Review of

'Assessing the robustness of Antarctic temperature reconstructions over the past two millennia using pseudoproxy and data assimilation experiments'

by F. Klein et al.

Recommendation: minor revisions

This manuscript presents data assimilation (DA) simulations for Antarctica for the past two millennia using a particle filter with the ECHAM5/MPI-OM and ECHAM5-wiso isotope-enabled GCMs. The simulations comprise pseudo-proxy experiments, which show that the DA can successfully capture the target oxygen isotope pseudoproxies, but that the skill in reproducing temperature variability is limited. It is also shown that this limited skill for temperature reconstructions is due to weak and temporally varying links between regional temperatures and oxygen isotopes, which also means that statistical reconstruction that rely on links fitted during a relatively short period are problematic.

The main objective of the study is to investigate the discrepancies with respect to the beginning of the anthropogenic warming and to the regional temperature trends between statistical temperature reconstructions for Antarctica and forced CMIP simulations. The simulations show an earlier onset and a more spatially homogeneous warming across Antarctica than the empirical temperature reconstructions by Stenni et al. (2017), which shows warming only in some parts including the Antarctic Peninsula and the West Antactic Icesheet. Potential reasons for this mismatch are an overestimation of the forced response in the models, or a dominant role of internal variability. Using assimilation of real-world oxygen isotope records it is shown in the manuscript that the DA simulations are consistent with the empirical temperature reconstructions and that there is therefore no evidence for a fundamental inconsistency between climate simulations and empirical regional Antarctic temperature reconstructions.

The methods applied are state-of-the-art and well explained, and the conclusions are mostly drawn in a sound way. The manuscript is very clearly written and provides an important contribution to palaeoclimate science. There is only one substantial point I would like to be discussed in more detail, which is the distinction of stationary vs transient offline DA methods and the implications on the conclusions. After this and a number of very minor comments have been addressed I fully support the publication of this very interesting and informative paper.

Specific comments

1)
There are two types of offline DA methods. In 'transient offline' methods the ensemble used for DA is time-dependent and generated by ensembles of forced simulations, and only the simulated ensemble at or around the time of the assimilation timestep is used for DA. In transient offline DA the ensemble size for DA is limited by the computational constraints on performing transient ensemble simulations. The ensemble size for DA can be substantially increased in 'stationary offline' DA methods by using all simulated timesteps as the ensemble for DA. The transient

offline approach has been used for instance in several studies by Goosse et al., and by Matsikaris et al. (2015); the stationary offline approach has to my knowledge been used the first time by Steiger et al. (2014) and has been applied in several other studies by Steiger et al.

Although it is made clear in the manuscript that a stationary ensemble has been used for DA, the difference between these approaches should be explicitly discussed in section 2.3. Note that the terminology transient/stationary offline is not established yet, but I believe it captures the key difference between the approaches.

Furthermore there should be a discussion on what type of conclusions can be drawn in the two cases if the DA simulations are in agreement with empirical temperature reconstruction. At the moment the conclusion is that there is no fundamental inconsistency between the models and the empirical data. However the question formulated in the introduction was whether the response of the CMIP simulations to the forcing is too strong, or whether internal variability is responsible for the discrepancies between the CMIP simulations and the empirical reconstructions, and the conclusions do not specifically address these two possibilities. In a transient offline approach an agreement between DA simulations and empirical reconstructions would imply that the superposition of forced and internal variability includes the empirically reconstructed states, and thus there is no indication that the forced signal is unrealistic. In contrast when using a stationary offline approach it would be possible to achieve agreement between assimilated states and empirical reconstructions even if the forcing signal was so unrealistic that the superposition of the forced signal and any realistic realisation of internal variability would not include the empirically reconstructed states, because the agreement could be caused by choosing simulated states from times with a different forcing than the actual forcing at a given time.

This shows the limitations of using stationary offline approaches for process studies. The authors' statement 'no fundamental inconsistencies' is fairly vague and a more specific discussion of what is meant by 'fundamental inconsistencies' should be provided.

2)
In section 2.3. it is said that online DA can outperform offline DA when the assimilated data involve a long-term trend. This is just one special case. In general information propagation in time does not have to imply slow changes, as fast changes might still be dynamically related. However, if the system shows slow changes it is clear that information is propagated forward in time. The explanations should be adjusted accordingly.

3)
Page2, line 17, replace 'signal' with 'change'

4)
Page 3, line 26, 'Our study being based … it is important'; wrong English

5)
Page5, lines 14/15, ' ... simulate similar … than another …', not well phrased, either replace 'than' with 'as' or reformulate.

6)
Page5, line 19, replace 'validating' with 'justifying'

7)
Page 8, line 1, replace 'of' with 'for'

8)
Page 8, line 9, replace ‘on’ with ‘to’

9)
Page 8, line 18, replace ‘pseudoproxy’ with ‘pseudoproxies’

10)
Page 11, line 8, replace ‘simulation’ with ‘simulations’

11)
Page 12, line 11, replace ‘model mean’ with ‘model mean correlation’ (if I understand correctly)

13)
Page 14, line 1, replace ‘of’ with ‘for’

14)
Page 14, line 25/26 ‘in the results with a last century …’, something is wrong with this sentence

15)
Page 14, line 28, replace ‘link between’ with ‘links of’

16)
Page 20, line 17, replace ‘hypothesis’ with ‘assumption’

17)
Page 26, line 9, delete ‘has potentially’