Clim. Past Discuss.,
doi:10.5194/cp-2015-192-RC1, 2016

Climate of the Past Discussions

Interactive comment on "Influence of proxy data uncertainty on data assimilation for the past climate" *by* Anastasios Matsikaris et al.

Anonymous Referee #1

Received and published: 19 February 2016

The authors aim to investigate reasons for the lack of regional skill which has been often found in experiments when large-scale temperature information is used to constrain or reconstruct temperature fields. Their main conclusion is that this is not due to data errors per se.

I have a major concern with the method being used. Selecting the best simulation out of 20 is not much of a data assimilation method (at least in the case of a high dimensional problem). I see no reason to expect that it will even give repeatable results. That is, a different set of initial random perturbations might give substantially different results, and thus the differences between the two experiments may be an artefact of the specific experiment. I don't see a good way round this other than to repeat the experiments and to see what happens. However, even in the case that the results are repeatable, the method is still very poor and it is debatable whether it really qualifies

as data assimilation.

The authors find no skill on small scales and conclude correctly that data errors were not the cause of this, since this result occurs even when near-perfect data are used. However, while I agree it is certainly true that continental scale temperature does not provide much small scale information, their experiments seem neither necessary nor sufficient to support this conclusion. Not necessary, because of course the information content of large temporal and spatial averages is far too small to distinguish between different climate states, in the same way that describing this page of text as 10% black and 90% white (I'm guessing) cannot tell you what is written on it. Not sufficient, because an equally possible explanation (which is certainly also true) is that even if the data used did have sufficient information, the method applied is almost certainly inadequate to take advantage of this. If the authors doubt my claim, a simple test of this would be to apply the same method while using the whole temperature field data in the assimilation process. That is, pick the ensemble member which fits the whole field best according to their preferred metric, etc. I am confident that such an approach with an ensemble size of 20 - even when the data indisputably does provide all the small scale information - would produce rather poor results, simply because the small ensemble will never contain a simulation which is anywhere near the data. This is essentially the same result that is well known in NWP for simple particle filtering methods (Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. L. (2008). Obstacles to High-Dimensional Particle Filtering. Monthly Weather Review, 136(12), 4629).

The method used in this manuscript does not, as far as I am aware, have any real justification in terms of solving the Bayesian updating problem implicit to data assimilation. The most that can be said for it is that it selects a simulation which is somewhat closer to the assimilated observations than a randomly-initiated simulation would be. Moreover, in high dimensional applications, the improvement may be very small. Again, the analogy with a page of text may be instructive. If you generate 20 random pages of text and choose the one that has closest to 10% ink, you would not expect the content

to relate any more meaningfully to my review, than the other 19 rejected pages. The most you could say is that it has the same amount of ink.

I have some more minor concerns. For example, the correlations reported for the results might not be as impressive as they appear at first. I suspect that at least part, perhaps a large part, of the correlation is due to the changing forcing applied over the 20th century. You can for instance see what appear to be strong correlations at regional and decadal scale in the famous figure SPM.4 in the 2007 IPCC report, in which simulations there was of course no data assimilation at all, the relationship being purely due to external forcing. In order for your statistics to be informative, it would be necessary to remove the forced element from them. Or perhaps as a simpler alternative, present the correlations achieved by simulations with no data assimilation.

Additionally, some of the comparisons between the methods seem unhelpful. For example, the cost functions differ in their scaling, so cannot be validly used for comparison between the methods (p6 l20). RMS errors could be directly calculated either with no scaling (it might be useful to know how accurately the results match the data, in standard units) or else with a variance scaling that is consistent across the methods. As it stands, there is no way I can easily tell whether the cost of 1.27 for DA-P actually represents a worse model-data mismatch than the 0.61 for DA-I.

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2015-192, 2016.