Second round review of

Signal detection in global mean temperatures after "Paris": an uncertainty and sensitivity analysis **by Visser et al.**

1 Comments of Peter Thorne, reviewer #1

The authors have undertaken a substantive redraft. The redraft has served to address many of the points raised by three substantive reviews. In particular, the removal of the future looking section and redrafting around defining pre-industrial are helpful. Redrafting has raised further queries which preclude a recommendation of acceptance without further revisions. I outline major and minor points to be addressed below.

Major points
1. It feels to me like the paper is not overly long and therefore I would query the value of SI over its incorporation into the main text which would presumably make it easier for the reader to understand the piece as a whole.

We agree with this comment and have incorporated the SI into the main text. We found the review on trend methods too long and placed it in the new Appendix A, directly at the end of the paper.

2. Related to the prior point there is an uneven degree of specificity given to the descriptions of the different statistical methods employed. This extends both across the main text and the SI. Why is the equation describing OLS methodology given in main text but the other (more complicated!) methods only described qualitatively? I'm not sure that I or a reader could repeat the analysis given the vague descriptions available. Section 2.1 should formally lay out mathematically the approaches employed in a consistent manner. Alternatively, Section 2.1 should lay out each method qualitatively and point to the SI (if retained) where each method should be laid out mathematically. This is necessary for reproducibility of the analysis and results.

Agreed. We have placed the mathematical description, including formulae in the main text (Section 2.2). See the new equations (1), (2), (3a) and (3b).

3. While sympathetic to the response that you are not trying to exhaustively describe the GMST datasets used I still find the current description inadequate for a reader to properly understand the (in)dependence issues and hence properly interpret your findings. I would suggest that you could add a table that clearly delineates key facets of each dataset that you could reference in place of current text. Such a table may have columnar headings:
a. dataset name and version
b. Land product used
c. SST product used
d. Interpolation method
e. Period of record
f. Key references
g. Website sourced
At a minimum, please specify the versions of the GMST products you have used in the text to enable replication. But, I think a table as suggested above would be far more helpful to the reader.

Agreed. We have added a new table 1 in the format suggested by the reviewer.

4. The editor is free to over-rule this but I, he and others authored Box 2.2 which you are using to support an expert judgement contention at line 56. I have carefully reread Box 2.2 and see no reasonable interpretation of the text therein that can support its use to contend expert judgement was involved in the derivation of either the trend estimate or its uncertainty. Please therefore remove this contention. Box

2.2 uncertainty range quoted pertains to choice of dataset and natural variability only. There is no role that expert judgement has in informing that range which formally must arise solely from the choices of dataset and natural variability. These were the only two factors considered in the calculation of the range being quoted and to suggest otherwise and that somehow expert judgement was factored into these numbers when it was not is a substantial misrepresentation of the process involved.
There is undoubtedly uncertainty in how the trend should be calculated as alluded to in Box 2.2 and expanded upon in the Chapter 2 supplement, but: i) this is not expert judgement; ii) this would inflate the range; iii) this is what your paper is getting at. Its fine to make all these points. What is not okay is to suggest that a range quoted in IPCC which is inferred solely and exclusively from the range of available products and the trend fitting uncertainty is somehow in addition fudged by an expert judgement factor. Sorry if I have labored the point, but it is really important to not imply something factually incorrect as to the IPCC process here which may yield issues down the line.

Okay.. We have removed the text 'and expert judgment'. It is interesting to note that any mathematical (climate) model has a lot of expert judgment in it, made by the designer/programmer. But this is not the topic of this paper.

5. The inclusion of the volcanic activity is currently a half-way house with the bulk of the analysis in the SI but in general poorly referenced from / discussed in the main text. If retained steps should be taken to better integrate the analysis more comprehensively into the text.

Agreed. We have incorporated the time-series results using the AOD indicator in the main text. Please see table 3 of the re-revised manuscript and references to eqs. (3a) and (3b).

6. The corrections outlined in lines 208 and 223-225 are introduced without sufficient justification for a reader. I'm unclear what these are myself. If you are applying corrections here for the dof issue then this should be incorporated into your methods section (see earlier comment ref. methods and reproducibility). If not then you need to be far more explicit what these are for and why you are justified in making them. Presently the basis is at best ad hoc to a reader.

Okay. We improved the text and added two important reference for this correction method (as applied in IPCC, 2013). Please see lines 173-177 of the new text, and lines 193-194.

Minor points

1. Not to belabor the point but the most common acronym in use across the literature for Global Mean Surface Temperatures is GMST and not GMT.

Okay. Changed throughout the whole text.

2. Lines 16-17 conclusive as regards methods

Done.

3. Line 23 leading observational GM(S)T products

Done.

4. Line 25 it is unclear what you mean by both sources of uncertainty are dominated by natural variability

Text improved.

5. Line 47 following the 21st

Done.

6. Three questions lines 89-95 are actually several questions in many cases and need some refinement to be much clearer to the reader. Line 89 to rather than as for? Third bullet could be simplified rather than 3 Qs.

We now made it 4 questions with each question consisting of one sentence. Lines 94-99.

7. Line 104 exhibit not show

Done.

8. Line 131-133 are a non-sequitor. If you mean to include the volcanic loadings in your analysis they need to be better integrated and I would question whether in terms of overall narrative this is the best place to bring this text in.

Agreed. We moved this text to lines 223-224.

9. Lines 143-144 should note that over the ten or so years that these are RCP scenario driven the RCPs themselves do not diverge substantively from one another. Rather, RCP scenario divergence grows later in the Century. The authors may consider whether it would be worth picking just RCP4.5 here for this reason which may serve to simplify the analysis?

Agreed. See also comment of Reviewer #2 as for double counting of simulations. We checked double counting in the 106 simulations. Reviewer #2 is right, there are simulations double. Therefore, we decided to follow the advice here made by Reviewer #1: we choose the 42 simulations with RCP45 with one member per model. This has been changed throughout the text. The new figure 5 now contains 42 increments in stead of 106 increments. The spread and values are almost identical to that made for the 106 simulations! This is not surprising given the double counting. In summary: we have substituted the 106 simulations to 42 simulations and inferences on the results stay almost exactly the same.

10. Line 168 (and same equation in SI) it is entirely unclear why the second equality applies. Why is there 125 in that term and why is it squared? If it is the time delta then it should be 137 and not 125.

Correct. This '125' must come from an older copy-past operation ;-[. The square comes from standard statistics: $\mathrm{var}(ax) = a^2\, \mathrm{var}(x)$ with parameter a any constant. We have added an extra step in equation (1) to make this more clear.

11. I would have thought that LOTI being both low and high estimate in lines 244-246 should be remarked upon. It is to me an unexpected result. I would not expect the same series to arise both the lowest and highest estimate for this term even across methods. Some further analysis and ensuing discussion of this result would be of interest to the reader.

A careful look into table 2 shows that three data-trend combinations have the high estimate of ±0.19 C where LOTI is one of them. Therefore, we changed the text a little bit here. See new lines 275-277. Also changed in the conclusion section.

12. Line 259-260 is unclear what is intended. Please expand for clarity.

Agreed. We made the text more clear on this point.

13. Line 313 assertion and associated precision is not justified by the prior text. Instead consider "This underestimation is uncertain but could amount to up to 0.1C". This would be consistent with the preceding text.

Agreed and done.

14. Lines 323-325 I do not follow the logic of the argument given here. Given that solar forcing is dominated by a cyclical component with a small linear aspect I doubt it would greatly confound any of the chosen techniques. If it does then you should show it to justify the decision or at least refer to one or more prescient references to justify the decision.

There is indeed a cyclic component in long-term reconstructions of the TSI series (SORCE data for example). However, the trend that is reflected in GMSTs, is not the cyclic component. See for example Haustein et al. (2017, their figure 1), Schurer et al. (2017, their figure S3) and many other studies. We have added an extra sentence for this point. Please see lines 350-354. Furthermore, we repeated our view on using non-stationary regressors in multiple regression models at the end of the new Appendix A. There is the danger of interpreting correlations as causal relations if we rely solely on statistical models. Any two series with a rising trend correlate high. Therefore, we prefer models with stationary regressors. Also see table A.1.

15. Lines 352-354. I don't follow the logic here. The RCP scenarios are "reasonable". Also, I don't see an argument for accurate simulations but rather accurate forcings up to 2005. The distinction matters.

We substituted the word 'simulation' into 'forcing', new line 381.

16. Line 355 Table (typo)

Done.

17. Line 394 in addition to instead of rather than

Done.

18. Line 399-401 Inconsistent with you earlier assertion of using the Hawkins et al. number which is smaller.

We changed the text here slightly, new lines 432-434. Hawkins et al. is on various choices for 'pre-industrial'. Here, we name the data for solar and volcanic forcing as explained in section 4.1.

19. Figure 2 requires further explanation in the panel in particular with regard to the middle panels

We find it beyond the scope of this article to explain all details of this graph and particularly the middle panels. We added the reference to Visser (2004) who gives a good and detailed explanation.

20. SI line 6 considerably not considerately

Done.

21. SI line 15 methodologically

Done.

22. SI lines 37-45 needs to make a clear association between the numbers and associated table for this passage to make any sense to the reader.

Agreed. Text highly improved: new lines 811-818.

23. SI line 39 resp?

Removed.

24. SI line 46 delete second in

Done.

25. Table SM2 not referenced anywhere as far as I can tell

This table is now in the main text as table 3. Reference is now.

26. Figure SM2 has only one panel in the submitted version

Agreed. Text removed (see caption of new figure 7).

2 Comments to the anonymous reviewer #2

1. Haustein et al. has just been published on a very similar topic and should be discussed: https://www.nature.com/articles/s41598-017-14828-5

Okay. We have named this new article on a number of places in the new text. New line 74, lines 401-403 and table A.1 as model #27 at the bottom of the table.

2. line 101 - 'corrected for natural variability' - I don't think this is correct?

Good comment. Removed.

3. line 104 - most GCMs are not tuned to the historical period, but to the present climate. Please ensure clarity in this sentence.

Improved: new line 107.

4. Are the 106 GCM simulations all different up to 2005? Some groups used the same simulation for the 1861-2005 period and branched off the RCPs in 2005, reducing the number of independent simulations available. Please check and clarify.

See our comment to Reviewer #1. The reviewer has a good point. There is some double counting in the 106 simulations. We checked that by making a 106 by 106 correlation matrix for the period 1861-2005. Therefore, we followed the hint of Reviewer #1 to use the 42 simulations with RCP4.5 only. See for example the new figure 5. Results differ only as for a few hundreds of a degree.

5. line 193 - please include the whole range of the AR1 parameter from the piControls, rather than discarding some, as indicated in the Author Response.

Done, see new lines 206-209.

6. lines 212-213 - this is not a scientific comment and should not be included.

Removed.

7. line 400 - your volcanic contribution is for a warming. Can you clarify why that is?

We have added in the new line 434 a reference to section 4.1. We suggest an explanation for this in lines 348-349.

8. Please ensure the version numbers for the observational datasets used are included.

Version numbers are given in the new table 1.