Climate
of the Past
Discussions

Interactive comment on "The influence of non-stationary ENSO teleconnections on reconstructions of paleoclimate using a pseudoproxy framework" *by* R. Batehup et al.

R. Batehup et al.

ryan130393@hotmail.com

Received and published: 17 November 2015

Dear Eduardo Zorita,

On behalf of the co-authors and myself, we are pleased to submit the revised manuscript. We have addressed all the reviewer's comments – reviewer comments are in bold, while our comments are in normal font. We have attached a manuscript with its difference to the submitted one highlighted using latexdiff, as well as an improved supplement. We believe that making these revisions has made the manuscript stronger so we would like to thank both reviewers for their constructive comments. Comments for reviewer #2 are on page C22.

Regards, Ryan Batehup (on behalf of the authors)

Reviewer Comments #1

Dear Oliver Bothe,

Thank you for your thorough and constructive comments on the manuscript. We have addressed each of the points raised. Your comments are in bold font, while our responses are in normal font. We have attached the revised manuscript (differences highlighted), updated supplementary, and an extra figure in this response.

Major comment

There are two possibilities, either I misunderstand the description of what has been done in the PNEOF1 "experiment", or the "experiment" doesn't represent what it is meant to show.

Maybe what you do is the following: you do the running correlations, you do the EOF, and you then select proxies from the regions deemed non-stationary and also having strong associations with EOF1. If this is the case, the method does what it is meant to do.

However, I understand the description of what you do as the following: you do the running correlations, you do the EOF, and you select proxies from the regions with strong associations with EOF1. Then, you do not sample from the non-stationary regions but possibly from the weakly correlated regions, which would explain the near-total lack of skill for some methods.

Let me rephrase: do you use the EOF to sub-sample the nonstationary regions for covarying regions; or, do you use the EOF to just organise all regions for covariability?

We performed the EOF on the running correlations (our measure of teleconnection strength) to organise all regions for co-variability. Out of extreme values (>abs(0.01))

of EOF1, 16-32% of grid points had a sufficiently high correlation (>abs(0.3)) to be used in the PNEOF1 experiment. We did not restrict the pseudoproxies to be non-stationary as this would have further reduced the available pseudoproxies to create reconstructions.

If my understanding is correct, do you assume, and if so why, that the EOF1 represents nonstationarity?

This is an incorrect assumption that we did make in the submitted manuscript. We thank the reviewer for picking up on this and we have revised the manuscript text to better reflect the analysis that was carried out (see page 19, line 11 of the revised manuscript).

"However, if pseudoproxies are selected from regions that demonstrate co-variability in the running correlations between TS and Niño 3.4 SST anomalies, reconstruction skill is devastated. To this end, an Empirical Orthogonal Function analysis (EOF) was used to 'organise' this co-variability, of which it is expected that non-stationarities are a major part. This is seen in the PNEOF1 experiment shown in Figure 7."

We have computed the percentage of non-stationary pseudoproxies in the EOF to be 9-15%. This suggests that regions displaying coherence in non-stationarities can cause problems with reconstructions even when selecting grid points that are not considered non-stationary. We have now discussed this briefly in the revised manuscript on page 19, line 29.

"The proportion of non-stationary grid points used in the PNEOF1 reconstructions was small, ranging from 9-15%. However, there was still a substantial loss of skill in these reconstructions even though the majority of grid points were classified as stationary by our statistical definition. This implies that a large and coherent change to the teleconnection exists in that region even if it considered mostly statistically stationary, and that was enough to degrade reconstruction skill. Thus, care should be taken to avoid the scenario where all constituent pseudoproxies used in a reconstruction lie in a region

where there are large, coherent variations in teleconnections, even if these variations are considered stationary."

The section in the discussion has also been modified (page 27, line 21):
"However, this skill improvement is affected by the degree of non-stationarity and tele-connection co-variability present in the reconstructions, with non-stationary proxy networks NSTAT_ntrop_ts, Fig. 8,red lines) and ``organised'' teleconnection co-variability (PNEOF1, Fig 7 a--d) reducing the degree of improvement in skill with increasing network size. Thus, where increasing network size would usually improve the reconstruction, non-stationarities and spatial coherence in variations in teleconnection strength can substantially temper this improvement. In extreme cases, where proxies are selected from co-varying areas (PNEOF1), Fig. 7), reconstruction skill may show no improvement with larger proxy networks. This further stresses the importance of ensuring that all constituent proxies utilised in a reconstruction are not affected by co-varying teleconnections. This is more likely achieved in spatially diverse, large multi-proxy networks."

Anyway, what does "EOF weighting < 0.1" [Line 18 on page 3871] mean? Do such values occur? (According to the color bar they don't.).

Thank you for pointing out this error. The value '0.1' in the manuscript text should have been '0.01'. This has been corrected in the manuscript.

Additionally, I wonder whether the trend in the PC1 suggests some problems in the control run.

There is little model drift in the GFDL CM2.1, being a 0.1°C rise in Niño 3 region sea surface temperatures (SSTs) per millennium (Wittenberg 2009). Further to this, the PC1 being discussed was computed from the running correlations, not the SST directly, so it does not necessarily indicate model drift as we can assume any drift is coherent across all grid points, which would not manifest in a change in correlations. This model is shown to "exhibit strong interdecadal and intercentennial modulation of

its ENSO behavior" (Wittenberg 2009), and we believe what is seen in PC1 is simply due to some of this intercentennial ENSO modulation.

Minor comments:

1. Could you please as soon as possible, i.e. before your final response to the reviewers, provide a version of Figure 10 including the MRV?

Figure 10 showed a few hand-picked reconstructions to show that there are some instances where variability (standard deviation) between proxy time series appears to be related to variance losses in the final reconstruction. We were unable to produce a new figure that accurately summarised the ideas of Figure 10 and removed the reviewer's confusion. Thus, we have decided to drop this figure and its corresponding paragraph in the manuscript. This manuscript change does not influence the final results of the manuscript.

2. I think the title should mention that you reconstruct of ENSO-variance. The introduction also should clearly state it. Similarly, on some/most instances where you write "reconstructions of ENSO" it would be more appropriate to write "reconstructions of ENSO variance". Alternatively you may state early on that "reconstruction of ENSO" implies "reconstructions of ENSO variance".

The following changes has been made:

The title has been changed to "The influence of non-stationary teleconnections on paleoclimate reconstructions of ENSO variance using a pseudoproxy framework".

The first sentence of the abstract's second paragraph has been changed to "This study examines the implications of non-stationary teleconnections on modern multi-proxy reconstructions of ENSO variance".

The first sentence of the abstract's third paragraph has been changed to "We find that non-stationarities can act to degrade the skill of ENSO variance reconstructions. However, when global, randomly-spaced networks (assuming a minimum of approximately

20 proxies) were employed, the resulting pseudoproxy ENSO reconstructions were not sensitive to non-stationary teleconnections".

In various places: "ENSO variability" has been changed to "ENSO variance".

Similarly at Page 3868 Line 20ff: The description of Figure 4 is, as far as I understand it, incorrect. Please be clear that it is the correlation between the running variance series and not between the Nino3.4-indices.

The sentence has been changed to: "the correlation between the pseudoproxy reconstruction of the Nino 3.4 running variance and the model Nino 3.4 running variance" (on page 16, line 22).

3. Methods:

a. Is the calibration window length implicitly meant to also be the length of the running variance windows? I ask, because you never explicitly mention the window-length for the running variances. It can't be the calibration window length, because you also use the full 499 years for calibration. So, what is the window length for the running variances?

The running variance window length is 30 years for all experiments. We have added the sentence: "All running variances were calculated using 30 year windows." to the revised manuscript to clarify (see page 14, line 4 of the revised manuscript).

b. Please clarify what calibration means in your setup.

Calibration is similar to the normal calibration between the paleoproxy time series and the instrumental record. In our experiment, calibration establishes relationships between the TS grid points and the model Nino3.4 index during a certain time period, which is the calibration window. We have inserted the sentence "The calibration window is the time period where relationships between the TS grid points and the model Nino3.4 index are established" on page 9, line 17.

c. You mainly consider skill in terms of correlation. Is correlation really the best skill measure in this case? Why?

We use correlation mainly for simplicity. We do also look at RMSE, but it is not within the scope of the study to examine the best skill measures for paleoclimate reconstructions. If it is an either correlation or RMSE choice, we would argue the correlation is the most important metric. For instance, what good is a small error if you have no idea whether the variance is increasing or decreasing? It is much more useful to know the relative changes, and to understand how what we are seeing now relates to what has happened in the past.

d. On page 3860 line 10 you describe the timeseries as "June-July"-averages. I assume you mean thirteen month averages. Please clarify.

This meant to mean 12 month averages, from July to June of the following year. This typo has been corrected.

e. You mention model-drift: Is it a problem in your simulation?

No, we do not think that model drift is an issue in this simulation. For instance, the study of Wittenberg (2009) who produced this data reports a drift of only 0.1°C per millenium in the GFDL CM2.1 eastern equatorial Pacific SSTs (Wittenberg 2009).

f. I wonder whether it would be better to convert all correlations to Fisher-Z-scores. I do not propose to do it, I only wonder whether it might clarify the Figures.

It would help to distinguish differences at higher correlations, but we feel leaving it as correlations is simpler and easier for the reader to understand. As such, we have left it as correlations.

4. There are references which should not be omitted. A paper on reconstructing ENSO-variance has to mention the work by Russon and colleagues (Russon et al., 2014, 2015). Furthermore, the general introduction of ENSO-reconstructions

appears to ignore the works of Watanabe et al. (2012), Cobb et al. (2013), and Li et al. (2013).

As a less important side-note I want to mention that there have been other studies dealing with non-stationarity of climate-modes in recent years.

A few sentences have been modified on page 4, line 15: "Tropical corals are the dominant proxy type in this region, and are known to provide very skilful reconstructions of the surrounding SSTs and ENSO. However, the addition of non-climatic noise to these proxies also complicates the estimation of the significance of changing in past ENSO variability (Russon et al. 2014; 2015), as does their limited life span (i.e., records are on average about 50 yrs in length, with the longest records less than two centuries) (Cobb et al., 2013; Neukom and Gergis, 2012)."

- Russon et al 2014 and 2015 has been inserted into page 4, line 19 (as above).

- Cobb 2013 has been inserted into page 4, line 22 (as above).

- Li et al. 2013 has been inserted into page 4, line 22.

- Watanabe2012 has been inserted into page 4, line 7.

5. On the better performance of MRV.

a. I think this would be an interesting test for what happens when noise is introduced in the proxies.

It would be interesting to explore this, but it is beyond the scope of the current paper. However, we do intend to examine this in future research.

b. You stress the overall better performance of MRV. One may argue that this is unsurprising as it uses the variance from the beginning, in this sense the comparison may be called biased.

We do not believe that this is a bias as the difference between the MRV and RVM method is simply due to the order of operations, which is not obvious.

However, on the other hand, I am not sure it is true. In a real world scenario: what is the more important skill-metric, the RMSE, where MRV performs consistently worse, or the correlation? Is there possibly another better suited skill-metric? You mention the potential need for re-scaling but don't do it. Why not, I would think implementing an MRVPS (MRV plus scaling) should be easy enough. There may be

reasons, but as you so far don't show the MRV series (e.g., in Figure 10), the reader is unable to assess this.

The question of which is the "correct" metric with which one should assess reconstruction skill has been widely debated in the literature for a long time. We do not intend to assess different metrics in detail here, as it is not the main point of the paper. But in response to the first question raised here, if it is an either-or choice, we would argue the correlation is the most important metric. For instance, what good is a small error if you have no idea whether the variance is increasing or decreasing? It is much more useful to know the relative changes, and to understand how what we are seeing now relates to what has happened in the past (from comment 3c).

Before moving onto the second component of this question, we would like to highlight that each of the reconstruction methods utilised, suffers from variance reduction. A feature that is clear in Figure 10 and Supplement Figure S6 of the revised manuscript, and is discussed on page 22, line 23.

"It is well known that all reconstruction methods result in a loss in ENSO variance, and this is clearly shown in Figure 10. In Fig.10a-d, we can see that, all reconstructions underestimate the model Nino 3.4 running variance (black line). However, this figure also shows that this variance loss is exaggerated with the MRV method (panels c, g), and this is also seen in Supplementary Fig. S6, particularly at the larger network sizes."

We expect, and will show here (and in the revised manuscript), that scaling the resulting running variance time series improves the RMSE of the MRV, but also note that it cannot be done well when limited to 31-yrs of data (the 31-yr calibration window). Thus, we have performed scaling on the 61 and 91-yr calibration window experiments. We have scaled using the average (calculated over the 1000 reconstructions) regression of the MRV and Nino3.4 RV within the calibration window to produce a scaled MRV – 'SMRV'. The proportions of the SMRV with lower RMSE than MRV is up to 70% and is indicated in the attached Figure E1, with the proportion getting larger as the number of proxies in the reconstructions increase. The median RMSE of each reconstruction method is also indicated on the plots on Figure S7.

We have also discussed this scaled MRV in a paragraph on page 23, line 25 and added a Figure S7 to the supplementary (while incorporating Fig. S3 into a new Fig. S2).

"In order to compensate for the variance loss of each reconstruction (Figure 10a-d), we rescale each method's resulting running variance time series (Figure 10e-h). Rescaling the running variance time series was carried out using the average (calculated over 1000 reconstructions) regression between the reconstructions and the modelled Nino 3.4 running variance within the calibration window. When the MRV (panel 10.c) is scaled to form the SMRV (panel 10.g), there is a jump in reconstruction variance (grey shading), such that the modelled Nino 3.4 index running variance is now encompassed by the grey shading. Using this simple scaling technique, we see a large reduction in the RMSE (see Figure S7) - up to a 0.1 reduction in the median (Figure S, panels c, g, black line) and no changes in the correlation (not shown). In fact, it is noteworthy that on average the scaled MRV has the smallest RMSE (significant to 99% level via a two sample t test) of all reconstruction methods."

Note that examples of the MRV reconstructed running variance are now provided in the new Figure 10 mentioned above, showing the range of the reconstructions compared to the Nino3.4 RV series they are attempting to reconstruct.

c. Similarly, page 3875 line 13ff: Isn't the damping an expression of especially large variance loss for MRV, at least if the term is used as commonly employed for reconstructions?

Yes, as described in our response to your point above, the MRV displays an exaggerated variance loss as the term is currently used. This is now discussed in the revised manuscript on page 23, line 4.

d. Page 3877 line 21: You write MRV excelled, but MRV also showed large RMSE. I think that should be mentioned. Can you estimate how noise/uncertainties in the proxies would affect this feature.

In regards to the latter point: as stated above, to examine this would require looking at the effects of noise in the pseudoproxies which is beyond the scope of this study. However, we do highlight the need for this to be examined in future research: "The compounding effects of noise and non-stationarities on the reconstruction method and hence, a reconstruction, should be the focus of future research efforts in this area" on page 26, line 18.

In regards to the former point: We have made some changes to better reflect the large RMSE: "However, the unscaled MRV method showed poor RMSE performance, meaning that it can only be used to provide useful information on the relative changes in ENSO variance." has been inserted into page 27, line 3.

The first line of the second paragraph of page 22 (line 17) has been changed to: "It is worth noting that although the MRV method shows the most consistently high correlations to ENSO and appears to be the least sensitive to calibration window position (smallest percentile ranges, Supplement Fig. S5), it has the highest RMSE (root-mean-square error)". A scaled version of the MRV, the 'SMRV' has been created - see previous comments.

e. Page 3879 line 16/17: MRV is the most robust in your perfect-proxy setting.

Again, we now discuss this possible limitation of our results. See page 27, line 3, where we added the sentence: "However, the unscaled MRV method showed poor RMSE performance, meaning that it can only be used to provide useful information on the relative changes in ENSO variance."

6. With respect to the discussion of CPS on Page 3873 Line 15ff. Why do you single out CPS here? It is not really worse than RMV or EOF.

We were actually highlighting the CPS as the next best method, due to this confusion we have now reworded this sentence from "It is noted that the CPS_RV method performs well, although mainly with longer calibration windows and for the random selection experiments" to "The CPS_RV method performs almost as well as the MRV, although mainly with longer calibration windows and for the random selection experiments" on page 21, line 26.

7. On the discussion of the tropical supposedly non-stationary grid-points on page 3865: Don't these mainly represent the wide range of internally varying differences in the evolution of ENSO-events which are potentially not captured by a simple stochastic process? I am not so much thinking of CP vs. EP ENSO, but different evolutions of one of these flavors. I think this could be discussed more extensively.

You make a similar point on page 3870, line 15ff. So, it's not only about different flavors but about the large variability in how events evolve. The statistical process captures not necessarily all dynamical variability.

We have discussed this more extensively on page 13, line 16:

"Of further note is a large non-stationary area in the equatorial Pacific; given this is the area surrounding our ENSO index it is debatable whether this should be considered as a non-stationarity. Rather, we expect the changing relationship in this surrounding region to be the result of complexities of ENSO that may not captured by the

simple stochastic model of stationarity. For instance, ENSO displays: i) significant non-linearities in its magnitude (An and Jin, 2004) and duration (Okumura and Deser 2010); ii) differences in the evolution of events with all La Nina's and most small to moderate El Nino's having SSTAs that propagate from east to west, while the SSTA of large El Nino events propagate from west to east (Santoso et al. 2013); and iii) changes in its spatial structure (CP-EP type events) which may be considered different flavours of events rather than non-stationarity teleconnections of the event (Gallant et al., 2013; Sterl et al., 2007)."

8. On Figure 10: The regressions and the correlations show rather weak relations between the two series. Anyway, why should the standard deviation of the running correlations of the proxies with the target and the variance reconstruction be related in any way. Put differently: what do we learn from the correlation of the variance of a measure with the square-root of the variance of running correlations of a measure with another measure?

We were examining the possibility of the variability between the proxies being related to abrupt changes in the reconstructed variance. As discussed in our response to this reviewer's first minor comment, we were unable to produce a new figure that accurately summarised the ideas of Figure 10 and removed the reviewer's confusion. Thus, we have decided to drop this figure and its corresponding paragraph in the manuscript. This manuscript change does not influence the final results of the manuscript.

The number of effective degrees of freedom of the time series appears to be small, are the correlations even significant?

Figure 10 has been removed from the revised manuscript - see minor comment 1 and the above comment.

10. In your plots you give explained variances for running variances. How much variance is captured by the stationary sample-variance-distribution?

We have looked at the distributions of the reconstructed running variances, and they closely fit with a normal distribution for all reconstructions, which suggests discussing the explained variance is valid. However, we are not sure if we understand the question correctly. We also would like to refer to Fig. 5 for a comparison of stationary and non-stationary reconstruction skill.

11. The discussion in the beginning of the last paragraph on page 3872 is not really relevant, is it? The case of a 500 year calibration window is not realistic, so the discussion should focus on the skill differences between 91yr and 31yr. Related: The second part of the paragraph (on page 3873) appears to be partially redundant.

The reconstruction skill is affected by both non-stationarities and reduced window size. Thus we believe it is fair to look at the impact of reduced window size alone, before making a judgement on the impact of non-stationarities. We have also removed the redundancies.

12. Page 3876 line 2ff: Please do not just point to panels but give some more information on what we see.

The following changes have been made to the surrounding text:

Page 24, line 24: "absence of a large spatially coherent region of correlations in the tropical Pacific Ocean (see Supplement Fig. S1e)" has been changed to: "absence of a large spatially coherent region of correlations in the tropical Pacific Ocean (compare tropical areas in Supplement Fig. S1b and Supplement Fig. S1e)"

Page 25, line 6: "The RVM method appears to perform better with precipitation than temperature in panels d, and h, with not much difference in panel l, which is consistent with the findings of McGregor et al. (2013)" has been changed to "The RVM method appears to perform slightly better with precipitation than temperature (Supplementary

Fig. S2, mainly at longer calibration windows), which is consistent with the findings of McGregor et al. (2013)."

13. It may help the reader if you extend a bit in line 22ff of page 3876 on what Wittenberg (2009) showed.

The sentence on line page 25, line 27 has been changed to: "Wittenberg 2009 discussed that such changes to ENSO behaviour could conceivably alter the teleconnections between ENSO and local climate and that these changes may not be represented in the historical record."

14. More a technical comment, but I put it here as well to emphasize it: I think Figure 8 is rather unclear, since I am not really able to distinguish the different hatchings.

Figure 8 has been simplified and clarified. The hatching has been removed and the remaining lines of proportion have been compressed into four panels. See the Figure 8 in the updated manuscript.

Technical Comments

General: I am surprised by the comma placement in the manuscript. However, as a non-native English-speaker, I do not annotate it.

The comma positions have been corrected where necessary.

My subjective impression is that the abstract could be shortened and could formulate the relevant points more concisely.

The abstract has been shortened and made more concise.

Page 3854 Lines 17-20: Please rephrase the sentence. (I do not really see how the "to which"-part relates to the previous sentence-structure.)

This has been changed to "Reconstructions of the variance in the Niño 3.4 index repre-

senting ENSO variability, were generated using four different methods. Surface temperature data from the GFDL CM2.1 were used as pseudoproxies for these reconstruction methods" on page 2, line 17.

Page 3854 Line 24: What do you mean by "uniformly-spaced"?

This has been changed to "randomly-spaced" on page 2, line 27.

Page 3859 Lines 21-24: I think this sentence could be clarified "It has also been shown that the model teleconnections, represented by correlations in 31yr windows between grid points and the Nino 3.4 index generated from the model, do exhibit variability between periods and compared to correlations calculated over the entire period (Fig.˜\ref{fig1}a, \citealp{Wittenberg2012})."

This sentence has been changed to "It has also been shown that the model teleconnections, represented by correlations in 31 year windows between grid points and the Nino 3.4 Index generated from the model, do change over time, and differ compared to correlations calculated over the entire period (Fig. 1a, Wittenberg, 2012)" on page 7, line 29.

Page 3860 Lines 21-22: "is used" : : : "are selected". Please clarify is/are.

This has been corrected to 'is' on page 9, line 3.

Page 3862 Line 3: "Fig. 1b" -> "(Fig. 1b)"

The correction has been made on page 10, line 10.

Page 3864 Line 22: you write of the "possible" range of running correlations. Is "possible" the correct word here

We believe it is: "A 95% confidence interval was generated at each grid point from the stochastic simulations and was used to represent the range of running correlations possible, assuming a teleconnection was stationary" on page 12, line 24.

Page 3865 Line 22: Do you examine the "likely" - as you write - or the "potential" effects of non-stationarities?

This has been corrected to 'potential effects' on page 14, line 2.

Page 3866 Line 22: "(2005); Hegerl" -> "(2005) and Hegerl"

This has not been changed as we believe that there are too many references to place an 'and' there. It is currently: "(Esper et al., 2005; Hegerl et al., 2007; Mann et al., 2007, and references therein)."

Page 3867 Line 16: I don't think you write in Sect. 2 how you calculate the running variance for the ENSO index (see above, what is the window-length of the running variance).

We have added a line at the beginning of 3.3 Reconstruction Methods: "This study examines the potential effects of non-stationarities on multi-proxy reconstructions of the running variance of the Nino 3.4 index (representing the variability of ENSO) using pseudoproxy data. All running variances were calculated using 30 year windows" on page 14, line 4.

Page 3867 Line 19: "dataset is available with larger proxy networks". My impression is, that there may be an "and" missing between available and with.

We have corrected this to: "when the entire dataset is available (and with larger proxy networks)" on page 15, line 24.

Page 3869 Line 5: "The skill metrics", I think you mean "the proportion of skill metrics"?

The blue and orange lines in Figure 4 are used to evaluate the skill of the methods. Although they are indicating the proportion of skilful reconstructions (skilful meaning explaining greater than 50% of variance), this proportion itself can be used as a skill metric for comparative purposes. See page 17, line 8 in the revised manuscript.

Page 3870 Line 1: "In all". Should that be "For all"?

Yes, that has been corrected on page 18, line 3.

Page 3872 Line 15ff: I think the second part of this sentence is incomplete.

This sentence has been removed.

Page 3872 Line 18: I don't mind "Fairly good chance" but I can imagine colleagues who are rather annoyed by such a phrase.

We have corrected this to: "using a minimum of 20 proxies gives a reasonable chance" on page 21, line 1.

Page 3872 Line 26: Is "would be" correct? Shouldn't it read "is" which may be qualified by "likely" or "potentially".

We have corrected this to "This decrease of skill is potentially due to some information loss in the relative datasets, and not necessarily due to non-stationarities" on page 21, line 8.

Page 3873 Line 7: I think the "However" is wrong.

This sentence has been removed.

Page 3873 Line 15: "It is noted"? Why passive construction? -> You note.

This sentence has been changed (page 21, line 26), see minor comment 6.

Page 3874 Line 3: Not the red line outperforms the other lines, but using 91 year calibration windows performs better than shorter windows.

Corrected to: "and displaying some sensitivity to calibration window length (91yr windows perform better than shorter windows)" on page 22, line 14.

Page 3874 Line 16ff: Is the sentence correct? Do you generally plot the "variance taken : : : of the correlations"?

Yes, this is the variance of the reconstructions. As the reconstructions are running variances themselves, this is the variance of the running variance. However, this sentence has been removed in the revised manuscript.

Page 3875 Line 18: Is "on this paper" correct?

This has been corrected to "Although not the focus of this paper, precipitation was also examined for all experiments" on page 24, line 18.

Page 3877 line 15: "highlight a case for considering" or just "highlight"?

We believe the existing sentence is more suitable: "the results presented here highlight a case for considering the influence of non-stationarities on real-world reconstructions and their underlying methods" on page 26, line 23.

Page 3878 line 21ff: I think the second part of this sentence is incomplete.

"Given the skilful reconstructions in ENSO variance that can be produced by neglecting pseudoproxies from the centre of action, as shown here, the utilisation of data solely from the eastern equatorial Pacific appears unnecessary" has been corrected to:

"Given the skilful reconstructions in ENSO variance that can be produced by neglecting pseudoproxies from the centre of action as shown here, the utilisation of data solely from the eastern equatorial Pacific appears unnecessary" on page 28, line 6.

Page 3879 line 17: "many various"?

We have corrected this to: "and there are many reconstruction methods in the literature" on page 29, line 7.

Page 3879/3880: My impression is that the second part of the conclusions is more or less redundant and repeats what the first part already said.

The redundancies have been removed (see page 29, line 12).

Page 3880 line 5ff: I think this perspective is not really necessary, but that's just

personal taste.

This line has been removed.

Figure 1: Please rephrase "is the correlation between of the entire 499 years of TS at each grid point and the model calculated Nino 3.4 index correlation coefficients"

This has been corrected to: "The shading is the correlation between of the entire 499 years of TS at each grid point and the model calculated Nino 3.4 index, both calculated from the GFDL CM2.1 data" (page 36).

Figure 2: Please provide labels for the color bars of panels b,d,f. verses -> versus

A new figure has been prepared – see attached updated manuscript (page 37).

Figure 3: You write "the pseudo-reconstructions running variance". Wouldn't it be more appropriate to write the "pseudo-reconstructions of running variance"?

Yes, it would be. I have corrected this to "between the pseudo-reconstructions of running variance and ENSO running variance" on page 38.

Figure 4 (and other Figure captions): You write "reconstruction's running variance" which is appropriate for CPS and EPC and to some extent for RMV, but shouldn't it be "reconstructed running variance".

Yes, thank you for pointing that out. This has been corrected in Figures 4, 5, 8 and 9.

"explaining greater than 50% of explained variance" -> "explaining greater than 50% of variance"?

This has been corrected to "explaining greater than 50% of variance".

Figure 6: Please rephrase "and this determines what values of proportion can be taken as larger groups have a wider range of possible non-stationarity proportions than smaller groups".

This has been changed to "determines what values of proportion can be taken, hence larger groups have a wider range of possible non-stationarity proportions than smaller groups" on page 41.

Figure 8: The hatching is, from my point of view, nearly unidentifiable.

The hatching has been removed in the new Figure 8.

Supplement: Please provide a clear copyright statement.

There will be a copyright statement on the front page of the supplementary material when it is processed by the editor.

Figure S1: Could you reshape the aspect ratio for final publication?

The aspect ratio has been fixed.

Figure S4: Caption, last line: "with with"

The repeated word was removed.

Reviewer Comments #2

Dear Anonymous reviewer,

Thank you for your thorough and constructive comments on the manuscript. We have addressed any issues that have been found. Your comments are in bold, while our responses are in normal font. We have also attached the revised manuscript (differences highlighted), updated supplementary, and an extra figure in this response.

Comments:

Introduction: The last paragraph leaves the reader thinking that teleconnection nonstationarity does not impact reconstruction skill. This is not the case and I would start this paragraph by noting that non-stationarity does degrade reconstruction skill. You can then follow that by noting that the impact of non-stationarity can be minimized by employing a large global network.

We have left the discussion of results out of the introduction as we decided to stick to the traditional report format. As our paper explores the impact of non-stationarities on the reconstruction skill, we have left these questions open ended in the introduction. However, we believe that the reviewer may be referring to the last paragraph of the abstract, in which case we have now reworded the opening sentence:

"We find that non-stationarities can act to degrade the skill of ENSO variance reconstructions. However, when global, randomly-spaced networks (assuming a minimum of approximately 20 proxies) were employed, the resulting pseudoproxy ENSO reconstructions were not sensitive to non-stationary teleconnections".

Page 3854, Line 15: add variance after ENSO. More generally, this is a problem throughout the manuscript. Make sure to be clear that these are reconstructions of ENSO variance. I will note places where this should be clarified but have likely missed some.

The fact that we are reconstructing ENSO variance has now been clarified throughout

the revised manuscript.

Page 3854, Line 27: This sentence seems tangential to the overall results and perhaps overly specific.

This sentence is intended to describe our result that the MRV reconstruction method appears to perform better than other methods. We have reworded this sentence to: "Different reconstruction methods exhibited varying sensitivities to non-stationary pseudoproxies, which affected the robustness of the resulting reconstructions" on page 3, line 4.

Page 3855, Line 11: Suggest removing: "Thus, these proxies are the essential tool for creating paleoclimate reconstructions."

This sentence has been removed.

Page 3857, Line 28: Suggest removing: "This places increasing stress on the assumption that teleconnections are stationarity. Further to this, it."

We believe this sentence is important to tie in the multiple results from different papers and show their relevance to our manuscript. This sentence has been tweaked to "This places increasing stress on the assumption that teleconnections are stationary. This raises the question as to whether non-stationarities have an appreciable influence on the robustness of past paleoclimate reconstructions" on page 6, line 9.

Page 3858, Line 4: I would replace variability with variance to be absolutely clear that that is what is being reconstructed.

This has been done.

Section 2: You are only using a single model and it will be important to note that PPE results have been shown to be model dependent, at least in the case of CFRs (e.g. Smerdon et al. GRL 2011 and Smerdon et al. Clim. Dyn. 2015). This is particularly important given that not all models have non-stationary teleconnec-

tions to the tropical Pacific (e.g. Coats et al. GRL 2013). Using a different model may provide different results, so making absolutely clear that the results are specific to the characteristics of this GFDL model and not necessarily applicable to the real world will be important (after all, a model with either stationary teleconnections or much more non-stationary teleconnections is arguably an equally plausible representation of the real world).

The reviewer is correct as the global number of non-stationary grid points may vary depending on the climate model. However, we believe that this is unlikely to affect the differences between reconstruction methods as these have been examined using only source pseudoproxies that can be considered non-stationary (see Fig. 5 of the revised and original manuscript). In order to address the reviewers concern, we do now discuss this potential caveat on in the conclusions on page 28, line 17: "These results make the implicit assumption that the modelled co-variability of the non-stationarities and relative proportions of non-stationary areas to stationary areas are realistic, which has not been explicitly tested here". The sensitivity of multi-proxy climate index type reconstructions to the climate model will be examined in the future though.

We have modified a sentence in the discussion: "We note that although we use the same model as in the Wittenberg (2009) study, the results are unlikely to be a product of the model configuration given that Gallant et al. (2013) identified nonstationarities in three different GCMs" has been changed to:

"Gallant et al. (2013) identified non-stationarities in three different GCMs. It is noted that while numerous models display non-stationarities, their regional existence may vary depending on the model used (Coats 2013). We do not expect our evaluation of various different reconstruction methods performance in the presence of non-stationarities to be affected by model configuration, however we intend to examine this in future research" on page 26, line 5.

Page 3860, Line 10: Does June-July imply a two month average? Based on the

rest of the sentence I assume you mean a 13 month average. Please clarify.

This was a typo, and has been corrected to "July – June" 12 month averages on page 8, line 17.

Page 3860, Line 11: The sentence on computational cost seems unnecessary.

This sentence has been removed.

Page 3861, Line 4: The reference to Lee et al. (2008) seems out of place because adding non-climatic noise at different levels is a relatively standard choice in PPEs. Perhaps restructuring to put the reference at the end of the sentence with a parenthetical note that Lee et al. is an example of this. The reference is also not listed in the references.

The sentence on page 9, line 9 has been changed to: "The pseudoproxies are not degraded by adding noise (e.g. Lee et al. 2008), as the effects of noise on the reconstructions are beyond the scope of the study".

The reference has been added to the reference section.

Page 3861, Line 11: Suggest removing "to some extent, making them at least partly relevant for reconstructing the ENSO signal."

As suggested, the sentence "This threshold is an arbitrary criterion that is simply there to ensure the pseudoproxies represent ENSO to some extent, making them at least partly relevant for reconstructing the ENSO signal" has been changed to "This threshold is an arbitrary criterion that is simply there to ensure the pseudoproxies at least partially represent ENSO" on page 9, line 14.

Page 2861, Second paragraph: Are 1000 random networks of each size from three to 70 used?

Yes, the reviewer is correct. We have now modified the sentence on page 9, line 23 to read: "To produce reconstructions of the model Nino 3.4 index variance, 1000 random

C2352

networks were selected per network size, calibration window length, and calibration window position. "

Page 3861, Last paragraph: It is important to note that a real reconstruction will only be able to calibrate on the observational record. The first sentence, however, seems to distract from this important point - perhaps try rewording.

"The correlation at each grid point over the whole time period (499 years) and ENSO is assumed to represent the true teleconnection strength, as its use for calibrating the proxies should result in more accurate reconstructions" has been reworded to "The correlation between ENSO and each grid point time series (i.e. Nino3.4 & TS) over the whole time period is assumed to represent the true teleconnection strength" on page 10, line 1.

Page 3863, Line 5: The last sentence is unclear and doesn't seem necessary.

The sentence "However, our experiments showed that this assumption also produces larger errors in the reconstruction (not shown)" has been removed.

Page 3863, Line 8: But you do show results from these experiments in the other figures. Perhaps remove this sentence.

We thank the reviewer for highlighting this discrepancy. The sentence "For the remainder of the paper, we show the second version of the experiments only, as it represents the most realistic case." has been changed to "As the second case is the most realistic case, we mainly focus on the second version of experiments for the remainder of the paper" on page 11, line 12.

Page 3865, Line 8: Do you mean segments or windows and not years? A year can't be non-stationary but the 31-year window, for instance, can. Or are you counting up the number of years within all the non-stationary windows? That would be much less intuitive. If it is the former I would suggest changing years to windows or segments and doing the same in the corresponding figure and

C2353

caption.

Yes, we do mean windows in this instance as each year we were discussing implicitly includes the 15-year period either side of the year in question. Thus we have simply changed 'year' to 'window' (see page 13, line 10). The caption of figure 2 has also been updated to be consistent with the manuscript text.

Page 3868, Line 21: I think that you mean the running variance of the Niño3.4 index.

This has been corrected in the manuscript.

Page 3869, Line 13: Perhaps provide a value in parentheses here (after larger network

sizes).

The mean percentage of skilful reconstructions has been calculated to be 68%. Using larger network size only (20 to 70), we get a value of 77% of all non-tropical reconstructions being skilful (as according to the manuscripts definition of skilful – explaining more than half the variance of Nino3.4). This value of 77% has been inserted into the manuscript on page 17, line 15 as was the definition of larger network sizes.

Page 3869, Line 17: Again change variability to variance.

This has been corrected in the manuscript.

Page 3870, Line 11-14: How are the psuedoproxies non-stationary if they display little variability in correlation, that is not intuitive.

This was an interesting find amongst the experiments. Since the non-stationary definition doesn't require there to be minimum magnitude change in teleconnection strength, the pseudoproxy time series just needs to deviate from the generated stochastic range earlier defined. This may be a flaw of the non-stationary definition, but rather we believe it is picking up the non-linearities of ENSO events in that tropical Pacific region. This

C2354

is now discussed in more detail in the manuscript on line 16 of page 13 in response to one of the other reviewer comments.

Page 3870, Line 19: I understood what you were saying after reading further but was confused by this sentence initially. Perhaps try to make the statement more clear.

The sentence "In regards to why non-stationarities do not seem to impact the high skill of random pseudoproxy selection of Sect. 4.1, we find that the likelihood of selecting non-stationarities is relatively low." has been changed to:

"The fact that we see a minimal effect of non-stationarities in the randomly selected pseudoproxy experiments may be because the likelihood of selecting non-stationarities is relatively low" on page 18, line 19.

Page 3871, Second paragraph: I found this confusing. I think what you are saying is that if the network consists of a large proportion of grid-points chosen due to spurious correlations in that calibration window, the reconstruction skill is very low. The statement: "non-stationarities at the same time" might be part of the problem. A grid point is either non-stationary or not, but if it is non-stationary and weakly correlated it won't always be eligible to be picked and that appears to be what you are getting after.

If a particular window's correlation lies outside the 95% confidence interval generated by the stochastic simulations, then we deem it as a non-stationarity. EOF analysis was employed to determine covariation in non-stationarities and we showed that this covariation negatively effects reconstruction skill (non-stationary grid points made up 9-15% of the PNEOF1 experiment). This paragraph has been changed substantially to address an issue of co-varying teleconnections in response to reviewer 1 (Oliver Boothe) Major comment on page 19, line 11.

Page 3873, Line 10: The result that increasing the length of the calibration win-

C2355

dow is less important for reconstruction skill as compared to the choice of method or the amount of non-stationarity is important and gets lost a bit in the manuscript.

Thank you, we have now tried to clarify this finding by adding a sentence in the discussion "The non-stationarities and reconstruction method usually had a greater influence on reconstruction skill than the calibration window length" on page 26, line 26.

Page 3874, Line 7: "correlations to ENSO variance." It might be worth changing everything to ENSO variance or everything to Niño3.4 variance throughout the manuscript for clarity.

This has been changed throughout the revised manuscript.

Page 3874, Second paragraph: The discussion of RMSE versus correlation for MRV seems unnecessarily drawn out (and slightly confusing). The takeway appears to be that the scaling of the variance for he MRV method is too low but the timing of variance changes are correct.

This paragraph has been changed in response to reviewer 1's (Oliver Boothe) major comment to include information about variance losses in the reconstruction method on page 22, line 23, along with a discussion of the effect of rescaling the pseudo proxy reconstructions. We have also tried to simplify the discussion to make the take home message clearer (see page 27 line 12)

Page 3877, Line 2: Coats et al. GRL 2013 showed model dependence of non-stationarity to ENSO so this statement isn't strictly correct.

That small section on page 26, line 5 has been changed to:

"Gallant et al. (2013) identified non-stationarities in three different GCMs. It is noted that while numerous models display non-stationarities, their regional existence may vary depending on the model used (Coats 2013). We do not expect our evaluation of various different reconstruction methods performance in the presence of non-

C2356

stationarities to be affected by model configuration, however we intend to examine this in future research"

Page 3877, Line 3: Suggest removing virtual.

"Virtual" has been removed on page 26, line 11.

Page 3877, Line 26: To this reader it was not obvious why the filtering produced these different interpretations.

We were trying to reconcile differences with previous studies here. We believe that the difference is due to the fact that the unfiltered time series contains decadal signals that are less likely to suffer from the signal cancelation of higher frequency variability, hence they act to enhance the skill of the RVM reconstruction. However, as this has not been thoroughly tested and added little value to the current manuscript, we choose not to elaborate on this further.

Page 3878, Line 1: The point here is not that multi-proxy networks will produce more informative reconstructions, it is that larger and more global networks will. Maybe flip the sentence structure so that in the back part you can explain that multi-proxy networks tend to be larger and more global.

The sentence "For reconstructions of large-scale phenomena like ENSO, networks will produce more informative reconstructions because the larger networks contain more information, including spatial information, compared to single site (Mann, 2002; Lee et al., 2008; von Storch et al., 2009; McGregor et al., 2013)" has been changed to:

"For reconstructions of large-scale phenomena like ENSO, larger more globally diverse networks will produce more informative reconstructions compared to those derived from smaller regions or single sites (Mann, 2002; Lee et al., 2008; von Storch et al., 2009; McGregor et al., 2013). The experiments conducted here support this hypothesis, as the proportions of skilful reconstructions increase as the number of source proxies increase for almost all reconstruction methods and calibration window lengths

C2357

(Figs. 8 and 5)" on page 27, line 12.

Figure 1, Caption: On Line 7 remove correlation coefficients.

The typo has been removed.

Figure 2: As noted above, years or segments. Segments would be much more intuitive. Colorbar has no label.

This has been fixed in the manuscript.

Figure 10, Caption: put tilda in Nino3.4.

This has been done in the manuscript. However, note that Figure 10 has been removed and replaced with another figure.

Figure 10: I find this figure hard to interpret in the context of the results. The blue line is the standard deviation of the correlations for a 30 year window, where the correlations are between a pseudoproxy and the actual Niño3.4 index. I don't see the relevance but perhaps I am misunderstanding what is shown. The red circles are just a very small subset of the plot, is this relationship consistent? In any case, this figure and the discussion would benefit from further clarification.
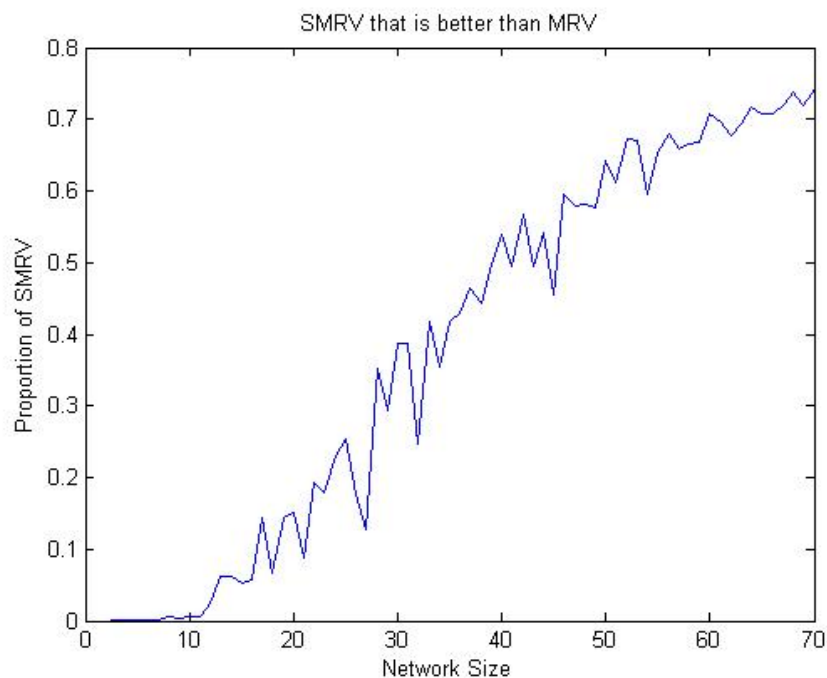
This figure has been removed.

Please also note the supplement to this comment:
http://www.clim-past-discuss.net/11/C2328/2015/cpd-11-C2328-2015-supplement.zip

Interactive comment on Clim. Past Discuss., 11, 3853, 2015.

C2358

Fig. 1. This is known as Figure E1 in this response.