Clim. Past Discuss.,
https://doi.org/10.5194/cp-2018-17-RC1, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Climate of the Past
Discussions

Interactive comment on "How wrong are climate field reconstruction techniques in reconstructing a climate with long-range memory?" *by* Tine Nilsen et al.

Anonymous Referee #1

Received and published: 13 April 2018

Thanks for the opportunity to review this interesting paper.

Please see the attached files for the content of the review. The first two files contain pp. 1 and 2 of the General and Specific Comments. The third file has further itemized notes for the Specific Comments.

Please note that "minor revisions" is selected for this article. This can potentially change to "major revisions" depending on how the authors respond to the question raised in the third paragraph of the Specific Comments, copied here.

"The most important alteration that may potentially be required concerns the nature of the hypothesis testing confidence intervals the authors utilize. In section 3.1, the

authors describe the Monte Carlo (MC) estimation of the theoretical confidence ranges they utilize for testing of the results against both fractional Gaussian (fGn) and AR1 null hypotheses. Since the actual tests themselves evaluate the mean power spectra of the ensemble experimental results in relation to these distributions, the question arises as to whether it is more appropriate to use estimated theoretical distributions of these processes directly as the authors do, or rather to use estimated distributions of the means of same-sized ensembles of the theoretical values. It is not within the statistical knowledge of this reviewer to resolve this question, but it is asked of the authors to evaluate whether it is applicable in this context."

Please note that the supplemental file is best used when the Comments bar is opened, to see all the comments, etc. as a sequential listing.

Please also note the supplement to this comment:
https://www.clim-past-discuss.net/cp-2018-17/cp-2018-17-RC1-supplement.pdf

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2018-17, 2018.

Referee Comment on:

How wrong are climate field reconstruction techniques in reconstructing a climate with long-range memory? Tine Nilsen, Johannes P. Werner, and Dmitry V. Divine (cp-2018-17)

General Comments

Overall, this paper provides an excellent addition to the literature concerning the characteristics of climate field reconstructions (CFRs). Specifically, its evaluation of the BARCAST CFR methodology provides an excellent isolation of how the spectral characteristics of spatial and mean field of reconstructions derived using it might/might not be biased by the temporal and spatial specification of the fundamental BARCAST statistical model. By way of context, it is worth noting that there has been interest concerning how the specification of the fundamental statistical model might affect the characteristics of reconstructions derived using BARCAST and related methods. The kind of well-designed, highly-specific experimental design the authors have implemented in order to clearly isolate fundamental characteristics of the method is a highly useful addition in this field.

Specific Comments

The results the authors present appear to be well-developed and without substantial technical issue, with the possibility of one exception mentioned in the third paragraph of this section. As the authors note, it does not appear surprising that BARCAST might tend to retain long-term memory information with better fidelity at the spatial mean scale than at the local scale, since the local disturbance term of the spatial covariance specification will tend to average out. Similarly, it is not surprising that the local reconstructions produced by BARCAST are generally of highest quality where there is co-located predictor information, due to the formal specification of the statistical model that necessarily relies on stochastic infilling based on all the model's estimated parameters for the away-from-predictor locations. In this way BARCAST differs from, as one example, CFR methods that reconstruct (typically) orthogonal components of entire fields directly, although such methods necessarily may introduce their own issues regarding the spatial and spectral fidelity of the reconstructions to the true fields they target. It is of interest to utilize a similar experimental design to that presented here to evaluate these, and other, CFR methods.

The authors appear to apply their evaluation criteria conservatively, notably by generally focusing on the fidelity of the entire temporal range of spectral characteristics for the reconstructions, rather than focusing to a significant degree on evaluation of results for specific temporal ranges. The use of metrics that compare cumulative distribution functions (CDFs) is interesting, and in this context it would be significantly helpful if the authors would describe in greater detail what the "Reliability" metric they derive and utilize indicates, beyond providing a

Fig. 1. This file is not a figure, but rather p. 1 of the the overall review text

single reference. This is particularly important because one of the key results indicated by these metrics is that there is good agreement with the true target data at co-located data and predictor locations, but that the associated confidence ranges do not show good agreement. This is a tantalizing indication that needs more description, in particular to understand the nature of the confidence range disagreement, which would be highly useful information.

The most important alteration that may potentially be required concerns the nature of the hypothesis testing confidence intervals the authors utilize. In section 3.1, the authors describe the Monte Carlo (MC) estimation of the theoretical confidence ranges they utilize for testing of the results against both fractional Gaussian (fGn) and AR1 null hypotheses. Since the actual tests themselves evaluate the *mean* power spectra of the ensemble experimental results in relation to these distributions, the question arises as to whether it is more appropriate to use estimated theoretical distributions of these processes directly as the authors do, or rather to use estimated *distributions of the means of same-sized ensembles* of the theoretical values. It is not within the statistical knowledge of this reviewer to resolve this question, but it is asked of the authors to evaluate whether it is applicable in this context.

Assuming that the above question is resolved in terms of retaining the existing hypothesis test structure, the most important request for revision is to add explanatory text, along with a smaller set of corrections and other additions, at a number of places in the article. These are all noted in the accompanying PDF document, using the PDF comment and editing capabilities. This request is made to help clarify and contextualize the descriptions in those places for the broader paleoclimatology science community, given the relatively mathematical nature of the article.

Perhaps the most significant of these is to consider adding thoughts about how BARCAST itself might be improved in the Conclusions section. Are there feasible changes to the fundamental temporal and spatial specifications of BARCAST that this research might suggest to consider going forward?

The reviewer thanks the authors and editors for the opportunity to evaluate this quite interesting article.

Fig. 2. This file is not a figure, but rather p. 2 of the the overall review text