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

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 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.

**We thank the reviewer for their effort evaluating our manuscript and the constructive comments. We have made substantial revisions to manuscript to address these comments, please see the revised manuscript and the individual responses below.**

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.
    a. 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

Calculating the median running variance of the 14 high-pass filtered ENSO reconstructions, where the running variance window length ranges from 16-yrs through to 60-yrs, yields a very similar result. For instance, in most cases the median variance in the most recent window is clearly higher than the median variance for any window during the interval 1600-1900 (see figure 1 below). As suggested we have also stated the results of this sensitivity test in the text in the revised manuscript (see lines 17-20 of page 13 of the revised manuscript).
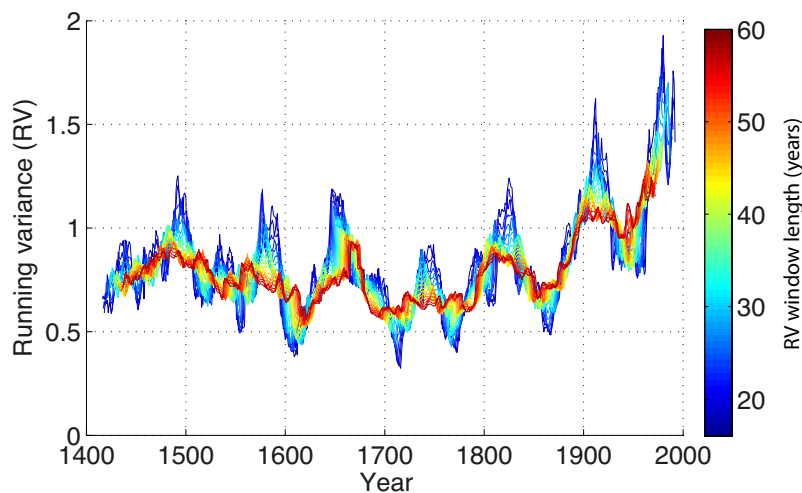


**Figure 1: Each line represents the median running variance of the 14 high-pass filtered ENSO reconstructions, where the lines color represents the length of the window the running variance was calculated in (see colorbar).**

The 10-yr high pass Butterworth filter was used to isolate the variability in the classical ENSO band of 2-8 years. In regards to use of the term 'cut-off period', this refers to the period where the variability is effectively dampened. The Butterworth filter response functions have been described in the literature in detail since they were first described in 1930 (Butterworth, S., 1930: On the theory of Filter Amplifiers, *Experimental Wireless and the Wireless Engineer, pp 536-541.*). So in this case variability with periods longer than 10yrs is removed from the time series. We have now clarified the text around the use of the filter (see lines 19-22 of page 3, 20-22 of page 9, 28-30 of page 9 and lines 31-32 of page 11 of the revised manuscript).

As suggested we have carried out sensitivity tests in regards to the choice of cut-off period, testing cut-off frequencies between 8 and 50-yrs. We find that the resulting median running variance time series display only very minor differences, suggesting that the results are insensitive to the choice of cut-off frequency. In fact, the median running variance of the high pass filtered data compares very well with the median running variance of the raw (unfiltered) data, indicating that the filtering process has only a reasonably small affect. However, we choose to retain the results utilizing the high pass filtered reconstructions to ensure the manuscript remains focused on the interannual variability of ENSO. As suggested we have also stated the results of this sensitivity test in the text in the revised manuscript (see lines 17-20 of page 13 of the revised manuscript).
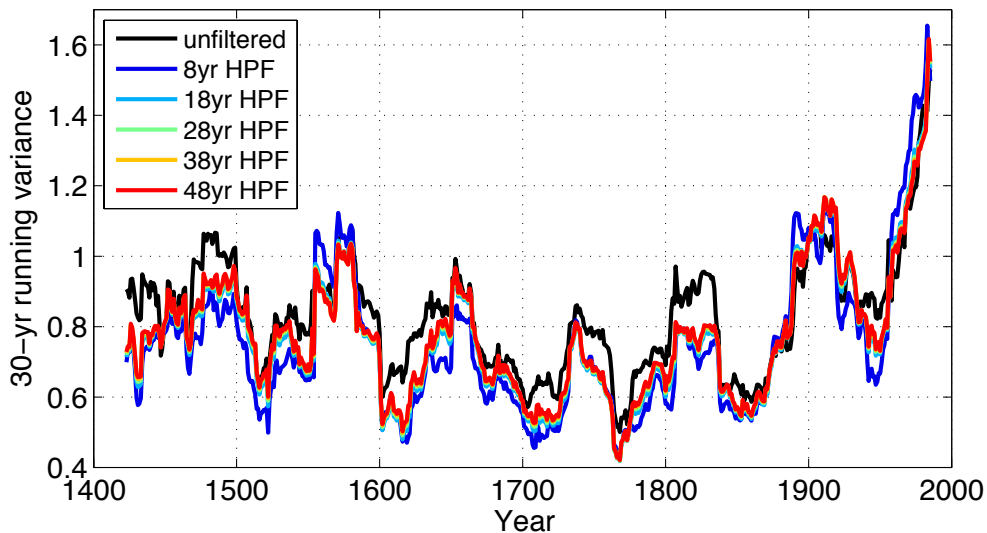


Figure 2: **Each line represents the median running variance of the 14 high-pass filtered ENSO reconstructions. Here, with the exception of the black line with is the median running variance of the unfiltered data, the line colors represent the cut-off frequencies of the high pass filter used on the data (see legend).**

c. My final example (and of least importance) is the choice of the period 1900-1977AD to perform the reconstruction to NINO3.4

This end date of 1977 was selected as it is the most recent year that all 14 ENSO reconstructions have data. The start date of 1900 was basically arbitrarily selected as all observational data sets and ENSO reconstructions have data. This ensures that all reconstructions and observations are being normalized over the same period. This start date could possibly be extended back to 1886, the first year that all observational data sets and proxies have data. However, checking the impact of this change in the normalization start dates reveals only very minor differences in the median running variance signal (Figure 3). As such we leave the normalization in the manuscript as is.
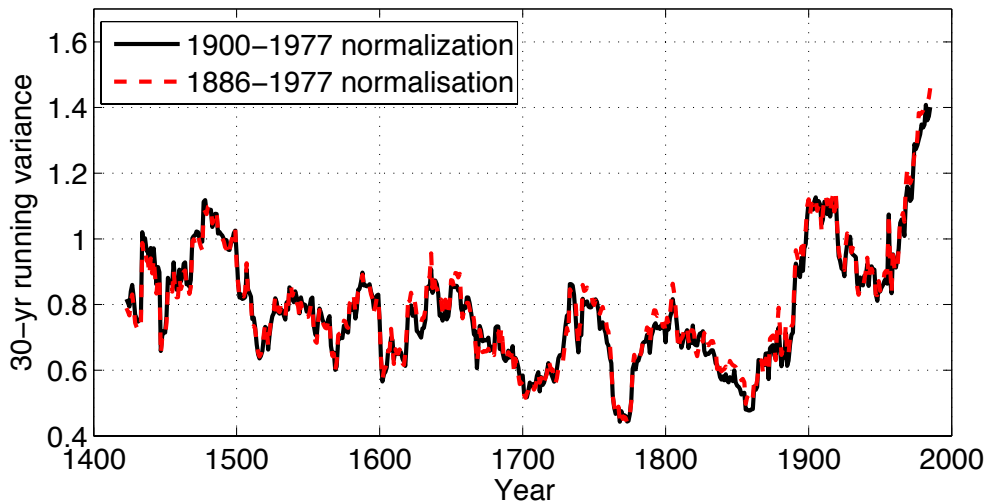


**Figure 3: the median running variance of the 14 high-pass filtered ENSO reconstructions normalized over the 1900-1977 (black) and 1886-1977(dashed red) periods.**

The reason an 1850 control simulation of CCSM4 was not used was because it was not available at the time of starting our analysis. However, the Landrum et

al. (2013) paper, which describes the CCSM4 last millennium simulation, concludes that the model's various modes of internal variability [i.e., the North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), and El Niño–Southern Oscillation (ENSO)] show little or no change to the imposed forcing when compared to a long 1850 control simulation with the same model. Thus, we expect the same results would have been found if the long control run of CCSM4 was utilized in this study. We have now added some text to this effect to detail the model experiment differences and what we expect to get from utilizing both CGCM simulations (See lines 23-30 of page 3 of the revised manuscript).

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?

We now discuss the behavior shown in these plots in more detail, including the large scatter, and we have also altered Figure 5 to better show the effect of dating errors. Please see lines 24-29 of page 7 of the revised manuscript.

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 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.
    a. 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 $\chi 2$ 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.

Point taken, we have now removed reference to the significance and simply discuss the correlation values. See revised manuscript on lines 26-29 of page 5 and lines 4-8 of page 6.

On the issue of precipitation signals in proxies: Yes it is hard to find direct precipitation reconstructions in absolute values. Here we wanted to demonstrate that in cases where ENSO signals information is carried to proxy sites via precipitation anomalies, the running variance signal may be less robust. How the proxy's response function transforms the precipitation signal into the proxy variable is a different question, which we cannot address here.

However, while many coral proxies are not explicitly calibrated to precipitation (or anything else for that matter), coral d18O anomalies are discussed in terms of the overlying anomalies of precipitation and SST. For example, the coral proxy of Bagnato et al. (2005) suggests that when the South Pacific Convergence Zone (SPCZ) shifts meridionally it creates precipitation and SST anomalies at the proxy location, as evidenced by maps of correlation coefficients calculated between an index of SPCZ location and the variables (precipitation and SST). The overlying precipitation and SST anomalies due to a meridional shift in the SPCZ are then implied to lead to coral d18O anomalies of the same sign. Hence leading the authors to conclude that coral d18O at that location is a good recorder of the regions SST and precipitation anomalies as well as the SPCZ location. Similar arguments are also presented in the studies of Charles et al. (2003) and Cobb et al. (2003).

Thus, in the context of the Bagnato et al. (2005) study, if it turns out that the magnitude of the precipitation anomalies is not correlated with the meridional location of the SPCZ, the estimated variance changes of the coral d18O record will be degraded by the inclusion of the precipitation signal and may have little or no relationship with the SPCZs meridional movement.

The addition of this constant to the running variance time series is equivalent to adding the effects of a constant white noise signal to the variance signal. In terms of simple linear regression between the original proxy and ENSO, this constant variance term would act to alter the scatter but not change the slope of the relationship. We believe that this is preferable to altering the slope of the relationship between the proxy and observations and it is consistent with the notion that the proxies record ENSO plus other phenomena not related to, or correlated with, ENSO. We now discuss this further in Appendix A in the revised manuscript (see lines 23-25 of page 16 in the revised manuscript).

The choice of product spanning 1800-1980 was simply to ensure that the proxy would add significant value to the manuscript. For instance, if we reduced the time domain such that the proxy only needed to span the period 1900-1980 the number of coral proxies would increase significantly. However, the number of proxies providing data prior to the observational based estimates of SST, which begin between 1854 and 1886, would be roughly the same. Thus, we do not feel the inclusion of these proxies would be adding value.

It is true that this choice does exclude the Palmyra record, however the Palmyra record is included in Section 4. Further to this, in Figure 4 below we show that including the Palmyra data does little to change the running variance of the combined single station proxies. This is due to the fact that it is only one proxy getting added to 21 others and the fact that it has only ~70-yrs of data in the ~500yr window prior to 1886. Thus, we do not feel that changing the criteria to include the Palmyra record is warranted. However, we have tried to alter the text to make the reasons for including this time period more transparent. See lines 3-5 of page 12 of the revised manuscript.
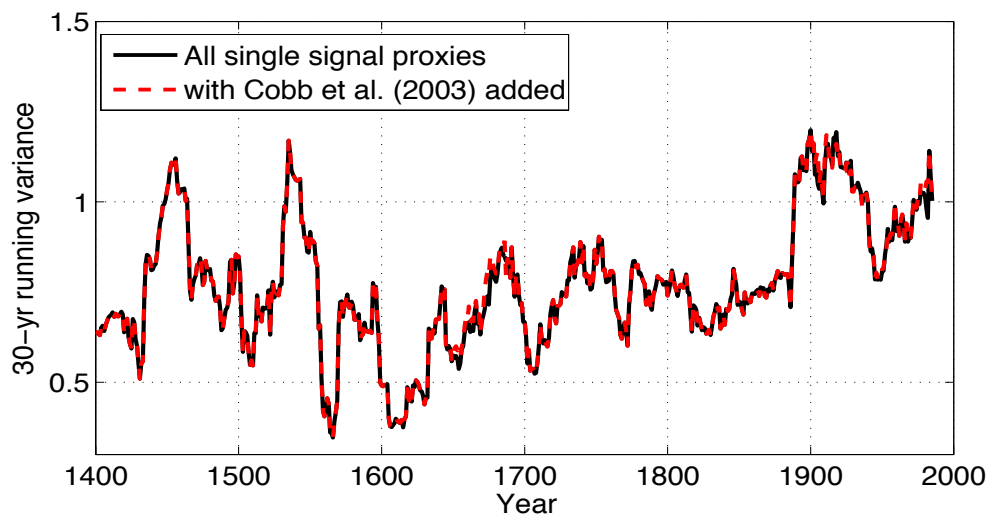
Figure 4: The black line represents the ensemble median of the 30-yr running variance of all single station proxies (Table 3), while the red line is the ensemble median running variance of all single station proxies with the Cobb et al. (2003) data included.

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.

We have now added a better description of the similarities and differences between the reconstructions. This additional text discusses how the past variability of the proxies varies in the context of the proxies most recent values, which brings out some interesting differences. This is now discussed in the third paragraph of the introduction of the revised manuscript.

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.

The text around the model descriptions has been changed as suggested. Please see lines 7-8 and 16-21 of page 4 in the revised manuscript.

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.

This text describing the spatial correlations has now been modified to provide more guidance. Also, we have removed the vague language around the discussion of this result. See the first and second paragraph of Section 3.1 in the revised manuscript.

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.

We have now combined these lists of caveats and present them in Section 6 of the revised manuscript. See lines 25-32 of page 14 and lines 1-19 of page 15.

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.

We have now rephrased this statement. Please see line 33 on page 7 and lines - 13 of page 8 in the revised manuscript.

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 alter- native 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. How- ever, 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?

The study of McGregor et al. (2010) utilized a PCA to identify the common time series and the common running variance time series. Here, we simplified their methods by using the median to identify the common signal. However, what our results suggest, and what was implied by the McGregor et al. (2010) study, is that working with running variance time series is preferable to working with the raw time series, if the goal is to reconstruct running variance time series. We have now tried to clarify what was meant by this on line 2-7 of page 7, lines 14-17 of page 8, and lines 3-8 of page 14 in the revised manuscript.

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.

We have changed the wording in response to the reviewer's comment. Please see line 21-25 on page 10 of the revised manuscript.

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?!

We have now removed the word apparent as suggested.

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?

This is true and has now been corrected. Please see section 5 of the revised manuscript. Thank you for picking this up.

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.

The caption has now been amended as suggested.

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?

We have now altered the reconstructions labels in the legends of Figs. 1 and 8, and Tables 1 and 2 to provide a more intuitive reconstruction title.

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).

This is clearly defined in the manuscript text just prior to referring to this figure, as such we do not believe it is also needed in the caption. See lines 29-33 of page 4 in the revised manuscript.

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

This has now been corrected.

We feel the manuscript order is appropriate and as such we have decided to keep the manuscript in its current form.

We have now included the mean correlation with the observations in Table 2 for those only interested in an average number. However we have decided to keep the rest of the table for those interested.

We think the better description of this comparison, seen in paragraphs 1-2 of Section 3.1 in the revised manuscript, helps clarify what should be compared to what. However, we also note that Figure 3 was generated entirely for this reason.