

Dear editor and reviewers, we would like first to thank you for your useful feedback and comments on our manuscript. You can find here below the Referee's comments in *italics* and our answers in blue. In **bold**, you can find the modifications that will be made to the manuscript.

Referee#1

General comments

The manuscript "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 Rezsöhazy et al. is a good example of using a process-based model MAIDEN as new tool in the assimilation-based climate (particularly, near-surface air temperature, precipitation and winds) reconstructions in the Southern Hemisphere over the last 400 years. I note that the Southern Hemisphere is poor investigated area in the context of climate reconstructions so far. Mostly the published long-term climate reconstructions are based on different regression approaches to make the link between the GCMs outputs and the proxy observations obtained from tree rings, ice cores, etc. In this work the authors made a considerable effort to use the process-based MAIDEN (that removes a number of limitations inherent in conventional regressions) as a proxy system model in a data assimilation procedure, using as a test the reconstructions of three climatic variables.

The paper is well structured and written. The introduction provided a comprehensive overview of the background information and the pertinent literature, and it demonstrated the need for the current study. They involved a wide range of statistical and modeling techniques, included original ideas, as well as several international global databases for testing their hypotheses and confirming their results.

We would like to thank the Referee for this positive general feedback, for the careful evaluation of our manuscript as well as for the useful comments that will be addressed in the revised version as specified here below.

But there is an issue which can be considered in the MS. The particle filter method described in Sect. 2.1 is not used often in research. Could the authors describe in details how were those climate particles used to adjust different climate reconstructions used as inputs of MAIDEN over the past 400 years (see the Section 2.4.1)? Could the authors produce some visualization (figure) how does this algorithm work?

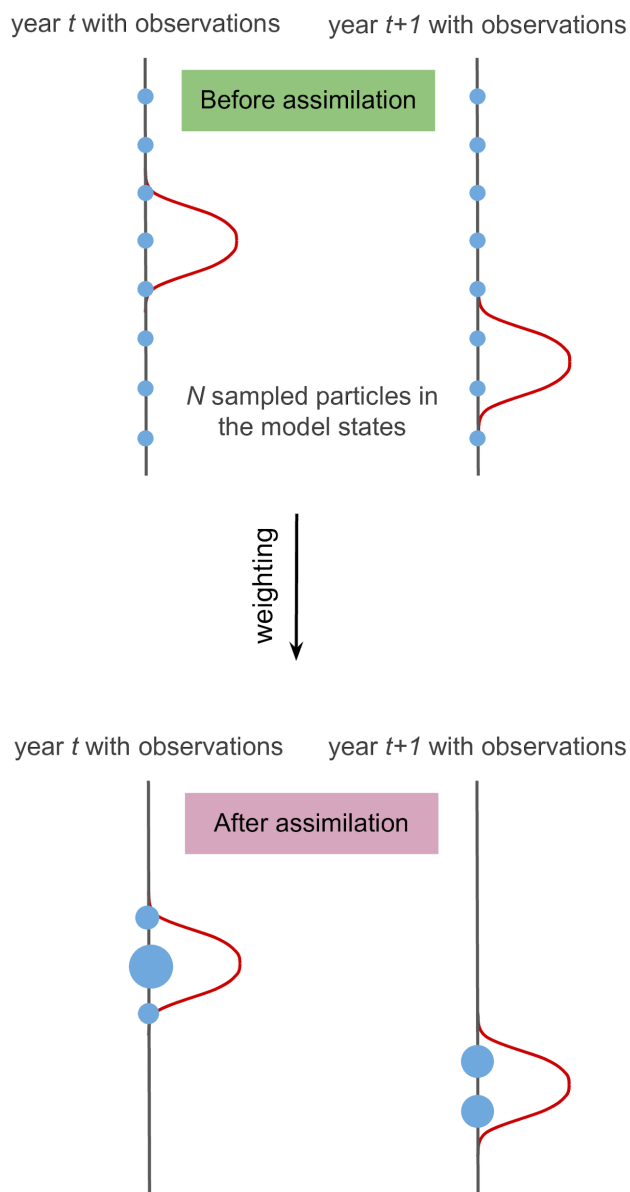
Previous studies have demonstrated that the particle filter is a valid method for offline data assimilation (e.g., Widmann et al., 2010; Dubinkina et al., 2011; Goosse et al., 2012; Klein and Goosse, 2018; Dalaiden et al., 2021). In particular, one of the advantages of this method compared to other methods is that it respects the climate system dynamics as represented in the climate model because only the weight of the sampled particles are modified by comparison with the proxy data during the assimilation. Additionally, it does not involve any linearization or inversion of the PSM. The method does not make any assumptions on the prior and posterior distributions of the state of the climate system, and it can be used without an a priori knowledge of the model.

The hypotheses that underline other filters such as the Kalman filter can be problematic when using a process-based model as MAIDEN for assimilating tree-ring series, especially

regarding the assumption of a Gaussian distribution of the tree-ring simulations (see the work of Steiger and Smerdon, 2017 with VS-Lite). This is one of the reasons to use the particle filter here.

A more in-depth explanation of how the particle filter works can be found in my PhD thesis (freely available here: <https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A258209/datastreams>), in the Appendix D.1 from p.159. This will be referred to as follows on I119, at the end of Section 2.1 : **“More details on the implementation of the particle filter in our experiments can be found in Rezsöhazy (2021).”**

In particular, the following illustration provides a schematic representation of the procedure of offline data assimilation with a particle filter:



Before the assimilation, all the N ($= 8$ in our example) particles (blue circles) sampled in the model states (vertical axis) have the same weight, which is proportional to the size of the circle, whatever the year considered (t and $t+1$ in the example) because the prior is fixed throughout the assimilation procedure. The distribution of the available observations with uncertainties for each year is represented by the red curve. After the assimilation, the particles have been weighted relatively to their associated likelihood, given the available observations for each year, and the size of the circle is modified accordingly. When the weight of a particle is zero, it does not appear anymore on the figure.

Is there some public domain where the code of the algorithm is located? The code is freely available upon request.

The principal idea of the MS is to use the certain process-based model as a proxy system model. What is a reason to use an additional proxy data (e.g. delta 18O) which was used in data assimilation procedure traditionally (through the linear regression)?

The idea behind the assimilation of other proxy data is to cover a vast area of the Southern Hemisphere (i.e., Antarctica), where no tree-ring proxy data can be found, in order to provide a comprehensive reconstruction of the Southern Hemisphere climate. By doing so, in addition to reconstructing the local climate where tree-ring proxy data are available, we also reconstruct the large-scale climate. The isotopic content from Antarctic ice cores (i.e., $\delta^{18}\text{O}$) is directly simulated in the climate model, so that no proxy system model is used. For the SMB, a simple assumption (i.e., the difference between precipitation and sublimation/evaporation ($P - E$) from iCESM1 is comparable to the snow accumulation observations) is made to assimilate the observations, and has been shown to be valuable particularly at large spatial scale (see Agosta et al., 2019; Souverijns et al., 2018; van Wessem et al., 2018 for instance).

This will be made clearer in the manuscript on l149: **"In addition to tree-ring proxy records which are the focus of this paper through the use of a process-based PSM (Sect. 2.4.1), we also assimilate other proxy data, i.e. ice core records, to cover a vast area of the Southern Hemisphere (i.e, Antarctica) where no tree-ring proxy data can be found, in order to provide a comprehensive reconstruction of the Southern Hemisphere climate."**

I would suggest to publish the MS after minor revisions.

Specific comments

Section 45: The term 'pseudoproxies' should be clarified.

The following sentence will be added on l48: "However, so far, they have never been used in a DA procedure with actual tree-ring proxy data, but only with pseudoproxies (i.e., **synthetic proxy data**; Dee et al., 2016; Acevedo et al., 2017; Steiger and Smerdon, 2017). **In these studies, the VS-Lite model (Tolwinski-Ward et al., 2011) is forced by climate model outputs in order to provide synthetic tree-ring proxy data. These synthetic proxy data are then incorporated in a DA framework, providing a controlled environment to evaluate the DA procedure and resulting reconstruction skill. In those experiments, using the non-linear VS-Lite model as a PSM for tree-ring data has the potential to**

improve the quality of the reconstruction compared to the traditional statistical linear approach of tree-ring models.”

Section 50: I would suggest to remove 'offline' because this term is introduced later in the MS. Done.

Section 75: The term 'June-July year' should be clarified because this is the first mention of it in the work.

The following words will be added to the manuscript on l77: **“To match the seasonality of tree growth in the Southern Hemisphere, all the analysis in this study are performed on a July-June year (e.g the year 1992 corresponds to July 1992-June 1993), instead of the usual January-December year, suited for trees growing in the Northern Hemisphere.”**

Section 80: A reference on the 'Bayes theorem' interpretation is needed!

The reference to van Leeuwen et al. (2009) will be added on l82: **“More specifically, the DA procedure is based on the Bayes theorem (van Leeuwen, 2009): (...)”**

Section 90: The statement "the predictability of the variables of interest is smaller than the temporal resolution of DA (i.e, one year in our study)" should be clarified with the corresponding reference.

The following reference **“Okazaki et al. (2021)”** has been added.

Section 105: Is the code of the particle filter approach available in some public depository?

It is available upon request. This will be stated accordingly in the manuscript.

Section 105: What does 'chosen frequency' mean?

This means that we can decide the frequency at which we sample a particle/year/model state in a model simulation. In our case, the frequency is one year, so that we take all the years in the model simulation as particles for the data assimilation. Therefore, we increase the probability to match the assimilated observations, which overall results in a better reconstruction. The manuscript will be modified as follows on l107: **“For instance, for a chosen frequency of one year, we sample a particle or model state in the model simulation every one year.”**

Section 125: The time span used in the work is the last 400 years. What was a reason to mention 'July 850 – June 2005 CE time period' as a time span 'used in this study'?

This corresponds to the years of the climate model used as particles for the assimilation, but not to the reconstructed time span. For each year of the reconstruction over the past 400 years, all the years of the climate model (July 850- June 2005 CE) are used as particles in the data assimilation framework to compare with the proxy observations, corresponding to a total of 3465 particles (3 members * 1155 “July-June” years) . This will be modified in the revised manuscript to make a clear distinction between the time span of the simulations and of the reconstructions, as follows on l128: **“This corresponds to the time span of the**

simulations used to sample the particles during the data assimilation procedure, by contrast with the time span of the reconstructions (i.e., the past four centuries)."

Section 145: What was a standardization procedure of TRW data used taking into account a non-climate noise (e.g., age-dependent trends, etc) in raw TRW measurements?

Here, we use the already standardized tree-ring series directly derived from the PAGES2k compilation without any modification. As the tree-ring series are from different studies (referred in the PAGES2k database metadata), the standardization technique used can differ from one tree-ring series to another.

Section 230: Why did the authors consider positive correlations only? Are the significant negative correlations worse?

Following Cook et al. (2013), and as stated in PAGES2k Consortium (2017), when tree growth increases under increasing temperature, they are "more likely to produce a reliable expression of past temperature variability compared to trees that respond inversely to temperature, for which the proximal control on growth is moisture stress". The same assumption has been made for precipitation.

Our decision is thus made on the assumption that a tree should grow better under increasing temperature (under temperature-limiting conditions) and precipitation (under moisture-limiting conditions).

The manuscript will be modified as follows on l232: **"Following Cook et al. (2013), and as stated in PAGES2k Consortium (2017), trees growing better under increasing temperature should be more reliable recorders of past temperature variability, because they respond to temperature-limiting conditions. The same assumption can hold for precipitation."**

Figure 2. The green dots should be highlighted stronger. Done.

*Figure 3. What does 'uncertainty' mean? This term should be clarified. This corresponds to the so-called "ensemble spread" as defined in Sect. 2.1. This has been modified accordingly: **"(i.e., the ensemble spread; Sect. 2.1)"***

*Section 370: What do 'Aus_002 and SAm_24' mean? Clarification is needed. Those are the names of the tree-ring sites as in the PAGES2k database. A few words will be added in the manuscript on l371: **"(i.e., the name of the tree-ring sites, as in the PAGES2k database; Aus_* stands for a Tasmanian or New Zealand site, while SAm_* stands for a South American site)."***