Review of

‘Using a process-based dendroclimatic proxy-system model in a data assimilation framework: a test case in the Southern Hemisphere over the past centuries.’

by J. Reszöhazy, Q. Dalaiden, F. Klein, H. Goosse, and J. Guiot

Recommendation: minor revisions

This manuscript evaluates whether using a process-based dendroclimatological proxy system model in the context of paleoclimate data assimilation provides better results than using simple regression-based tree growth models. This is a relevant research question which has not been addressed yet. For the specific case study presented in the manuscript it is found that the performance of the two methods is similar, with some differences depending on regions and performance measures. The analysis is technically sound and the study is a valuable contribution to the further improvement of data assimilation methods. However, there are some general points and some details that could have been discussed more clearly. I recommend publication after the specific points listed below have been addressed.

Specific comments

1)
Lines 25-28, 36-46: The problem with statistical models are not only assumptions on linearity and stationarity, but also that they are ‘inverse models’. It should be made clear that multiple climate states may lead to the same response in proxy data, and that this can be taken into account with PSMs, whereas inverse models assume invertibility of the relationship, which may not be the case. It should also be mentioned that the PSMs are a specific form of the ‘observation operator’ in the general DA framework.

2)
Lines 30-31: ‘impact’ should be replaced by ‘contribution’, and ‘natural’ with ‘natural, random’ or ‘natural, internal’, because ‘impact’ describes an external influence on a system, and ‘natural’ variability includes natural forced variability.

3)
Lines 46-50: The setup of the pseudoproxy studies and of the role of VS-Lite in them should be better explained, so that the main aspects become clear without reading the references.

4)
Lines 56-58: A very brief explanation of the setup of the calibration and validation of the MAIDEN model in Reszöhazy et al. (2020, 2021) would be good. For instance, what are the inputs and outputs?

5)
Lines 62-64: Has MAIDEN not been evaluated in the Northern Hemisphere or did it not perform well?

6)
Lines 66-68: This is the first time oxygen isotopes data are mentioned. It is explained later that these are available from the isotope-enabled GCM simulations, but it would be good to briefly mention this already here.

It is unclear why the different types of proxies are linked to different spatial scales. Is the argument that by using only TRW data large areas would not be covered at all with proxy data, or that isotope and/or snow cover data are linked to larger spatial scales than TRW data. If it is the latter, why is this case?

7)
Lines 116-118: The discussion of dynamical consistency should be more precise and avoid overselling.

The individual particles for a given timestep are dynamically consistent climate states, but in an offline DA there is no dynamical consistency in time. This is not a problem if the timesteps are so long that the atmospheric states are almost independent. However, the dynamics of the ocean and cryosphere components of the climate system involves also very long timescales, and the ocean and cryosphere states influence the atmospheric states.

Moreover, dynamical consistency in space and between variables involves non-linear equations and the weighted ensemble mean is therefore not dynamically consistent.

8)
Line 146: 'anomalies … are subtracted from the TRW timeseries' seems wrong. Please rephrase.

9)
Line 153: Genitive s after reference should not be there.

10)
This is the first time the 'observation operator' is mentioned. The fact that PSMs are one example for observation operators in DA should have been mentioned already in the introduction (see also comment 1).

11)
Lines 174-176: As far as I understand the annual quantity of carbon that is added to a tree (Dstem) is proportional to the added tree volume. i.e. proportional to r * delta_r, with r the radius of the tree and delta_r the tree ring width. It seems therefore problematic to compare Dstem only with delta_r. It would be good to add a comment on this in the text.

12)
Line 186: A correlation of 0.3 means that less than 10% of the variance is explained. A comment on how this may affect the DA would be helpful.

13)
Line 225: The standard meaning of ‘validation’ is quantification of skill, not demonstration that the skill is high. The statement should be replaced by something like ‘only two locations satisfied the same selection criterion as MAIDEN’.

14)
Lines 310-312: The meaning of CE and the reason for using it should be better explained. For a linear prediction model the correlation includes the complete information about the amplitude of the predictions, because the squared correlation is the explained variance. The reason this is not sufficient in the context of the study is that no linear models are applied to correct the simulated variables. The CE compares the variance of the residuals with the variance of the predictand, with CE = 1 associated with perfect predictions, CE = 0 with a residual variance identical to predicting the observed mean, and CE < 0 with larger residual variance than when predicting the mean.

15)
The underestimation of variance in the reconstructions is discussed in several places and attributed to weak constraints on the prior through the proxy data, and thus similar weights for a large number of particles. This is in principle correct. However, this is a fundamental issue with the particle filter, and potentially with other data assimilation methods, and not all relevant aspects become clear. The paper would benefit from a more systematic discussion of the reasons for variance underestimation, including at least the following points:

- Individual members of ensemble climate simulations or sequences of selected timesteps from individual members have in the ideal case realistic temporal variability on all spatial scales.

- Any averaging of random, non-forced variability will reduce variability. This means that any ensemble mean will always have unrealistically low variability, regardless of how it is constructed. However, the extent of the reduction depends on whether members with more similar or more different variability are used for calculating the weighted ensemble mean.

  In contrast to unweighted ensemble means the Particle Filter gives high weights to ensemble members that match the observations. If the empirical information strongly constrains the particle selection, the particles will be more similar than in a less constrained case, and there will be less underestimation of variance.

- The similarity measure is determined at the locations of the proxies, but the weights given to each particle are independent of location (more detail on this in the manuscript would be helpful). Particles that have similar states at the proxy locations don’t necessarily have similar states at other locations, and the reduction of variance in unconstrained locations is therefore likely to be larger than in constrained locations.

16)
Lines 351, 394: replace ‘constrain’ with ‘constraint’.

17)
Line 354-359: These statements are very unclear. A TRW PSM does not assess ‘errors in the simulated variance of climate signals’. Please clarify the argument.

18)
Line 361: ‘fully verify our DA procedure’ is not well phrased.

19)
Lines 362, 396: ‘if’ should be replaced with ‘whether’.

20)
Lines 386-387: The comment on the potential influence of the different number of tree ring records for reg and MAIDEN is helpful, but this potential issue should be already mentioned in the introduction and/or method section, and the experimental setup using different numbers should be justified.