Clim. Past Discuss.,
https://doi.org/10.5194/cp-2017-61-AC4, 2017

Climate of the Past Discussions

Interactive comment on "Pseudo-proxy evaluation of Climate Field Reconstruction methods of North Atlantic climate based on an annually resolved marine proxy network" *by* Maria Pyrina et al.

Maria Pyrina et al.

maria.pyrina@hzg.de

Received and published: 21 June 2017

Answer by M. Pyrina et al. to: Anonymous Referee #3

We would like to thank Referee #3 for the constructive comments. We will try to be clearer in the introduction about the broader significance of the work and expand our analysis including more validation statistics as the reviewer suggested. Also, we would like to thank the reviewer for the technical corrections.

Specific comments:

1) Why choose calibration periods that are earlier than the validation period? Given

that every real-world climate reconstruction would be calibrated during the later period during which instrumental data is available, wouldn't it make more sense to calibrate the PPEs using the 1850-1999 period and then validate on the earlier time periods (Medieval, LIA, etc)? To me, that seems more intuitive and would still address the stationarity issue.

We agree with the reviewer that every real-world reconstruction would be calibrated in recent times in order to reconstruct past times, but we conducted the experiment in this way in order to be able to additionally compare our model-based pseudoproxy results with the pseudoproxy results based on the reanalysis data. Furthermore, this approach gives us the opportunity to perform tests on calibration periods that are considered to be different in their climatic background state. Although from a practical point of view the choice of the real-world calibration period is constrained by the availability of meteorological observations, the climate models allow us to circumvent these issues and carry out sensitivity experiments in the context of the virtual world of the climate model.

2) Is there any reason why these two particular CFR methods (PCR and CCA) were chosen, while other common CFR methods (e.g., RegEM-TTLS and RegEM-ridge) were excluded? I'm not necessarily suggesting that the authors need to redo the analyses with additional CFR methods, but I would at least like to see a little justification for why these methods were chosen while others were excluded.

We agree with the reviewer and therefore plan to further justify the selection of the two methods in the introduction. These methods are widely used in paleoclimate reconstructions. Therefore, we used these two methods in order to check how sensitive the results are depending on the analysis we choose.

3) I think it would be useful to include other complementary validation statistics, such as mean bias, coefficient of efficiency (CE), reduction of error (RE), and/or root mean squared error (RMSE). I'm not sure that only correlation and standard deviation ratio

are enough for a robust assessment of model performance. A spatial assessment of mean bias and either CE or RMSE could add important information to this study.

We agree with the reviewer that complementary validation statistics could add important information to this study and therefore we plan to include the Reduction of Error (RE) and the Root Mean Squared Error (RMSE).

4) How is it possible to have a standard deviation ratio greater than 1 for the CCA results? As I understand it, all parametric reconstruction methods will result in at least some variance loss unless the proxy is perfectly correlated with the target climate variable (McCarroll et al. 2015), which would only be the case at the particular grid cells with the noise-free pseudoproxies. Even in other studies that used noise-free pseudoproxies (e.g., Smerdon et al. 2010, 2011), at least some variance loss was observed. I therefore don't see how the reconstructed SSTs could have grid cells with greater variance than the "observed" (or in this case, model-simulated) SSTs.

Thanks to the reviewer we identified a small error in the scripting that does not change the conclusions of this manuscript, but due to that the SD ratio regarding only the CCA results was slightly overestimated in the regions lacking proxies. In the revised version the necessary corrections will be done.

5) The authors state that they chose to retain the first 10 PCs in PCR and the first 5 EOFs in CCA. How sensitive are the results to this choice? How were these thresholds chosen? Why were different thresholds chosen for the two CFR methods? As it stands, these seem like arbitrary choices. Did the authors consider more objective criteria for determining these thresholds, such as the "estimated noise continuum" approach used by Mann et al. 2007 or an optimization approach similar to Smerdon et al. 2010?

For the PCR method we retained 10 EOFs because they represent more than 90% of the spatial co-variance of the North Atlantic SSTs and in this way we capture most of the NA SST covariance. In the CCA method we retained 5 EOFs, as 5 is the maximum number of EOFs that we can keep in the case of the proxy field, because this number

depends on the number of proxy locations. We will add some lines of explanation in the method section.

6) I would like to see some more discussion about the stationarity of the PPEs. Specifically, there is an implicit assumption when creating the pseudoproxies that the proxy response to SST variability is stationary (since the pseudoproxies are just SST+noise). Is this necessarily a realistic assumption for real-world A. islandica, or is it possible that the response of this species to SST variation could be non-stationary (similar to the well-known "divergence problem" in high latitude tree-ring widths)?

It might indeed happen that the response of a living animal changes through time, but so far there is no known "divergence problem" in Arctica islandica. However, there may be other potential sources causing non-stationarity in the response of bio-physiological proxies (turbidity, salinity, food availability) or basic changes in the ecosystem functioning which are not accounted for in our approach testing for stationarity. These questions are difficult to model statistically, as the non-stationarity may arise with very different character and has not been clearly characterized in real proxies, as the dendroclimatological divergence problem illustrates. We will briefly discuss how a pseudo-proxy test could address some non-stationarities, e.g. by introducing AR1 coefficients that are themselves random variables.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2017-61, 2017.