An Analysis of the Steig et al (2009) Antarctic Temperature Reconstruction
O'Donnell et al JCLI-3656

Recommendation: Accept with minor revisions.

This paper deconstructs the methodology used by Steig et al (2009) to produce a reconstruction of Antarctic surface temperature since the 1950s, performs a critique of their method, then suggests and implements improved methods for achieving the same aim. The subject is an important one and the authors' careful analysis helps greatly to clarify the issues.

It is, taken as a whole, quite well written and produces very credible results backed up by extensive supplementary material. I found the additional material to be useful and interesting. However, as it took me more than one attempt to understand the paper and I only managed it once I had fully read the supplementary material, I think there is some material which needs to be brought in to the main text to allow the reader to completely follow the arguments from that alone. The Guide to Authors states that the main paper should stand alone and I recommend that the authors think carefully about the relationship between it and the supplementary information.

There are also a couple of things that I think could be discussed, which aren't currently.

Therefore, my comments are aimed largely at improving the paper's flow with just a few remarks on the results and their interpretation.

Please note that I have not read, reviewed or used the code which is provided, as I believe this is outside of my remit. However, I do support the authors' efforts in providing it in this manner to those who wish to explore the issues further.

1. Items to think about:

1.1 The differences between the authors' proposed alternative reconstructions are interesting and significant (according to Figures S15 and S16), yet they are not discussed in the concluding remarks. Given that these two reconstructions are similar, but different, and the authors comment that they are also similar, but different, from the work of Monaghan et al (2008) and Walsh and Chapman (2007), I would think there was an interesting discussion to be had about what this tells us about the uncertainty arising from analysis differences between these four. The authors may believe this is outside the scope of the paper,

but I think it would provide the reader with useful information.

1.2 I would be very interested to know how faithfully the new reconstruction reproduces the considerable trend experienced by the Peninsula over the full period. Reconstructions often contain lower trends than the observations they are based upon. A co-located average comparison, i.e. averaging Peninsula station temperature anomalies and reconstructed temperature anomalies only at the locations of those stations and then comparing the temporal evolution of these two averages, might show something interesting and be considered for inclusion somewhere if it does.

1.3 Having concluded my review I am now wondering about the title of the paper and its organisation. The paper is much more than an analysis of someone else's work and I'd like the authors to rethink the title. I'm also now wondering if it should be a paper in two parts, which would allow the authors to make more of the very good material currently in the Supplementary Material. If it were split into two parts, the first part would be quite short and be an analysis of the Steig et al work, but the second part could be longer and provide a much clearer documentation of the proposed new analysis (I assume here that the authors would recommend the RLS over the E-W reconstruction). Could the authors please consider whether or not they think that reorganisation would be appropriate? It shouldn't be a big job, as the Supplementary Material is well-written. If that is undertaken, I'd be happy to take another look at it.

2. Items to provide greater clarity:

2.1 The authors discuss the erroneous temporal information in the AVHRR data set. I presumed this was mainly a high frequency problem after reading the main paper, but I see from the Supplementary Information that a residual trend is found. Perhaps this could be clarified in the main text?

2.2 I did not understand what the term "on-grid" meant when reading the main paper and I'm still not sure I understand it now. Could the authors please clarify this term on first use?

2.3 Introduction, second paragraph: I suggest maintaining the same style in a) - e), i.e. using "augmentation of" and "estimation of".

2.4 Whilst I understand what the authors mean by "reconstituting the extracted PCs with their corresponding spatial eigenvectors", many readers will not. I like to think of this as building up a temperature

field from a weighted sum of eigenvectors - I think painting that kind of picture might clarify the concept.

2.5 Section 2: "We restrict our replication of ... S09 ... to steps that follow ..." according to the earlier summary of the S09 process, these are all their steps. Without having re-read S09 at this point, I didn't understand the point of this sentence.

2.6 Section 3a: "These factors all highlight a need to calibrate the AVHRR ... (or vice versa)" Not vice versa, I think, given what the aforementioned "factors" are.

2.7 Section 3b: I believe the authors use "x" and "X" in Equations 1 to 6 to describe rather different types of things. I suggest that it is confusing to the reader to use the same letter (albeit in a different case), because they naturally assume that they might be similar quantities. Could the authors think of an alternative to one of them?

2.8 Section 3b: "X=(AlB)" is used at the start of this section, whereas a variant of "X=(A B)" is used in Equation 6. I found the latter more straightforward so, if it is correct to use this, please do in both cases.

2.9 Section 3b: The equations in the Supplementary Information are generally better explained than the equations in the main text and so are easier to follow, e.g.:

- Please define what n and p are when introducing the "nxp matrix of observations"

- Please define U, Lambda and V.

- What is xk in Equation 4?

2.10 Even after re-reading the paper and reading the Supplementary information, I don't understand why the AVHRR PCs would be temporally incomplete. I don't see any evidence of AVHRR data gaps in Figure S3. I can see that there would be missing values in some locations due to cloud, but I wouldn't expect that to preclude calculation of a PC for the field as a whole, unless the data gaps were very numerous. Could the authors please explain that?

2.11 Section 3c: When "suggests the possibility of mutual reinforcement" first appeared in the text, I found it rather opaque. I understand the concept now, but perhaps its meaning could be clarified earlier in the paper?

2.12 Section 3c: "These observations present a major difficulty in ascribing a calibration function to RegEM". Because of the use of words that might be more commonly intended to mean something different, this sentence is hard to understand. I suggest something along the lines of "These features present a major difficulty in using RegEM for calibration." I also found the rest of this paragraph difficult to follow, so could the authors please consider rewriting it?

2.13 Section 3c, last paragraph: suggest change "components" to "PCs" to be consistent.

2.14 Section 3c, last paragraph: "truncation parameter k will be less effective" because if errors are random, they are found in the lower order modes and so removed. Without stating that, it is assumed the reader understands this.

2.15 Section 4: this section was hard to understand, particularly the last sentence in the first paragraph. I recommend that it be rewritten.

2.16 Section 4: Does "i" indicate a station location?

2.17 Section 4: I wondered if because the authors had neglected the contribution of a series to itself, this explained any of the discrepancy found.

2.18 Section 4: "less than half are weighted similarly" this is subjective - can the authors be more specific?

2.19 Section 4: I understand the last sentence after having read the Supplementary Information, but didn't without.

2.20 Section 5: I suggest replacing "boundary conditions" by "shape of the boundary"

2.21 Section 5: Suggest replacing "the statistical authority cited by S09 as their source for determining k" by "North et al (1982)", as I feel the former comes across as a bit argumentative.

2.22 Section 5: I noted here that a picture of the EOFs would be useful to refer to, so perhaps part of S4 could be reproduced in the main paper? I also regretted not having S2 in the main text to aid interpretation of Figures 2 and 3.

2.23 Section 6b: "we address this .. by simply infilling a matrix" How?

2.24 Section 6c: I don't understand the first sentence.

2.25 I wanted to see whether or not the spatial trends shown in the main paper were significant. I think it helps to understand the importance of the differences. I was pleased to find it in the Supplementary Information, perhaps that information could be added to Figures 4 and 5?

2.26 Why are the diagonal elements of Table 1 not 1?

2.27 I wondered if the verification results were sensitive to the choice of which 28 stations were in the subset?

2.28 It might be useful to state in the caption for Table 6 that the full reconstruction is verified against 24 withheld stations.

2.29 Figure 1: could the black and blue time series be plotted on different axes? It looks to me like there is a constant offset in the blue line, but because of the scale, I can't see whether or not there is and what size it might be.

2.30 Figure 2: blue and black are indistinguishable here. Please state in the captions that the locations are the locations of the stations.

2.31 Figure 3: does this show different locations to Figure 2?

3. Comments related to Supplementary Material

3.1 Section S2d: is the difference in trend of 0.08 C/decade statistically significant too? It is hard to detect this trend in the plot.

3.2 Section S2d: the authors discuss the "dramatic change in NOAA-11 ICT variability and the mid-2000 jump in NOAA-14 variability" however, I don't see a significant effect of the latter in the bottom panel of Figure S3, so I'm not sure it's worth mentioning.

3.3 Section S2d: "NOAA-9 demonstrates" should be "NOAA-7 and NOAA-9 demonstrate"

3.4 Section S2d: for completeness, the authors might recognise here that the station data is also likely to contain some errors (I note this is done later, so a reference forward might be made).

3.5 Section S3: "Due to the vastly larger number of data points in the Peninsula" I think this is overstating it. The number is much larger,

but "vastly" is over-doing it, I think.

3.6 Section S3: I understood here why the authors had used the West Antarctic stations alone in one of the verification statistics in the main paper, but I think that needs to be explained better there, as it wasn't clear to me when reading that stand-alone.

3.7 Section S3: the issue with the Peninsula stations swamping the West Antarctic stations discussed here is not explained clearly in the main paper, perhaps a few more words could be added on this?

3.8 Section S5: I got to the end of the last paragraph and wondered eagerly if the spatial structures were incompatible. Later on I found the answer, but it might be helpful to look forward the outcome here.

3.9 Caption for Figure S7: Is the sentence "For the AVHRR data ..." essentially repeating the previous sentence?

3.10 What do italics denote in Table S3?

3.11 Section S6, second sentence following equation S1: should "estimation-maximization" be "expectation-maximization"?

3.12 Figure S10: why is there a separation in results after k=5?

3.13 Section S9b: I am worried by the fact that the authors have selected c=0.1 as the optimum parameter, yet it is the lowest value tried and there is no evidence from Figure S14 that the verification is converging at c=0.1. Why not test lower values, given that this is the case? I am not convinced that c=0.1 is optimum here.

3.14 Section S9c: it is very interesting that the covariance results are more sensitive to kgnd than the correlation results and this is shown nicely by Figures S18 and S20. I would like to understand why. Any ideas?

3.15 Section S9c, last para: Please rephrase the sentence starting "Most importantly, trends ..." because I don't think the authors intended to imply that the trends for all locations are similar

3.16 Section S9d, last para: I didn't understand why the truncation parameters in the RLS and E-W reconstructions should not provide filtering.

3.17 There is no text to go with S15 and S16. I think they are worth discussing, particularly in the context of my first comment.

3.18 I liked figures S21-S26. I felt they should be more prominent.

3.19 using "R squared or CE" in the column headings in Table S5 is confusing. It suggests they are interchangeable, yet footnote d in the Table suggests they are not.

3.20 Figure S27 is very helpful.

4. Typographical errors:

4.1 Should there be a subscript j on the A tilde in Equation 6?

4.2 Section 5: "influenced based visual similarity" should be "influenced by visual similarity"

4.3 Section 7c, last para: "Table 6" should be "Table 7"

4.4 Weddell Sea is mis-spelled more than once.

4.5 Section S4: "TIR" here "IR" should be subscript.