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

Peter Thorne

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.
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.
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.
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 re-read 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.

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.
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.

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.
2. Lines 16-17 conclusive as regards methods
3. Line 23 leading observational GM(S)T products
4. Line 25 it is unclear what you mean by both sources of uncertainty are dominated by natural variability
5. Line 47 following the 21st
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.
7. Line 104 exhibit not show
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.
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?

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.
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.
12. Line 259-260 is unclear what is intended. Please expand for clarity.
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.
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.
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.
16. Line 355 Table (typo)
17. Line 394 in addition to instead of rather than
18. Line 399-401 Inconsistent with you earlier assertion of using the Hawkins et al. number which is smaller.
19. Figure 2 requires further explanation in the panel in particular with regard to the middle panels
20. SI line 6 considerably not considerately
21. SI line 15 methodologically
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.
23. SI line 39 resp?
24. SI line 46 delete second in
25. Table SM2 not referenced anywhere as far as I can tell
26. Figure SM2 has only one panel in the submitted version.