Clim. Past Discuss., 7, C2365–C2370, 2012
www.clim-past-discuss.net/7/C2365/2012/

Climate of the Past Discussions

Interactive comment on "A novel approach to climate reconstructions using Ensemble Kalman Filtering" *by* J. Bhend et al.

J. Bhend et al.

jonas.bhend@csiro.au

Received and published: 16 January 2012

We thank Reviewer 4 for very valuable comments to the manuscript. The responses to the specific comments follow below. Paragraphs from the original review are marked with '$$'; our comments follow immediately after the respective paragraphs.

$$ The authors need to greatly expound on the description of the initial ensemble. I have no idea how the 30 ensemble members differ from one another. If I understand correctly, all 30 members are identically forced. How does the unconstrained ensemble spread vary through time? Is the model so strongly forced that the ensemble lacks the diversity necessary for an ensemble method? Without this, I cannot interpret Figures 2,3. If the spread collapses from 1850 to 1899, the results are dominated by the choice of the initial ensemble.

The individual ensemble members differ in their initial conditions (spliced off of a control run reflecting conditions around 1600 AD). As the climate model has been run with identical (but temporally varying) boundary conditions for all ensemble members, the spread of the ensemble characterises internal (unforced) variability of the atmosphere. The magnitude of internal variability in any given season depends on boundary conditions, however, we do not find strong evidence of varying ensemble spread (boundary conditions over the period from 1850 to 1899 are relatively stable except for the effect of major volcanic eruptions). Unsurprisingly, the intra-ensemble variability is very similar to the interannual variability of the individual ensemble members. The same holds true for the spread of the analysis: the spread does vary through time but no clear tendency towards lower spreads is noticeable. Collapsing spread is not expected as we do not cycle the procedure.

To clarify the above issues in the manuscript, we expand the section on the model simulations.

$$ I can only infer that only 30 ensemble members existed to be used for initial conditions. If possible, I would like a rationale that there is no practical benefit to a larger ensemble size. Ensemble methods for state vectors of size 10e6 and order one hundred ensemble members are computationally tractable. Since the variance of the ensemble is fundamental to the method, and the estimation of variance is facilitated with larger sample sizes; it stands to reason that a larger ensemble (constructed properly) will generate a better result – limited by the computational effort, naturally. Show me the point of diminishing returns. Figure 5 stops short of any ensemble size greater than the one used.

We are aware that larger ensembles will lead to a better representation of the ensemble error covariance and thus enhance the accuracy of the method and reduce the need for localisation. However, as you correctly infer, we only have an ensemble of 30 members available. In the context of paleoclimatology, even ensembles of that size are at present very rare. Furthermore, the change of skill in dependence of ensemble size indicates that the skill varies only slightly with ensemble sizes larger than 10 or so in areas where proxy information is assimilated. Taking into account the need for stronger covariance localisation with smaller ensembles, we are also aware that larger ensembles will be mainly beneficial far away from the assimilated information. We would really like to be able to quantify the effect of increasing ensemble size way beyond the 30 members present. Unfortunately, this is not possible at the moment.

$$ The localization/cutoff had a profound impact on the study, yet only one value was used and rationalized (pg 2843:14) "... to reduce inter-hemispheric influence.". The sensitivity to the choice of L=5000 km begs to be explored.

This will be done. The introduction and motivation for covariance localisation has been expanded and the effect of different cutoff lengths is documented.

$$ There is repeated evidence that the analysis procedure does the right thing; the en- semble spread has been reduced in the vicinity of the observations in a way such that the ensemble is more like member 30 (the 'truth'). In the language of the data assimilation community, the error of the posterior (the analysis) is smaller than the error of the prior (the unconstrained ensemble). BUT - since the posterior is never used as an initial condition (i.e, the assimilation is not cycled), it is exceedingly difficult to know if the posterior ensemble spread is so greatly reduced that it no longer captures the uncertainty. Figure 3 c,d show that the posteriors in the regions with observations have half (roughly) the ensemble spread of the prior. Is this good/bad/sufficient/appropriate? I don't know. The metric described in equation 6 does not account for overfitting, indeed – if all the analyses were identical to xËĘref – the result would be the highest possible score with perfect certainty and Figure 3 c,d would have 0 percent in the vicinity of the observations. A rank histogram goes a long way toward describing if the ensemble is under- or over-dispersed. In short, you DO have to worry about what LEADS TO filter divergence, even if you are not cycling the procedure.

Thank you for this very valuable comment. We adjust the analysis accordingly and

include discussion of potential overfitting in the present setup. The rank histogram indeed points to overfitting with a localisation with a cut-off length at 5000 km (as used in the study). Lower cut-off lengths help to mitigate that effect at the cost of less skill (RE) in areas that are further away from the assimilated information.

$$ The authors need to rationlize the form of the observation error being added to the synthetic obsertions. Why red noise and a bias? This study is a proof-of-concept to assimilate paleo data - so show me how the synthetic observations resemble paleo observations. In the absence of that, just give me gaussian noise with some mean and variance. This would be more consistent with the statement (pg2841:19) "... and the variance of the disturbance is known exactly." than generating red noise with AR(1) (why AR-1?) with a coefficient of 0.7 (why?) scaled by 1.5 standard deviations (why?). It is not obvious to me that this results in a variance that is known exactly or resembles the properties of the intended paleo observations. Figure 1 shows the values of the correlations. A correlation of 0.5 is "pretty weak". All I can glean from this is that there are 5 or 6 sites out of 37 that have a different correlation between summer and winter and that more than half are weakly correlated at best. I'd rather see something like a signal-to-noise ratio ...

As there is no standard approach to generate pseudo-proxies, we decided to add a simple disturbance to the original temperature data from the reference series to generate pseudo-proxies following earlier work (Mann and Rutherford, 2002; von Storch et al. 2009) using correlations between 0.2 and 0.8 characteristic for proxy data as a target (Jones and Mann, 2004). As a consequence of the scaling of the normalized noise series by 1.5 times the standard deviation of the local temperatures, the signal-to-noise ratio is constant for all proxies (SNR of 0.66). We will clarify these issues in the manuscript.

$$ I am left wondering if the results are sensitive to several things. There are 411 simulated years to choose from – why was the period 1850 to 1899 chosen? Surely the forcing files are better constrained for a more recent period. Without knowing anything

about the initial ensemble details ... would the results be the same for a different ensemble member defining the 'truth'? A different time period?

We have used all the different simulations as reference in turn to assess the influence of sampling errors on the procedure (not shown in the manuscript). We will add a short paragraph summarizing these findings.

We agree with the reviewer that the choice of the exact period is somewhat arbitrary. We tried to keep the period reasonably short to facilitate computations but sufficiently long to avoid severe sampling issues. Furthermore, we have chosen a period with strong volcanic eruptions and more variability in the boundary conditions (due to the fact that the reconstructed boundary conditions rely on more observations / proxies than in earlier periods). We have also analysed the results for a different time period and find that the results are qualitatively robust to the time period chosen. We expand the discussion of sampling issues and the choice of time period accordingly.

$$ Figures 4,5 use boxplots to convey the fact that the posterior is closer to the truth (ensemble member 30) than is the prior. Since the state vector is so small and calculations can be done offline (pg 2838:line 29) it should be straightforward to cross-validate these results using several different ensemble members as the "truth". This brings me to the next point. Using the full state vector is easy to justify and should be done. Using a subsampled portion of the state vector is ad-hoc and prevents it from being used as initial conditions – which prevents any claim of dynamical stability. Since the stated goal (pg 2839:2,3) is to assimilate climate proxy data into a coupled atmosphere-ocean GCM, the full dynamical state will be needed (presumably the posteriors will be used as initial conditions for subsequent model advances in this situation).

We have carried out the cross-validation of the results (not included in the paper) – see comment above. As we newly introduce covariance localisation for the composite indices as well, we do not see much merit in doing the computations with the full state vector. Therefore, we refrain from performing the computations on the full state vector

and stick to the procedures as outlined in the manuscript.

$$ Specific comments:

$$ Figure 1 needs to be split into two figures. The hollow circle surrounding a solid circle is terribly confusing and does nothing to set up the subsequent graphics. Make two panels side-by-side with Nov-Apr on the left and May-Oct on the right and Figures 1,2,3 all benefit.

Figure 1 has been changed.

Interactive comment on Clim. Past Discuss., 7, 2835, 2011.