This paper offers an analysis of the Antarctic temperature reconstructions published by Steig and others in *Nature* in 2009 (S09), based on the combination of AVHRR satellite data and ground-based weather station data, and proposes several methodological changes for building on that work. The paper suffers from serious shortcomings. The paper attempts to reach conclusions that are simply not substantiated by the underlying analysis. Indeed, the main thesis of the paper is that the conclusions reached in S09 are in error, whereas in fact the results largely confirm each of the conclusions in S09. Moreover, while most of the technical details appear sound, the discussion is misleading and confusing, and at least one critical aspect of the methodology is flawed. Although the analysis presented is thorough, and has the potential to make an important and useful contribution to the literature, the problems are serious and must be addressed, and another round of reviews conducted before the paper should be considered further for publication in the *Journal of Climate*.

Problems with the paper (hereafter RO10) fall into five basic classes:

Problem 1) RO10 present several different reconstructions of Antarctic temperature that differ from those in S09. Discussion of the differences between the RO10 and S09 reconstructions is misleading, generally emphasizing the differences and ignoring or obfuscating the similarities.

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).

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.

Problem 4) The choice of 'optimum' parameters to use for the reconstructions presented in the main text appears flawed, and at the very least creates a highly misleading picture of the differences between RO10 and S09's reconstructions.

Problem 5) Discussion of the implication of the results in RO10 for understanding of Antarctic temperature variability reflects a misreading of the literature on the subject, and includes several misstatements about what was actually said in S09. In particular, the reconstructions in RO10 are said to show greater conflict with climate model results than S09 does. If anything, the opposite is the case.

I address each of the above-listed items in turn, and follow with some additional technical questions and editorial suggestions for improving the paper.

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

There are three main issues here.

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

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.**

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. 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.

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.]

Problem 2) *Several reasons for differences between S09 and RO10 are presented, but no clarity is provided as to which of these differences dominates.*

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.

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. 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). 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.

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.

Problem 3) *Discussion of the technical details of what was done in S09 is misleading or incorrect in several instances.*

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.

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.

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.

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.

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.

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.

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.

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

RO10 uses several different truncation parameters for different aspects of their calculations. As already noted, the effective number of spatial patterns retained in the satellite data (determined by the combination of parameters k_{sat} and k_{RegEM} or c) is the dominant cause of differences between the reconstructions in RO10 and S09. However, the results in RO10 are also quite sensitive to the truncation parameter, k_{gnd} used for the infilling of missing data in the ground weather stations, which (unlike in S09) is done separately in RO10. RO10 find that the optimal choice of k_{gnd} for this infilling is 6, when considering the statistics of infilling the ground stations alone. When considering the overall reconstruction statistics, however, they find that, depending on the method used (covariance vs. correlation, E-W vs. RLS), a value of either k_{gnd} =6 or k_{gnd} = 7 is optimal.

For the reconstructions presented and discussed in the main text, RO10 find that using the values of k_{gnd} that optimize the reconstruction as a whole results, *uniquely*, in an insignificant warming trend in West Antarctica. While RO10 acknowledge that the sensitivity to k_{gnd} indicates that they have likely underestimated the amount of warming in West Antarctica, they do not explore why *particular* choices of k_{gnd} result in dramatically different results. Yet the answer almost certainly lies in Table S3 in the Supplementary Information, where it is shown the using $k_{gnd} > 5$ in a correlation setting, and $k_{gnd} > 4$ 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 k_{gnd} = 4 or 5 for *all* West Antarctic stations. This means that k_{gnd} >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 k_{gnd}.

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 k_{gnd}; 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 k_{gnd} 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 in

* Note that if the station records were substantially complete, this would not be a problem, but the critical ground stations in West Antarctic (e.g. Byrd and Mt. Siple) are substantially *incomplete* there is no question that the verification statistics are based on comparison with largely infilled data.

the 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'.

RO10 state that "the verification statistics for West Antarctic stations are still superior with $k_{gnd} = 7$." 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 k_{gnd}, but those same choices has a significant impacts on the reconstruction of the ground station data alone, why does RO10 use a value for k_{gnd} that is clearly inferior for the latter?

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.

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. 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 k_{gnd} values used in the main reconstructions in RO10 are suspect.

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.**

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 k_{gnd}) and as much as 0.25 °C/decade for RLS correlation, using $k_{gnd} = 5$, 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 ¼ to ½ 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.**

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. 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.

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 multi-decadal 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.

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.

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.

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.)

**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.

**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.

**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.

**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.

Concluding Remarks and Recommendations.

RO10 have done some useful and potentially publishable work. But the chief thesis of the paper is that the results in S09 are in error, whereas virtually all of the conclusions reached in S09 are fully supported by RO10, even using their preferred reconstructions. These include (taken directly from the text in S09):

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

*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].

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

*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 …).

*Virtually all areas warmed between 1957 and 1980.

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

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.

It will be especially important that the following specific items are addressed:

1) 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.

2) 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.

3) 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.

4) The revised version should eliminate extraneous and speculative discussions, such as the section regarding Chladni patterns.

5) 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.

6) 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) 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.