Short Form Response to Review A

Due to the length of Review A, we felt it would be convenient to provide a paraphrased short form response – which summarizes the major issues in 2-column format – and a separate long form response, which addresses each of the specific issues in detail. The short form response paraphrases the five primary problems noted by the reviewer and references the applicable detailed response paragraphs. The detailed response specifically addresses all problems noted by the reviewer, in the same order as the review. Brief descriptions of manuscript changes are included after the short form.

Problem 1	
a) RO10 discusses only one of several reconstructions presented by S09.	a) S09 present the primary T_{IR} (along with a detrended variant) and AWS reconstructions in the main text. They give little credence to the detrended T_{IR} results, and present a logical reason why the results would be suspect. We therefore do not address this variant. We address the AWS reconstruction implicitly, as all of our reconstructions begin with an AWS-style reconstruction. S09 also present a PCA based reconstruction in the SI, but present insufficient information to make a quantative comparison possible. However, our paper addresses the choice of its primary parameter (¶1.A-B).
b) The RO10 reconstructions are not directly comparable to the S09 main reconstruction and should be compared to the detrended reconstruction.	b) The RO10 reconstructions were <u>not</u> performed using detrended AVHRR data as asserted by the reviewer. For the E-W reconstructions, the AVHRR data was used as-is. For RLS, the reconstructions were performed following a row-centering step. The row-centering step does not affect either the magnitude or distribution of trends, as the RLS reconstructions do not use any temporal AVHRR information. The RO10 reconstructions are directly comparable to S09's primary reconstruction (¶1.C – I).

c) The regional averages and statistical significance of those averages reported in RO10 is misleading because RO10 did not adjust regional boundaries based on reconstruction trends, leading to their regional averages including both warming and cooling areas.	c) The regional averages and statistical significance of those averages as reported in RO10 are accurate. Moreover, matching regional definitions *post hoc* as the reviewer suggests invalidates significance calculations (¶1.J – O).
d) The seasonal patterns of change are more similar to S09 than what is represented in the RO10 text.	d) None of the seasonal comparisons between S09 and RO10 made by the reviewer to support his argument are accurate (¶1.P – U).
Problem 2	
a) The RO10 text does not describe the relative importance of the modifications proposed.	a) This is a valid concern and the text has been amended accordingly (¶2.V).
b) The primary difference in spatial patterns between S09 and RO10 is due to the new truncation parameters and not due to calibration or improper infilling.	b) While the largest contributor to the difference in spatial patterns is the additional AVHRR eigenvectors, this dependence is less than the reviewer implies. The RO10 pattern of trends can be replicated with as few as 5 AVHRR eigenvectors, but to do so requires correcting the calibration and infilling issues with S09. Simply including the optimal number of eigenvectors *without* correcting the calibration and infilling issues does not reproduce the RO10 pattern of trends. Both of these, when corrected, have a significant impact on the spatial distribution of trends (¶2.H – U).
c) The primary difference in overall trends between RO10 and S09 is due to detrending the AVHRR data and extracting the modeled PCs.	c) We do not detrend the AVHRR data, so the contribution from detrending is null. The contribution to overall difference in trends by {correctly} using the modeled PCs accounts is only 0.02 deg C/decade – or approximately 33% of the difference between S09 and RO10 (¶2.A – G).

d) RO10 do not show that any transfer of trends from the Peninsula to the continent have occurred.	d) If transfer of trends did not occur, then use of the modeled PCs would account for the entire difference in trend magnitude between S09 and RO10 (see detailed response). However, the actual transfer of trends due to an insufficient number of AVHRR PCs and incompatible AVHRR and ground station spatial structure accounts for approximately 2/3 of the difference in magnitude between RO10 and S09. This effect is also corroborated by large-scale contamination of seasonal behavior in S09 by the Peninsula (¶1.R, 2.W – 2AC).
Problem 3	
a) RO10 refer to the incorrect source for determining truncation parameters in S09.	a) We have corrected the text. The remainder of the discussion concerning the validity of this procedure, however, is still applicable (¶3.B - C).
b) RO10 incorrectly state that visual similarity was used as a criterion for eigenvector retention.	b) We have corrected the text (¶3.D).
c) RO10 mischaracterize the regression model of the infilling algorithm used by S09.	c) The regression model stated by S09 would require infilling actual values using missing ones. Our characterization is accurate and can be confirmed both by the author of the algorithm and by inspection of the algorithm itself (¶3.G – I).
d) RO10 do not show that allowing the AVHRR data to influence the infilling of ground stations is improper, as this depends on the assumed source of error	d) As stated in our text, this is a calibration issue and is independent of the source of error. S09's method implicitly assumes that AVHRR data and ground data are interchangeable quantities. Since they are not (as shown in our text), and since RegEM cannot perform the required calibration, allowing the PCs to influence the ground station prediction is invalid regardless of the results (¶3.J, ¶3.N).
e) The use of the modeled PCs in RO10 depends on the assumed source of error.	e) The use of the modeled PCs is a necessary consequence of a proper calibration and an entirely separate issue from the assumed source of error (¶3.K – N).

f) The discussion in RO10 concerning eigenvector patterns and mathematical artifacts is speculative and inadequate.

f) This discussion is not crucial to our conclusions. As the reviewer has clarified our misconceptions concerning eigenvector retention criteria in S09, the relevance of this is greatly reduced. We have amended the text and SI to remove this discussion (¶3.O).

Problem 4

a) RO10 verification statistics are computed to infilled values, leading to inappropriate selections of k_{gnd}, particularly in West Antarctica.

a) This assertion is inaccurate. All verification statistics were calculated by comparing reconstructed estimates to original, withheld station data (¶4.A – B).

b) The suspect values of k_{gnd} are the only ones that fail to result in statistically significant warming in West Antarctica.

b) As stated in both S9.a and S9.b, the panels showing the k_{gnd} sensitivity tests are done using settings that result in maximum trends, without optimizing the other parameters for those particular choices of k_{gnd}. They also do not reflect the set of reconstructions with the next highest verification statistics. When the other parameters are optimized for those values of k_{gnd}, the overall trends and trends in West Antarctica are lower and comparable to the optimum settings for k_{gnd}. The reviewer does not note that, in Table 6, it is clear that the higher trends for the other values of k_{gnd} result in degraded verification statistics – both in West Antarctica and across the continent. Testing by infilling ground stations using ridge regression with the ridge parameter for each time step determined via generalized cross validation yields validation statistics, spatial patterns and West Antarctic trends (~0.11) comparable to our main results, and performing RLS reconstructions after offsetting stations based on periods of mutual overlap (i.e., no infilling) yields validation statistics, spatial patterns and West Antarctic trends (~0.10) also comparable to our main results (¶4.C – U, 4.Y – AE). These additional tests – which form the basis for a future work – have now been incorporated into the main text.

c) S09's method is less sensitive to the quality of ground station information because it utilizes AVHRR information during infilling.	c) This claim is not correct. Use of the AVHRR data does not make the ground station infilling less sensitive to the quality of data (¶4.W – X).
d) Monaghan reports warming at Byrd using an updated reconstruction method that includes nearby AWS stations and is in excellent agreement with the AVHRR data, which further indicates the RO10 reconstructions are suspect.	d) Unlike Monaghan *et al.* (2008) and contrary to the implications of the reviewer, our reconstructions *do* show warming at Byrd and *do* include nearby AWS stations. Additionally, our reconstructions match the station data at Byrd (and across the continent) significantly *better* than the either the raw AVHRR data or S09, even when the station in question has been <u>entirely withheld</u> from our reconstructions (¶4.Y – AE).
e) RO10's reconstructions are less accurate in West Antarctica than S09's, and in virtually all their reconstructions, RO10 find the same results for significant warming in the Ross region, which is nearly identical to S09.	e) Both the RLS and E-W reconstructions outperform the S09 reconstruction (and the raw AVHRR data) in West Antarctica as across the entire continent. The reviewer incorrectly cites our sensitivity study in the "virtually all" statement without noting the degradation in verification statistics for those variants. Even were the reviewer correctly to have cited our sensitivity study, the pattern in West Antarctica (minimum warming on Ross) is substantially different from S09 (maximum warming on Ross). If the reviewer meant magnitude and not pattern, the Ross warming in our results is outside the 95% CIs for the S09 results. This is, indeed, substantially different (¶4.W –AD).
f) The West Antarctic trends of RO10's reconstructions should be compared to the S09 detrended variant. If this is done, certain variants of the RO10 reconstructions show warming between 25% and 150% of S09.	f) The reviewer's belief is incorrect. Our reconstructions do not utilize detrended AVHRR data (RLS has a row-centering step that does not affect the resulting trends) and are directly comparable to S09's main reconstruction. Additionally, the reviewer (here and elsewhere) cites results from sensitivity tests without fully understanding the context provided by the SI or examining the degradation in verification statistics for these variants. We have provided clarification of this in the text and SI (¶1.C – I, ¶4.AF).

Problem 5

a) RO10 attempt to use their reconstructions to cast doubt on GCM results.	a) Our text makes no statement on the broader context of GCM results. The brief note in our text is specific to the *single* model result shown by S09 in their own text. However, we agree that providing additional context is valuable and have amended the text accordingly (¶5.B, ¶5.E – F).
b) RO10 do not mention that their results are consistent with S09's sea ice hypothesis.	b) Our paper is not on sea ice. We make no statement one way or the other. The request for us to do so is extratopical (¶5.C, 5G – I).
c) RO10's results are more consistent with coupled GCMs included in AR4.	c) While we had no statements on this in the original text, we agree that some context is useful and have amended the text (¶5.E – F).
d) RO10 do not mention how well their results fit Thompson and Solomon's hypotheses.	d) Our paper is not concerned with these hypotheses. Our paper is concerned with correcting the deficiencies in the S09 method. If future authors wish to make this comparison, we would be delighted. The request by the reviewer for us to analyze/propose/compare possible physical causes with the magnitude and pattern of temperature change provided by our reconstructions is extratopical (¶5.H – I).

As the "Additional Technical and Editorial Comments" and "Concluding Remarks and Recommendations" are largely summarized forms of the above, these are addressed in the detailed response.

AMENDMENTS TO THE TEXT BASED ON THIS REVIEW:

1. Abstract:
    a. Clarified that we find statistically significant warming in West Antarctica adjacent to the Peninsula and that we conclude the regional average is statistically significant
    b. Added that the overall trend for both the continent and East Antarctica is positive

2. Section 1:
    a. Clarified how criticisms relate to the 3 different reconstructions presented by S09 (T_{IR}, AWS, and standard PCA)

3. Section 2:
    a. Moved Figure 1 to the Supporting Information

4. Section 3:
    a. Clarified the meaning of A, b, U, lambda, and x_k
    b. Separated the calibration issues of 1) allowing the PCs to influence the ground station data and 2) use of the modeled PCs to facilitate explaining the relative contribution of each in the Results

5. Section 4:
    a. Clarified the difference between the RegEM coefficients and the AVHRR spatial eigenvectors

6. Section 5:
    a. Properly referenced the S09 truncation procedure as Mann *et al.* (2007)
    b. Removed reference to *RealClimate*
    c. Added a discussion on the limitations of the procedure described in Mann *et al.* (2007)
    d. Removed discussion of physical meaning and nodal (Chladni) patterns
    e. Updated reference list to reflect the above changes

7. Section 6:
    a. Clarified that use of the modeled PCs is a calibration concern
    b. Moved information from the SI to the main text to clarify that the ground-only reconstructions were for screening purposes
    c. Removed the back-reference to Section 5 concerning physical meaning and visual inspection of the AVHRR spatial eigenvectors

8. Section 7:
    a. Clarified the discussion concerning validation to indicate which portions were screening and which portions were used to select parameters
    b. Clarified that, unlike Monaghan *et al.* (2008), we find warming at Byrd station
    c. Added a table of seasonal trends and clarified the differences between S09 and RO10
    d. Added results from ridge regression infilling and offset RLS without infilling, (concerning k_{gnd} sensitivity)
    e. Added statement comparing CEs for our reconstructions to the average explained variance for the raw AVHRR data and S09
    f. Added a statement that the combined infilling of PCs and ground stations did not significantly affect S09's reconstructions
    g. Added discussion and a table showing relative contribution for the proposed modifications

9. Section 8:
    a. Removed reference to the ModelE simulation

10. Supporting Information:
    a. Added Figure 1 from the main text
    b. Removed discussion and references to Chladni patterns
    c. Clarified that ground station cross validation was for screening purposes only
    d. Clarified that the random withholding tests were informational only and were not used to determine any truncation parameters
    e. Clarified that the k_{gnd} sensitivity panels were not optimized for those particular values of k_{gnd}

Detailed Response to Review A

While extensive, Review A contains some key misunderstandings of our methods that lead the reviewer to a number of erroneous conclusions and incorrect criticisms of our study. The reviewer also makes several important factual errors. These will be discussed in detail. As several of these misunderstandings are a result of unclear wording in our text, we agree that minor revisions of the main text and Supporting Information are helpful. For clarity, statements extracted from the review will be *italicized* and enclosed in brackets {}. Statements were quoted from the review in the same order they appear, and comprise all material statements from the review.

{*Problem 1) Discussion of the differences between the RO10 and S09 reconstructions is misleading, and in some cases simply wrong.***}**

{*First, S09 presented several reconstructions, but RO10 discusses only one of them.*}

1.A S09 present only two reconstructions in the main text. One is the T_{IR} reconstruction (including a detrended variant), and the other is the AWS reconstruction. The AWS reconstruction is a reconstruction that infills the ground station data without including satellite data. Since both the RLS and E-W reconstructions in our text start with essentially an AWS reconstruction (i.e., they use a very similar station set as the S09 AWS recon and an identical infilling algorithm), we address the S09 AWS reconstruction implicitly, and statistically demonstrate that the S09 choice of $k = 3$ for the truncation parameter is suboptimal. We have clarified this point in the text.

1.B While S09 do present a third reconstruction version using conventional principal component analysis, it is mentioned as a byline in the Methods and is shown only in the SI. It is not accompanied by verification statistics, a list of predictors, or any tabulated trends. Without tabulated trends or verification statistics, quantitative comparison to this reconstruction is impossible. Since the reconstruction utilizes the same number of AVHRR PCs as the T_{IR} (which we show to be suboptimal), our paper addresses the primary parameter choice for this reconstruction as well. Precise replication of this reconstruction is redundant and (due to the absence of key information) not possible.

{*RO10 argue that the magnitude of the trends in the AVHRR data are suspect, because they are much larger than in the ground-based observations from manned weather stations. This is a valid concern, but it is directly addressed in S09. Specifically, S09 included a reconstruction based on a fully detrended AVHRR data, which results (of course) in smaller trends than in the main reconstruction, a fact completely ignored by RO10. RO10 also detrend the AVHRR data in their reconstruction, but without citing S09 for the origin of this idea. Comparison of the main S09 reconstruction (AHVRR not detrended) with the RO10 reconstructions is then said to show trends that are only 25% to 50% of those in S09. But this is apples and oranges. An apples to apples comparison shows that the RO10 trends are more than 35% to more than 80% those of the S09 AVHRR-detrended results. For example, S09 detrended gives 0.13/decade for West*

Antarctica, 0.06 for East Antarctica; RO10 reports 0.05 to 0.1 for West Antarctica and 0.05 for East Antarctica.}

1.C This comment reflects a misunderstanding of our methods. To avoid the same misunderstanding by readers, we have amended the text to make this clear.

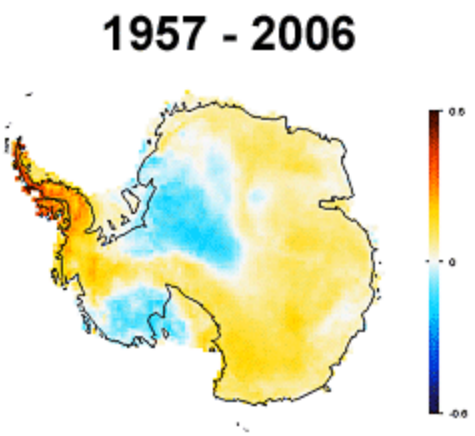
1.D The nature of the detrending in S09 – and the resulting impact on the reconstruction – is not applicable to our methods. In S09, because the first three AVHRR PCs are used as is, linearly detrending the AVHRR data results in a near-zero trend for the 1982 – 2006 period and significantly affects the 1957 – 2006 trends. This has no physical justification. We did not criticize this reconstruction variant because S09 gave little credence to the detrended results, emphasizing that "in general, detrending of predictand data lowers the quality of the reconstructions…" With respect to the S09 reconstruction, we agree.

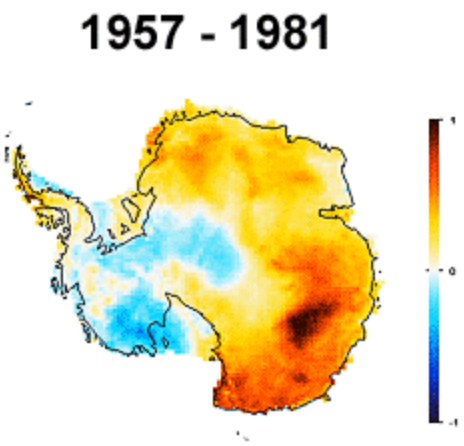
1.E In RO10, contrary to the reviewer's belief, no detrending of the AVHRR data was performed for the eigenvector-weighted reconstructions. Additionally, the only pre-processing performed for the RLS reconstructions is limited to row-centering the entire data matrix by the mean temperature for each time step. This no more removes spatial covariance information than converting raw temperatures to anomalies removes temporal evolution information.

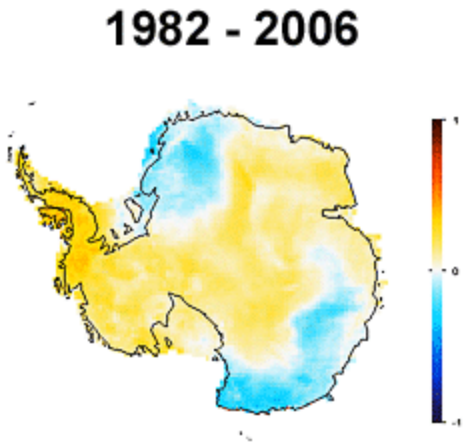
1.F As explained in the SI, this step provides an elegant means to remove the effects of splicing errors and measurement drift (which are inhomogeneities that affect the entire grid simultaneously) from the eigenvector determination without affecting the relationship between points on the grid. This results in more stable results – particularly when *subperiods* of the AVHRR data are used to perform reconstructions. It minimizes changes in the V matrix between subperiods following SVD of the AVHRR data.

1.G Because the RLS method does not utilize the AVHRR PCs, the effect on reconstruction trends is nil. This is true regardless of other parameter settings. Reconstruction trends and spatial patterns of change with this step removed are shown below:

	West Antarctica	Peninsula	East Antarctica	Continent
Row-centering	0.05 +/- 0.08	0.29 +/- 0.10	0.05 +/- 0.09	0.06 +/- 0.07
No row-centering	0.05 +/- 0.08	0.29 +/- 0.11	0.05 +/- 0.09	0.06 +/- 0.07







1.H Given that the eigenvector-weighted reconstructions utilize the AVHRR data *as is* and the row-centering step in RLS has no effect on reconstruction trends, our results are directly comparable to S09's main results.

1.I Finally, it is clearly inappropriate to cite S09 (or any of the many earlier papers on detrending and removal of low-frequency information for calibration purposes) as we did not perform the implied operation.

{ Second, RO10 choose to report average trends only for (roughly) the same geographic areas as reported in S09; that is, East Antarctica, West Antarctica and the Peninsula. While such comparisons should certainly be made, the choice to separate these regions in S09 was based not only on geographic delineations but also on the spatial pattern of the reconstructed trends. In RO10, a somewhat different pattern is obtained, particularly in West Antarctica: areas to the west roughly of Byrd Station (that is, areas towards the Antarctic Peninsua) are found to be warming, whereas areas to the east (towards the Ross Sea) are found to be either neutral or cooling. By averaging these together, the net trend is of course smaller than in S09. But this hides the fact that when these areas are considered separately, RO10 actually find greater warming in a large portion of West

*Antarctica. This careless averaging of opposing trends leads to the remarkable statement in RO10 that "the Peninsula is the only region that consistently demonstrates a statistically significant trend." **This is patently false.** As shown clearly in RO10's Supplementary Information, Figure S15, there are statistically significant positive annual mean temperature trends for virtually all of West Antarctica west of Byrd Station, as well as for an extensive area in East Antarctica (North Victoria Land), including the eastern Ross Sea region.*

*Note that delineating areas of West Antarctic east and west of Byrd Station, as I suggest should be done, is not arbitrary. Byrd lies near the main West Antarctic ice divide, which represents a clear meteorological boundary (for example, accumulation rates are much higher on the west versus the east side of the divide). Furthermore, it is of fundamental interest to glaciologists whether the area of Pine Island Bay – where rapid ice discharge is occurring – is warming at the surface. Not only does RO10 confirm the results in S09, but RO10 actually shows somewhat greater warming in this area! **This is one of the most important findings in S09 and it is fully supported by RO10, yet scarcely gets a mention in the paper.**}*

1.J This comment is puzzling. The reviewer seems to be implying that the regions chosen should be based on the resulting reconstruction trends rather than on the standard basis of natural geophysical and/or geographical delineations, and additionally implies that we should have chosen different regions than the paper we are criticizing. This is illogical. We chose geographical delineations because we are analyzing a paper that also (approximately) chose geographical delineations and used region names that have specific geographical meaning. We also note that if we were to choose regions based on the reconstructed trends, using the reviewer's logic, we could easily argue for including part of West Antarctica in the Peninsula. Since the warming in our reconstructions is clearly an extension of the Peninsula warming, this would be more logical than carving up West Antarctica. The resulting West Antarctic trends would then be *very much* below that reported by S09.

1.K In addition, the reviewer has misconstrued the statement "the Peninsula is the only region that consistently demonstrates a statistically significant trend" as implying that the other geographical regions – East and West Antarctica – are entirely devoid of grid cells that show statistically significant trends. We consider this interpretation to be unusual, and believe that most readers would grasp that the statement applies to regional averages. However, to avoid confusion, we have clarified the abstract and the main text to prevent this confusion.

1.L With respect to the reviewer's statement that the regional boundaries in S09 were chosen based on the resulting reconstruction trends as well as geographical delineations, this explanation is not present in the main S09 text, the SI, or any personal communication between the S09 authors and the present authors. If the reviewer's claim is both true and germane to the S09 analysis, then the S09 text is incomplete. It is unreasonable to suggest that the present authors should use regional boundaries that differ from traditional geographic boundaries based on reasoning that does not appear in the S09 text. As a side

note, the S09 boundaries do not appear to be a good match to the resulting trends regardless (Fig. S1).

1.M The reviewer goes on to state that he would propose a different delineation of West Antarctica based on a geophysical criterion in connection with the Pine Island Glacier. We consider a criterion based on factors important in glaciology not to be relevant to the present work – especially given that the paper we are criticizing did not use this criterion. The purpose of the paper is to deconstruct the S09 method, demonstrate how it can be improved, and show how the improvements change the results. Discussions of glaciers (and, hence, delineating areas based on glaciology) is clearly extratopical. Expressing the "importance" of the differences between our reconstructions and S09 in terms of glaciers (or sea ice, or precipitation, or any other physical process) is likewise extratopical.

1.N As noted in our abstract and throughout our paper, we are concerned with *what changes* when the deficiencies in S09 are corrected – not how important those changes are to various hypotheses or physical processes. The reviewer's insistence (here and elsewhere in the review) that we express the differences between S09 and our work in terms of how important those changes are to glaciologists, S09's sea ice forcing hypothesis, the theories of Thompson and Solomon, or any other related topic is not reasonable. We show what changes. We leave interpretation of those changes to others.

1.O Finally, the reviewer seems to imply that stating an average for a region is illegitimate if the region contains both positive and negative trends. This is an unusual and suspect interpretation of "area average", as matching the area definitions to one's results is a *post hoc* decision that would invalidate the regional statistical tests of significance. The area average does, indeed, show mixed results for trend significance in West Antarctica. However, as our sensitivity testing indicates the trend is likely to be significant, we have amended the abstract to be clear.

{Third, the comparison of seasonal reconstructions in S09 and RO10 is highly misleading. For example, the abstract says that "while the seasonal patterns of change for the Peninsula region are similar, there are substantial differences in the patterns for West Antarctica and the pole." Nowhere is it mentioned that both reconstructions show a) maximum warming in Winter and Spring in most areas; b) minimum winter in Summer; c) maximum warming in Fall; and warming of virtually all areas in Spring. All of these points are made in S09 and are important to the climatological interpretations of the results in that paper. All of these findings are also supported by RO10 in their preferred reconstruction, yet this is nowhere given more than a passing mention. RO10 also emphasize the annual mean cooling in the Ross Sea region in their reconstruction, but fail to mention that in their results, the trend is positive in this region in Spring it is comparable to the rate of warming on the Peninsula. This is a critical point in S09 due relationship with sea ice, and because this appears to contradict earlier results that attributed recent Antarctica temperature changes to changes in ozone.}

1.P The reviewer asserts agreement of our reconstructions with seasonal patterns in S09. Actual seasonal trends for our reconstructions are listed below, with the two seasons demonstrating the most warming highlighted in red (or three seasons, in the case of a tie):

Method	Region	Winter	Spring	Summer	Fall
RLS	West	-0.08 +/- 0.05	**0.15 +/- 0.04**	**0.08 +/- 0.03**	0.05 +/- 0.04
	Peninsula	**0.40 +/- 0.07**	0.17 +/- 0.04	0.18 +/- 0.02	**0.40 +/- 0.04**
	East	**0.09 +/- 0.05**	**0.07 +/- 0.03**	**0.07 +/- 0.03**	-0.02 +/- 0.01
	Continent	0.07 +/- 0.04	**0.09 +/- 0.03**	**0.08 +/- 0.03**	0.01 +/- 0.04
E-W	West	-0.01 +/- 0.03	**0.11 +/- 0.03**	**0.06 +/- 0.02**	0.02 +/- 0.03
	Peninsula	**0.41 +/- 0.05**	0.19 +/- 0.03	0.20 +/- 0.02	**0.34 +/- 0.03**
	East	0.04 +/- 0.05	**0.08 +/- 0.03**	**0.06 +/- 0.03**	-0.01 +/- 0.01
	Continent	0.04 +/- 0.04	**0.09 +/- 0.04**	**0.07 +/- 0.02**	0.01 +/- 0.04

1.Q The reviewer's claims (obvious typographical error corrected) and the actual results compare as follows:

a) Maximum warming in Winter and Spring for most areas. Only East Antarctica in RLS shows this pattern, but only barely, as the Spring trend is only 0.002 higher than the Summer trend.

b) Minimum [warming] in Summer: All but the Peninsula in our reconstructions have Summer as one of the two most rapidly warming seasons.

c) Maximum warming in the Fall: This is untrue for any region in our reconstructions, though the Peninsula trends for RLS in Winter/Fall are nearly indistinguishable.

d) Warming of all areas in the Spring: We agree that all areas warm in the Spring, but also note that all areas warm with comparable magnitudes in the Summer. This is at odds with S09, who show a marked difference in Spring and Summer warming.

e) West Antarctica warming in Spring is comparable to the Peninsula: This is true of RLS, but not of the E-W reconstruction, which shows Peninsula warming in the Spring to be nearly double that of West Antarctica.

1.R In short, our actual results disagree substantially with S09's, the reviewer's bolded assertions notwithstanding. Indeed, the table above provides additional support that S09's seasonal trends are contaminated by the Peninsula, as the S09 seasonal trends largely follow the behavior exhibited by the Peninsula (we also note this in our SI).

{*S09 write: The simulations [of Gillett and others] show warming in austral summer and autumn, restricted to the peninsula, whereas in our reconstruction the greatest warming is in winter and spring.*
This is a major point made in S09 and is fully supported by RO10.}

1.S The reviewer asserts that our reconstructions fully support this; however, we show a *minimum* of Peninsula warming in the Spring and a *maximum* in the Fall – which is precisely the opposite behavior in those seasons as compared to S09. We do, however, show maximum warming in the Peninsula in the Winter, but unlike S09, this is *restricted to the Peninsula*.

1.T We agree with the reviewer that seasonal trends should be set out more clearly in the text, and we have amended the text to include the above table and additional discussion of the differences.

1.U Lastly, as before, we make no attempt to evaluate whether a particular difference between S09 and RO10 is important to any physical process or theory.

{*In summary, RO10 generally compare their minimum temperature trends with the maximum trends calculated by S09. This is extremely misleading. A revised version of RO10 will need to make clear that the differences are actually much smaller than currently implied, and will need to detail the validation of the major seasonal patterns. [I suspect, furthermore, that when the uncertainties in each method are taken into account, the trends in an apples-to-apples comparisons probably overlap completely. I do not mean here the trend uncertainties (i.e. comparison of the trends with the variance) but the fraction of unexplained variance based on the r^2 verification statistics, e.g. as described in the published Corrigendum article accompanying S09.]*}

1.X The summary conclusion in this paragraph has been shown to be inaccurate, as we do not detrend the AVHRR data. Our results are directly comparable to S09's main results. The differences in seasonal patterns are quite significant, and the text has been amended to more clearly delineate them as indicated in ¶1.T.

1.Y With respect to the fraction of unexplained variance, S09 calculated verification statistics to the AVHRR data, which by their own definition is the response variable – not the explanatory variable. Though we do not specifically mention this issue in our paper, our calibration discussion deals with it implicitly. Given that the explanatory variables are the ground stations, the proper verification statistics would have been to the ground stations. The statistics S09 calculate show the difference between the calibrated model (which S09 discard) and the raw satellite data in separate early and late subperiods of the AVHRR data. It makes no statement on the accuracy of the reconstruction with respect to the explanatory variables; it merely highlights that the AVHRR data and ground data show different temporal behavior. Discarding the model PCs is one of our major criticisms of the S09 method and is dealt with specifically in ¶3.J – M.

1.Z Since we do not use the AVHRR temporal data at all in RLS and use a properly calibrated model in the E-W reconstructions, this step is not applicable to our

reconstructions. Though it would provide a spatial map of error estimates, this would be an estimate of the temporal error between the ground data and the satellite data assuming the satellite spatial structure is accurate. It is not an estimate of the uncertainty in the reconstruction. Because the explanatory variables are not present throughout the grid, only point estimates of reconstruction error can be directly obtained. This information is available in Tables S5 and S6.

{Problem 2) Several reasons for differences between S09 and RO10 are presented, but no clarity is provided as to which of these dominates. This leads the reader to the impression that multiple assumptions and in/or errors in S09 each has a large impact on the results, whereas the differences are overwhelmingly the result of just one assumption (the number of retained EOF spatial patterns).

RO10 proposes several improvements to the S09 methodology regarding a) the selection of the number of principal components retained and the related regularization parameter, k; b) inhomegeneties in the AVHRR satellite data; c) the choice to infill missing values in the ground station data using information from the AVHRR data; and d) whether to replace the raw PCs of the AVHRR data (1982-2006) with modeled PCs derived from a linear combination of the ground station data.}

{Although I largely agree with the methodological improvements suggested, RO10 do not articulate their relative importance, leaving the reader with the mistaken impression that S09 makes a series of egregious errors. Yet it is easily shown that differences in the spatial patterns of warming and cooling between the results of S09 and RO10 are overwhelmingly due to (a), and that the only significant impact of (d) is on the magnitude (not the pattern) of temperature change. There is virtually no impact of either (b) or (c) on the results. (Indeed, RO10 do not actually utilize (b), other than a simple detrending of the AVHRR data, which was also done in S09 as discussed above). The impact of (d) is important, but this is not as objective a decision as RO10 claim, as I will discuss in the context of Problem 3, below.}

2.A Before answering this comment some clarification is required. Our abstract states that the three primary deficiencies of the S09 method are (reordering to be more consistent with the reviewer's order):

(i) Insufficient number of satellite PCs
(ii) Improper determination of spatial structure during infilling
(iii) Improper calibration

2.B The methodological improvements motivated by these issues are:

Mod 1: Determine all truncation parameters via cross-validation

- Motivated by (i)
- Refers to the reviewer's (a) above
- Assumed by the reviewer to be the primary driver for spatial pattern differences

Mod 2: Weight the stations by the AVHRR eigenvectors during regression

- Motivated by (ii)
- We believe this to be the intent of the reviewer's (b), though the reviewer refers to a detrending step that does not exist (Problem 1)
- Assumed by the reviewer to have negligible impact

Mod 3: Use the modeled PCs for all times

- Motivated by (iii)
- Refers to the reviewer's (d) above
- Assumed by the reviewer to be the primary driver for overall trend magnitude differences

Mod 4: Infill ground stations without including AVHRR data

- Motivated by (ii) & (iii)
- Refers to the reviewer's (c) above
- Correctly assumed by the reviewer to have negligible impact

2.C Modification 4 is discussed in more detail in ¶3.J – M. We agree that this correction to the calibration procedure has a negligible impact on the reconstruction (however, this is not true of the other calibration correction, Mod 3). The text has been amended to make this clear. We will defer further discussion of Mod 4 until Problem 3.

2.D All of these methodological improvements are applicable to the E-W reconstructions. For RLS, modifications 2 and 3 take an alternate form, wherein the AVHRR spatial eigenvectors are used directly and the PCs are discarded. Separating the effects of modifications 2 and 3 is therefore not possible in RLS, and for that reason, we will address the reviewer's concerns using the E-W reconstructions.

2.E We first address the reviewer's claim that magnitude (not spatial pattern) differences are primarily the result Mod 3.

2.F The result of performing S09-style reconstructions with only Mod 1, only Mod 3, and a combination of Mods 1 & 3 is shown below and compared to the RO10 E-W reconstruction:

	Original S09	Mod 1 only	Mod 3 only	Mods 1 & 3	RO10 E-W (Mods 1, 2, & 3)
West	0.20	0.13	0.16	0.08	0.04
Peninsula	0.13	0.28	0.11	0.26	0.29
East	0.10	0.08	0.08	0.07	0.04
Continent	0.12	0.10	0.10	0.07	0.05

2.G Mod 1 alone contributes 0.02 of the ~0.07 difference in continental trend between S09 and RO10. Mod 3 alone contributes about 0.02. When performed together, the modifications contribute 0.05, leaving 0.02 as a result of Mod 2. Although the reviewer

asserts otherwise, it is clear that the differences in magnitude are a nearly equal combination of the three primary deficiencies of the S09 method outlined in our abstract.

2.H Next, the reviewer claims that the resulting spatial patterns of change are almost entirely due to Mod 1, with negligible contribution from the other modifications. This is a crucial point, as the reviewer's subsequent comments concerning geographical trend translation depend on this claim being true.

2.I Given that the spatial information is carried by the AVHRR spatial eigenvectors, one would fully expect that eigenvector retention would have a significant impact on the resulting spatial patterns. This is obvious, and we do not see the value of modifying the text to explain it. Assuming that the calibration and resulting set of regression coefficients used to predict the PCs are compatible with the coefficients used to recover the gridded estimates, then it follows that Mods 2 and 3 would have a negligible effect. If the coefficients are *incompatible,* then Mods 2 and 3 would be significant. S09 clearly falls into the latter case.

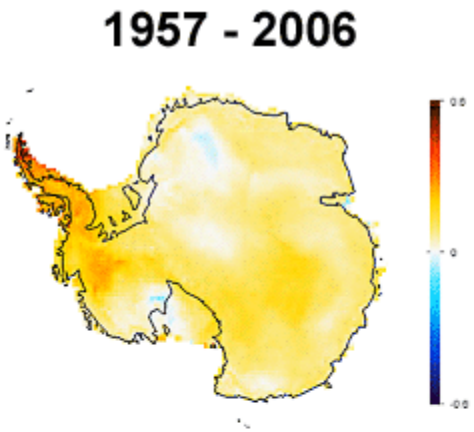
2.J To demonstrate this, we will show the spatial patterns of change for 4 separate cases:

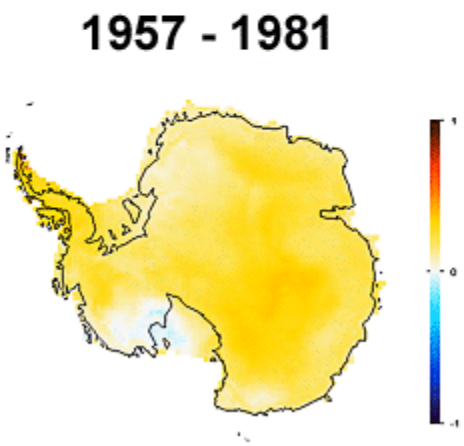
- Variant 1: Modification 1 alone
- Variant 2: Modification 1 and 3
- Variant 3: Only 5 retained eigenvectors and modifications 2 and 3
- Variant 4: Only 5 retained eigenvectors and modification 3

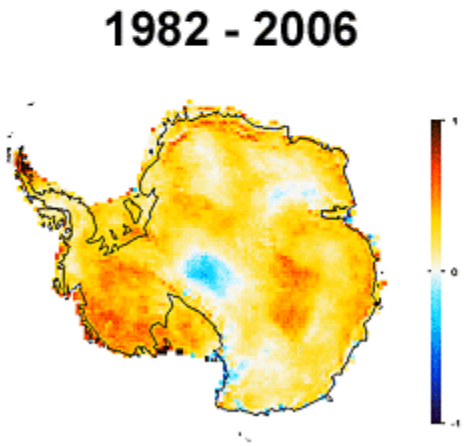
2.K If the reviewer's belief that Mod 2 is negligible is accurate, then Variants 1 and 2 should not significantly differ from either each other or the RO10 reconstructions. This applies to both the overall trends and the subperiods.

2.L If the reviewer's belief that the primary driver of spatial pattern differences is the *number* of retained eigenvectors, then Variants 3 and 4 should significantly differ from the RO10 reconstructions, as the RLS and E-W reconstructions utilize 80 and 100 eigenvectors, respectively, and this test retains only five. If the claim is correct that Mod 2 is negligible, they should additionally be consistent with each other. This applies to both the overall trends and the subperiods.

VARIANT 1: MODIFICATION 1 ALONE





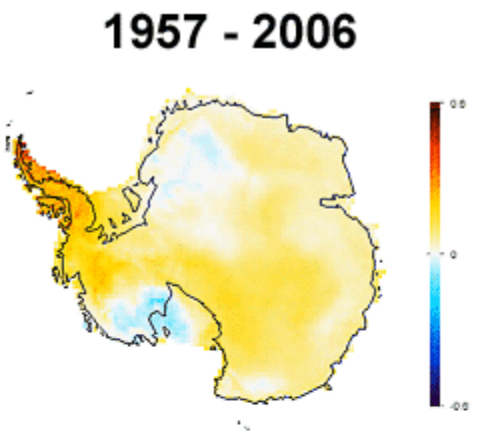


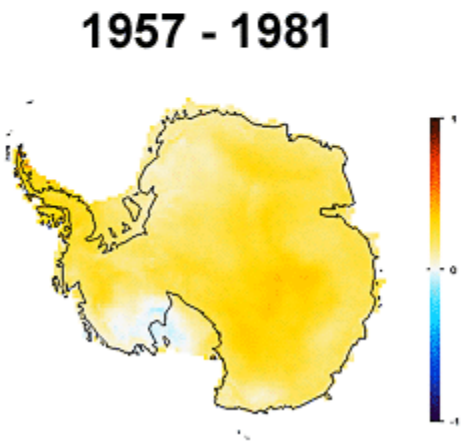
2.M In the full period, this variant captures some (but not all) of the features in the RO10 reconstructions. Those features captured are the reduced warming in the Ross region as compared to S09 and better localization of the Peninsula trends. Absent are the prominent Ross, South pole, and Weddell area cooling. Additionally, the continent-wide trend is much closer to S09 (0.010) than RO10 (0.05).

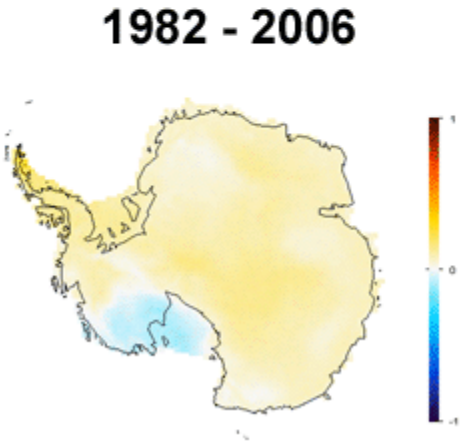
2.N More significant differences are apparent in the subperiods. The 1957 – 1981 plot looks far closer to the equivalent S09 subperiod than the RO10 reconstructions. The Ross cooling is reduced, the pole is warming instead of cooling, and the strong warming in Victoria/Wilkes Land is absent. Given that the latter two features are in well-observed regions of the continent and match ground records, their absence is significant.

2.O The 1982 – 2006 plot is also substantially different, as it is merely the truncated, but otherwise unaltered, AVHRR data. This is a crucial observation. *If* the regression coefficients are directly compatible with the AVHRR eigenvector weights, then using the modeled PCs (Mod 3) could not greatly alter the *patterns* of this subperiod.

VARIANT 2: MODIFICATIONS 1 AND 3





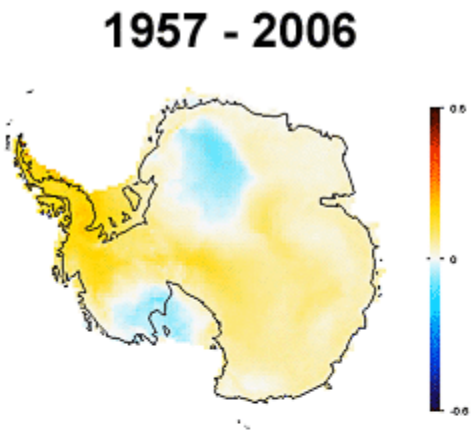


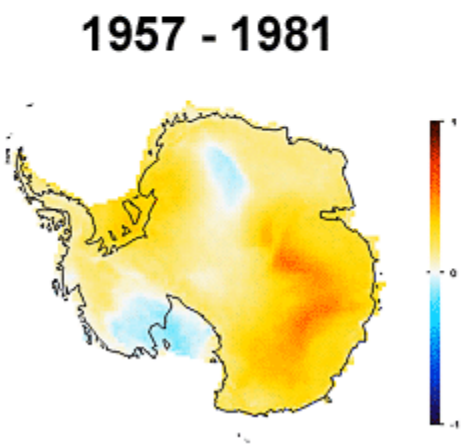
2.P In the full period, adding Mod 3 provides patterns that are very similar to the RO10 reconstructions. Most of the essential features are captured, albeit with the Weddell and South Pole areas showing less cooling than RO10. While the spatial patterns are reasonably well represented, as noted in the response to the previous problem, this does not extend to the overall magnitude. This variant captures only 2/3 of the difference in the continental trends, leaving a substantial portion unaccounted for.

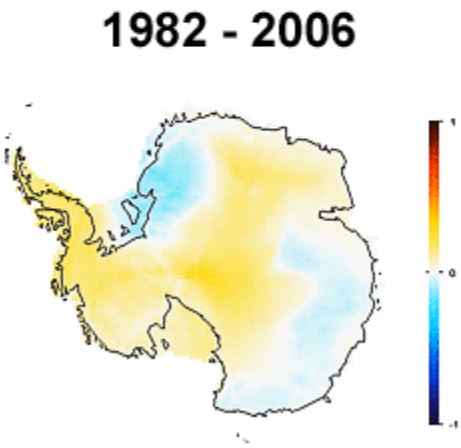
2.Q In the subperiods, the patterns remain significantly different from RO10. As Mod 3 only affects the 1982 – 2006 period, the 1957 – 1981 plot is unchanged and retains the same deficiencies noted in Variant 1. The 1982 – 2006 plot, on the other hand, looks substantially different from both Variant 1 and the RO10 reconstructions. It is clear that Mod 3 has a *significant* impact on the spatial distribution of trends. This confirms the statements in our text that the coefficients used to predict the PCs differ materially from the weights used to recover gridded estimates, and shows the reviewer's belief that use of the modeled PCs has little impact on the spatial patterns is not correct.

2.R Furthermore, the 1982 – 2006 plot is missing *all* of the essential features of the RO10 reconstructions. It shows a visibly apparent loss of variance, displays a large cooling region in the Ross area, and is missing the Victoria/Wilkes Land and Weddell area cooling. Therefore, the implication that Mod 2 has a negligible effect is similarly inaccurate.

VARIANT 3: 5 EIGENVECTORS, MODIFICATIONS 2 AND 3





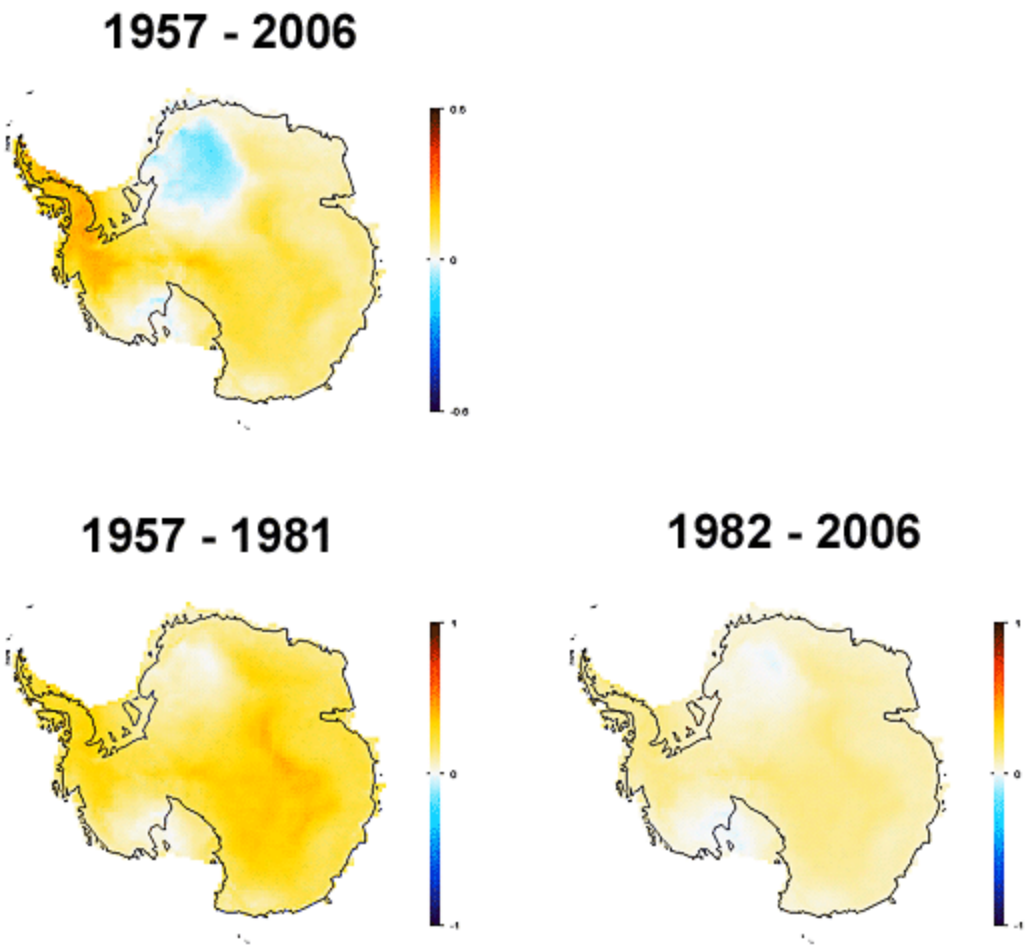


2.S With only 5 PCs but including Mods 2 and 3, most of the essential spatial features of the RO10 reconstructions are present, both in the 1957 – 2006 period and in the subperiods. Though there is visually apparent variance loss between these reconstructions and RO10 – and the warming in Victoria Land near Cape Adams is significantly reduced in the 1957 – 1982 period – the overall pattern in Variant 3 is close to the RO10 reconstructions.

2.T It is clear that Mod 1 alone cannot account for the spatial differences between S09 and RO10. The same is true of the combination 1 and 3. Furthermore, the dependence on the

number of retained eigenvectors is less than implied by the reviewer, as most of the essential features of RO10 are reproduced with as few as 5 retained eigenvectors.

VARIANT 4: 5 EIGENVECTORS AND MODIFICATION 3



2.U This variant demonstrates the significant impact of Mod 2. While most of the full period features are captured in this reconstruction, the subperiods are clearly different from both Variant 3 and the RO10 reconstructions. In particular, without Mod 2, the 1957 – 1981 and 1982 – 2006 subperiods are virtually identical, with the exception that the latter displays muted trends.

2.V To address the concern that the contribution of each modification is not documented, we have amended the text to include the table at the beginning of this discussion. We do not feel it is valuable, however, to add the spatial pattern plots to the main text. These can be easily incorporated into the SI if the reviewer and/or editor desire.

{A closely related issue is the claim that use of too few EOFs by S09 leads to a 'transfer' of trends in the Peninsula station to other locations in the continent, creating an

artifactual warming – particularly in West Antarctica. RO10 write that: "Due to the vastly larger number of data points in the Peninsula, the regression results necessarily will be determined primarily by Peninsula stations". RO10 illustrate this by showing that S09 uses a "different spatial pattern" in the earlier (1957-1981) part of their reconstruction than the later (1982-2006). Referring to this as a "different spatial pattern" is incorrect, because in fact S09 are using precisely the spatial EOFs in both the early and late part of their reconstruction.}

2.W This comment reflects a misunderstanding of Section 4 and the SI, as the reviewer has conflated our discussion on the recovery of the grid with the regression performed by the RegEM algorithm. We understand perfectly that the reconstruction is recovered by multiplying the AVHRR spatial eigenvectors by the reconstructed PCs *for all times* and, in fact, we explicitly state this in both Sections 1 and 4. Nowhere do we say that S09 recover the gridded estimates using different spatial structures. However, we have amended Section 4 of the text to ensure this misunderstanding will not occur among readers.

2.X What we do say is that the spatial structure of the V matrix in RegEM during infilling assigns coefficients to the ground stations that have a different relationship than what is present in the spatial eigenvectors for the corresponding AVHRR grid cells. This is graphically depicted in Figure 2. The reviewer even agrees with this assessment:

{What RO10 mean is that the spatial weighting that relates the ground station data to the AVHRR spatial EOFs is different. That is, in the matrix relationship Ax = b, x winds up being slightly different between the early and late parts of the reconstruction. This is because the late part of the reconstruction is simple taken directly from the AVHRR data, while the earlier trends are from a linear combination of the ground station data (since no AVHRR data exist before 1982).}

2.Y Although the reviewer agrees, he does not seem to have thought through the implications. The estimated values for each PC are simply linear combinations of the non-missing values in the augmented matrix. Unless both the absolute and relative magnitudes of the coefficients in *x* yield weights that match the corresponding AVHRR eigenvector grid cells weights, the gridded estimates *must* differ from the original and infilled data. Except in the unique case where all regression coefficients are inflated by the same scalar multiplier, this difference will manifest itself in *both* a change in trend magnitudes and a redistribution of trends. This is easy to test: simply extract the *x* matrix for a given time step, modify one or more of the coefficients, recalculate the PCs, and observe that temperature estimates across the entire grid are affected.

2.Z The question as to whether this effect is significant has already been answered in the discussion concerning the contribution of modifications 1, 2, and 3 to the reconstructions. Modification 2 constrains the relationship of the regression coefficients to the pattern present in the spatial eigenvectors and modification 3 ensures this relationship is used for all time periods. Because these modifications (even in the absence of a significant number of additional eigenvectors) result in significant changes to both the overall trends

and spatial distribution of trends, it follows that there is a very real and significant issue with the S09 method.

{In any case, a simple look at the data shows that the greater warming in the eastern vs. western part of West Antarctica is inherent in the raw satellite data. Greater warming West Antarctica than on the Peninsula is simply a consequence of the smaller number of EOFs retained, not to the alleged 'transfer' of trends. This is easily seen in a straightforward calculation of trends in the raw AHVRR data, based simply on the number of EOFs retained and using no ground station data at all. That is, if one does the linear decomposition on the AVHRR data, and then recomposes the AVHRR data while retaining just 1, then 2, then 3.. etc. EOFs, it is readily seen that any number of EOFs retained (>2) still shows greater warming in the Ross Sea than on the Peninsula, that 6 EOFs are required to capture the cooling at the pole (which is spatially very restricted), and that at least 4 or 5 are required to capture the full magnitude of warming on the Peninsula. This was recognized in S09, where it is stated that:

> *A disadvantage of excluding higher-order terms (k>3) is that this fails to fully capture the variance in the Antarctic Peninsula region. We accept this tradeoff because the Peninsula is already the best-observed region of the Antarctic. }*

2.AA With respect to the trend distribution in the AVHRR data and the number of eigenvectors needed to recover certain spatial features, we agree with the reviewer. However, the reviewer conflates the patterns present in the AVHRR data (which appears in its raw form only in the 1982 – 2006 period) with what happens when the PCs are estimated with and without the spatial eigenvectors being used as a constraint in the regression. Simply because the raw AVHRR data demonstrates one pattern does not mean that the relationships between the regression coefficients in RegEM demonstrate the same pattern. They do not, a fact to which the reviewer agrees prior to making the above statement. The fact that this difference exists necessarily results in a redistribution of information when the grid is recovered by the AVHRR eigenvectors. The reviewer's observations *strengthen* our argument that ensuring the regression coefficients and eigenvector weights are compatible before recovering the grid is a crucial step.

2.AB The fact that S09 acknowledge that they do not capture Peninsula variance does not change the results.

{ While RO10 may be correct that the alleged 'transfer of trends' also occurs, this is not demonstrated in the paper. In either case, it should be made clear that this actually amounts to at most a very small effect, rather than a critical problem with the reconstruction method in S09, a the current version of the manuscript insinuates.}

2.AC As shown above, we *are* correct that the "alleged 'transfer of trends'" occurs; it *must* occur. The transfer of trends is a necessary consequence of the difference in spatial structure between the AVHRR data and the regression coefficients determined by RegEM, which in turn is due to a lack of spatial constraint during PC estimation in the S09 method. If it were not, then modifications 2 and 3 would have negligible effect on

the reconstruction. Since they have a substantial effect, this effect is quite real and is, indeed, a key shortcoming of S09.

{Problem 3) Discussion of the technical details of what was done in S09 is confusing or incorrect in several instances. For example the claim that S09 uses a 'different spatial structure' in the satellite era and the pre-satellite era is wrong.}

3.A We note that the spatial structure comment was made under Problem 2, and has already been discussed.

{The major problem here is the discussion of the choice of 3 for the number of eigenvectors to retain, and for the choice of the TTLS truncation parameter. RO10 assert that S09 use "a combination of physical meaningfulness and statistical separability" to obtain the value of 3. This confuses two different issues. What is in fact stated the paper is that Steig et al. "use an adaptation of RegEM in which only a small number of significant eigenvectors are used" and reference is made to the paper Mann et al. (2007) that describes the methdology used in the choice of the number of eigenvectors to retain (and similarly, the truncation value to use). The cited paper states that both the number of eigenvectors to retain and the truncation parameter to use in RegEM is chosen by "the number leading eigenvalues of the calibration period data matrix that lie above the estimated noise continuum ... estimated by a linear fit to the log eigenvalue spectrum." Also discussed in Mann et al. (2007) is the linear fit to the log eigenvalue spectrum can itself be poorly constrained due to sampling error. In the AVHRR data, all eigenvalues > 3 are indistinguishable from the next higher order eigenvalue (that is, 4 overlaps with 5, 5 with 6, etc.). Even though there is an apparent visual break in slope between eigenvalues 10 and 11, and 13 and 14, which could be used to justify retaining higher-order eigenvectors, doing so requires a subjective decision as to which eigenvalues to include in the linear fit. In contrast, there is a sharp, unambiguous break in slope between eigenvalue 3 and 4, meaning that eigenvalues <4 (and only eigenvalues <4) are unambiguously outside the noise.}

3.B We thank the reviewer for correcting our misunderstanding and have amended the text to state Mann *et al.* (2007) as the reference for determining the RegEM truncation parameter and retained AVHRR eigenvectors. However, as this is an identical (or nearly so) procedure as that described in North *et al.* (1982), the cautions from North about using such a procedure to determine the truncation point fully apply and should remain.

3.C Furthermore, the effectiveness of the procedure described in Mann is *et al.* (2007) was tested using separate noise realizations for each pseudoproxy, which does not admit the possibility of nonlocal correlation between predictor and predictand residuals. This assumption is violated in Antarctica, as discussed in Section S2 and depicted graphically in Figure S3. Mann *et al.* (2007) also note that the procedure is too conservative at high SNRs. The coefficient of determination between the AVHRR data and ground data (0.45) indicates an SNR of roughly 1.0, which is the highest SNR tested by Mann *et al.* 2007 (excepting the perfect pseudoproxies). S09 do not show that this SNR is below the unspecified threshold from Mann *et al.* (2007) and do not show that the procedure is

effective when the residuals show significant nonlocal correlation. The results from our cross-validation testing are strong evidence that it is not. Finally, Mann *et al.* state that the procedure is heuristic, describing it as a "conservative choice that works well in practice," and even suggest using cross-validation as an alternative tool. The revised text reflects this.

{*Furtherer, nowhere is it stated in S09 that the choice of retaining 3 eigenvectors depends on the interpretation of the spatial patterns in terms of the atmospheric dynamics, nor is there any claim about 'visual similarity' as RO10 claim. What is stated is that these eigenvectors "can be meaningfully related to important dynamical features of high latitude Southern Hemisphere atmospheric circulation" and reference is made to the extensive treatment of this in an earlier paper (Schneider et al., 2004). This observation gives additional confidence – albeit subjective -- that the correlations between station data and the principal component time series associated with each of the first three eigenvectors will be meaningful, but that in no way implies that this was criterion used in choosing to retain only 3.*}

3.D We appreciate this clarification and have removed any suggestion that physical interpretation of the eigenvectors affected the choice of truncation parameter or that this interpretation was motivated by visual similarity.

{*Related to this, the discussion in RO10 of the likelihood of 'mixed modes' due to the overlapping error bars on eigenvalues 2 and 3 is misplaced. Indeed, it is strange that RO10 spend so much time on this, and then proceed to retain much higher order modes. One cannot claim that including eigenvectors 2 and 3 is problematic in S09 due to the overlapping error bars of their associated eigenvalues, and then proceed to justify the use of even higher order eigenvectors, for which all of the error bars overlap. Note that RO10 do suggest an iterative method for choosing the number of eigenvectors to retain. This is both novel and valid, and almost certainly represents an improvement to S09, and is hence a valuable contribution to the literature. This is a separate issue, though, from accurately discussing the methods actually used in S09.*}

3.E We thank the reviewer for making this point and agree that the focus on eigenvectors #2 and #3 is of no value and is confusing. It gives the false impression that S09 did not follow their own referenced procedure. We have removed that portion of the text.

{*It is clear then, that the choice of 3 eigenvectors in S09 was based on established procedures, as clearly stated in the paper, and there is no reason to suggest or imply otherwise. Indeed, such speculation as the motivations of the authors has no place in a scientific journal article.*}

3.F We appreciate the reviewer's clarification of the reference for the retention procedure. However, our text made no statement on the motivations of S09, nor was any implication intended.

{*Another area where RO10 mischaracterizes what was done in S09 is in the discussion of the mathematics used to infill the mixing ground station data. RO10 claim that "S09 define an augmented matrixX=(A|b) , where A is said to represent the ground station data (predictor) and b is said to represent the AVHRR principal components to be estimated (predictand)." They then claim that this S09 misstate what was done and that actually "the PCs appear in matrix A (not b)". This is simply more speculation about whether Steig et al. really mean what they say. There is no reason to think that the PCs are anywhere but in (sub) matrix b, as stated in S09, and the station data are in matrix (sub) matrix A. It appears that RO10 to be confused by the use of the terms "predictor" and "predictand", and assume (wrongly) that any missing values in the station data must be in matrix b, since this is referred to as the "predictand" matrix. But these terms are clearly used in S09 to provide a qualitative description of what is being done. The RegEM algorithm used is an iterative one, applied to a single matrix, and there is nothing that requires that missing values be in any specific part of that matrix.*}

3.G This statement by the reviewer is inaccurate. The regression model in the TTLS algorithm is $Ax = b$. In this model, A is *defined* as the set of available values and b is *defined* as the set of missing values to be estimated *for each time step*. These definitions are clear from both the commented Matlab code and Schneider (2001). While it is true that regularization is only performed once per iteration (truncated eigendecomposition of the correlation matrix), a *separate regression* is performed for each combination of available and missing values. The TTLS code calculates the spatial eigenvectors for the combined (A b) matrix – in which the columns corresponding to A and b may appear in any order – and extracts a new spatial matrix V_{11} for each time step. The V_{11} matrix contains the eigenvectors from 1 to k arranged in columns and the components of the predictor (matrix A) data series arranged in rows. The pseudoinverse is then computed, pre-multiplied by the vector of available values (matrix A) for that time step, and post-multiplied by the first k eigenvectors corresponding to the b matrix for that time step. This yields a vector of estimates for that particular time. If the PCs have available values, they appear in A. If they do not, estimated values for them appear in b.

3.H The regression model implied by both S09 and the reviewer would require estimation of available values from missing ones, which is indeterminate. It also implies that the PCs are never used to estimate ground data. This gives the misleading impression that RegEM can be used to calibrate unlike quantities.

3.I The Matlab TTLS module used by S09 is available at:
http://www.gps.caltech.edu/~tapio/imputation/index.html

{*Another issue is the discussion in RO10 regarding the choice to allow the satellite data to influence the infilling of the ground station data. Is this is in fact problematic, as RO10 suggest? This depends on whether one believes all of the spatial-structure errors are in the AVHRR data, and not in the ground stations, but this is an assumption, not an objective fact. RO10 suggest a method for resolving the issue of differing spatial structures in the two data sets – their "eigenvector-weighted" (E-W) approach – and this is a valid method. It does not follow this is necessarily the best approach to use, nor that*

it has a "material impact" on the final reconstruction results. And indeed, RO10 themselves imply it may be unimportant when they say that "Even if the mutual interaction between PCs is negligible, the situation is at best calibrated PC estimates (expressed as functions of ground data) from 1957-1981 with uncalibrated PCs (not expressed as functions of ground data) spliced on the end." This again relates to point (d) from Problem 3 above, regarding whether to replace the raw PCs of the AVHRR data (1982-2006) with modeled PCs derived from a linear combination of the ground station data. Whether this is appropriate or not depends on what one assumes about the potential sources of error. As discussed in more detail below, there is good evidence to suggest that the trends in the station data – as used in the preferred reconstructions of RO10 – are in substantial error in West Antarctica.}

3.J The issue of whether the satellite PCs can be allowed to directly influence the ground station infilling is a calibration question and does not depend on the source of error. RegEM implicitly assumes that all series are equivalent quantities, and uses each of the variables to predict every other variable based on the pattern of available and missing data. Unless the explanatory variables are wholly complete, portions of the explanatory variables will be expressed as a linear combination of the response variables. This results in the response variables being partially calibrated to the explanatory variables, and partially calibrated to linear combinations of themselves. If they are already interchangeable quantities, this is not an issue. If they are not, it is most certainly an issue. The mutual exchange of information between response and explanatory variables invalidates the calibration. Whether this materially affects the answer *in this particular case* does not mean that future applications will be as forgiving. Documenting this deficiency in the peer-reviewed literature is a worthwhile contribution to help prevent misuse of the algorithm in cases where it may matter. We agree with the reviewer that the effect on S09's reconstruction is negligible, and we have modified the text accordingly.

3.K The question of whether to replace the raw AVHRR PCs with modeled PCs is also a calibration issue and is independent of the assumed source of error. Calibration is by definition placing the response variables in terms of the explanatory variables. In S09, the station data are the explanatory variables and the PCs are the response variables. Unless the explanatory and response variables are shown to be interchangeable (and our text clearly demonstrates that they are not), then the response variables must be replaced by the modeled response. If the response variables are never expressed as functions of the explanatory variables, no relationship between them is ever established and valid predictions cannot be made.

3.L If S09 believed the ground stations have greater error and wished to retain the AVHRR data, then the appropriate calibration would have been to regress the station data against the corresponding AVHRR grid cell data. As long as the spatial sampling is representative of the entire grid (which is a separate consideration), the modeled stations could be used to extrapolate the PCs back in time. Retention of the raw AVHRR data in the 1982 – 2006 timeframe would then be perfectly justified.

3.M Either way, the response variables must be modeled as functions of the explanatory variables, or the response and explanatory variables must be shown to be directly interchangeable. The discussion under Problem 2 shows that they are *not* interchangeable, and that the result is significantly affected when the calibration is properly performed.

3.N Lastly, given that small regularization parameters for RLS yield the highest verification statistics, the likelihood that the systemic differences in AVHRR vs. ground data are a result of error on the ground stations is small. Otherwise, larger regularization parameters (i.e., more filtering) would have provided better predictions. The extremely small regularization parameters used in our study indicate that the AVHRR data is, indeed, the primary source of error for temporal evolution.

{*Finally, the discussion in RO10 of the possible artifact of the eigenvector patterns, and their relationship to Chladni patterns is misplaced in this paper. The use of eigenvector decompositions to examine the relationship between different climate fields is well established in the literature, and its limitations well known. If the authors wish to write a manuscript about this, they should do so, but the discussion in RO10, based on simple visual comparison, is purely speculative and wholly inadequate and should be removed.*}

3.O While the possibility that eigenvector patterns are mathematical artifacts is well established in the peer reviewed literature, we agree that we spend insufficient time developing this as it might apply to S09. Since the reviewer has corrected our misunderstanding that physical interpretation affected the choice of truncation parameter, the relevance of this discussion is greatly reduced. We have amended the text per the reviewer's recommendation.

{Problem 4) RO10's choice of kgnd = 7 for the results to emphasize in the main text is suspect.}

{For the reconstructions presented and discussed in the main text, RO10 find that using the values of kgnd that optimize the reconstruction as a whole results, uniquely, in an insignificant warming trend in West Antarctica. While RO10 acknowledge that the sensitivity to kgnd indicates that they have likely underestimated the amount of warming in West Antarctica, they do not explore why particular choices of kgnd result in dramatically different results. Yet the answer almost certainly lies in Table S3 in the Supplementary Information, where it is shown the using kgnd > 5 in a correlation setting, and kgnd > 4 in a covariance setting, results in CE scores less than zero in two of the weather stations in West Antarctica, Byrd and Mt. Siple. Furthermore, the maximum CE score is found either for kgnd = 4 or 5 for all West Antarctic stations. This means that kgnd>5 results in estimates of the missing data in West Antarctica stations that is further from climatology (which would result, for example, from an artificial negative trend) than using lower values of kgnd.

There appears to be a critical flaw here. As described in the Supplementary Text, the ground station data are infilled using an initial choice of kgnd; then, the infilled ground

stations are augmented by the AVHRR PCs and the PCs are then infilled; then, the process is repeated after withholding selected stations from the ground station matrix that has already been infilled to provide additional verification targets. Verification statistics are then calculated and based on this, a new value kgnd is chosen.

What is happening here? Evidently, the final choice of k_{gnd} is based on a how well the infilled station data are matched in the overall reconstruction, not on how well the original raw station data are matched. This means that the optimal value of k_{gnd} chosen in this iterative way can be a value that provides the best verification statistics with bad data! There appears to be nothing in the algorithm that prevents unacceptable values of k_{gnd} from being choisen as 'optimal' with respect to the full reconstruction that includes the satellite data. Inspection of Table S3 shows that the 'optimal' values of k_{gnd} used inthe main reconstruction of RO10 are not acceptable choices for the infilling of the ground stations, particularly in West Antarctica. To put this another way, RO10 have chosen values of k_{gnd} that minimizes the difference between two different reconstructions – the reconstruction of the AVHRR PCs on the one hand, and the reconstruction (infilling) of the ground station data on the other. This is fundamentally flawed in the approach, and cannot be considered a valid verification exercise. Indeed, RO10 note that the values of k_{gnd} chosen with the ground stations alone should be "similar" to the value chosen in the full reconstruction, but they use values that are clearly in no practical way could be considered 'similar'.}*

4.A In the above paragraphs, the reviewer makes three factual errors. One is relatively minor; the reviewer claims we do not explore the dependence on k_{gnd} when we have an entire section in the SI devoted to just that (a section the reviewer quotes from later). The second, and by far the most important, is that the reviewer claims that the reconstruction verification statistics are calculated using infilled data, and states that the choice of k_{gnd} is determined by minimizing the difference between the *infilled* ground stations and gridded reconstruction temperature estimates. This is incorrect.

4.B Section 7.c is explicit that all verification statistics are calculated versus *actual station data* that has been *entirely withheld* from the reconstruction. This fact is repeated in both Tables 6 and 7 in the main text, Tables S5 and S6 in the Supporting Information, and Figures S22, S23, S25, and S26. The withheld stations that are used for verification are neither used in the gridded reconstruction nor included in the infilled ground station matrix. The supposition of the reviewer that k_{gnd} was in any way determined based on calculations to infilled values is inaccurate.

4.C The reviewer's third factual error is assuming that the truncation parameter that best reproduces withheld station data during the ground-only infilling will also provide the input matrix that best reproduces withheld station data following RLS and E-W regressions. This is not categorically true (though, in certain circumstances, it can be), and this erroneous assumption by the reviewer is closely related to the misconceptions he expressed in Problem 2.

4.D First, however, let us consider an alternate scenario.

4.E Let us assume that, instead of calculating any verification statistics at all following ground station infilling, we performed our cross-validation with the following steps:

- Infill ground station matrix, withholding specified stations for verification
  - Note that this is not an early/late verification test . . . the verification target station records are simply omitted from the matrix in their entirety
- Perform the RLS or E-W regressions to generate PC estimates
- Recover gridded estimates by multiplying the PCs by the spatial eigenvectors
- Calculate verification statistics to original, withheld station data
- Repeat for all combinations of adjustable parameters (k_{gnd}, k_{sat}, k_{RegEM}, c)
- Select the combination of adjustable parameters that yields the highest verification statistics

4.F This is, of course, a perfectly legitimate means of performing cross-validation. We run the entire reconstruction and compare the results to original, withheld ground station data. Based on the verification results, we select our parameters. We believe the reviewer would agree that this procedure would be both perfectly valid and quite common.

4.G Performing the above procedure, *without* the intermediate step of calculating infilling-only verification statistics, will yield precisely the same combination of parameters as chosen in our text. Doing so, however, requires an exceedingly large number of reconstructions be performed (covariance and correlation settings, 2 different infilling algorithms for ground-only, 14 station sets, 10 values for k_{gnd}, 99 {RLS} or 5 {E-W} values of k_{sat}, 2 different infilling algorithms for E-W, and 12 or 15 values of either k_{RegEM} or c, in both full {fewer withheld stations} and verification {additional withheld stations} settings):

Total RLS recons: 2 * 2 * 14 * 10 * 99 * 15 * 2 = 1,663,200 permutations
Total E-W recons: 2 * 2 * 14 * 10 * 5 * 2 * 12 * 2 = 134,400 permutations

4.H This is computationally prohibitive. Therefore, we chose to initially *screen* for a range of station combinations and values of k_{gnd} that provided good intermediate verification statistics, under the assumption that these would be *close* to what would prove to be optimal for the reconstruction cross-validation. This reduces the total permutations to a more manageable number:

Total RLS recons: 2 * 1 * 1 * 4 * 99 * 15 * 2 = 23,760 permutations
Total E-W recons: 2 * 1 * 1 * 4 * 5 * 2 * 12 * 2 = 1,920 permutations

4.I Following selection of our parameters, we sampled values outside of the screened ranges. In no case did the sampled values exceed the cross-validation statistics for our optimal parameters. The reviewer's concerns with our validation method are misplaced, as they are due to a misreading of the text, tables, and figures.

4.J Having established that our cross-validation method is both perfectly legitimate and quite common, we turn to the reviewer's third misconception. The reviewer assumes that the optimal truncation parameter for a ground station only infilling will produce the optimal matrix for input into the RLS and E-W regressions, but provides no evidence that this should be so. In truth, it is highly unlikely that the reviewer's assumption would hold.

4.K As already stated by the reviewer in Problem 2, the set of weights produced by the RegEM infilling are not the same as the set of weights at the corresponding locations on the AVHRR grid. This is due to a number of reasons, including uneven sampling of the grid, differing errors on the station data and AVHRR data, the fact that station values are point values and AVHRR data are grid cell averages, and others. Also important is that the effect of an early/late verification test is to greatly increase sampling error in the correlation matrix in RegEM, especially for short record length stations, because half the station record is omitted. Correlation tests between stations in Antarctica can give substantially different results when only half the data is used. In the gridded tests, however, the station is either completely included or entirely omitted. This difference, along with uneven sampling of the grid, are the most critical, as they will affect not only the spatial coefficients, but also the order and magnitude of the eigenvalues, and the degree of mixing between modes. This necessarily results in changes to the regression coefficients.

4.L Given that this can result in an over-emphasis of a station in a cell that has a low AVHRR eigenvector weight and a de-emphasis of a station in a cell that has a high AVHRR eigenvector weight, the spatial makeup used to determine the PC can result in substantial error when the PC is multiplied by the AVHRR spatial eigenvector. As shown under Problem 2, this error does not just affect the reconstruction value at the station location, it affects the reconstruction value across the entire grid. In fact, in the case where the eigenvector weight is *low*, the effect at points *distant* from the station is greater than at the station itself.

4.M This means that when the PCs are estimated using the RegEM coefficients, a *different estimate* is achieved than if the AVHRR data were available for all times. When this estimate is recombined with the corresponding spatial eigenvector, the temperature estimates at each grid cell change from what exists in the ground station matrix. If the temperature estimates change, then the verification statistics necessarily change. If the verification statistics change, then it is a distinct possibility that a *different* truncation parameter provides more accurate results in the full reconstruction than in the ground-only reconstruction. And since the spatial distribution of the stations is substantially unrepresentative of the overall AVHRR grid, this is an unsurprising result. This behavior is evident from comparing Table S3 with Tables S5 and S6:

TABLE S3:

Station Metadata				Minimum CE Early/Late Cross Validation Experiments (SATELLITE INFORMATION EXCLUDED)							
				Correlation k_{gnd}				Covariance k_{gnd}			
Name	Lat	Lon	Record Length	4	5	6	7	4	5	6	7
Byrd (AWS)	-80.0	240.6	187	0.07	0.06	*-0.03*	*-0.00*	0.04	*-0.46*	*-0.45*	*-0.83*
Erin	-84.9	231.2	62	0.22	0.18	0.14	0.04	0.22	0.21	0.04	*-0.19*
Mount Siple	-73.2	232.9	140	0.17	0.09	*-0.14*	-0.20	0.14	0.16	*-0.12*	*-0.17*

TABLE S6 (optimal values for k_{gnd} of 7 for correlation and 6 for covariance)**:**

	RLS		E-W			
	Correlation	Covariance	Correlation	Covariance	S09[b]	Monte Carlo[c]
Station Name	CE (Station Data Not Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Byrd (AWS)	0.60	0.31	0.48	0.51	0.44	-0.61
Erin	0.60	0.69	0.55	0.57	0.38	-0.15
Mount Siple	0.33	0.18	0.41	0.29	0.43	-0.47

4.N We reiterate that Table S6 is calculated to *original station data* that was *entirely withheld* from the ground station matrix and subsequent RLS/E-W regressions. Note the vast improvement in verification statistics for the gridded reconstruction – where the solution is constrained by the eigenvector weights and the resulting estimates are recovered in *gridded* instead of point fashion – as compared to the entirely unconstrained ground-station only infilling. As the tables above indicate, this effect is of substantial significance.

4.O Nor is this property unique to RO10. It is also exhibited by S09. For example, using S09 parameters, the manned Byrd station (not AWS, as S09 do not use AWS data in their main reconstruction) demonstrates a minimum CE of 0.13 in early/late withholding experiments in the PC-augmented ground station matrix. However, it then demonstrates a verification CE of 0.26 to the gridded results *when it is entirely withheld* from the augmented matrix during combined infilling of the ground stations and PCs.

{*RO10 state that "the verification statistics for West Antarctic stations are still superior with kgnd = 7." As noted above, this is hard to reconcile with the data in table S3, but in any case, RO10 state that "the differences are minor." This is a critical point: Even if I am missing something in the above discussion (I don't think I am), it still raises the question: If the differences in the overall reconstruction statistics are minor between choices of kgnd, but those same choices has a significant impacts on the reconstruction of*

the ground station data alone, why does RO10 use a value for kgnd that is clearly inferior for the latter?}

4.P The value of k_{gnd} chosen for the optimal reconstructions is *superior*, not only to our own screening step, but also to the S09 reconstruction. This is demonstrated by the substantially improved verification statistics for the gridded reconstruction. As explained above, it does not follow that the value of k_{gnd} that gives the best verification statistics for the ground station infilling will also yield the most faithful gridded reconstruction. In cases where the regional sampling is subject to significant bias (as in Antarctica), it is *highly unlikely* that the reviewer's assumption would hold.

*{**The suspect values for k_{gnd} in the reconstructions used in the main text in RO10 for comparison with S09 are the only values that fail to result in a significant warming trend in West Antarctica.** The authors recognize this, stating that "we cautiously conclude that the trend in West Antarctica is likely significant and may be closer to 0.10 °C/decade than 0.05 °C decade." Yet this critical point is largely buried in the text; it is barely given mention in the abstract, which merely states that "mixed results are obtained for trend significance in West Antarctica." **Furthermore, RO10 fail to note that** not only do more optimal values of k_{gnd} results in greater trends in West Antarctica than presented in their main reconstruction, **this also results in a difference in the pattern of trends.** As shown in their Figure S17, using either $k_{gnd} = 5$ or $k_{gnd} = 6$ results in warming over all of West Antarctica, including the Ross Sea region, in very good agreement with S09.}*

4.Q As the reviewer's discussion of what is "optimal" for k_{gnd} has already been shown to be inaccurate, we will address the implication that we do not fully address the uncertainties in the West Antarctic results, or that these are "largely buried in the text". On the contrary, the largest section of our results in the main text is devoted solely to this issue, and we provide a table (Table 6) of trends *specific to West Antarctica* that clearly outlines these uncertainties, along with verification statistics. We also include an entire section in the SI to discuss this, along with panels showing the results of our sensitivity tests that the reviewer has used extensively.

4.R The reviewer seems to have the impression that Figures S17 – S20 represent the *next closest* reconstructions to the optimal parameters. They do not. We could have shown, for example, the 50-PC, $k_{RegEM} = 6$, and $k_{gnd} = 5$ eigenvector-weighted reconstruction that yields West Antarctic trends of 0.06 +/- 0.07 and demonstrates better verification statistics (CEs of 0.42 in West Antarctica and 0.62 overall) than both the $k_{gnd} = 5$ and $k_{gnd} = 6$ reconstructions the reviewer cites. However, we found it simpler and more informative to show what happens as one varies k_{gnd} without optimizing the remaining parameters for that particular choice of k_{gnd}. Recall that our cross-validation testing required performing over 25,000 reconstructions. There are (quite literally) hundreds of reconstructions that show lower West Antarctic trends, spatial patterns very similar to our optimal reconstructions, and display higher verification statistics than the panels in the SI.

4.S We would like to direct the reviewer to S9.a and S9.b, where we discuss sensitivity related to the other parameters. For E-W reconstructions, note that high values of k_{RegEM} and k_{sat} yield higher trends in West Antarctica and across the continent. The reconstructions shown in the SI have k_{sat} set at the maximum of 100 and k_{RegEM} at 9. For RLS reconstructions, note that low values of c produce the highest trends. The reconstructions we show in those panels have c set to the lowest value that consistently prevented computational singularities during inversion of the eigenvector weights matrix, and therefore represent the high end for trends in the reconstructions we performed.

4.T Evidence reinforcing the notion that the k_{gnd} = 5 solution is inaccurate may be found with the Peninsula trends. The ground station trend at the 7 stations included as Figure 6 in our revised text is 0.41 +/- 0.16, which matches very well with our optimal solution at those same locations (0.40 +/- 0.14). In the k_{gnd} = 5 solution the reviewer consistently cites, the reconstructed trend at those locations is 20% higher: 0.50 +/- 0.12. We also note that the overall Peninsula average in the optimal solution – which matches ground trends – is ***outside*** the 95% CIs for the average in the k_{gnd} = 5 solution. These are clear indications that the k_{gnd} = 5 solution produces excessive trends. The reviewer has taken the results of a *sensitivity test* to question our conclusions without understanding the context provided in the SI, and without examining any other aspect of the sensitivity test.

4.U In addition to the sensitivity testing already noted by the reviewer, we performed two additional sets of tests not presently incorporated into the text. The first is to infill the ground stations via ridge regression, with the ridge parameter for each time step selected via a generalized cross validation function. This yielded spatial patterns of change comparable to our main results, with a West Antarctic trend of 0.11 +/- 0.08 and a reduced area of cooling on Ross. The second test was to perform RLS reconstructions *without* first infilling the ground stations. The stations were offset based on periods of mutual overlap. The resulting West Antarctic trends were 0.08 to 0.12 +/- 0.07, depending on the length of overlap used. Both tests are the subject of a future work currently in progress. In terms of the present work, two results that do not depend on k_{gnd}, both of which produce the similar magnitudes and patterns of trends as our optimal reconstructions, gives us confidence that our selected results are, indeed, the most accurate. A brief discussion of this has been incorporated into the main text and the SI.

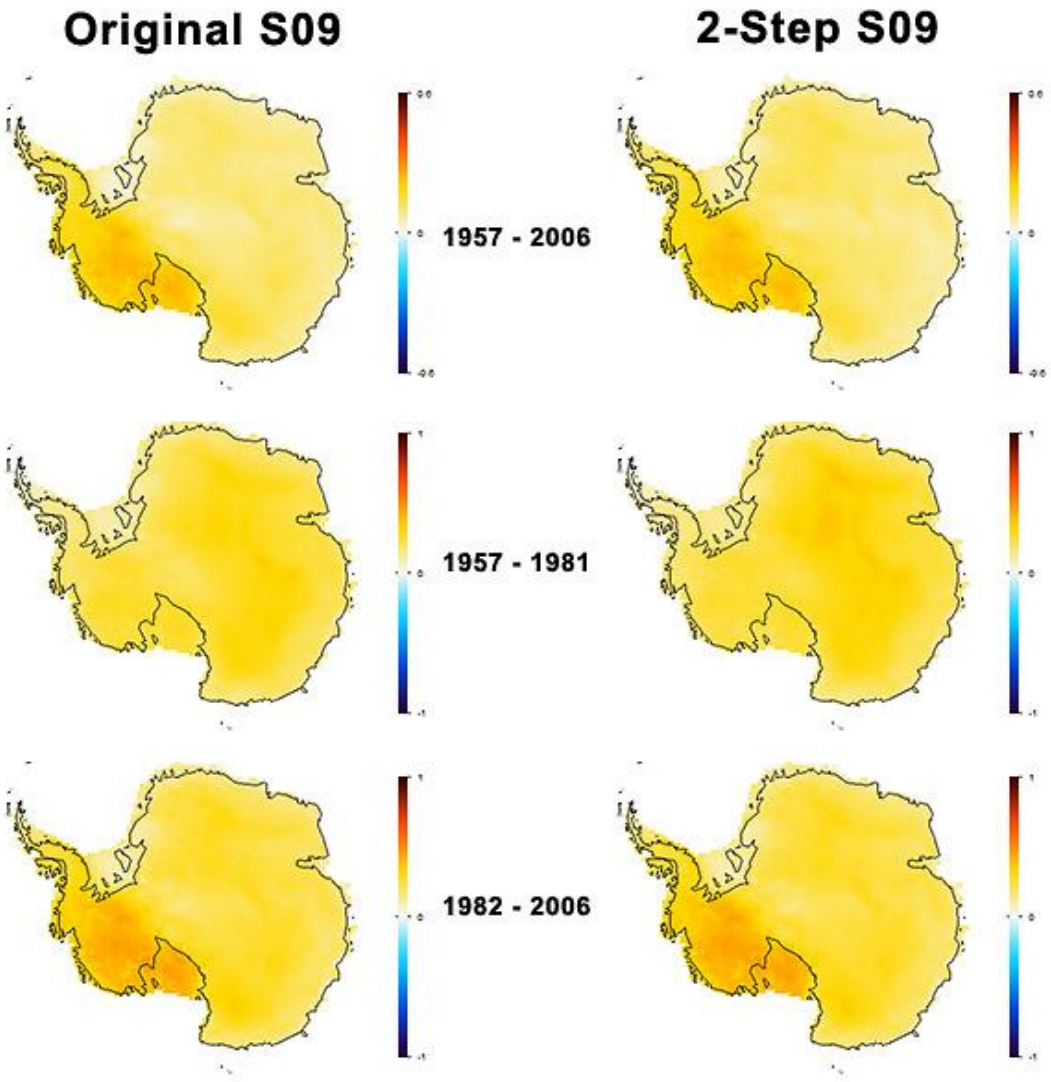
4.V Lastly, we respectfully disagree with the reviewer's opinion concerning the pattern of trends. S09 shows maximum West Antarctic warming on Ross; *all* of our reconstructions (whether k_{gnd} is optimized or not) show ***minimum*** warming on Ross. This pattern does not change with k_{gnd}. Assuming the reviewer meant magnitude and not pattern, we point out that the Ross warming in the examples cited by the reviewer is 0.08 +/- 0.13 and 0.07 +/- 0.12, respectively, and half the land area of West Antarctica shows a similar or *lesser* average trend. The point estimate of Ross warming in S09 of 0.22 is outside the 95% CIs for these values (and vice-versa). This cannot objectively be called agreement.

{*Clearly, the results for West Antarctica are highly sensitive to the quality of the ground stations used in West Antarctica. Importantly, S09's methodology is less sensitive to errors in the ground stations than is RO10's because the former uses information from*

the AVHRR data when infilling the ground stations, while RO10's does not. This does not necessarily mean that S09's methodology is superior: RO10 provides a sound argument as to why the use of the AVHRR data to help infill the ground-station data may possibly be problematic. The point, though, is that the choice to fill in the ground station data using information only from other ground stations makes the full reconstruction much more sensitive to errors in that calculation.}

4.W The claim by the reviewer that S09's methods are less sensitive to the quality of the ground data is not accurate, and the reasoning given is also inaccurate. One can test this by splitting S09 into 2 steps: first infill the ground stations, and then add the PCs. If this is done, the following results are obtained:

	West Antarctica	Peninsula	East Antarctica	Continent
Original S09	0.20 +/- 0.09	0.13 +/- 0.05	0.10 +/- 0.10	0.12 +/- 0.09
2-Step S09	0.19 +/- 0.09	0.13 +/- 0.05	0.10 +/- 0.10	0.12 +/- 0.09



4.X The results are nearly identical. The assumption that including the AVHRR PCs in the ground station infilling stabilizes the regression is accordingly not correct. As discussed previously, doing so violates the integrity of the calibration. However, as the number of included PCs approaches or exceeds the number of stations, the effects of the improper calibration become more significant. This places a practical limit on the number of PCs that can be used in S09's method and is discussed in Section 4 of our text.

{*It is germane to this point that RO10 find that their results are more consistent with those of Monaghan et al. (2008) and Chapman and Walsh (2008). These studies, like RO10, reconstruct missing values in the Byrd station location – the only station of any length in West Antarctica – using a linear combination of other, mostly distant weather stations. Monaghan et al. (2008) noted in their paper that Byrd shows cooling in recent decades. Yet Monaghan et al. have since shown at several meetings that their Byrd station reconstruction was problematic and in conflict with nearby AWS stations that were not included in their original work. Monaghan's updated reconstruction, combines information from these other nearby stations – that are highly correlated with Byrd. The results – reported at the MOCA meeting in Montreal in July 2009 – shows significant annual warming at Byrd (concentrated in winter and spring), in excellent agreement with the AVHRR data. These problems in Monaghan et al. (2008) was pointed out in S09 but are ignored entirely in RO10. Although the updated work of Monaghan remains unpublished, it is at least another red flag that the choice of kgnd values used in the main reconstructions in RO10 are suspect.*}

4.Y The reviewer mentions "excellent agreement" between the unpublished Monaghan reconstruction and the AVHRR data. This is left unquantified. Regardless, both the RLS and E-W reconstructions show *better* agreement with the Byrd data than does the AVHRR data. Specifically, the RLS reconstruction demonstrates an R^2 value of 0.76 and an r^2 value of 0.84 to the Byrd AWS ground station data. For E-W, those values are 0.55 and 0.67. The full-rank AVHRR data displays values of 0.46 and 0.45. S09 displays values of 0.45 and 0.45.

4.Z Secondly, the reviewer has taken "more consistent with" to imply "identical to". While Monaghan *et.al* show cooling at Byrd in recent decades, we do not. His criticisms of the cooling in Monaghan's reconstruction are not applicable to ours. The RLS reconstruction trend at Byrd from 1982 – 2006 is 0.15 +/- 0.32. The E-W trend is 0.02 +/- 0.23. For the latter half of the satellite period (1995 – 2006), those values are 0.16 +/- 0.87 and 0.15 +/- 0.66, respectively. As a minor aside, the reviewer implies that we do not use nearby AWS stations. This is untrue. We do, indeed, use nearby AWS stations, as shown in Table S1.

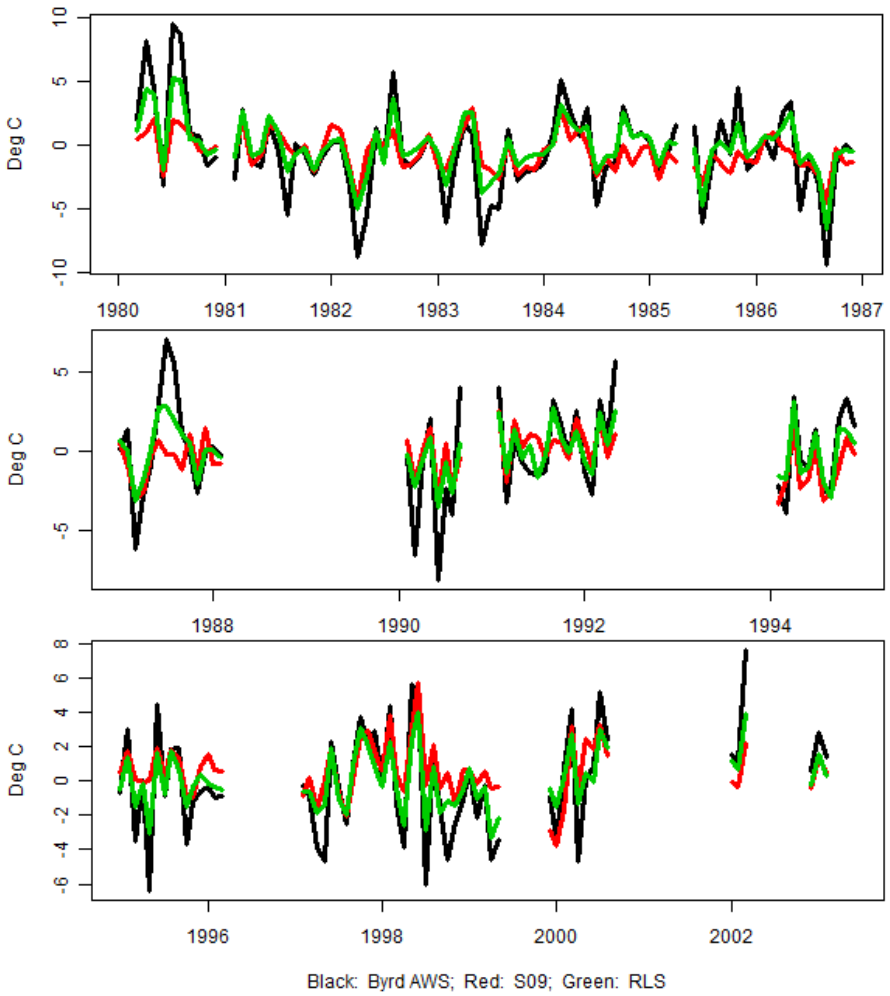
4.AA Finally, because of the incomplete temporal coverage at Byrd, the calculated trend at the station itself changes markedly with the inclusion/exclusion of only a few points. In particular, dropping 1980 from the series increases the trend from 0.23 +/- 0.41 to 0.69 +/- 0.38. Truncating in 2001 (after which annual data is less than 33% complete) drops the trend to -0.02 +/- 0.59. The sensitivity of the trend to addition or removal of a small number of points was noted in particular by Shuman *et al.* (2001), whom S09 reference

(Ref. 13) without mentioning this sensitivity. We compare this below, using months that correspond to existing data from Byrd AWS. The reconstruction that shows the largest deviation from the Byrd trend is bolded and underlined:

	1980 - 2003	**1980 - 2002**	**1980 – 2001**	**1981 - 2003**	**1981 - 2002**	**1981 - 2001**
Byrd AWS	0.23 +/- 0.51	0.17 +/- 0.56	-0.02 +/- 0.59	0.70 +/- 0.48	0.63 +/- 0.52	0.46 +/- 0.56
S09	**0.56 +/- 0.25**	**0.55 +/- 0.29**	**0.56 +/- 0.31**	0.72 +/- 0.27	0.71 +/- 0.29	0.73 +/- 0.32
RLS	0.02 +/- 0.30	-0.01 +/- 0.32	-0.10 +/- 0.35	0.28 +/- 0.28	0.26 +/- 0.31	0.17 +/- 0.33
E-W	-0.09 +/- 0.21	-0.12 +/- 0.23	-0.18 +/- 0.25	**0.07 +/- 0.20**	**0.05 +/- 0.22**	**-0.01 +/- 0.24**

4.AB In spite of the vast differences in the calculated trends, none of the reconstructions are outside the 95% confidence intervals for Byrd.

4.AC Additionally, our reconstructions (RLS in particular) reproduce monthly data more accurately than either S09 or the AVHRR data:



4.AD Unlike Monaghan *et al.* (2008), both of our reconstructions show positive trends at Byrd, with RLS showing a point estimate of 0.15 since 1982. Given the uncertainties in the Byrd AWS trend and the fact that our reconstructions match the monthly data better than either the AVHRR data or S09's reconstruction, the implication that our reconstructions are somehow inferior is unjustified.

{ The parsimonious conclusion here is that RO10, by choosing to effectively assume that the AVHRR data are erroneous, but that the ground station data are reliable, have created a reconstruction that is less accurate in West Antarctica than that of S09, which assumes either reliable trends (in their main reconstruction) or zero trends (in their 'detrended' reconstruction) in the AVHRR data. ***When one considers that significant warming in West Antarctica, and a pattern of warming that includes most of West Antarctica, is a key result in S09, it would seem appropriate to highlight the fact that in virtually all their reconstructions, RO10 find the same result.*** *}*

4.AE As Tables S5 and S6 and Figures S21 – S26 demonstrate, both the RLS and E-W reconstructions outperform the S09 reconstruction in all regions, including West Antarctica. Our reconstructions also outperform the raw AVHRR data – even at locations where the data for the corresponding ground station had been omitted from the reconstruction. The reviewer cites our sensitivity study in the statement "virtually all", which we previously showed to be in error. The West Antarctic results shown in our reconstructions are, indeed, significantly different than S09.

{It is further important to note that the comparison RO10 makes for the magnitude of trends in West Antarctica greatly overstates the disagreement. While RO10 say that the overall trend for West Antarctica is "closer to 0.1 °C/decade", in the supplemental text they report values of 0.13°C/decade (covariance RLS, using their stated optimal value of kgnd) and as much as 0.25 °C/decade for RLS correlation, using kgnd = 5, which as noted above is superior for the infilling of the critical ground station data. Moreover, even if one accepts the lower value, the relevant apples to apples comparison (as discussed under Problem 1 above) is with the detrended AVHRR reconstruction in S09, which obtains 0.13°C/decade for West Antarctica. None of these comparisons are consistent with the statements in RO10 that "this is still to " that reported by Steig et al." and "the magnitude of West Antarctic trends in the S09 reconstruction are larger than would be supported by either explanation [i.e. overfitting during the regressions or masking of a robust feature with noise]." To the contrary, ***RO10's results show that the West Antarctic trends – using parameters acceptable or superior by their own verification statistics -- vary between values 25% and nearly 150% those of S09.*** *}*

4.AF The conclusions and "apples to apples" comparison point here have all been addressed and shown to be incorrect.

{*Problem 5) Discussion of the implications of the RO10 results for our understanding of Antarctic temperature variability reflects a misreading of the literature.}*

{*Apart from the four groups of technical and presentation issues above, RO10 also suffers from a mischaracterization of the broader implications of the results with respect to our understanding of Antarctic climate change. Especially problematic is the emphasis placed on the alleged "discrepancy between the coupled GISS ModelE runs cited by S09 and Antarctic temperatures." The authors appear to interpret S09 as having claimed that their results can be used to validate the GISS ModelE results and hence any changes to that reconstruction must now cast doubt on the GCM results.*}

5.A We devote only 3 sentences to this, and our comparison is limited solely to the GISS ModelE run cited by S09. Given that the purpose of this paper is to quantify the deficiencies of the S09 method, this comparison is appropriate. Unlike what is stated by the reviewer, we do not say that S09 made the claim that their results can be used to validate GCM results (or vice versa). We simply point out (accurately) that the model result S09 *cited* deviates further from both our reconstruction and the ground information than it does from the S09 reconstruction. S09 used agreement with this particular model to lend credence to their results. Making the observation that our reconstructions (which possess superior verification statistics in both West Antarctica and across the continent) show a poorer match is relevant and appropriate. It does not reflect any misreading of the literature.

5.B We thank the reviewer for pointing out that the ModelE runs were atmosphere-only, and we have amended the text accordingly.

{*What is actually said in S09 is that :*

> *When driven by observed sea-surface-temperature (SST) and sea-ice boundary conditions, the model reproduces many of the basic features of our reconstruction, with warming over most of the continent and persistent in West Antarctica.*

S09 also state that:

> *It is noteworthy that both in the reconstruction and in the model results, the rate of warming is greater in continental West Antarctica, particularly in spring and winter, than either on the peninsula or in East Antarctica.*

While it is true that – if RO10 are correct that the Peninsula trends are larger, one aspect of the latter statement may be wrong – the other points would remain entirely valid. That is, greater warming in West Antarctica than in East Antarctica, and particularly in winter and spring (and also in fall in both S09 and RO10), as would be expected if sea ice – the hypothesized critical factor in S09 – plays a role.}

5.C We agree that the reviewer's quotes accurately reflect the S09 text. However, nothing in our text implies that S09 said anything else. The only point we make – and it is perfectly accurate – is that our reconstruction and the historical ground station information indicate greater and more consistent warming on the Peninsula than in West Antarctica. We make no statements on whether S09's sea ice hypothesis is correct.

{*The simple point made in S09 is that getting the trend pattern right requires getting the sea ice pattern right. The point is not that the S09 results can be used to validate the*

model runs, as RO10 seem to think. Indeed, S09 say very clearly that one does not expect a perfect match because:

> *The details of the comparisons obviously depend on the accuracy of the SST and sea-ice observations (the latter are not generally considered reliable before 1979), and multidecadal internal variability in the model is substantial.*

Further, RO10 incorrectly refer to 'coupled' model runs, whereas what are shown in S09 are atmosphere-only runs with various given boundary conditions. S09 do refer to the coupled runs, but note that:

> *The same model, when run in coupled mode (that is, with a dynamic ocean) fails to reproduce the strong trends observed in West Antarctica and the peninsula.*

RO10 get the point entirely backwards here, in claiming that their new results are in conflict with the models. In fact, if their reconstruction is better than that of S09, then they have actually demonstrated better agreement with the coupled model runs, as reported e.g. in the 2007 IPCC AR4.}

5.D Again, we do not make any statements concerning S09's sea ice hypothesis. We simply show that, when the deficiencies in S09's methods are corrected, the magnitude and spatial distribution of temperature change is different than S09. Whether this contradicts or reinforces the sea ice hypothesis is left to others.

5.E The reviewer seems to believe that we have made a general statement concerning the fit of our reconstruction to "models". We do not. We refer only to the model run presented in the S09 text, and our observation is accurate. We refrain from generalizing this statement to include additional models.

5.F However, we agree with the reviewer that some larger context is useful. We have amended the text to indicate that our Peninsula results are quite comparable to the overall patterns in the 11-model composite reported in Chapman & Walsh (2005) and the winter patterns in the 19-model average shown in Connolley & Bracegirdle (2009) – all of which were used in AR4 – though we note that in both cases, the model averages fail to reproduce the Ross, pole and Weddell area cooling demonstrated in both the overall and Winter trends in our reconstructions.

{*Also in this context, RO10 should refer to the recent paper by Turner et al. (2009), GRL. Turner et al. argue that in the GCMs (in this case, the Hadley center model) this pattern of sea ice change probably reflects the strengthening of the Amunsden Sea Low, which creates sea ice convergence along the coast of western West Antarctica, while producing divergence in the Ross Sea. Increasing the low pressure in the Amundsen-Bellinghausen Seas is similarly implicated in S09 to explain the warming in West Antarctica, entirely consistent with Turner et al.'s modeling results. While Turner et al. do not specifically address temperature changes, their results would suggest that warming should be concentrated to the west where sea ice losses are greatest, and that there should be smaller trends, or perhaps even cooling, in the Ross Sea region. Thus, again RO10's results would appear, if anything, to provide greater support for GCM results than does S09.*}

5.G We do not hypothesize any physical explanation for our results anywhere in the text. Attempting to reconcile modeling results for quantities other than surface temperature with our reconstructions is outside the scope of this work. If future authors wish to make this comparison, we would be delighted.

{*Finally, RO10 entirely ignore the critical point in S09 that warming in Antarctica is greatest prior to the 1980s, which supports the hypothesis of Thompson and Solomon (various papers starting in 2002) that recent Antarctic cooling over East Antarctica is due to an increase in the Southern Annular Mode index in response to the development of the stratospheric ozone hole. RO10's results fully support this conclusion of S09, showing (e.g. Figure 4) that most of Antarctica warmed from 1957-1981. They also show that the substantial cooling in East Antarctica as a whole is only reproduced for a fairly narrow range of time periods (1970s to early 2000s). S09 point out that the Thompson and Solomon work that sought to explain this cooling was based on observations (1969-2000) that were clearly not representative either of long term trends nor of recent trends. In particular, S09 show in their Figure 3 that their results duplicate Thompson and Solomon's East Antarctic cooling for 1969-2000. In contrast RO10 simply say that "The reconstructions also provide evidence of cooling in various parts of Antarctica for all time frames analyzed." This is highly misleading because it implies that S09 et al. show otherwise (they do not), and because RO10's results actually fully support the finding that cooling is greatest in this intermediate period. They do not show, as implied, that the periods of greatest cooling are random in time. Rather, cooling appears when expected – during the time (mid-1970s to early 2000s) during which the Southern Annular Mode index increased.*}

5.H We agree that we ignore this point. It is not relevant to our analysis. We make no attempt to reconcile sea ice forcing, the behavior of the SAM, the ozone hole, or any other physical mechanism with our results. As above, if future authors wish to do this, we would be delighted.

5.I Furthermore, the reviewer states that our finding of cooling in Antarctica for all time periods analyzed – and the implication that S09 find differently – is somehow inaccurate. We do, indeed, find significant regions of cooling for all time frames analyzed; S09 do not. However, we have added a qualifier to our comparison, stating that S09 also find cooling in East Antarctica over the 1969 – 2000 period. Whether our results or S09's results (or both) are consistent with the work of Thompson and Solomon is extratopical. We make no claims one way or the other.

{*Additional Technical and Editorial Comments*}

{*Reference is made in the supplemental text to verification tests done with 'random removal' of 5% of the data. If only single months or years are removed, then serial correlation is a problem, and this does not constitute a valid test. Random withholding of at least 5-years of data in sequence would be more appropriate. (Note that most RO10's verification results are based on more appropriate 25-year long data withholding exercises. Nevertheless, this problem should be addressed.)*}

6.A As stated in the caption for Table S2, the random withholding was done solely for the purpose of evaluating the deterioration in the regression results when withholding large quantities of data. We have now clarified in the text that no parameter decisions were made based on these results – especially since this was performed only at the screening stage. As explained in Problem 4, parameter decisions were based on gridded reconstruction cross-validation experiments, in which the verification targets were entirely excluded from the reconstruction.

{*The seasonal comparisons between S09 and RO10, like the annual comparisons, are quite misleading. For example, on page 22 it is stated that "While S09 report .. greatest warming in winter and spring, our results yield greatest warming in winter and fall." This implies that they do not find significant warming in spring, but of course they do find this, as is obvious in Figure 5. The only difference is in whether it is fall or winter that shows greater warming; that the order of importance switches (marginally) can hardly be considered a major difference in the results. Similarly, on page 23 it is stated that RO10's study "resolves general cooling of East Antarctica during the fall," implying that S09 gets a different result. But S09 also reports East Antarctic cooling in fall. Hence, there is no disagreement.*}

6.B S09 specifically claim greatest warming in Winter and Spring on both the Peninsula and West Antarctica. Our results show the greatest warming in Winter and Fall for the Peninsula and the portion of West Antarctica adjacent to the Peninsula, and additionally show greatest cooling in the Ross region and the portion of West Antarctica adjacent to Ross during Winter. This is, indeed quite different from S09, who show the greatest *warming* in the Ross area during this season. We fail to see how this implies that we do not find significant warming in Spring for these regions, especially as plots of the results are depicted right next to each other in Figure 5. In their text, S09 stated maximums; we compared our results to those maximums. They are different. Whether this difference is crucial to S09's sea ice hypothesis is, again, extratopical.

{*On page 21 it is stated that the results show "mild cooling to neutral trend in the Ross region" whereas as discussed above, RO10 themselves show that anything from cooling to significant warming may be valid, including trends greater than shown by S09. Moreover, in Spring, RO10 shows significant warming in the Ross Sea region even in their preferred reconstructions. Spring is emphasized in S09's results because it is highly relevant to sea ice changes (the greatest changes in sea ice have occurred in spring). RO10's are clearly in complete agreement in both sign and magnitude during this important season.*}

6.C The reviewer's misunderstanding of the results of our sensitivity tests was explained in Section 4. The importance of Spring warming to the sea ice hypothesis is extratopical.

{*The reproduction of S09's results (Figure 1 in RO10) could be eliminated to save journal space. It is appreciated that this was shown for the review process, but it is probably not necessary if the paper is published.*}

6.D We have moved Figure 1 from the main text and into the Supporting Information.

{*The Supplemental figures showing areas of insignificant trends in gray is hard to read. It might be better to use cross hatching, or to place contours around those areas with significant trends to help delineate the boundaries.*}

6.E We have amended figures S15 and S16 to be easier to read.

{*Concluding Remarks and Recommendations*}

{*Significant warming extends well beyond the Antarctic Peninsula to cover most of West Antarctica, an area of warming much larger than previously thought.*}

7.A We agree that the Peninsula warming extends into West Antarctica. We agree that portions of West Antarctica show statistically significant warming. However, both the overall pattern of temperature change and seasonal patterns are significantly different than S09 and are closer to previously performed reconstructions. The reviewer's suggested modifications to the text (other than those agreed to earlier in this response) are unnecessary, and, in general, would inaccurately represent our results.

{*Although this is partly offset by autumn cooling in East Antarctica, the continent-wide average near-surface temperature trend is positive [but not necessarily statistically significant, as is stated clearly in S09]*.}

7.B We have amended the abstract to state that we find a positive (albeit statistically insignificant) trend for East Antarctica and the continent from 1957 – 2006. We note that this was already stated in our results.

{*Simulations using a general circulation model reproduce the essential features of the spatial pattern and the long-term trend.*}

7.C The simulation shown by S09 does not reproduce the essential features of our reconstruction.

{*Cooling over much of East Antarctica did occur in recent decades, but was strongest during the short time interval considered in earlier studies (1969–2000 …).*}

7.D We never state otherwise, though we have amended the text to indicate that the S09 reconstruction shows cooling over this period.

{*Virtually all areas warmed between 1957 and 1980.*}

7.E While true of S09's reconstruction, this is not true of ours. An examination of Figure 4 shows significant areas of cooling during this timeframe, especially for the RLS reconstruction. The smaller area of cooling shown for the E-W reconstruction is

comparable in size and greater in magnitude than the Fall cooling in S09, which the reviewer has previously deemed significant. By the standards the reviewer has applied to S09's work, we feel the claim of "virtually all" is inaccurate.

{*…the greatest warming is in winter and spring [and fall], and in continental West Antarctica as well as on the peninsula.*}

7.F Stating that 3 out of 4 possible seasons show the "greatest warming" is arguably a misuse of the word "greatest". Regardless, even this unusually wide definition of "greatest" is contradicted by our results, as West Antarctica shows cooling in Winter and has the single season not mentioned by the reviewer (Summer) as one of the two most rapidly warming seasons.

{*The improvements to the Steig et al. methodology suggested by O'Donnell and colleagues could help to provide useful insights about the spatial details of Antarctic temperature trends, such as the cooling at the South Pole, and the probable greater warming trends on the Peninsula vs. elsewhere on the continent. The authors are to be congratulated on this aspect of their work. But in its present form, the manuscript is extremely misleading, giving the reader the false impression that multiple and serious errors exist in S09 and furthermore that such purported errors lead to dramatic differences in the results. The results are not very different at all with respect to all key conclusions in S09, and the differences are overwhelmingly due to a different choice of the number of retained eigenvectors. It bears repeating that the authors have chosen to emphasize results that are in greatest disagreement with S09, without convincingly defending the parameter choices required to obtain those results; and furthermore, that even with those questionable choices, each of the main conclusions in S09 are unaltered. A revised version of the manuscript should focus on the suggested methodological improvements, and should honestly and clearly report the impact of these changes, making it clear where the results and conclusions of Steig et al. are supported, as well as where meaningful differences may occur.*}

7.G As shown, multiple and serious errors *do* exist in S09 and these *do* lead to dramatic differences in the results. Our parameter choices are fully justified and objective. As shown previously, the reviewer has reached his conclusion both by inaccurately comparing our reconstructions to the detrended S09 reconstruction and by taking the sensitivity tests out of context. If by "main conclusions in S09" the reviewer is referring to whether our reconstruction supports the sea ice and other hypotheses, those conclusions are outside the scope of this work and we make no comment on them. If by "main conclusions in S09" the reviewer is referring to overall magnitude and patterns of temperature change, then we do, indeed, find substantial differences.

{*Elimination of vague and misleading language. A particularly egregious example is the statement that "only the Peninsula" shows significant warming, whereas all of RO10's results show significant warming in western West Antarctica, and also over large areas of East Antarctica. Various other examples are given above.*}

7.H We have clarified the text to indicate that the Peninsula side of West Antarctica shows statistically significant warming, and that points within regions can display results that differ from the regional average. However, it is a true statement that the area averages for West Antarctica display mixed significance and the averages for East Antarctica and the continent are not significant. This is neither vague nor misleading. We note that S09 make similar statements concerning East Antarctica and continent-wide trends in their own text.

{*The problematic and likely invalid cross-validation method that results in suspect values of k_{gnd} being used. Even if the currently-used values can somehow be justified, it still needs to be emphasized that any other reasonable choices result in reconstructions that show substantially better agreement with Steig et al. Given the highly questionable nature of the main reconstruction shown in the current version of the text, at least one other example needs to be clearly shown and described in the main text, not merely in the supplementary text.*}

7.I We have explained these misunderstandings in detail. Based on that, we disagree that any additional results need to be shown in the main text. If we did show additional results, we would naturally choose the other combinations of parameters that display the highest verification statistics for both West Antarctica and the continent. These generally show equivalent or lower trends than the reconstructions we have already selected (even for $k_{gnd} = 5$), as noted in ¶4.R.

{*Discussion of the broader context and comparison with modeling results needs to be corrected. In the current version of the manuscript, the implications regarding model (GCM) results are precisely backwards.*}

7.J There was no discussion of the broader context of agreement with models. The short discussion was limited to pointing out that our reconstructions disagreed with the model result S09 chose, which is an accurate observation. However, we agree that providing some context is useful and have amended the text as described in ¶5.F.

{*The revised version should eliminate extraneous and speculative discussions, such as the section regarding Chladni patterns.*}

7.K As noted in ¶3.N, this discussion has been removed from both the main text and the SI.

{*The text should accurately reflect the S09 methodology, and this section of the text could be made much more concise, since the methodology is already fully described in S09 and cited references. The confusing statements about the (A|b) matrix need to be corrected, and speculation and insinuation as to the reasons for Steig et al's choice to retain 3 eigenvectors should be avoided. Reference to statements on the RealClimate blog should be eliminated; RealClimate is not a journal of record. The claims about the "transfer of trends" from the Peninsula to the rest of Antarctica needs to be demonstrated or eliminated. If it can be shown to be valid, it still needs to be made clear that it is of minor or negligible impact, compared to the number of retained eigenvectors.*}

7.L

- We have amended the text to correct the references S09 used for eigenvector retention and removed statements indicating that similarity to physical processes motivated the decision
- The detailed description of the methodology is crucial in explaining the calibration issues with S09
- The S09 description of the augmented matrix is mathematically incorrect
- The reference to *RealClimate* has been removed
- The transfer of trends was demonstrated in Problem 2 of this response
- The impact of this transfer is not minor or negligible, and the text has been amended as indicated in Problem 2 to show this

{*Averaging over geographic regions that are both warming and cooling (in the reconstruction) is misleading and should be avoided. To average over all of West Antarctica simply because Steig et al. did so is not a good justification, and obscures the fact that RO10 find comparable or greater warming than S09 over large areas in both East and West Antarctica, even if the preferred reconstructions in RO10 were acceptable.*}

7.M Area averages which include warming and cooling regions are both commonplace and acceptable. We have added clarification that this occurs in our reconstructions. However, defining areas *post hoc* based on the resulting reconstruction trends is neither commonplace nor statistically valid. Confidence intervals calculated from such *post hoc* decisions are meaningless. We also respectfully suggest that the reviewer displays inconsistency with this requirement, as his comparisons to S09 seasonal patterns require averaging areas that include warming and cooling.

{*The relative contribution of different suggested methodological changes needs to be clearly and honestly presented. As already discussed above, the overwhelming difference between the reconstructions in S09 and RO10 is obviously the choice of eigenvectors retained. It is almost certainly not due to "improper calibration", or "improper determination of spatial structure", both of which are given equal weight in the Abstract of the paper.*}

7.N As indicated in ¶2.V, a table has been added to document the effect of these changes on trend magnitudes. Contrary to the assertions by the reviewer, all three modifications are of approximately equal importance.