Review of Choblet et al. “A continental reconstruction of hydroclimatic variability in South America during the past 2000 years”

General Remarks

This study represents a first attempt to reconstruct regional climate change in South America with a focus on multi-decadal to centennial timescales. It introduces a new approach to reasonably include proxy data of lower than annual resolution by explicitly assimilating the models and observations at different temporal resolutions. As such it is a very welcome and important contribution to the field, as particularly for spatially explicit reconstructions, a multi-frequency method has hitherto not been applied in a real-world continental scale study. I am pleased to see that that earlier attempts to reconstruct climate fields in South America are now being improved in terms of methods, data handling and spatial coverage.

Naturally, such a reconstruction comes with the substantial challenge to verify its results, as no instrumental data are available to test the performance on multidecadal and lower timescales. The authors have addressed this issue by cross validation with individual records and several sensitivity tests.

While I think from a methodological perspective this study deserves publication (although I am not a DA expert), I am not fully convinced by the quality of the reconstruction product and thus the climatological interpretations.

The reconstruction shows little to no skill in the internal validation in the key (SASM) area. The same is true for the instrumental validation (Section B2). The latter is not surprising as the signal at higher than multidecadal frequencies comes from remote proxies. However, it worries me that even the longer-term fluctuations in the instrumental data Fig. AF B3 are entirely missed by the reconstruction. Unfortunately, I cannot provide a clear recipe on how to improve this situation and I acknowledge the amount of work the authors have invested to assess the robustness of the results. Maybe more clearly flagging this limitation throughout the paper already helps a lot. In the following I provide some further suggestions. I do not expect the authors to do all of this, it should rather be seen as a collection of options. I am sure the authors have a better understanding than myself about which are the most reasonable things to do.

- The reconstruction seems to perform best outside of the range of the speleothem data in southern South America, where other reconstructions already exist. Maybe it is worth checking if the multi timescale approach improves the reconstructions in this area by comparing the high-resolution (trees) vs. all-proxies reconstructions (there are a few low-resolution data from this area) with the multidecadal fluctuations in the instrumental records and the SADA?
  Similar in the NW-corner of the area in Central America, which seems to be the only region with reasonable COE values. I understand this is not the region of interest, but to validate the method it may still be useful?
- The key problem seems to be that the speleothem records disagree quite strongly (as suggested by Figs. 2 and AF19). Are they out of phase due to dating problems or weak climate signals? Or is it a true signal with small scale variations in hydroclimate, which the models fail to reproduce? Given the lack of skillful validation it seems necessary to elaborate more on this. Partly this is done in the manuscript e.g. on page 13, but for example it may be clarified if the mentioned dipole is entirely missing in the model data

or show some more local time series from the recons to identify potential issues. Some use of and reference to the nice video material may also help.
Or: which frequencies are responsible for the low COEs? Fig .5 suggests that the signal is coming from the speleothems at frequencies below 150 years. If the problem is mainly at the higher frequencies, other records may be responsible.

- It is mentioned several times that the reconstruction compares well with local proxy record studies (even in the abstract), but this is not illustrated in the paper. Maybe a more explicit comparison including illustrations would help understanding how the reconstruction incorporated and weighted the low-frequency information from these records? This may also help to identify the most reliable bits of the reconstruction in time and space.
- I am obviously biased here, but I miss the inclusion of the documentary records in this study. South America has the best collection of hydroclimate-related records in the Southern Hemisphere and it is almost a shame not to use this beautiful dataset. They could be included as proxies as in the Neukom et al. reconstructions, but they may also be used for further validation. They are clearly high-resolution datasets so may be of limited use here, but some records include multi-year droughts or pluvials, which may be compared with the results. At least it should be stated why these records are excluded, given that they are even mentioned in the introduction. Some other records from earlier collections are also missing and it is not clear why (see below).

In summary, a revision of the paper may benefit from a more careful treatment of the largely unknown reliability of the reconstruction in the monsoon area, be it by changes in focus and/or wording, a more explicit treatment of the issue as suggested, or otherwise.

Specific / minor suggestions

Line 13-14: I don't understand the part of the sentence starting with "with exceptions in ..". Are the regions an exception to the non-uniform pattern? How? I also did not find this statement to be described in detail in the main text.

L. 14: As stated above it is not really shown how the reconstructions align with local studies. So this should either be included or the statement removed.

L. 16-17: To be precise, Neukom et al. 2014 have used a speleothem record in their field reconstruction. This is picky, but I anyways think the key advancement here is that the lower resolution records can now be adequately treated in the multi-timescale approach in contrast to the very simple way they have been used earlier. So this is what should be emphasized in this sentence in my opinion.

L. 24: I am not sure about the "relatively stable external forcing", considering the volcanic eruptions and their large influence in triggering cold periods. Maybe rephrase.

L. 37: "Widely unknown" is a pretty strong statement. I do not disagree, but at least it should be recognized that there is existing work on this. Currently the statement reads as if there wasn't. Please change or include some reference to acknowledge this, it can also be the regional chapter from the latest IPCC report.

L 43ff: I think it is worth including the key review paper by Prieto and García Herrera here (doi: 10.1016/j.palaeo.2008.07.026)

L 51-52: Could cite Neukom & Gergis 2012 (Southern Hemisphere proxy data review) here (doi: 10.1177/0959683611427335).

L58: Also include Boucher et al. 2011 (doi: 10.5194/cp-7-957-2011) here.

L60ff: I missed the Vuille et al. (2012) reference in the section about speleothems (although I saw it was used later on in the manuscript).

L93ff: This section contains quite some details about the methods and should largely be moved into the methods section. A brief description of PaleoDA and what it does and where it was used is sufficient here.

Section 2.1: As mentioned above there are some records that were used in earlier collections or hydroclimate reconstructions (Boucher et al. 2011, Neukom and Gergis 2011, Neukom et al. 2014) that were excluded, and it seems to me some of them would fulfill the criteria mentioned on page 6.:

Documentary records (https://www.ncei.noaa.gov/access/paleo-search/study/8703)

Lake Sediments: Puyehue (Boës and Fagel (2008) used in Boucher11 & Neukom14), Potrok Aike (Haberzettl et al. (2005) used in Boucher11), El Junco (Conroy et al. (2009) used in Neukom14)

Marine Sediment 106KL East Pacific-Peru (Rein (2007), used in Boucher11 & Neukom14)

L228: Maybe clarify, why a selection of 100 years was used and not the full collection. I understand it is due to computational limitations, but this may be stated here.

L242ff: Move this paragraph describing the history of DA up to the beginning of the methods description.

L254ff: The procedure to combine the low- and high-resolution data is not so easy to understand. An illustration similar to Fig. AF2 may be quite helpful.
Also, as far as I understand, the 10-year blocks are non-moving. Please clarify in line 256.

L297ff (C): There are PSMs for tree rings and coral existing. Please clarify why statistical calibration was preferred here.

L318: For better understanding I would move the statement of what values were used for the SNR up here right after the Smerdon citation.

L358ff: Validation: The COE is based on a validation period, and I do not understand, which periods were selected here. The full period that each withheld dataset covers? Please clarify. This may also be helpful to understand why the values are so low.
It is good to see that the DA improves the values but in fact as long as they remain below zero the reconstruction performs worse than a flat line, which seems to be the case over the key area of interest. In the text, this may be misunderstood.
Is it possible to show the development over time? Potentially this would help identifying periods, where the reconstruction is more reliable, because the proxies show better agreement?

L440: I find this interesting. Is it a correct interpretation that the models with the largest variations tend to also lead to the reconstructions with the largest variability? This may be clarified.
In general with DA it is a clear sign of a noisy reconstruction if the results just show small

variability (at least in my experience). This may also be stated somewhere, for example to explain why the tree-only (and partly No-speleo) and LMR hydroclimate recons are basically a flat line in many instances…

L450: I think it should be AF11 not AF12.