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.

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.

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.

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.

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

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.

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.

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.

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.

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

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

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

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.

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

13)
Line 254, it should be 'principal' not 'principle'

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'

15)
Line 195, replace 'lost' with 'lowest'

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

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.