Clim. Past Discuss., 8, C175–C178, 2012
www.clim-past-discuss.net/8/C175/2012/

Climate of the Past Discussions

Interactive comment on "Identification of climatic state with limited proxy data" *by* J. D. Annan and J. C. Hargreaves

Anonymous Referee #3

Received and published: 27 March 2012

The manuscript presents an analysis of the performance of a particle filter data assimilation method using a pseudo-proxy setup based on the LOVECLIM model. Similar methods have already been used in several studies for the assimilation of paleoclimatic proxy-data (e.g. Goosse et al. 2006, 2010, Dubinkina et al. 2012) and this study complements the previous studies in a very useful way. The approach is scientifically sound, the results are informative, and I enjoyed reading this paper. It is in general well written but some points could have been explained more clearly, partly to make the paper better accessible for non-specialist readers. These points are listed below and after they have been addressed I am happy to recommend publication.

Specific comments

1.) The study should be better linked to other studies.

In the introduction it would be useful to discuss how similar the methods used by H. Goosse are to a particle filter. As far as I am aware the ensemble member selection used in Goosse et al. (2006, 2010) can be regarded as a degenerated particle filter.

Do the results presented in the manuscript have implications for the interpretation of H. Goosse's results? A short comment on this in the conclusions would be good.

In the introduction it is mentioned that the study complements Dubinkina et al. 2012, but no information is given what these authors have found and to what extent the present study is consistent with Dubinkina et al. and whether substantially new insights have been gained. It would be good to have a few comments on this in the conclusions.

2.) As far as I am aware the cost function used in Goosse et al. (2006, 2010) is based on the similarity of simulations and proxy data over multi-decadal periods. In contrast the particle filter used in the present study only compares the ensemble simulations and the pseudo-proxies for each year. The similarities, differences, advantages and disadvantages of the two approaches should be briefly discussed.

One point that could be mentioned is the reduction in sample size of the ensemble if similarity over multi-decadal periods was used for the weighting in the particle filter. Comments on the physical consistency or even the existence of trajectories in state space would also be helpful. If I understand correctly the particle filter used here yields only posterior PDFs that cannot be used for constructing physically consistent trajectories, whereas some of H. Goosse studies slightly perturb the best-fitting ensemble member to define the initial conditions for the ensemble for the next multi-decadal period and thus yield trajectories that are to a good approximation physically consistent.

It should also be clarified whether in the current study annual means have been used to define the cost function. As there are no related comments, I assume this is the case, but in principle it seems possible to take individual months during a given year into account, so some clarification is needed.

3.) A few more comments on the Boer and Lambert (2008) definition of potential predictability would be good. Why is it not possible that predictable k-year means have the same contribution to the total variance as unpredictable k-year means? What statements on potential predictability can be made for a given variance ratio?

4.) For readers not familiar with particle filters it would be good to explain in the method section more explicitly how the weights are used to calculate the posterior PDF from the prior PDF.

5.) In section 2 it is said that the simulations are 140 years long. In section 3 it is said that 20 years are considered as spin-up and that this leaves 100 year long simulations. There is an inconsistency of 20 years.

6.) In section 3 it should be mentioned that defining the pseudo-proxies as temperature plus noise is a special case. In a general data assimilation framework simulations and proxy data would be related through the 'observation operator'. The question of scaling of the proxies brought up by another reviewer is linked to this as choosing a scaling factor can be interpreted as specifying an observation operator, but this does still not cover the general case.

7.) A reference in section 3 for the convergence of the likelihood weighting method for large ensemble sizes towards the exact solution of Bayes' Theorem would be good.

8.) The statement at the beginning of section 4 on the four epochs could confuse some readers. It should be clarified that only the number and location pseudo-proxies is linked to these epochs.

9.) Can more be said about the reason why quite clear improvements in correlations for denser proxy-datasets are only accompanied by very modest improvements in RMSE? Has this to do with unrealistic variances?

10.) I found parts of section 4.3 not well explained. Which forcings are considered? In the context of paleoclimate reconstructions for the last millennium solar and volcanic

forcings would be the most relevant, but here greenhouse gas forcing is considered. I do understand that the purpose here is only to investigate to what extent any forcing affects the performance of the particle filter, but this should be clarified.

Interactive comment on Clim. Past Discuss., 8, 481, 2012.