

Reviewer 2

Review of ‘**The importance of input data quality and quantity in climate field reconstructions – results from a Kalman filter based paleodat assimilation method.**’ by J. Franke V. Valler, S. Brönnimann, R. Neukom, and F. Jaume Santero

Recommendation: minor revisions

This manuscript analyses the effect of the choice of input data in a Kalman filter based assimilation method for climate proxy data. Given the increasing use of data assimilation methods in paleoclimatology this is a practically relevant question. The climate reconstructions derived by assimilating tree ring width data for the period 1901 – 1970 AD are evaluated against gridded observations for near surface temperature, sea level pressure and precipitation. The best reconstruction skill is found when using input data that are a compromise between good spatial coverage and selecting only proxies that are well linked to local temperature and precipitation.

Overall this is a good paper that provides relevant information for the paleoclimate community, and in principle I support publication. However, several points should be explained better or should be modified. They are listed below.

We appreciate that the reviewer considers this study to be of interest to the paleoclimate community.

Specific comments

1) The period which is analysed should be stated in the abstract. At the moment the first time this information is given is in line 130. The abstract should also briefly mention the type of assimilation method.

This information on the method and time period will already be added to the abstract.

2) Line 21, ‘improved’ relative to what? This does become clear later in the text, but the abstract needs to make sense on its own.

Will be changed to “...but fail to provide information for other regions and other variables”.

3) The paper links the results in several places to terms in the Kalman filter, namely to the proxy system model, the observation error covariance matrix and the background error covariance matrix. These are important comments, but readers who are not experts in data assimilation will probably not understand them, because the Kalman filter equation that is used is not given in the paper. I do appreciate that these details are given in previous publications, but with respect to its core elements a paper should be self-contained. Please add the equation to the method section and discuss there how the terms are calculated and how information is spread by the various terms from the proxy data to the different reconstructed meteorological variables. When presenting the results please refer back to this discussion where appropriate.

We will add more details about the method in the revised version to allow reader to understand this paper without reading previous publications about the method.

4) Line 92, 'we need a forward model that simulates them in the model state vector'. This is not well phrased.

Will be rephrased.

5) There should be clear comments to what extent the findings can be transferred to other data assimilation methods used in paleoclimatology. It is likely that methods with a similar structure, i.e. using PSMs and variations of Kalman filters, will have similar sensitivities to the selection of input data, while others, for instance particle filters, may not.

We think that the quality/selection of input data has similar consequences not only in other data assimilation methods but in statistical reconstruction, too. However, there will be method dependent differences. We will add a paragraph on this topic to the discussion.

6) Line 57-59. The statement on the similarity between the method used in this paper and the method used in the last millennium project is misleading. The method used in the paper uses a 'transient offline method', in which the background state is time-dependent due to the signal of the external forcing. This aspect is actually highlighted by the authors in lines 120-124 and lines 231-232. In contrast, the last millennium project uses a 'stationary offline method' in which the background state does not depend on time. This crucial difference should be mentioned.

We will add this information already here in the introduction to avoid confusion.

7) Lines 121-122. The comments on low-frequency variability should include a discussion of the setup for the simulations that provide the background state. Why is sea surface temperature mentioned as a forcing? Are the simulations done with atmosphere-only GCMs? If so, which sea surface temperatures are used? These comments should also discuss the role of random, internal, low-frequency variability.

As mentioned above, we will extend the methods section that it allows the reader to understand the entire study without reading previous publications on the method before.

It should also be clarified that the validation measures are calculated from annual values and are thus dominated by inter-annual variability. This fact and the short evaluation period imply that an evaluation of low-frequency variability is not possible in this study.

Yes, as explained in the answers to the first reviewer's comments, our methodology is not suitable to draw conclusions on the proxy data sets' influence on centennial-scale climate variability. This study is complementary to the results found by Tardif et al. (2019) on low frequency variability effects caused by the input data selection.

8) Line 125, it is not clear from which data the running mean is calculated and what 'model' refers to.

As mentioned before, we will explain the method in greater detail in the revised manuscript.

9) Line 131, ‘just at correlation itself’ sounds strange.

Will be corrected to “Instead of analyzing absolute correlation coefficients, we analyze correlation improvements ...”

10) Lines 137/139/264, ‘punishes’ should not be used in scientific writing.

Will be replaced by “penalize”.

11) Line 138-140, please include a more detailed justification of why the evaluation is based on ensemble means rather than on individual ensemble members followed by averaging of the skill scores. This should include explicit statements on the effect of the reduced variability in ensemble means on the RE; the current statement is unclear.

Thanks for this suggestion. You are right that we can expect reduced variability in the ensemble mean compared to the ensemble members and compared to the validation data, too. Nevertheless, most user will be interested in the ensemble mean and its skill. One would expect that the ensemble mean of the transient simulations before assimilation has little variability and that the assimilation would lead to an increase in variability bringing it closer to the observations. This is also what we observe with the generally positive RE skill scores. This way, a perfect skill score of one could only be reached, if the ensemble would have no variability left and perfectly matches the observations. We do not expect that the pattern or sign of the skill score would change in case of averaging the skill of the single members. However, the skill scores of the single members and hence also their mean could be expected to be larger than the skill score of the ensemble mean because variance is not systematically underestimated (see Fig. 6 in Bhend et al., 2012). We will test this and justify in the revised text why we present the results in one way or the other.

12) Line 227, ‘TRW limitations remain the same’ is not clear.

We will clarify this by writing: “TRW limitations are time-independent”.

13) Line 254, it should be ‘principal’ not ‘principle’

Will be corrected.

14) The use of hyphens is inconsistent and often wrong. Adjectives that are constructed from two words should usually be hyphenated. Examples are ‘temperature-sensitive’, ‘regression-based’, ‘time-dependent’ (which is better than ‘time-variant’ used in line 113), ‘low-frequency’ (if used as an adjective), ‘inter-annual’, ‘multi-decadal’, ‘multi-variate’ etc. In some case it is also correct to combine the two words, e.g. ‘multivariate’

Thanks for the explanation, we will check the entire document.

15) Line 195, replace ‘lost’ with ‘lowest’

Will be corrected.

16) Lines 212-213, This is not a proper sentence.

Will be rephrased.

17) The paper Matsikaris, A., Widmann, M. and Jungclaus, J., 2016. Influence of proxy data uncertainty on data assimilation for the past climate. *Climate of the Past*, 12(7), pp.1555-1563. addresses similar questions and should be included in the introduction and/or the discussion.

We will add this reference and report/discuss their findings.