

## Reviewer #2

**This manuscript identifies the best choices from a number of different spatial and temporal localization approaches and from different inflation techniques for the background error covariance matrix in Ensemble Kalman Filters used in paleoclimatic applications. The optimization of these technical details in data assimilation is important for the growing paleoclimate data assimilation community. The results are systematically derived and the manuscript is in general well written. I support publication after the points listed below have been clarified or corrected.**

We thank the Reviewer for the careful revision of the manuscript. We will follow his/her recommendations and add further details at the indicated points to provide better understanding how ensemble Kalman filter methods work, how we set up the blending experiment and how covariance inflation techniques can help to better estimate the background-error covariance matrix. We are also thankful for improving the English of the manuscript.

### Specific comments

**Page 1, Line 2, replace ‘of the assimilation system’ with ‘of some assimilation systems’**

We will replace it.

**Page 1, Lines 17/18, 24/25, ‘boundary conditions’ are specifications of state variables at the boundaries of a model domain, and thus not the same as ‘forcings’, which are external influences on the system. The two should be distinguished throughout the text. It seems that here the statement are about forcings. If so, reformulate avoiding the use of ‘boundary condition’.**

Will be replaced as:

Climate models constrained with realistic, time-dependent external forcings provide fields that are consistent with these forcings and the model physics.

If information from observations is optimally blended with climate model simulations, the result is the best estimate of the climatic state, given the observations, given the external forcings, and given the known climate physics.

**Page 2, line 6, ‘linear models’ of what? I think it should be ‘linear dynamical systems’. A short comment on why KFs are used with non-linear systems, including GCMs, would be good. ‘Gaussian distributions’ of what? The state variables?**

Yes, the KF is optimal for linear systems (linear dynamical model and linear observation operator) with Gaussian error distributions (model, background, and observation). In atmospheric data assimilation the most commonly used KFs are the ensemble-based Kalman filters that can handle some non-linearity in the dynamics and in the observation operator. Moreover, in our approach we use an offline implementation of the ensemble square root filter, i.e. we only adjust the precomputed model simulations with the observations and never deal with the question of model dynamics. Nevertheless, we will remove this sentence because it is not necessary here and just caused confusion.

**Page 2, line 13-16, I suggest using ‘stationary offline’ and ‘transient offline’ for the two approaches.**

Thank you for the suggestion. We will use the terms “forcing-independent offline” for “stationary offline” and “forcing-dependent offline” for “transient offline” to better distinguish between the two methods.

**Page 2, lines 17/18, ‘The true climate state is not known, therefore it has to be estimated’. Does ‘it’ refer to ‘the true climate state’, as the sentence suggests or to ‘the uncertainty of the background state’, which would link better to the first sentence in this paragraph? This sentence should be clarified, or it could simply be deleted (which I think is the better option).**

We will delete the sentence.

**Page 2, line 20, What is a non-simplified KF? A KF with a ‘true’ background error covariance matrix? If so, how can this exist? The background error covariance has always to be estimated somehow. Please clarify the statement.**

We will replace ‘Ensemble-based KFs are simplification of the KF’ to ‘Ensemble-based KFs are approximation of the KF’.

In the Kalman filter, the background-error covariance matrix is given as

$$P^b = \overline{(x^b - x^t)(x^b - x^t)^T}$$

the superscripts b and t indicate the background state and the true state, while the overline marks the expectation value.

In practice, as was said, the background-error covariance matrix has to be estimated. In ensemble-based Kalman filter techniques, the ensemble covariance will provide the background-error covariance matrix and the true state is estimated by the ensemble mean.

$$P_e^b = \overline{(x^b - \overline{x^b})(x^b - \overline{x^b})^T}$$

**Page 2, line 23, It seems that the sampling error for the background error is not only a random error, but leads to a systematic underestimation of the background error, otherwise inflation would not be a suitable approach. Please explain better.**

Ensemble-based filters with flow-dependent covariances can diverge due to sampling error. Our small 30 member ensemble may result in an error underestimation, because we do not assess the uncertainties in SST boundary conditions and external forcings. In an online DA system if the error of the analysis is underestimated and it is forwarded to the next time step, it will result in giving less and less weight to the observations in the following time steps. This should not be a problem in offline approaches. However, we conduct this experiment for completeness because as we say in the manuscript, there are two commonly used methods to reduce the negative effect of sampling error: inflation (e.g., Anderson and Anderson, 1999), and localization (e.g., Hamill et al., 2001) of the background-error covariance matrix.”

**Page 2, lines 25/26. The statement on distribution of ensemble members refers to online approaches, but the approach used by the authors is an offline approach. This is confusing. Please briefly explain how the ensembles are generated in an online KF, and that in offline approaches the ensemble is given, but that the background error covariance still needs to be inflated.**

To avoid any confusion, we will change the sentence to: "Two methods are commonly used in online KF approaches to reduce the negative effect of sampling error:" How the ensembles are generated in an online KF is already described on page 2, line 8-11. This remains the same in ensemble-based KFs: the ensemble members are updated with the observation when they become available and the updated state, in case of ensemble-based KFs, the updated ensemble members and their uncertainties are propagated forward by the model to the next time step. We will add a short comment here that these steps are valid for ensemble-based KFs as well. Finally, with regard to the need to inflate the background error covariance, please see our answer to the last comment on page 2, line 23.

**Page 3, line 14, replace ‘other method’ with ‘additive method’**

We will replace it.

**Page 3, line 25, replace ‘form’ with ‘from’**

We will replace it.

**Page 3, line 28, Don’t use ‘forced by boundary conditions’, as forcings and boundary conditions are different (see comment above). If I understand correctly for all ensemble members the same greenhouse gas, solar and volcanic forcings have been used, as well as the same SST boundary conditions. Please clarify.**

Yes, the same forcings and boundary conditions were used in the 30 members.

Will be replaced as:

The 30 ensemble members were forced and driven with the same external forcings and with the same boundary conditions.

**Page 3, Lines 29-31, The SST reconstruction can be expected to strongly influence the results of this data assimilation approach with an atmosphere-only GCM. There should be some comments on how the SST reconstructions have been made, what is known about their uncertainties, and why this approach is taken rather than data assimilation with a coupled atmosphere-ocean GCM.**

The main reason why we use the SST reconstruction by Mann et al. 2009 is that it was the only available reconstruction at the time when we ran the simulation ensemble, like we already explained in the paper: "For sea-surface temperatures (SSTs), which have a particularly large effect on the simulations, the reconstruction by Mann et al. (2009) was used. This is the only global gridded SST reconstruction that dates back till 1600."

We already describe limitations of the reconstruction and how we tried to improve them in the discussion paper, too: "The SST reconstruction by design captures interdecadal variations (Mann et al., 2009), hence intra-annual variability dependent on a El Niño/Southern Oscillation reconstructions (Cook et al.,2008) was added to the SST fields." We will add a reference to the paper of Bhend et al. 2012, which explains the simulations in a little more detail.

There are multiple reasons why we do not use a coupled atmosphere-ocean model:

1. We understand a paleo-reanalysis to be a product that best describes the atmospheric states of the past globally. 400 years of just externally forced coupled atmosphere-ocean

simulations would not be in close agreement with the true state of the ocean in the past and e.g. not produce El Niño events in the years when they actually occurred. Therefore such simulations would not capture related teleconnections patterns in the corresponding years, either. We cannot gain the same information that currently comes from the SST boundary conditions with our data assimilation setup because we only assimilate absolutely dated tree-ring proxies on land and mostly at mid to high latitude. It would be an option to assimilate coral data in the future but there are few records that back to 1600 and their assimilation has not been tested yet. That is why coral data assimilation is not an option for the presented sensitivity experiments. The focus of this study is to show, how the paleo data assimilation setup published by Franke et al. 2017 can be further improved without introducing further changes that would blur the effect of the improved covariance estimation.

2. If we would like the ocean to be in closer agreement with the observations in a coupled model, we would need to do an online assimilation scheme that could take the longer memory of the ocean into account whereas our atmospheric states have no memory that extends beyond a months or year when the next proxy observation becomes available. Hence, running a coupled ocean-atmosphere model would not allow us to work with our offline assimilation system and do all the sensitivity experiments, which we present in this study.

3. Finally, an ensemble of coupled simulations plus the additional spin-up time for the ocean to reach equilibrium would have simply not been computationally affordable.

**Page 3, line 30, I think ‘till’ should not be used in formal writing and should be replaced with ‘until’ (also later in the text).**

We will replace it, also in page 4, line 10-12

**Page 4, line 1, replace ‘boundary conditions’ with ‘forcing’, and make a separate statement on land-surface boundary conditions, including which variables are prescribed.**

Will be replaced as:

Further forcings include solar irradiance, volcanic activity, and greenhouse gas concentrations (for more details see Bhend et al., 2012; Franke et al., 2017). The land-use reconstruction by Pongratz et al. (2008) was used to derive the land-surface parameters.

We would suggest only mentioning that the land-use reconstruction by Pongratz et al. (2008) was used in the paper because we would not like to give very different levels of detail for each forcing or boundary condition. The reconstruction by Pongratz is the standard land-use reconstruction used in basically all ECHAM model simulations since the reconstruction was published. For the Reviewer interest, the 14 land-cover classes defined by Pongratz et al (2008) were mapped to land surface parameters closely following the procedure of Hagemann et al. (1999).

**Page 4, line 16, ‘CCC400’ has not been introduced**

In the title of section 2.1 we try to indicate that the model simulation is called CCC400.

It is not an abbreviation but the name of the model simulation.

To make it more clear we will replace page 3, line 25-26 as:

It uses a 30 member ensemble of atmospheric model simulations (CCC400) as background to reconstruct monthly climate states between 1600 and 2005.

**Page 4, line 23, If I understand correctly the deviations of the ensemble members from the ensemble mean are updated in the online EnSRF according to equation 2, but not in the offline EKF, which uses an existing ensemble. Please clarify.**

In both online and offline EnSRF methods, the mean and the deviation from the mean are updated, according to Eq. (1) and Eq. (2). However, in an online application the analysis and its uncertainty are propagated forward to the next time steps. In case of an offline approach the process stops after all available observations at the current time steps have been assimilated.

**Page 4, lines 28-30, the statements on the 6 month periods are partly redundant.**

We will rewrite it.

**Page 5, lines 5-7, How have the error variances been chosen? Should  $\sigma^2$  be  $K^2$ ? If so, why is the error for documentary data smaller than for instrumental data? Which multiple regression?**

The error variances are rough estimates that include for instance measurement uncertainties, temporal inhomogeneities, and the fact the a station is not representative for a grid cell (see Franke et al. 2017 and Frei 2014).

The error variance of documentary indices without units is set to 0.25 standard deviations<sup>2</sup>. Because in the experiments described in this paper, we did not assimilate documentary data, we will delete the information regarding to this source in the revised paper.

Will be replaced as:

We set the error variances of instrumental temperature observations to  $0.9 K^2$ , and of instrumental pressure data to  $10 hPa^2$ . The errors of tree-ring proxy data are calculated as the variance of the multiple regression residuals of the H operator.

With regard to “multiple regression” we write in section 2.2 that the H operator was designed as a multiple linear regression model to transform the model state to proxy records. We did not explain it in details only referred to the paper by Franke et al., 2017 (page 5. line 2-3). We will add further information about the H operator.

**Page 5, lines 16 – 19. The notation is not clean. In line 16 it is said that R is a diagonal matrix, in line 19 that R is a scalar. The problem is that the same notation is used for an equation using the full set of observations (where R is a diagonal matrix) and for the equation when the individual observations are assimilated sequentially. Please reformulate.**

It is true that the same notation is used for describing the whole system or the serial approach, but in other papers (including the original paper by Whitaker and Hamill, 2002) the notation remains unchanged. Therefore, we would like to follow the common usage of the notation and keep it as it was in the submitted paper.

**Page 6, line 1, replace ‘localization function’ with ‘the localization function’.**

We will replace it.

**Page 6, line 19/20, replace ‘additive inflation’ with ‘the additive inflation’, and ‘hybrid’ with ‘the hybrid’.**

We will replace it.

**Page 6, line 23-25, The explanation is confusing. One can select ensemble members for the whole period or some or all ensemble members for some time steps; how exactly are the climatological state vector and the associated error covariance matrix calculated? The simulations have already been performed; why are there substantial computational costs for using a large number of ensemble members?**

The climatological state vector is created as, 1) we define how many members  $x^{\text{clim}}$  should have. We used 100, 250, and 500 members to create  $x^{\text{clim}}$ . 2) If the climatological state vector should have e.g. 250 members then 250 years are randomly selected from the whole model simulation, between 1601 and 2005. 3) Every year has 30 members from which one is randomly selected and kept. 4) The chosen members are combined in the climatological state vector.

The climatological background-error covariance matrix is calculated from this ensemble, by using the ensemble perturbations.

The climatological state vector is randomly resampled after every second assimilation cycle. The increased computational cost partly comes from the creation of the climatological state vector. The other time consuming part comes from the updating of the climatological part after each observation is assimilated. (The standard way when observations are assimilated serially).

**Page 6, line 29,  $H^T$  is at the end of all terms in the equation. Can it not simply be deleted?**

Yes, it is possible to delete them.

**Page 7, line 4-5. It is not clear how  $x^{\text{clim}}$  is calculated and updated, what  $n$  is, and what ‘propagated’ means in an offline assimilation scheme.**

We will give a better explanation how  $x^{\text{clim}}$  is obtained and define  $n$  as the ensemble size used to create a climatological state vector.

We assimilate observations serially, that is observations are processed one at a time. After the first observation is assimilated and the analysis is obtained, this analysis field will become the background state for the next observation. On Figure 2 the arrows pointed from the analyses to the background state vectors are meant to indicate this process. Propagating and updating also refer to this process, that the information from the observation is also incorporated into  $x^{\text{clim}}$  and not only in  $x^b$  in most of our experiments (see Table 2).

We will reformulate the sentences to avoid confusion.

**Page 8, line 7, replace ‘isotropic localization’ with ‘an isotropic localization’.**

We will replace it.

**Page 8, line 15, ENH is not defined.**

In page 8 line 2, we introduced ENH, which stands for extratropical Northern hemisphere.

**Page 9, line 3, Explain why  $P^b$  does not have full rank, why this is a problem, and what this has to do with inflation.**

In the Introduction (page 2, line 20-29), some applications of covariance inflation were mentioned and how they help to avoid the negative effect of sampling error. Inflation techniques were discussed later again in section 3.2.

In page 9, line 3, we will emphasize once again the negative effect of the small ensemble size and referring back to section 3.2 how inflation techniques can help. We will reformulate the sentence using the term 'sampling error' because it has been introduced previously and it will make a better connection.

**Page 9, lines 5-6. This is not well phrased; it is not the covered model space that is multiplied with the inflation factor.**

Thank you for pointing out the mistake. We will reformulate the sentence by indicating that only the deviations ( $x'$ ) were multiplied.

**Page 9, line 16, It has not yet been mentioned in the main text in section 3 that 250 members have been chosen and how they have been selected (see also earlier comment on this).**

In most of our experiment, we indeed used 250 members to create the climatological state vector. We wrote that the ensemble size of the climatological part ranged between 100 and 500 members in the experiments. We will clarify for the readers, how the climatological state vector is obtained.

**Page 12, line 2, replace 'verification' with 'validation' (this is used elsewhere in the text) or 'performance'**

We will replace it.

**Page 12, line 6, missing full stop after '2'.**

We will correct it.

**Page 12 and also in main text, Add a comment on whether the skill differences found are substantial and practically relevant.**

We agree that this statistic is missing and we will provide some statistical evaluation in the revised paper.

**Page 14, line 11, 'kalman' should be upper case.**

We will correct it.

**Page 14, line 16/17, typos and missing spaces**

We will correct it.

**Figs. 1, 3, 8, 10 should be bigger**

We will provide well readable figures in the final version.

## References

Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, *International Journal of Climatology*, 34, 1585–1605, 2014.