different climate model forcings leads to better agreement with the proxies.

Although correlation maps between tree ring records and temperatures similar to those given in Fig. 1. (see next comment) are a useful guidance for the choice of area I encourage the authors to give a conceptually and if possible statistically more sound discussion of the question of which area size to choose.

8.)

page 2643, first paragraph: basing the correlation maps on first differences seems fundamentally wrong and is not consistent with the argument given. It is true that correlation are influenced by strong trends and it is advisable to remove this effect. This can be done in a straightforward way by de-trending the data. Although using first differences also removes the effect of trends the resulting time series are in principle the derivatives of the original series, and the correlations measure the link between the derivatives. It remains to be seen whether the standard correlation maps look similar, but even if so there is no justification for using first differences.

9.)

page 2468, lines 13-15: This comment is a bit surprising as GHG and orbital forcing are both small for the period 1000 AD to 1850 AD.

10.)

Page 2652, line 29: There should be no question mark as this sentence is a statement not a question (the fact that even native speakers sometimes use a question mark in this sense doesn't make it better).

11.)

Comment 1.) applies here again, i.e. why is there no autocorrelation in  $\eta$  ?

12.)

The comments on the forcing are confusing. If I understand correctly the effect of the forcing is contained in  $\delta$  and  $\eta$ , so these terms are not uncorrelated with the forcing as in SUN12.

13.)

page 2655, part on correlation statistic: I did not fully understand this part and it seems several important details in linking the different equations are not properly explained. Please check this part carefully and explain intermediate steps in sufficient detail so the line of argument can be followed by the typical reader of Climate of the Past.

14.)

Page 2657, part on  $D^2$  test statistic: Again the explanation of this part is rather short.

15.)

page 2658: why are there two types of outer brackets (curly ones and big standard ones)?