Clim. Past Discuss., 10, C1825–C1842, 2014
www.clim-past-discuss.net/10/C1825/2014/

Climate of the Past Discussions
Open Access

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.

A. Matsikaris et al.

axm368@bham.ac.uk

Received and published: 3 November 2014

Response to anonymous Referee #1

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.

We thank the reviewer a lot for his/her very useful and constructive comments. The points mentioned by the reviewer are addressed in our revised manuscript.

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.

CPD
10, C1825–C1842, 2014

We acknowledge that the study is only a first step in the characterization of the interest of the on-line and off-line approaches. However, as the reviewer correctly points out, the big amount of computing power required by simulations performed with a general circulation model did not allow us to perform tests with different assimilation timesteps. Additionally, the 10-year resolution of the North American proxy-based reconstructions did not allow us to use annually resolved proxy data, since we aimed for a complete NH reconstruction. We performed though an additional test regarding the information propagation (section 3.3, paragraph 3), as well as a significance test to evaluate the role of random sampling effects (section 3.2), as suggested by the referee #2 (see response to referee 2). A wider discussion of the limitations of the case study and the fact that the study is only a first step in the characterization of the interest of the on-line and off-line approaches has also been made, in our revised discussion section (section 3.3, paragraphs 4 and 5).

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).

All references have been checked in the revised text and the above mentioned mistakes have been corrected.

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 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.

This is a very good suggestion, and the number of plots has now been reduced from 10 to 6 in the revision (while a seventh one regarding the MOC was added). This forced us to change the structure of the paper as well, with no longer separated sections of on-line and off-line validations.

a. We show both the direct average of the NH and the NH mean, as the first is directly comparable to the proxy datasets, which are only available as continental means, whereas the second one is the usual mean temperature given in most climate studies, but not the direct equivalent of the proxy-based reconstructions (mentioned also in the text).

b. Figures 1 and 4 and Figures 2 and 5 have been merged in the revision (new Figures 2 and 3 respectively), with the individual simulations being replaced with shading for the range of the different ensembles, to facilitate the comparison between on-line and off-line methods. Figures 3 and 6 were also summarized in one figure in the revision (new Figure 4), with the ensemble mean and analysis being represented with different symbols and colours for both the on-line and off-line ensemble. Figure 7 now includes the results of the off-line simulation too (new Figure 5). Figure 9 became

CPD
10, C1825–C1842, 2014

indeed redundant and was removed. We also included the results of the off-line simulation in Figure 8 (new Figure 6).

As for the final comment, we do not show the individual ensemble members in the on-line experiment anymore, as indeed they are partly influenced by the initialization but also show a strong component of random, internal variability. Following a suggestion by reviewer 2, we have investigated the influence of the initialisation by comparing the standard deviation of the ensemble during the decadal sub-periods in the on-line and off-line case. We found that the ensemble spread in the on-line case is smaller, which is evidence for the influence of the initialisation during the entire decadal assimilation time-step. We have included a remark on this in the paper in Section 3.3, paragraph 3.

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.

It can refer to both, normally to the on-line, but also to the off-line as for example by Annan and Hargreaves, 2012, who given the limited predictability of the system, followed an off-line technique by simply reverting to the climatological prior for each individual year. We clarified this in the revision.

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

The sentence was rephrased in the revision.

Page 3453, line 20. I would not say "at the DA" (furthermore "use" and

CPD
10, C1825–C1842, 2014

"using" are present in the same line). Please rephrase.

The sentence was rephrased in the revised text.

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?

The long control simulation for 1850 is not relevant and should not be mentioned; therefore we removed the phrase from the revised manuscript.

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

We changed this phrase because indeed it did not make much sense.

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

We mean both. This was clarified in the revision.

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

We thank the referee for their remark. This was added to the text.

Page 3459, line 21. There are more recent and more comprehensive references on the past changes in ENSO compared to Jones and Mann (2004).

We removed this sentence since it is not the focus of this paper.

CPD
10, C1825–C1842, 2014

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

Sentence rephrased in the revision.

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.

Indeed, the ensemble mean has a higher ratio of forced to random variability and thus a higher correlation than the individual members, but because the random components of the individual members partly cancel each other, the total variance of the ensemble mean is much lower than the individual members. We mention this in the revised text.

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.

This is a true statement. The fact that the RMS error of the ensemble mean is lower than the error of most of the individual members might either indicate the influence of forcings, or can be simply due to the lower variance of the ensemble mean compared to the individual members, which might bring it closer to the proxies. This is corrected in the revision.

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.

This is a fair point from the reviewer. We removed this discussion from the revised text.

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".

We believe that 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, as they would represent the true climate better and the correct information would be propagated when applying the on-line approach, whereas the off-line one would not be benefited to the same extent, as it is an a posteriori selection. We now explain this in the revised text.

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

Change made in the revision.

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

As above, we now mention that the RMS error of the ensemble mean is lower than the error of most of the individual members either due to the influence of forcings, or simply due to the lower variance of the ensemble mean compared to the individual members.

Figure 8. The colors are not clear. I would recommend to use one color

CPD
10, C1825–C1842, 2014

for all the individual members as there is no need to identify them and clearly contrasted colors for the ensemble mean and the analysis.

Changes have been made to this figure, with the off-line results also being added.

Response to anonymous Referee #2

General comments:

The paper presents a comparison of two closely related on-line and off-line data assimilation approaches in paleoclimatology. This is – to my knowledge – the first study that has specifically addressed this problem. The authors find very little difference between the two methods in large-scale metrics and conclude that either their assimilation approach offers too little control of the slow components of the climate system (the ocean) or there is very little information propagation on decadal timescales. Due to the specific choices of the approach and due to the wide range of alternative choices (and approaches), the conclusion drawn in the paper may not apply in general. Nevertheless, the manuscript offers a valuable starting point for further discussion.
The manuscript is generally well written, but there is quite some redundancy in the discussion and presentation of the results. Also a few issues require further attention (see comments below). Therefore, I suggest to accept the paper with major revisions.

We thank again the reviewer for his/her very useful comments.

CPD
10, C1825–C1842, 2014

Lack of information propagation

The authors argue that the similarity of the on-line and off-line approach points to either a lack of control of the slow components, or a lack of information propagation on decadal scales in the model. While the first hypothesis is impossible to test without additional data (that may not exist), the information propagation on decadal timescales could be tested. In the on-line assimilation all ensemble members are initialized with the same ocean state at the beginning of each decade. Therefore, if there is information propagation, one would expect less spread in the on-line ensemble at the end of the decade than in the off-line ensemble (starting from different ocean initial states). I strongly suggest to analyze and discuss this and the consequences in the paper. To be able to better interpret the similarity between the on-line and off-line approach might greatly strengthen the conclusions drawn from this paper.

This was a very good point; we implemented the additional test and discussed it in paragraph 3 of the revised section 3.3. We tested the information propagation on decadal timescales by calculating the standard deviation of the ensemble spreads for the on-line and off-line methods for the different continents. The results are shown in table 4. For the NH direct average, we computed the standard deviation of the ensemble spreads for the whole period (for every year of the simulation period), as well as for the last year of each decade, as suggested by the referee, and then computed the mean of these standard deviations. The standard deviations were 0.25 for the on-line compared to 0.30 for the off-line in the yearly test, and 0.28 compared to 0.31 respectively for the final year test. For the NH mean, the differences were a bit smaller. The standard deviations were 0.19 for the on-line compared to 0.23 for the off-line in the yearly test, and 0.22 compared to 0.23 respectively for the final year test. All the results show that the members are slightly closer together in the on-line experiment, a fact that is also in agreement with Figures 2 and 3. No significant difference is observed between the all-year ensemble spreads and the spreads of the

CPD
10, C1825–C1842, 2014

last year of each decade. The smaller spread in the on-line ensemble than the off-line ensemble (which starts from different ocean initial states) is a hint for information propagation. However, we note that it is not clear from this analysis whether the information propagation is strong enough to lead to substantially higher skill of the on-line DA method.

Uncertainty analysis

Sampling effects may affect many of the aspects discussed in the study due to the limited ensemble size and relatively short time period analyzed. Therefore, sampling uncertainty should be more thoroughly addressed where possible. For example, resampling might be used to illustrate the distribution of skill metrics (correlation, rmse) when randomly sampling a best model in the off-line method. This may help to interpret the significance of the differences between the on-line and off-line method. Also a resampling approach could be used to illustrate the dependency of skill of the off-line ensemble member selection on ensemble size (i.e. from 1 to 10 members) instead of justifying ensemble size based on a study with an atmosphere-only GCM.

This was also a very good idea which we implemented as an additional test and describe in Section 3.2 of the revised manuscript. We applied a resampling method to illustrate the distribution of the skill metrics (correlation and RMS error) when randomly sampling a best model in the off-line method. We calculated the correlations for the NH direct average for 100 random analyses in the off-line experiment, after randomly selecting one member as the best for each of the 10 decades. The mean correlation of the randomly sampled distribution with the proxies was 0.48 (with a standard deviation of 0.21), ranging between negative values and 0.8. These correlations are very low compared to the 0.94 correlation found for the off-line DA analysis. For the NH mean the correlation of the resampling was 0.63 (with a standard deviation of 0.15). The same resampling experiment was performed for the RMS error of the NH direct

CPD
10, C1825–C1842, 2014

average. The mean RMS error was 0.62 (with a standard deviation of 0.13), ranging between 0.3 and 0.95. The RMS error found for the off-line DA analysis was only 0.11, falling well outside the above range. Similarly, for the NH mean the RMS error of the resampling was 0.51 (with a standard deviation of 0.10). The results show that the skill of the DA analysis is significantly different from the skill obtained from the random sampling.
Regarding the dependency of skill of the off-line ensemble member selection on ensemble, we are preparing another study which includes this subject.

Redundancy in presentation and discussion of results
I strongly urge the authors to shorten the presentation and discussion of results in the manuscript. This concerns in particular the presentation of results in figures. Figure 3 and 6 for example could be summarized in one figure with the spread in ensemble results represented by a bar (or boxplot), and the ensemble mean and analysis as symbols for both the on-line and off-line ensemble. Similarly, Figures 1 and 5 might be superimposed (replacing the individual simulations with shading for the range of the different ensembles). In my opinion, superposition of the results of the on-line and off-line approach would be highly beneficial to the reader as the main point and novelty of the manuscript is the comparison of the two approaches. Therefore, the results should also be presented in a way that facilitates this comparison. Similarly the discussion of the RMSE and correlation of the individual approaches may be reorganized to better reflect the similarity of the methods and focus more on the comparison while at the same time improve readability.

Thanks again for the suggestion, the changes have now been implemented in the revision and four plots have become redundant and removed, as discussed above at the response to the referee 1.

CPD
10, C1825–C1842, 2014

Specific comments

Title: The title is a bit long and very specific. Maybe "On-line and off-line data assimilation in paleoclimatology: a case study" or something along those lines could help to shorten the title.

We are grateful that this remark was pointed out. The title has now been changed according to the suggestion.

P3451L11: Large-scale and spatial field reconstructions are contrasted without motivation why this is relevant for the manuscript. Please clarify.

The motivation lies to the fact that data assimilation has a target similar to that of the spatial field reconstructions. We mention this link in the revised text (Section 1, paragraph 3).

P3451L15and22: errors vs. uncertainties. Please either clarify why you use different terms to describe model and proxy errors/uncertainties or use the same term twice.

We call them both errors in the revision.

P3451L14+: Please add references to substantiate the specific limitations of proxy series mentioned in the following two sentences.

References added.

P3452L19: The authors mention that Goose et al. have performed on-line and off-line data assimilation already. Please clarify that the novelty of this

CPD
10, C1825–C1842, 2014

manuscript is the focus on the comparison of the on-line and off-line approach.

We clarify this in the revised text in the introduction.

P3453L6: I do not fully understand the meaning of involved in this sentence. Do you want to say that the off-line assimilation is difficult to apply or that the method of Annan and Hargreaves is an off-line assimilation based on a simple likelihood weighting algorithm? Please reword.

We meant the latter. The sentence has been rephrased to make this clear to the reader.

P3454L10: First mention of decadal, please better motivate why multi-yearly (decadal) is important for climate reconstructions. If proxies with yearly and seasonal resolution are available for assimilation, long memory (as in decadal) may be less important.

The experiment design with decadal assimilation is motivated by a number of reasons. Firstly, since we aimed for a complete Northern Hemisphere reconstruction, the 10-year resolution of the North American proxy reconstructions didn't allow us to use annually resolved proxy data for the assimilation. Additionally, the annually resolved proxies include substantial noise, which is cancelled out with the decadal averaging. Finally, in a climate change context, the yearly changes are in general of less interest compared to the decadal variability. This is now discussed in the revised introduction.

P3454L13: Please provide reference for "up to decadal predictability in the North Atlantic"

CPD
10, C1825–C1842, 2014

Reference has been provided in the revised text.

P3456L4: I would love to see evidence (a figure) of the growth of ensemble spread in ocean variables such as AMOC. Understanding of how quickly the ensemble diverges would also help the readers to better understand the particularities of the DA method (see general comment on information propagation).

A plot showing the growth of ensemble spread in the AMOC has been added to the revised manuscript, showing how quickly the ensemble diverges in the first 100 days of the simulation (new Figure 1).

P3457L6: is the 10 to 30-year resolution of the North American proxy reconstructions the reason for the experiment design with decadal assimilation?

This is one of the reasons, see comment above.

P3458L14: double negative, please rephrase.

Change made in revision.

P3459L15: Consider moving the first two paragraphs to the introduction and methods sections to improve readability as these are clearly not your results.

Change made in the revision, while the structure of the manuscript has changed in general.

P3460L15: I suggest to add a table with correlations for the different experiments (as for rmse) to simplify discussion and comparison.

CPD
10, C1825–C1842, 2014

A table was added in the revised text.

P3461L13: please indicate the range of the correlation for the individual proxies (maybe also in table).

We guess that the referee here means individual members, not proxies. If so, we would prefer not to do this as we believe that this information is provided with the ensemble mean, and the on-line members do not have a clear signal due to their construction.

P3461L23: This is not true. The lower RMSE of the ensemble mean is simply due to the lower variance of the ensemble mean (Annan and Hargreaves, 2011). If you were to find that the ensemble mean is more highly correlated with the proxy series than most of the individual members then you might argue to find evidence for the importance of forcing.

This is a fair point. We discussed this in the response to the referee 1 and changed the revised text.

P3462L1: as presented above.

Change made in revision.

P3463L11: on the basis of

Change made in revision.

P3463L11: If you mean "is not expected to be more skillful than ALL the

CPD
10, C1825–C1842, 2014

individual Members" this may be true, but maybe beyond your point. I would disagree with the general statement, however, as we expect the analysis to be more skillful on centennial, hemispheric scales even if we train the analysis with decadal, continental data (not least due to dependence of the continent and hemispheric averages).

We partly understand the expectation of the reviewer, however our expectation was different as nothing in the cost function assured a more skilful analysis on the centennial scale. This was what showed up in the results too.

P3464L13: insufficient control of the ocean state only affects the on-line method. The off-line method is an a posteriori selection for which the ocean state is irrelevant (as long as the ensemble is sufficiently large).

We thank the reviewer for spotting this mistake; it has been corrected in the revised text.

P3464L6: missing period. 1640-49 AD.

Change made in revision.

P3465L1: We know whether the ocean has predictability in this experimental set up (see general comments).

This has been changed after conducting the additional test that the reviewer suggested.

P3473: Please specify what the dashed (dotted) grey lines mean. Also please consider moving to a standard line plot to be able to better differentiate the lines

CPD
10, C1825–C1842, 2014

(if you decide to show all the individual members, see general comments).

The plots have been changed in the revision as explained above.

P3480: Please use different symbols to distinguish the ensemble mean and analysis from the individual members (and see general comments on presentation).

Changes to the figure have been made.

Please also note the supplement to this comment:
http://www.clim-past-discuss.net/10/C1825/2014/cpd-10-C1825-2014-supplement.pdf

---

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