

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

Anonymous Referee #4

Received and published: 7 December 2011

The study explores the methodology of creating an atmospheric analysis in a paleoclimate context from an ensemble of low-resolution atmospheric model states and sparse, infrequent, synthetic observations. The Ensemble Square Root Filter (EnSRF) algorithm is introduced and the usefulness of using an ensemble to express uncertainty and variability are demonstrated.

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

C1983

of the initial ensemble.

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 10^6 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.

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.

There is repeated evidence that the analysis procedure does the right thing; the ensemble 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 x^{ref} – 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.

C1984

The authors need to rationalize the form of the observation error being added to the synthetic observations. 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 ...

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?

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).

C1985

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.

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

C1986