Clim. Past Discuss., 9, C1338–C1345, 2013
www.clim-past-discuss.net/9/C1338/2013/

Climate of the Past
Discussions
Open Access

Interactive comment on "Inferred changes in El Niño-Southern Oscillation variance over the past six centuries" *by* S. McGregor et al.

Anonymous Referee #1

Received and published: 2 July 2013

General Comments

This study principally investigates whether existing proxy reconstructions of ENSO are consistent with one another in terms of their inter-decadal changes in variance. The motivation for this approach stems from the well-founded claim that most studies have focussed on time-domain changes, but if the primary interest is in overall variability of ENSO, then robustly determining the variance of some relevant metric may be sufficient. Furthermore, such an approach may plausibly be less sensitive to age model errors than the time-domain reconstructions. The substance of the study involves two principle methodological investigations; firstly, into the merits of different approaches to extracting the common changes in variance from multiple records, or reconstructions, using GCM-derived pseudo-proxies. Secondly, the uncertainties associated with the use of single-site and/or composited records to reconstruct changes in variance. Following these, comparisons and syntheses are then made between existing reconstructions, in both a single site and multi-site sense.

I found the topic of the paper to be interesting and certainly well within the scope of the journal. The MS is generally clearly written on a paragraph level and many of the arguments are persuasive. I found the pseudo-proxy analysis and age-model error components to be thorough and interesting. However, there were several aspects of the paper that I found to be unclear and there is a clear need for greater sensitivity testing of what is a quite parameter heavy data processing method. On balance of these factors, I would be happy to recommend publication, if these comments can be met. On a more general level I also found the MS, especially in its use of figures, somewhat unwieldy to read and I would suggest (although not insist) that the eventual impact of the work may be much improved by some substantial restructuring of both the text and the figures. I have included a few final structural suggestions to this affect at the end of the review.

Specific Comments

1) Method sensitivity issues: There are various points in the method where certain parameters are chosen, without particularly clear justification and without statement of the effect that these have on the results. The clearest example of this is the choice of the 30year sliding window for the running variance calculations Page 2933 Line 25. I appreciate the need for this value to be at least several multiples of the characteristic ENSO cycle duration, but why 30years rather than 25years, or for that matter 50years? Some sensitivity tests on this should be easy to carry out and the results could be stated in the text and/or an appendix. It would be interesting to see whether this choice affects the final conclusion of the study, as stated in the abstract, that the common variance of the reconstructions over the period 1600-1900AD is 'considerably lower' than during the last 30year window (1979-2009AD). A second, similar, case is the choice of the 10year highpass filter prior to the calculation of running variances on the

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

model and reconstruction data. As before, what difference would other choices make here? Again, I cannot see any obvious rationale for a 10year choice, so some form of sensitivity analysis should probably be undertaken. Whilst I can accept that this is unlikely to have much impact on the model data, in my own experience of working with proxy data, such choices can be surprisingly important. Also, from a method replication point of view it would be good to know what kind of filter design is used (i.e. what exactly is meant by the cut-off period?). My final example (and of least importance) is the choice of the period 1900-1977AD to perform the reconstruction to NINO3.4 'normalization'. It is clear from Figure 1 that at least one of the records ('proxy 5') may behave differently were a sub-set of this interval to be have been used.

2) Choice and use of climate model simulations: The study uses two specific model simulations. One is a forced last millennium experiment with CCSM4 and the other a pre-industrial control simulation with the GFDL-CM2.1 model. I did not find it clear from the text why this combination of models was selected. If the authors seek to separate the effect of the climate forcings included in the former, then surely the equivalent control experiment should be from the same GCM? Alternatively, if the authors wish to compare behaviour across GCMs to establish robustness of results to model inadequacies in ENSO realisation, then surely the same experiment should be used? In this latter case, many more last millennium and picontrol experiments than the two used here are now available on the CMIP3/CMIP5 archives. At the very minimum, I think more clarity is needed on why these models were selected.

3) Interpretation of Figure 5 It is not cleat to me that Figure 5 supports the claim in the text (Page 2937 Line 2 onwards), which seems to be that the 'no dating errors' plots show close relationships and those with errors less so? Rather it appears that a and b have reasonable (although still with a lot of scatter) unitary relationships, whereas c and d clearly lie off the unitary line?

4) Statistical methods I am not qualified to give a detailed critique of the methods the authors employ here. However, I found several things to be unclear and thought I would

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

raise these as points of discussion. Whether these should be viewed as requests for alterations in any revised MS should depend on how the authors/other reviewers view them.

4A) When comparing the correlation coefficient 'r' values of the model running variance data with those of the original Ts (or precip) data (Section 3.1), the authors make statements of what constitutes significant levels (e.g. r2 >0.1 is given at Page 2935 Line 3) of the running variance r2 values. I can see how these significance levels are derived for the 'raw' data, but cannot easily see how the same levels can apply to the running variance data. The latter must have much reduced degrees of freedom (presumably a de-correlation time somewhat equivalent to the sliding window length) and also will not be normally distributed (if the raw data was so distributed, then the running sample variance will be χ2 distributed and it is very unlikely the error terms in a linear regression through the variance data would then be normally distributed). I appreciate the authors can still calculate an r value in this case, but it is not clear (at least to me) what these values mean in the context of either strength or significance. That said, the form of Figure 3 demonstrates the qualitative relationship they seek to establish between the raw data and running variance r2 values perfectly well. 4B) The authors conclude that the use of precip data (at least from single sites) leads to lower matches between the r2 values of the raw and running variance data than for the Ts data. This seems unsurprising as the NINO3.4 Ts time-series is drawn from roughly normal Ts data (so that the sample variance of that will be roughly χ2 distributed), whereas the raw precip data is typically more closely Γ distributed and so its running variance will also follow a generalized Γ distribution and is therefore a-priori unlikely to be well correlated to that of NINO3.4. Many precipitation sensitive proxies (e.g. North American tree rings) are calibrated to a normalized precipitation index (such as the PDSI, SPI, or more trivially relative changes in local precipitation, all of which transform the Γ variate to something near-normal) for broadly these reasons and I suspect if the authors were to try their analysis on such a transformed field, it may somewhat change the outcome. Alternatively, if the authors are aware of existing proxy studies using calibrations to precip itself, rather than such a transform, then that would help justify the utility of the comparison they present. 4C) I do not follow how the authors adjust the overall variances of the reconstructed records (section 4) to that of NINO3.4 by adding a constant variance term. What would the physical meaning of such an additional noise term be and wouldn't such an approach alter the relative changes in variance through time that are present in a given reconstruction? It would seem to me that were the original proxy to NINO3.4 relationship to be some form of simple linear regression, then this approach entails adding the 'extra variance' to the noise term, which neglects the fact that the slope of those relationships may not be unitary. Would it not make more sense to multiply the reconstructions by a given factor to make this normalization correction? I am open to persuasion, but would certainly welcome more information on this, perhaps a slight expansion of the mathematical basis in Appendix A to cover this.

5) Choice of single site proxy data records I was curious why the authors opted for the requirement of coral records to continuously span the period 1800-1980 AD in order to be considered for the single record data-sets. The 'calibration' interval used for the reconstruction variance corrections is 1900-1977 AD and the period plotted on Figs 8/9 is 1400-2000AD and therefore the new interval of 1800-1980AD seems somewhat arbitrary. This choice has the effect of excluding the Palmyra (Cobb et al., 2003) corals from the single site coral exercise and were this to not be the case, this would change the current result of only having SW Pacific corals in that composite?

Technical/Minor Comments:

Page 2931 Line 21: Whilst Fig 1 does indeed show uncertainty, I think it would be fair to state here that there is also some commonality (for whatever reason, possibly non-independence of the reconstructions) between at least some of the plots on Fig 1. A little more description of where the records do and do not agree would be welcome.

Page 2933 Line 6-10: The terms 'robust ENSO' and 'quite realistic' are imprecise and either need to be better specified, or re-phrased. It is, to my mind, not a well resolved question that any given GCM well replicates all aspects of the real world ENSO phenomenon. In particular, I do not see how we can be confident that the multidecadal fluctuations in amplitude are representative of reality, given the small sample available in the latter? I would be more comfortable with saying that they were 'consistent' with the observations, rather than realistic.

Page 2933 Line 26: please clarify what is meant by the 'correlation between the two maps'. I take this to mean a correlation between the spatial r2 values derived from the raw data and running variance calculations, but a little more guidance in the text would be helpful. I appreciate it is hard to assess the significance of these 'spatial' r2 values, but I did find the language of 'reasonably well' at Page 2934 Line 5 to be unsatisfactorily vague.

Page 2935 Line 10: The authors correctly note the multiple climatic controls on proxy systems as a limitation to their approach. I think another necessary caveat here is the assumption that the spatial patterns of ENSO behaviour remain stationary through time, either in the real or model climates considered. I note that this point is indeed given as a limitation in the conclusions section and this makes me wonder if these two 'lists of caveats' chould be merged somehow.

Page 2937 Line 6: 'roughly equivalent' is imprecise and does not necessarily seem justified to me, given the amount of scatter seen in all the panels on Figure 5.

Page 2937 Line 16: I was unclear whether the McGregor 2010 reference proposes both of these methods (MRV and RVM), or whether it proposes MRV as a better alternative to an existing method RVM. In either case, it seems to me that the RVM method is somewhat of a straw-man, in that one would not have expected such an approach (running variance of the median of multiple signals) to have performed well in the age-model uncertainty cases they consider. This does not make the current analysis or conclusions uninteresting and it is good to see that the intuitive result emerges. However, in the conclusions, the authors state that RVM is widely used in the literature, so perhaps they could give other examples of such use? Also, are there still other methods for calculating the common variance that could be at least mentioned here as further alternatives? What about calculating the principle components of the available data and then looking at their variance?

Page 2939 Line 20: I found it (very) confusing as to what exactly is meant here in the claim that the most recent 30year interval is 'significantly higher than the median error bars' in the context of Figure 7. I think (based on the abstract, where it is clearer) what may be intended is that the recent value lies outside the black lines for the whole period 1600-1900AD, but this is not the same as the statement here of 'during the past 400 years' as that implies either the period 1600-2000AD or maybe 1579-2079AD. This conclusion (and when it is reiterated in the conclusions section) must be reworded and also preferably made clearer how it relates to Figure 7.

Page 2940 Line 8: I think the word 'apparent' needs to be removed, as these reconstructions are indeed not independent in any strict sense. If it were really the case that the method specific processing could render them so, even in the absence of shared input data, that would surely be a very worrying conclusion?!

Page 2942 Line 11: Throughout the paragraph starting 'Calculating the . . .', I think the figure references are intended to be to Figure 9, not Figure 8?

Fig 1: the caption should be more explicit as to what both the terms 'normalizing' (adding a constant variance to match that of the period 1900-1977AD, see comment above) and 'ENSO' (NINO3.4 Ts, but that does not have to be the case) mean here.

Fig 1/8: Flipping between the figures and tables to check the legend for the different reconstructions is frustrating. Can a truncated form of the reconstruction name not be included on the legends for these figures?

Fig 3: The caption should re-define what is meant by the 'running variance of ENSO' (the 30year running variance of the HPF NINO3.4 Ts time-series).

CPD

9, C1338–C1345, 2013

Interactive Comment

Fig 8 caption: there is a misspelling of the word ‘recent’.

A few structural suggestions:

The authors go to some lengths to consider whether existing multi-proxy reconstructions of ENSO can be considered independent of one another (Section 4.1). Whilst their arguments seem largely persuasive, I might suggest that a more obvious place to ’start’ is to return to the single site records first? i.e. move the single station analysis (Section 5) in front of the application to existing (overlapping) reconstructions (Section 4)? However, given that in many cases the two sections yield similar results this is a largely presentational choice.

Table 2: Whilst the rigour is laudable, I wonder if it is necessary to give all of these results explicitly, at least in the main body of the MS.

Fig 2: I think I see what this figure intends to achieve, but it is incredibly opaque to initial inspection. I wonder if there might be any alternative forms for this information?

Interactive comment on Clim. Past Discuss., 9, 2929, 2013.