Clim. Past Discuss.,
https://doi.org/10.5194/cp-2018-168-RC1, 2019

Climate of the Past Discussions

Interactive comment on "Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation" *by* Veronika Valler et al.

Anonymous Referee #1

Received and published: 4 February 2019

Review of CP-2018-168

This paper tests several different methodological choices that are typically made (or could be made) in paleoclimate reconstructions using DA. I think this presents a good and valuable presentation and discussion of these choices. The findings and suggestions for future reconstructions are very helpful to the community performing these types of reconstructions.

Section 3.1: Is there any specific justification for the choice of the ratio of L_z and L_m being 2:1? Could this, or some other ratio, be justified by looking at the correlation length scale in observational data?

Section 3.1: Is there any justification for the specific localization values that you chose for each variable that was reconstructed? Are these values data-driven or just educated guesses? Were any experiments done to test on optimal localization value? I would assume that if these values were used based on weather DA experiments, they might not apply on the longer paleo time scales where one would generally expect the correlation length scales to be larger.

Section 4.2.1: When you are comparing the distributions, you say that for example, the most skillful reconstruction is obtained from the 100c_PcL experiment. What is the basis for saying it's the best? What aspect of the distribution are you comparing? The median or some other specific value(s)?

Many of the distributions shown in the figures look very similar so it was hard for me to feel confident about the statement that one particular set of reconstruction choices was better than another. Are the distributions statistically distinct? Instead of comparing the distributions, would it be possible to show the differences compared to the "original" reconstruction (i.e., you'd compute the difference in the skill score for each location and then summarize this distribution of differences in the plots)? I'm wondering if this, or something similar, might make the differences more clear. Because currently when I look at the distributions, many of them look very similar and perhaps even statistically indistinguishable.

Fig 8 & 10: It would be very helpful to give a little more explanatory information/labeling on each panel, such as was done in Fig 3.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2018-168, 2018.