Clim. Past Discuss.,
doi:10.5194/cp-2016-129-RC2, 2017

Climate of the Past
Discussions

Interactive comment on "Reconstructing past climate by using proxy data and a linear climate model" *by* Walter A. Perkins and Gregory J. Hakim

Anonymous Referee #2

Received and published: 30 January 2017

Summary: Perkins and Hakim demonstrate climate field reconstruction using an "online" data assimilation technique. In their methodology, time slices of the reconstructed climate field are tied together with a Linear Inverse Model (LIM) approximation to a coupled climate model (which are themselves far too computationally expensive to assimilate proxy data in a climate reconstruction context). Perkins and Hakim implement the approach for estimating surface temperature, using LIMs calibrated on several different coupled climate models, and compare results to an offline approach. For certain choices of a blending parameter, the authors are able to achieve improvements in global mean temperature with all of the LIMs compared to the offline case. Changes in field reconstruction skill compared to the offline case are non-uniform across space, but tend to improve on average for some of the LIMs.

General Comments:

This paper is a very nice contribution to the field of paleoclimate reconstruction. Data assimilation approaches have been of interest to the paleoclimate community for a long time, and an "online" version represents a real advance- in my opinion this is much more satisfying than the offline reconstruction approaches. The authors have implemented a non-trivial experimental design, and it's a cherry on top that the new approach even yields improvements in skil over past methods.

The presentation of the work in the paper feels incomplete in several ways. The figures show comparisons of metrics of skill, and comparisons of estimated climate fields to a benchmark estimate, but no visualization of the reconstructions themselves, or comparisons to the actual target (the GISTEMP field and/or GMT time series). Though the interest of the authors is primarily on the increases in skill compared to previous reconstruction approaches, the paleoclimate community will want to see that estimates from this approach are reasonable compared to the actual target field. I appreciate that it's not reasonable to provide such figures for each of the many reconstruction experiments, but perhaps show some visualization of the GMT time series and error of reconstructed field relative to GISTEMP for one LIM experiment at the optimal value of the blending coefficient "a", and put the others in a supplementary document. (If the total number of figures is a concern, the authors might consider putting figures 1-3 into supplementary documentation and keeping just figures 4 and five in the main text, as all 5 are a bit redundant).

The authors have also neglected to describe of tabulate the computational expense associated with their reconstruction exercises. In addition, I wonder if they plan to make code for carrying out any of the reconstructions publicly available. Sharing code is perhaps the very best way to get other researchers to use and build on (and cite!) the advances in your work. One such place the authors might consider archiving their code would be the NCDC NOAA Paleoclimate Software Library: https://www.ncdc.noaa.gov/paleo/softlib/

Finally, more commentary on interpretation of the model and results are needed in the discussion. In general I am a fan of combined results & discussion sections. In the present case though, the authors can provide much more "discussion" along with the delivery of the "results." Additional interpretation, and speculation as to reasons for observed results, will make the paper much richer, more interesting, and scientifically valuable.

Specific Comments:

Title: Suggest changing the title to assert what is novel about this paper: paleoclimate fields are reconstructed using an *online* data assimilation scheme. How about something along the lines of "Reconstructing paleoclimate fields with an online data assimilation methodology"? (Note also that "paleoclimate reconstruction" automatically implies proxy data are used.)

Abstract: pp 1, Line 5: LIMs have been shown to have comparable skill to CGCMs in what sense? This statement currently seems to vague. pp 1, Line 15-17: The last sentence may need to be revised or made more specific, to address the meaning of the "dynamical evolution" to which the authors attribute improvements in skill. When I think of "dynamics," I think of the description of the underlying physical mechanisms driving changes in time, where the term is used in contrast to a "statistical" description. The LIM is purely statistical though, so I think the authors mean the term in the sense of using the model forecast as a prior for each subsequent timestep.

Introduction: pp 1, Line 23: The approach to CFRs described constitutes a whole class of techniques, so change "This technique provides. . .. But also has inherent limitations" to "these techniques provide. . .. But also have. . ." pp 2, lines 1-3: It seems the physical consistency issue due to use of EOFs is also a limitation of the method presented in this paper though, right? Seems a bit disingenuous to list this here as if it's a limitation the present approach will address. pp 3, line 2: The authors might expand upon what they mean by "dynamical" at the first use of the word here, to make the precise nature of their contribution more immediately accessible to a wider audience. After reading the

paper thoroughly, I see they mean simply the use of a forecast from the LIM from the previously assimilated state as the prior for the next state. However on my first read, I thought they were claiming the use of physics-based information in the forecasts. line 5: Even a very quick (1-2 sentence) overview of the gist of the "offline method of Hakim et al. (2016)" would be useful here for the reader unfamiliar with this previous paper. Briefly summarize the difference between the online and offline approaches to be compared. pp 4, lines 16-19: Doesn't the implementation of the LIM in an EOF basis make this methodology subject to the same limitations as regression-based CFRs as described on pp 2, lines 1-4? Line 26: How many modes are retained in this study? (This detail is sufficiently important to be moved from the appendix to the main paper). What's the justification for the choice based on e-folding times of a year or greater? Are results sensitive to number of retained modes?

Data and experimental configuration: pp 5, line 24: "For the prior, we used . . .. the CCSM4 last-millennium simulation": Do the authors mean this is the model used for the climatological prior used for the blending used to prevent the collapse of the ensemble as described at the end of the previous section? If so: I would expect the EOFs of the prior and the CCSM4-based LIM to be the same, but different for the other CGCM-based LIMs, thereby perhaps giving the CCSM4-based LIM an advantage, or at least somehow controlling the divergence of that ensemble differently than for the three other CGCM-based LIMs? Line 25-26: The linear observation models for proxy data" should be described in enough detail to enable reproducibility. Are the proxy data simply linear in temperature of the gridcell containing each proxy location? Or a collection of gridcells representing the regional signal of each record referred to in the next sentence? Also, is there any particular justification for the choice of the GISTEMP product for calibrating the proxy models? Finally, it's interesting the authors use several proxy types with known differences in their spectral signatures. Is there any difference in the construction of linear observation models for the lower versus higher frequency proxies? Even if not in this work, the authors might acknowledge this issue and note it as an area to expand on in future work. pp 6, line 4: Optimal in what sense? What is the criterion used to determine the optimum? Line 18 and forward: Consider using the term “validation” rather than “verification”, here and throughout the paper (and in future climate reconstruction papers!) “Verification” is rooted in the latin word for truth. Of course, there is no ground truth to compare against in paleoclimate reconstruction, and estimates can only be shown to be valid in light of the known data and uncertainties, rather than true! Line 26, eqn 10: Provide a sentence or two of insight into how to interpret the CRPS statistic for readers unfamiliar with it. For example you can describe here where you introduce it that lower CRPS is better, or describe limiting cases of its value.

Results and Discussion

Pp 7, line 6: Verification → validation Line 10: provide interpretation of the steep drop in skill as the parameter a goes to one. Line 15: be more specific than writing the CRPS and CE results are “generally consistent.” Do you mean the rank of models is the same as measured by both statistics? Line 18-20: Similar to preceding comment: it’s imprecise to say there are “slight differences in results... when comparing CE and CRPS.” These are two different metrics that measure different things in the first place. I wonder again if the authors mean to make a statement comparing the rank of models as measured by the two different metrics? In all figures, the authors show central estimates across ensembles, but no measures of uncertainty. Once it has been established in the results that the skill varies with the blending coefficient, it might be interesting to show some analysis of estimates and uncertainty across ensemble members for fixed “a” (probably at the value that optimizes one metric or another). For example, I’m curious about the spread in trend values across reconstruction ensemble members in figure 3 for fixed values of “a”. I’m also curious how these compare to the uncertainty in estimates of the trend as measured by GISTEMP. Pp 7., line 31- pp. 8, line 3: I would speculate that this underestimation of trend in combination with skillful match of phase and amplitude of GMT variability might be interpreted in terms of the paleoclimate proxies as high-frequency bandpasses of the climate signal, that do not

tend to preserve the low-frequency signal. This is a well-known feature of many dendrochronologies, for example, although there do exist “standardization” methodologies to prepare tree ring time series to preserve the low-frequency signal. It would be interesting to know whether the proxy time series used in this study have been prepared using methods aiming to preserve low frequency climate variability.. Subsection 4.1 is missing figures and reporting of the estimated GMT time series compared to the target GMT time series. Pp. 9, line 21– seems odd not to show some spatial measures of skill against the target, rather than just against the offline case. Line 23-25: The authors should change “All LIM-forecasting cases show improvements to CE, most notably in the same North Atlantic to Barents Sea area” to “All LIM-forecasting cases show improvements to CE in the same North Atlantic to Barents Sea area.” As written currently, this seems to falsely state that CE improves everywhere compared to the offline case for all LIM-forecasting reconstructions, rather than just in the region noted. Is there climatic significance to this North Atlantic/Barents Sea area that might explain why the LIM forecast- based reconstructions seem to improve skill there compared to the offline case? Or can the authors speculate as to why this region has low skill in the offline case to begin with to explain the near uniform improvements there under forecasting? Pp 10, line 3-4: “There is a clear distinction between LIMs calibrated on data from the shorter instrumental era, and the millennium-scale climate simulation data”– This is an interesting point. Remind readers explicitly at this point which are which, so that readers can easily reference what you’re talking about in the figures. Also, can you describe the distinction you mean clearly and precisely? Looks to me like the millennial-scale ones have fewer regions of large-amplitude degradation in CE relative to the offline case.

Conclusions: Before stating conclusions in terms of improvements relative to the offline case, authors should state conclusions about the basic ability of the methodology to estimate the target. (Note this will require adding another set of analysis comparing reconstructions to GISTEMP target to the results.)

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2016-129, 2016.