Dear Anonymous Referee #1

For clarity, we repeat the reviewer's comments in blue italic font and the replies are in black.
Since comments 2, 3, 4 are similar to the comments 3, 2, 1 by the anonymous referee #2, some of the replies are used in common.

Data assimilation in paleoclimatology is a rapidly growing field. The present paper addresses the model-data comparison step that is critical in every data assimilation scheme. Up to now, proxy records are generally first transformed to obtain a reconstruction of simulated variables such as temperature or precipitation before being assimilated. Simulating the measured quantity using proxy system models and performing the comparison directly for this variable provides in theory many advantages. The present study analyses those advantages and the potential limitations of the methodology based on both idealized and realistic experiments. It demonstrates the ability to directly assimilate isotopic composition of several proxies thanks to the application of forward proxy models. The study also identifies the regions/variables where the skill is already satisfactory and the promising ways of improvement. The authors thus provide very interesting results for methodological developments and the application of data assimilation techniques in paleoclimatology. The study thus deserves publications in Climate of the Past but some modifications are required in the experimental design and in the discussion to reach conclusions that are easier to be interpreted and to be compared with recent work as detailed below.
Thank you very much for the positive and valuable comments.

1. *Several groups are currently working on the direct assimilation of proxy records. The authors could not be blamed for not discussing all the very recent publications in the submitted version but a comparison of the conclusions reached here with the ones of Dee et al. (2016) must at least be included as the latter study is focused on a very close subject. In particular, Dee et al. (2016) compare a direct assimilation of isotopes using an isotope enabled atmospheric model with the assimilation of temperature derived from the proxy records, as in the present paper. The publication of those recent papers also requires to modify some sentences like lines 80-81 and 116-117 where it is said that it is the first time that proxy data are assimilated directly (see also Acevedo et al. 2016).*
   Acevedo W., B. Fallah, S. Reich, and U. Cubasch (2016). Assimilation of

PseudoTree-Ring-Width observations into an Atmospheric General Circulation Model. Clim. Past Discuss., doi:10.5194/cp-2016-92, 2016. Available at http://www.clim-pastdiscuss.net/cp-2016-92/

Dee, S.G., N.J. Steiger, J. Emile-Geay, and G.J. Hakim (2016): On the utility of proxy system modeling for estimating climate states over the Common Era. Journal of Advances in Modeling Earth Systems. doi:10.1002/2016MS000677. Available at http://onlinelibrary.wiley.com/doi/10.1002/2016MS000677/pdf

We included the Acevedo et al. (2016) and Dee et al. (2016) in Sect. 1 and modified the corresponding sentences. Also, we included Dee et al. (2016) in Sect. 5.1 to discuss the comparison between proxy DA and reconstructed DA.

2. *I was surprised that the data assimilation method was not described at all in section 2.1. If I am right an ensemble Kalman filter is applied but this is only stated in the conclusions (the word Kalman is mentioned first line 528). A long description of the method is not required but its main characteristics should at least be mentioned in section 2.1.*

The description of data assimilation method is included in the revised manuscript (L133-137). We used EnSRF (Houtekamer and Mitchell, 2001) with slight modification following the previous studies (Bhend et al., 2012; Steiger et al., 2014).

3. *The interpretation of experiment T2-ASSIM and its comparison with CTRL are not straightforward to me as the conclusions strongly depend on the signal to noise ratio selected and it is not possible from the information given in the paper to compare this signal to noise ratio with the error used in CTRL. One option would be to use the model results to estimate the impact of an error of 0.5 per mil on the isotopic composition, as imposed in CTRL, on a temperature reconstruction based on those isotopic records using simple statistical methods (for instance a regression as often done in paleoclimate reconstructions). Then, additional sensitivity experiments can be performed with such a temperature reconstruction derived from the isotopic composition (and not using the temperature simulated by the model) or alternatively assimilating temperature using the signal to noise ratio of this reconstruction that would be compatible with the error imposed in CTRL.*

Thank you for the comments. We modified the experimental setting for T2-Assim following your suggestion. In the modified experiment, temperature is reconstructed from the isotopic records which is used in CTRL by simple regression-based method. Proxies whose correlation with local temperature during calibration period (1871-

1950) is not statistically significant ($p < 0.10$) are removed following Mann et al. (2008). This screening process reduced the available data from 94 to 81 grid points. Based on the correlation between isotope ratio and local temperature, SNR can be estimated through the equation (Mann et al., 2007):

$$\mathrm{SNR} = \sqrt{\frac{r^2}{1 - r^2}}$$

where r is the correlation. SNR is shown in Fig. 8. Subsequently, this reconstructed temperature (T_r) is assimilated. The assimilated result is shown in Fig. 7. The result is slightly degraded in T2-Assim compared with CTRL due to relatively large error in T_r (Fig. 8). As shown in Dee et al. (2016), the reconstruction skill is somewhat compensated by the structure of Kalman gain. Figure S1 shows the correlation scale length to show the difference in the structure between CTRL and T2-Assim. The correlation scale length was found by computing point correlation between the prior (temperature) and the prior-estimated observation (temperature and $\delta^{18}O$ for T2-Assim and CTRL, respectively) for the observation grids, binning these correlations by distance, and computing the mean of each bin. The correlation is consistently high in T2-Assim, which means that the observation information is more effectively used to update the analysis. To sum up, the accuracies are not substantially different among proxy DA and reconstructed DA. However, we should note that this is only the case as long as the relation between temperature and isotope remain the same.

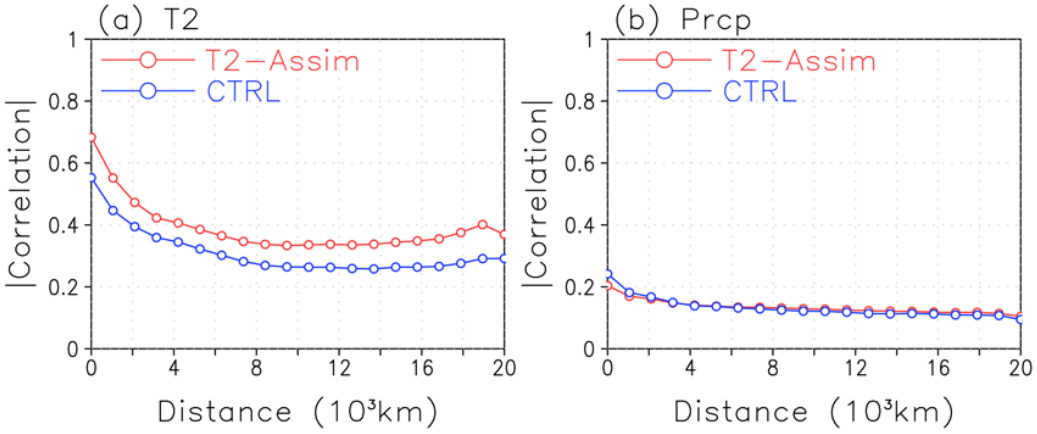

Figure S 1 Mean correlation scale length for T2-Assim (red) and CTRL (blue). The prior is (a) temperature and (b) precipitation. The prior-estimated observation is temperature and

$\delta^{18}O$ for T2-Assim and CTRL, respectively for the both panels.

4. *The low skill of experiment REAL can have many origins: biases in climate models, limitations of proxy system models, non-climatic noise in the data, local signal in the records not represented in large-scale models, etc. The present study does not address the relative contribution of each of those elements and this is perfectly fine for me as it is not the goal of the present study. Nevertheless, some recommendations like line 51, line 497, line 502, line 506 on the improvement of models seems relatively vague and not really justified by the results. I would thus recommend to be more careful and to focus on the main results of the study.*
   Thank you for the comments. We understand that there are multiple factors other than model errors for the low skill in REAL experiment and that we do not know their relative contribution. Thus, we carefully modified the abstract and Sect. 6 in which we avoided arguing vague explanation.

Specific points

1. *Abstract, line 42-43. This sentence is not clear without reading the main text. Please rephrase (see also general comment 2).*
   Thank you for the comments. We omitted the sentence for better readability.

2. *Line 100. The data are not erroneous, this is the interpretation that is questionable.*
   We reworded that part as “such questionable reconstructed data”. Thank you.

3. *Line 143. The ‘simplification’ is valid for some variables but not for others that change more slowly such as oceanic temperatures.*
   We clearly mentioned that the simplification is valid at least for atmospheric variables in the revised manuscript (L150-153).

4. *Line 150-151. What is meant by ‘changing the algorithm’. The text should be more explicit and provide a reference if available.*
   We rephrased the sentence as “the proxy DA could address non-stationarity if one uses temporally varying background ensemble”.

5. *Line 176. A few words should be given on the version of MIROC5 applied as the reference is not available yet. In particular, it should be stated if only the atmospheric*

component is applied (as suggested lines 214-215) or if it is coupled to an interactive ocean.

Thank you for the comments. The version of the model is five (hence MIROC"5") and we used only the atmospheric component of the GCM. To make it clearer, we changed the sentence as "we used a newly-developed model based on the atmospheric component of MIROC5".

6. *Line 189. Why is the deep ocean composition needed for corals that live in shallow waters?*

Thank you for the comment. The isotopic ratio in the upper layer of the ocean is determined by the balance of precipitation, evaporation, and vertical mixing from deeper water, not deep water. We modified the term "deep" to "deeper" in the revised manuscript.

7. *Line 250. I guess the four sensitivity experiments has to be compared to experiment CTRL. This should be already stated at this stage.*

Two of them (i.e. CGCM and VOBS) were conducted to explain the difference among CTRL and REAL and the experimental settings were changed in a stepwise manner, from idealized way to more realistic way. Thus, CGCM were compared with CTRL, and VOBS were compared with CGCM. The other two were compared with CTRL. We included sentences explaining what experiment was used to evaluate each sensitivity experiment.

8. *Line 322. Is it just a repetition of line 318 with a different sign or new information?*

No, it is not. The first sentence described the reconstruction skill for temperature and precipitation by comparing the analysis and the truth. On the other hand, the second sentence explained how the high reconstruction skill was achieved by comparing the assimilated variable (δ) and the reconstructed variable (temperature and precipitation) at the site.

9. *Line 333. Why using 'on the other hand' here?*

The closely correlated area was limited around the observation site for $\delta^{18}O$ in tree-ring cellulose, but the high correlation was not limited around the observation site for $\delta^{18}O$ in coral. Thus, we used the "on the other hand" here. To make the context clearer, we modified the sentence in the revised manuscript (L345-350).

10. *Line 336. The results for temperature should be discussed too.*

    The results for temperature were included in the revised manuscript (L350-351). Thank you.

11. *Line 348. Is this increase noticed in simulation results or in observations? Please be more precise.*

    The temperature has been increased both in observations and simulations. In the manuscript, what we meant was observation. We modified the sentence and put a reference in the revised manuscript (L364-366).

12. *Line 411-412. I would suppress this sentence as it does not bring new information.*

    Suppressed.

13. *Line 415-419. I may miss something but I do not see how the low reproducibility of corals could play a role in the perfect model framework of CTRL as it is assumed that the climate and proxy models have no systematic bias (see also line 496).*

    In this chapter, we compared VOBS and REAL, where VOBS is a perfect model experiment assuming that the climate and proxy models have no systematic bias and REAL is not a perfect model experiment. In the REAL, we assimilated observed data in the real world. Thus, models do have biases.