**Climate
of the Past
Discussions**

Interactive comment on "An assessment of climate state reconstructions obtained using particle filtering methods" *by* S. Dubinkina and H. Goosse

Anonymous Referee #2

Received and published: 25 February 2013

The authors compare three data assimilation methods for climate state reconstruction in the Southern Hemisphere. They find that the two variants of particle filters out-perform nudging especially for variables that are less closely related to near-surface temperature. The paper is novel in that the authors compare different methods for climate state reconstruction using data assimilation. The manuscript is generally well written and comprehensive, the setup of the study and some of the conclusions drawn however, require further discussion.

General comments:

- The title announces a comparison of climate state reconstructions from different particle filtering methods, however, the study focusses on a comparison between reconstructions based on paricle filtering and nudging. I suggest adjusting the title to better reflect the contents of the title.

- The setup of the experiment complicates the comparison between the different approaches and the interpretation of the results. Also, the motivation to chose this particular setup is not clear to me. Especially the choice of nudging with near-surface temperature over oceans seems particular in the light of recent studies. Keenlyside et al. 2008 have used this simple approach, but Swingedouw et al. 2012 note that this choice (with strong nudging) is problematic and 'optimising' the strength of the nudging parameter is crucial. You don't motivate your setup further, but the results indicate that either the nudging parameter used is too strong, or nudging doesn't work for the SH. More discussion on the motivation for the current setup and the implications is needed (e.g. should nudging be discarded for reconstructions?).

- You chose correlation and RMSE to assess the performance of the different assimilation methods. These metrics are simple and easy to understand, however, they do not allow the reader to quickly grasp the added information through assimilation. A measure of skill that relates the performance of the simulations with data assimilation to the performance of the unconstrained ensemble would be much more informative. At the very least, the correlation and RMSE of the unconstrained ensemble should be mentioned in the text and shown in the figures along with the correlation and RMSE of the data assimilation results.

- Please rewrite the abstract to better summarise the most significant findings and their implications.

- The manuscript would benefit from copy-editing to improve clarity and readability.

Specific comments:

p44l7: 'twin experiment' has not been introduced previously and may be difficult to understand. Consider rephrasing.

p44l8ff: "The net of..." is this sentence needed here?

p45l7ff: Consider rephrasing to "... are biased in the sense that the analysis is linearized and they thus assume gaussian distributions." as it is not only the prior that is gaussian. Whether the gaussian assumption represents a serious limitation in paleoclimatological applications with generally fairly aggregated data, however, remains to be seen.

p45l21ff: Similarly, EnKF suffers from spurious off-diagonal covariances when the ensemble is small, with the concesquence of filter divergence and unreliable (underdispersive) probabilistic estimates.

p47l8: not 'used by'. Better 'used in' or 'used with'

p47l16: '... which allows us to ...'

p47l26ff: You perform two sets of experiments, one with information for assimilation available everywhere, which may be thought of indicating the upper limit of skill through assimilation of near-surface temperature, and the second experiment with information where observed temperature is available. Arguably, the spatial coverage of proxy information would be even coarser. Some discussion as to how the skill of the assimilation might be affected in a more 'paleo' and thus data sparser context would be interesting.

p47l26ff: A figure illustrating the two different cases (spatial locations used for assimilation) would be very helpful. As the sparser case (using HadCRUT3 locations) has varying temporal density, you may want to show average coverage in an early and late period (say 1850-1900 vs 1950-2000).

p49l2: omit 'following'

p53l3: How does the nudging parameter compare to nudging in Keenlyside et al., 2008 and Swingedouw et al. 2012?

p53l4ff: I do not understand what 'taking into account the instrumental surface temperature records HADCRUT3 ...' means in this context. Do you project HadCRUT3 on model-derived EOFs to construct the stochastic error? Please clarify.

p53l12: The model error covariance is assumed to be diagonal. This almost certainly overestimates the degrees of freedom in model errors considerably as there is significant spatial correlation to be expected. Please justify your choice and discuss potential biases resulting from this.

p53l26ff: How well do the different data assimilation methods compare with the unconstrained ensemble. That is, how much of the correlation is due to external forcing and how much is due to internal variability. Such a comparison would facilitate the interpretation of differences between the different assimilation methods. Especially for the case with nudging south of 66S, such a comparison might be interesting as it would highlight the importance (or lack) of correlation between mid- and high-latitude weather.

p54l26ff: Consider rephrasing "We obtain that ... smallest mean RMS error." to remove redundancy and increase readability.

p55l14: From Fig. 2, ocean heat content may need more than 15 years to adjust. What

about spinning up with perpetual 1850 conditions (and for the different assimilation methods) to overcome the potential bias?

p55l19f: replace 'providing with' with 'resulting in'

p55l20ff: 'Even a free model run...' please provide quantification (see comment above).

p56l13: What about the other variables and what about the unconstrained ensemble in this case?

p56l16: Do you mean " ...close to degeneracy for the larger domain."

p56l16: I don't agree with your statement on how degeneracy affects correlation. If the ensemble collapses to the pseudo-observations, we would expect higher correlations with degeneracy, but very small ensemble spread. Please clarify.

p57l14ff: I do not understand what the prior distribution is. Please rephrase or clarify what you mean by prior distribution in Sec. 2.

p57l20: Is the nudging applied at HADCRUT3 locations only? I assume so from comparing Fig. 2 and 6, but please clarify.

p57l25: Annan and Hargreaves (2012) have also ...

p58l11: Do you suggest here that you can estimate the forced change with data assimilation without changing the forcing of the model over time? I.e. are you arguing that the assimilated information provides constraints that are strong enough to override changes in forcing? Please clarify.

p58l13: please show the five different initial condition experiments in Fig. 2. This may clarify your point.

p59l10ff: What is the correlation of the unconstrained ensemble?

p59l10ff: The statement that the correlation is different between the nudging and the particle filters for the end of the 19th century is not backed by the plot. Nudging performs (significantly?) worse from 1907-1948, but not earlier.

p60l6ff: consider replacing with " ...linked to the pseudo-observations such as surface air temperature and sea ice concentration, but also variables such as geopotential height and sea surface salinity."

p61l1: This statement is interesting. The assessment of skill for variables that are less closely related to the assimilated information provides a stricter test in that the use of physically ill-conditioned levers are exposed (as in the case with nudging and ocean temperatures at various depths). The extremely efficient particle filter does not seem to suffer from such severe deficiencies, but it is also not able to outperform the sequential importance resampling for variables that are less closely related to the assimilated information. This is somewhat unexpected and it would be worth discussing the strengths and limitations of the extremely efficient particle filter and the sequential importance resampling in more detail.

p61l13: estimates

p61l13: 'reliable' might be misunderstood to mean not over- or underdispersed. You do not discuss these issues here, therefore I suggest to rephrase the sentence. Also, you need to clarify why you think the reconstructions of geopotential height and salinity are not 'reliable'.

p61l20: "Past4Future contribution no. X" placeholder?

Fig. 1: Consider reorganising the plots with only 1 panel per location with the different assimilation methods and the truth superimposed for better comparison. Furthermore, you may increase the readability of the plots by freeing up space for the main plot through a reduction of redundant axis labels where possible (e.g. one common axis across multi-panel plot).

Fig. 3,4,7-10: Please add the 'No data assimilation' case for all figures.

Fig. 7-10: Please clarify that these plots relate to the sparse pseudo-observations.