Clim. Past Discuss.,
doi:10.5194/cp-2017-28-RC2, 2017

Climate of the Past Discussions

Interactive comment on "Multi-century cool and warm season rainfall reconstructions for Australia's major climatic regions" *by* Mandy Freund et al.

Anonymous Referee #2

Received and published: 30 May 2017

With appropriate corrections, this paper will be a useful contribution to the literature related to the character and causes of variation in Australian hydroclimatology. Much of what is done is interesting, and the sub annual approach is great to see, but there needs to be some additional attention to detail, particularly related to the rationale and specifics of the research methodology, and a more critical approach to the results presented would be ideal. The paper has the potential to be very good, but I think it has some way to go to get there.

Below, I discuss aspects of the paper sequentially, explicitly highlighting what I consider critical points that I think must be addressed and major points that should be. Trivial points are collected at the end.

Printer-friendly version

Discussion paper

Abstract The abstract reads well, but minor changes will be required if the authors accept some of the criticisms that follow (e.g. the statement [p1, 15] that the rainfall reconstruction aligns well).

1 Introduction

[p2, 15]. State when instrumental data collection started. More generally, make sure that you are not assuming your readers are Australians when it comes to what may seem to be common knowledge.

[p2, 28–33]. This is a useful paragraph, but it would it be useful to extend it slightly with a comment on the relevance of palaeoclimate reconstructions under future conditions of changed boundary conditions.

Major. [p3, 11–17]. It might be useful to rephrase “process-based methodology” to more clearly capture the atmospheric dynamical aspect of what you are doing. Also, you need to explain why this approach will maximise skill and utility. If I recollect correctly, the advocates of the Cook approach of point-based regression would argue that this achieves the same. You need to justify your claim here.

Major. [p3, 11–17]. I was surprised to see the analysis based on NRM regions. The approach is contrary to what seems the more common and sophisticated approach of examining relationships at finer spatial resolution, so I would like to see some rationalisation for the choice here. A key criticism is that the spatial scale is too coarse for some regions to adequately capture the character of spatial hydroclimatological extremes and risks conflating contrasting regions into an unhelpful whole. My concerns here returned when I encountered Figure 4, where it is clear from the instrumental data that the regionalisation approach has some undesirable consequences. For the millennium drought, the bipolar R region pattern cancels out; for the WW2 drought, widespread drought in the west is lost; for the Federation drought, the centres of drought are displaced east. I do appreciate that you are not in a position to revise the analysis, but think you should give a more convincing rationale for the approach taken, and follow up with a paragraph in the discussion to discuss the implications and outline if you think an alternative approach would be useful (or not).

[p3, 23]. The NRM regions cannot be clearly distinguished on Figure 1. Figure 4 is much better.

2.1 Instrumental data

[p4, 8–10]. Some expansion of the description of the AWAP data would be good. For example, it would be useful to state what homogeneity analysis has been undertaken (by BoM).

Major. [p4, 11–18; Table 1]. Insufficient information is provided on the climate drivers. For example, the metrics for the intensity and position of the subtropical ridge [over Australia] are not common knowledge, ditto blocking, and there are multiple indices for the SOI. All of this can be simply solved by adding an appropriate descriptor to Table 1. SAM appears to be missing from the table. The IPO is mentioned later but not used in the analysis and there is no equivalent west-pole Southern Oscillation index (you have one for SST, I presume that is what NWP is). Perhaps a little more rationalisation would be appropriate.

2.2 Palaeoclimate data

Critical. [Section 2.2]. Overall, Section 2 seems too superficial. The reader needs a better understanding of this fundamental data in order to interpret the subsequent results. See following for specific details.

Major. [p4, 20; Figure 1]. A cross-reference to details in the supplement is needed here. Also, the mapping is not up to the task of showing the spatial distribution (need zoomed in insets for high density areas) – e.g. I can only see one of the five speleothem proxy locations. It would also be useful to colour-code the symbols to show the spatial degradation back in time. Also, is it possible to distinguish those proxies actually used? Table S1 indicates numerous proxies that were not used for any region (all zeros).

Critical. [Missing details – proxy data pre-processing]. It is common practice to pre-process proxy data in ways that unavoidably affect the frequency response of any climate reconstruction. It appears (and you should state) that you do not re-process the data to ensure consistency, but it is essential that you comment on what has been done by the original workers (or subsequently). Without this information, your readers may incorrectly assume that Australian hydroclimatology is characterised by essentially no centennial-scale variability, when in fact the case is that it has been removed. Although a critical omission, the solution is very simple – you just need to state what frequency information is credible in the reconstruction. A related paragraph in the discussion would also be appropriate.

Critical. [Missing details – proxy dating fidelity]. Similar to the above, you are assuming that the dating of the proxies is accurate. That is fine, but a comment to the effect that dating is not revisited here may be appropriate. However, Table S1 indicates that you have used a number of non-annual proxies, yet I see no comment on how these are meaningfully included in an annual-resolution reconstruction. The rationale, the explicit methodology (interpolation?), and the implications should be mentioned.

3.1 Reconstruction

[p5, 3–15]. Good to see this focus on stationarity. Looking at only linear relationships and ignoring lag relationships is simplified but acceptable. But can something more be said about the interquartile range approach? i.e. where exactly does this come from and has it been tested for this purpose? I presume this analysis relates to the binary scores in Table S1 (the table caption does not provide the relevant information).

[p5, 9, and relevant to multiple other places]. Statistical significance is mentioned here for the first time. Why 0.1 (seems a fairly weak test) and how are significance levels adjusted for autocorrelation?

Critical. [p5, 17–24]. This section describes the reconstruction methodology. The credibility of the work rests on this, so the reader needs to thoroughly understand the details of what has been done. There is not sufficient detail for me to be sure I completely follow what has been done. While it is appropriate to lean on other references for comprehensive treatment (but relevant cited important references are missing from the references) the onus is on the authors to present sufficient details here. The Tierney et al level of detail is a useful model in this context.

[p5, 24]. How spliced?

[p5, 27–28]. 52, 33 years. At face value 1934–1984 & 1900–1933 gives 51, 34. Missing something?

[p5, 30]. "...not entirely independent..." could be interpreted as mostly independent, which is incorrect.

3.2 Analysis

[p6, 7–9]. Rationale for this analysis? I don't know what normalized trends means in this context.

[p6, 14–16]. Detail redundant here (provided in Table 1).

[p6, 18]. Deciles need a time interval (e.g. 36 months).

[p6, 19–21]. Can this be rephrased for clarity?

4.1 Regional climate driver influences

[p7, 13]. ENSO "stands out" only in the warm season. The cool season map is mostly red, but this is misleading when the more nuanced bar graph results are considered. See later comments on Figure 2.

[p7, 14]. 44% < "most", so presumably you mean something else.

[p7, 24]. SSWF has the only warm season yellow (IOD) bar.

4.2.1 Reconstruction skill

[p8, 3, 5]. Figure 3 panel labels are given in the text, but are not shown on the figure.

Major. [p8, 5–8]. Some clarification of this earliest year comment is required. First, although it doesn't say so in the text, the Figure 3 caption indicates that statistics relate to calibration rather than verification statistics. Wouldn't the latter be better? Second, why is half the maximum calibration variance explained an appropriate metric here, rather than a fixed R2 threshold? Third, given that you can only assess skill based on comparing with observations, I assume that the early dates relate to how reduced data sets (corresponding to nests) perform against the instrumental data. If this is incorrect then some additional explanation is required. Whether correct or not, have you taken into account degraded proxy performance with time outside of the calibration/verification period? Loss of sample depth, and thus signal, is characteristic of the tree ring data, so there is more to reduced performance than simply the number of proxies. Probably nothing much you can do about this, except to note that the early dates will be inflated (too early), but to an unknown degree.

4.2.2 Reconstruction time-series

[p8, 22]. Probably best to delete "…and past centuries (Fig. 5)", because all comments in this paragraph relate to Figure 4.

[p8, 29]. Define "low-frequency". I am struck by the lack of it.

[p8, 26–27]. Perhaps I am missing something here, but doesn't your rescaling methodology force this? If so, then this is not a relevant comment.

Major. [p9, 15–27]. This is an interesting approach, but I am unconvinced by the interpretation. Because 30 years is a fairly short window, I suspect that analysis of serially-correlated random numbers may give similar results to what you see here. If so, then the patterns identified cannot realistically be interpreted in the manner done, although the conclusion would be the same. I am not convinced that it amounts to "…an additional verification measure".

4.3.1 Contextualising recent rainfall trends

[p10, 7–22]. Apart from the apples vs. oranges caveat (see discussion of Figure 6), this seems OK, but it does beg the question why 30/50 year trends are a key metric, rather than, say, 30/50 year means and variance. See previous comment about providing the rationale for this aspect of the methodology.

4.3.2 Contextualising the spatial extent and intensity of past droughts

[p10, 31–32]. Surely two droughts are not enough to make such a relatively bold statement, especially since the reconstruction gets the significance of the two droughts around the wrong way (gridded AWAP shows WW2 drought is more significant, but reconstruction indicates the Federation drought).

Major. [p11, 11–19]. Figure 7 is nice, but here are confusing elements to the results that require explanation. Recon (1900-214) shows central region (R) below average for both the WW2 and Federation droughts. Recon (1600-2014) has WW2 average and Federation very much below average. While I appreciate that deciles are a moving target, drilling down into the results is needed to make sense of what is going on. At face value, Recon (1600-2014) lacks credibility, because it relegates arguably the most significant drought of the instrumental record, based on the instrumental data, to relative insignificance!

4.3.4 Extreme years in a long-term context

[p12, 27–28]. This is pushing the envelope, but I am not convinced that you have actually shown that the reconstruction is actually up to this rather demanding task. I would need to see that the instrumental extreme years are captured in roughly the appropriate order.

[p13, 2–16]. Some of this material may be better in the discussion.

[p13, 18–28]. Results in this section have to be taken at face value because the tabled presentation is not well suited to “seeing” the claimed patterns.

4.6 Comparing our reconstruction with previously published

[p13, 33–34]. Please be explicit about the degree of overlap (%).

[p14, 4]. Is linear correlation against the PDSI appropriate? I don't recall if the PDSI scales linearly and it is also a water balance approach, so has significant memory. My point here is that you might be short changing yourselves by an overly simplistic inter-comparison.

[p14, 8–10]. I don't understand where you are going with this the last sentence. It reads like a criticism of the PDSI, but I suspect that is not your intention. The temperature dependence targets evaporation, making PDSI arguably a superior drought index. And the spatially unresolved parts presumably relates to the point-based approach, which is also arguably superior (you certainly have not convinced me otherwise).

Major. [p14, 10]. The poor warm season agreement with the PDSI analysis, except for one region, is quite alarming, especially the near-zero relationships in regions containing the cities where most Australians live. Given that this affects the perceived credibility of Australian drought reconstruction, it might be appropriate to follow up on this here, or in the discussion.

[p14, 14–17]. The cool season SE results are encouraging (water resources implications), but not so the dry season. Coupled with the poor agreement with the drought atlas, and the unconvincing relationship with the coastal records, I'm left doubting the credibility of the reconstruction.

5 Discussion and conclusion

[p15, 6]. "Eastern Australian" is too broad a phrase – agreement is much more spatially restricted. Personally, I think "high-level" is overselling things. Given that you are reconstructing the same thing (drought) from significantly similar data sets, I was expecting to see most variance in common, and you are well shy of that.

[p15, 7–8]. I suggest you limit the "compared well" comment to the cool season.

[p15, 11]. Interesting comment about highlighting the quality, because to me they highlighted the limitations.

[p15, 14]. This is a reasonable statement. But not picked up is some notable evolution in patterns for some regions. For example, MN & R in Figure 6 appear to have increased variability in the late 20th c. Is this real, or a splicing artefact?

[p15, 14–]. I remain unconvinced by this regression slope analysis approach. It can tell you about the rate of change and its significance, but is that really the important metric in terms of the process explanations you then mention? It also misses important cumulative impacts. For example, the SS and SSWF results show a cumulative decline to a mean substantially lower and with the most extreme droughts all relatively recent. MN and R show the reverse. A different type of analysis would be required to assess the significance of these changes.

Major. [p16, 2–3]. Comparison of instrumental vs. reconstructed trends can only reasonably be made with relevant caveats associated with the pre-processing of the palaeo data. Pre-processing has likely reduced supressed multi-decadal trends, so your histograms in Figure 6 will be pulled in at the tails, which clearly will affect your assessment of how the instrumental data trends (which have not been similarly treated) compare. Note though that recognising this actually reinforces your conclusion about recent trends being within the range of natural variability.

Major. [p16, 9–10]. The discussion in this paragraph follows on and emphatically restates earlier comments about the quality of the reconstruction of historical droughts that I think can reasonably be challenged (see [p11, 11–19], above). First, surely you only have two droughts. The millennium drought is outside your proxy data period, so it is essentially spliced instrumental data, is it not? If so, then agreement of spatially-averaged instrumental data with the original gridded data is meaningless, although it does point to issues with spatial units that are too large (a paragraph discussing this spatial scaling issue would be appropriate). For the other two droughts, you can only really claim good agreement for the Federation drought. As previously stated, I think the WW2 drought reconstruction is severely awry, and suggests to me that the methodology may only be suitable for capturing some types of drought (perhaps some additional forcings are not captured by the proxy network). The credibility of the reconstruction is challenged by the poor representation of what the instrumental data shows to be the most extreme drought in the instrumental period (Figure 7, left column).

[p16, 27–33]. This is an interesting point. Can you relate it back to the drivers?

Major. [p17, 3–4]. I don't disagree with this, but it does presuppose that teleconnection patterns will remain stable in a future warming world. The flip side of this is that the reconstructions extend back into a globally cooler period. If teleconnection were different then (and there is evidence to suggest they were for some of your proxies), what then are the implications for your reconstruction (because the transfer function will be wrong)? Moreover, drought is not just rainfall. Australian researchers have shown that droughts have intensified in response to increasing T (and thus evaporation), have they not? So a rainfall-only analysis is only part of the story. Surely worth some serious commentary.

Table 1. SAM is missing. Additional details of indices would be useful (e.g. I assume NCT and MWP are SST based). A sentence or two describing each index would be useful. Surprising you have not included a west pole pressure index.

Table 2. The caption could usefully be reworded for clarity. Is the information for "Instru" the reconstructed data for the instrumental period, the same but with instrumental data spliced on the end, or the instrumental data? This seems a rather ineffectual way of presenting the information – visualisation would highlight temporal patterns, temporal clustering, and inter-regional patterns in a way that tabled numbers do not.

Figure 1. See previous comment about inability to resolve the proxies and the regions on this map. Also, given that many proxies were eliminated, would it not be more useful

to limit the map to those proxies that actually end up being used. Moreover, it would be interesting to see this broken down by region in the supplement. Without a laborious process of extracting the relevant information from Table S1 and remapping, this useful information about contributing proxies is unavailable to the reader.

Figure 2. It appears that the SOI is generally superior or comparable to the other three ENSO indices. That point could be made in the text and this figure simplified. There is a wealth of information in the bar charts, but the maps are unsophisticated in the treatment of this, and I think counterproductive in oversimplifying matters. I don't recall comment in the text related to the logic of pooling SAM and BLK. It would be useful to include the region codes along with their long names on the bar charts (also on the maps). Consider adding a horizontal line separating the two parts of the figure. Because you don't have axis labels, you need to explicitly state in the caption that the bar graphs show correlations. See previous comment about uncertainty about whether autocorrelation has been allowed for in the significance levels cited (see [5, 9] above).

Figure 3. It would be useful to have the years corresponding to the plotted statistics shown.

Figure 4. There are several instances where the reconstruction is outside the ensemble range. Having gone to the trouble of calculating the ensembles (a good thing), why isn't a mean/median (or other measure of central tendency) used for the reconstruction? Doing so would "fix" some of the points of difference with the instrumental data (e.g. in MB, MN, WT). It would introduce other issues, but the net benefit may be positive, and a transfer function based on the full data rage may be more robust. Just a thought.

Figure 7. I presume that the millennium drought is "missing" for Recon (1900–2014), because you only have instrumental data. If that is true, I don't get why it appears in Recon (1600–2014) – its not reconstructed, its spliced instrumental data isn't it? Need to specify the time periods for decile calculation (12/24/36 months?).

Figure 8. Please expand the caption to better explain exactly what is being shown here.

How is the starting point for each drought determined?

Minor points

[p2, 8]. Delete “a” at the end of the line.

[p2, 10]. This style of referencing with a list of references at the end of a paragraph is unfortunate. They presumably don’t all relate to the last point, and if they do then there are missing references in the body of the paragraph.

[p3, 27]. Do you mean “Compile”?

[p10, 32]. New paragraph (millennium drought)?

[p11, 11]. [somewhat] similar?

[p11, 33]. provide[s] insight.

[p12, 3]. Suggest you change “many” to “several”? 4–5/8 and only cool season.

[p12, 14]. Seems to [be] a result.

[p12, 24]. Breaking up paragraph into smaller ones would help readability.

[p12, 26]. Expand “Black Thursday” for benefit of non-Australian readers.

[p14, 5]. Reconstruction[s]?

[Table 2, 4]. referred [to] as.

[Figure 8, 3]. Delete “a” (3rd last word).

[Figure 9]. Fenby et al should be Fenby and Gergis.

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2017-28, 2017.