

## ***Interactive comment on “On-line and off-line data assimilation of palaeoclimate proxy data into a GCM using ensemble member selection” by A. Matsikaris et al.***

**Anonymous Referee #1**

Received and published: 2 September 2014

The paper compares two approaches applied in data assimilation focusing on the last centuries: the so-called off-line approach and the on-line one. As they have been both applied in recent studies but no formal comparison has been performed up to now, the study is timely and the result that no major difference in the performance between the two approaches has been found in the case studied is interesting. Nevertheless, only one test is performed using a specific experiment design. It is thus not clear if the conclusion that the authors has reached could be generalized to other cases or not. This limitation is (briefly) discussed in the text but, if possible, additional tests would be more than welcome, as discussed below. To my point of view, several relatively minor modifications in the text are also required before publication.

C1419

### General points

1/ I guess that the interest of the on-line method decreases as the time between two assimilation steps increases. For instance, in weather forecast, with an assimilation step of one day, the on-line approach has clear benefits. Even though the initial conditions in the ocean may have an impact several years ahead, the limited effect on a 10 year timescale is maybe not surprising. The simulations are performed with a general circulation model, so every test requires a lot of computing power. This may be too demanding for the present study but a test using an assimilation step of one year would be very interesting, for instance. As this is straightforward for the off-line method, this would already be interesting to see if the performance of the method is changing a lot when different length of assimilation step are selected in that case. At least, I would like to see a wider discussion of the fact that the study is only a first step in the characterization of the interest of the on-line versus off-line approach. The one given in page 3465 is too short. The differences between the two approaches may be specific to the target selected for the evaluation of the performance, the period investigated, the variable assimilated, the number of members in the ensemble, the frequency of assimilation, the assimilation method, etc.

2/ I have not checked all the references but still have found that several of them are not cited correctly. I have listed a few of them below but I suggest that the authors verify all of them to be sure that no additional error is present.

Page 3451, line 13. Crowley and Lowery, 2000 is not a spatial field reconstruction. Page 3452, line 28. Crespin et al. (2009) used an on-line method. Page 3454, line 16. The main goal of Steiger et al. (2014) is not to discuss the decadal predictability of the climate but rather to present a data assimilation method adapted for past climates (and should thus be introduced earlier).

3/ The number of figures could be strongly reduced to focus on the main points (or to save space for the additional experiments, see point 1). a/ Is it useful to show the direct

C1420

average of the Northern Hemisphere, as well as the Northern Hemisphere mean in Fig. 2. b/ On-line and off-line results devoted to the same variables should be grouped on the same figure. This means that Figure 1 and 4, Figure 2 and 5, Figures 3 and 6 should be merged. The results of the off-line simulation can be included on Fig. 7. Figure 9 would then become useless. I would not show the individual members in the on-line experiment as the signal present in those individual members is not clear to my point of view. If there is no impact on the initialization, they have similar characteristics than in the off-line approach and have thus no specific interest. If they are different, then the authors have to justify why this has no influence on the performance of the method.

#### Specific Points

Page 3453, line 7-8. It is not clear if the sentence beginning by "In the particle filter approach" refers to an on-line or off-line approach.

Page 3453, line 15. The sentence is a bit heavy with "include" and "including" in the same line. Please rephrase.

Page 3453, line 20. I would not say "at the DA" (furthermore "use" and "using" are present in the same line). Please rephrase.

Page 3455, line 21. I do not see why the long control simulation for 1850 is mentioned. How is it used in the present framework? Is it the initial state for the spin-up with constant 850 boundary conditions?

Page 3456, line 14. What are the "unobserved forcings"?

Page 3458, line 13. What is meant here by biases: an error on the mean state or also on the variance?

Page 3459, line 18. Volcanic forcing likely had also a role in the cooling.

Page 3459, line 21. There are more recent and more comprehensive references on

C1421

the past changes in ENSO compared to Jones and Mann (2004).

Page 3459, line 25. "found that to find"; please rephrase.

Page 3461, line 14. The correlation with the ensemble mean is higher than for the individual members but I guess this comes at the expense of a much lower variance for the ensemble mean. If it is actually the case, this is a bias of the ensemble mean that can be mentioned.

Page 3461, line 23-25. I disagree that "The fact that the RMS error of the ensemble mean is lower than the error of most of the individual members indicates the influence of forcings in some continents". A lower RMSE can be simply due to the lower variance of the ensemble mean. A time series with a constant zero anomaly (and thus no response to the forcing) may also have a lower variance than individual members. This should be discussed.

Page 3462, lines 7-14. The discussion of the changes in the Southern Hemisphere is very speculative and does not include many of the recent references on the subject. I would remove it as it is questionable and is not the main focus on the study. If the authors wish to keep it, a deeper discussion is needed.

Page 3465, line 17. I do not understand why "The use of proxies with the minimum possible noise would give a better chance to the on-line approach to capture the true climatic state".

Page 3465, line 24. "some of the weaknesses"

Page 3466, line 5. See the remark above on the same topic.

Figure 8. The colors are not clear. I would recommend to use one color for all the individual members as there is no need to identify them and clearly contrasted colors for the ensemble mean and the analysis.

---

Interactive comment on Clim. Past Discuss., 10, 3449, 2014.

C1422