An Analysis of the Steig *et al.* (2009) Antarctic Temperature Reconstruction

Ryan O'Donnell
Mattawan, Michigan

Nicholas Lewis
Bath, United Kingdom

Steve McIntyre
Toronto, Canada

Jeff Condon
Chicago, Illinois

Corresponding author address: Ryan O'Donnell, 23522 Finch Avenue, Mattawan MI 49071.
E-mail: ode3197@yahoo.com

Abstract

A detailed analysis is presented of a recently published Antarctic temperature reconstruction method that combines satellite and ground information using a regularized expectation-maximization algorithm. Though the general reconstruction concept has merit, the method as described is susceptible to spurious results for both temperature trends and geographic distribution. The deficiencies include: (a) improper calibration of satellite data to ground data; (b) improper determination of spatial structure during infilling; and (c) use of an insufficient number of satellite-derived principal components. This study proposes two different methodologies to resolve these issues. One method utilizes temporal relationships between the satellite and ground data; the other combines ground data with only the spatial component of the satellite data. Both improved methods yield similar results, and these results disagree with the previous method in several important aspects. Rather than finding warming concentrated in West Antarctica, the improved methods find warming over the period of 1957-2006 to be concentrated in the Peninsula ($\approx 0.3^{o}C\ decade^{-1}$). The improved methods also show average trends for the continent, West Antarctica and East Antarctica that are approximately half that found using the unimproved method. While mixed results are obtained for trend significance in West Antarctica, the average trends for the continent and East Antarctica are not significant at the 5% level. Finally, 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.

1. Introduction

In a 2009 study published in *Nature*, Steig et al. (hereafter S09) present a novel reconstruction technique to extend Advanced Very High Resolution Radiometer (AVHRR) infrared satellite observations back to 1957 using manned ground station temperature information as predictors. Rather than providing only point-estimates of past temperatures, the method allows high-resolution, gridded estimates to be obtained for the entire Antarctic continent. Previous Antarctic gridded reconstructions (Chapman and Walsh 2007; Monaghan et al. 2008) relied on interpolation or kriging methods to estimate temperatures at non-instrumented points. In Chapman and Walsh (2007), interpolation was guided by correlation length scales calculated using the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) for ocean and coastal areas, and station-to-station pairs for the Antarctic interior. In Monaghan et al. (2008), the ERA-40 reanalysis data was utilized to provide the kriging field. In contrast, S09 perform multiple linear regression of satellite temporal data against ground data, and then *directly* recover gridded estimates using the satellite spatial structure – obviating the need for interpolation.

The S09 method involves the following major steps: a) decomposition of the cloud masked AVHRR data into principal components (PCs) and spatial eigenvectors; b) augmenting a matrix of station data starting in 1957 with the first three AVHRR PCs; c) estimating missing data in the augmented matrix with an infilling algorithm; d) extraction of the completely infilled PCs; and e) estimation of temperatures at all grid points by

reconstituting the extracted PCs with their corresponding spatial eigenvectors (Steig et al. 2009; Steig, personal communication).

Our approach to this topic begins with demonstrating replication of the S09 results. We discuss the S09 choice of infilling algorithm and inability of the algorithm to provide the necessary calibration function in Section 3. In Section 4 we show that the method used by S09 results in a different spatial structure being used for infilling than is present in the satellite data, which distorts the spatial distribution of temperature trends. Section 5 closes out the first half of the article by arguing that the choice of 3 principal components is suspect.

In the second half of this article, we present alternate reconstructions that address our concerns with S09. We outline the corrections to the methodology in Section 6. In Section 7, we discuss the primary features of our result, similarities and differences as compared to S09, and cross-validation statistics. Recommendations and conclusions are contained in Section 8. Additional details not covered in the main text are provided in the Supporting Information.

2. Replication of S09

We restrict our replication of the S09 process to steps that follow decomposition of the Advanced Very High Resolution Radiometer (AVHRR) data into 3 principal components (PCs). For replication, we utilized the archived READER data set (Turner et al. 2003) for ground data and the cloud masked AVHRR data set, both published by Steig

on his university website. Prior to use in the reconstruction, the READER and AVHRR data are converted to anomalies and scaled to unit variance.

For the period of 1957-2006, our replication yields linear trends in ^{o}C decade^{-1} of 0.12 for all grid cells, 0.10 for East Antarctica, 0.13 for the Peninsula and 0.20 for West Antarctica. These values are all within 0.01 of those obtained using the published T_{IR} reconstruction, with identical spatial and seasonal patterns of temperature change. The results of the replication are shown in Figure 1 immediately below those from the T_{IR} reconstruction. The reader should note that to allow broader comparisons, the S09 trends listed above were computed using traditional geographic boundaries rather than the *ad hoc* definitions used by S09 and therefore differ slightly from the trends reported in that study. The minor changes to geographic definitions do not impact any of our conclusions. The geographic boundaries used for this study may be found in the Supporting Information.

3. Calibration via infilling?

a. Sources of systematic error in the AVHRR data

The AVHRR instrument is carried aboard the NOAA Polar Pathfinder series of satellites. It is a multichannel sensor designed to provide imaging at both visible (channels 1 & 2) and infrared (channels 3-5) wavelengths as described by Fowler et al. (2009) at the National Snow and Ice Data Center. The form of the AVHRR data used by S09 is a set cloud masked in similar fashion to Comiso (2000), regridded to 50km by 50km resolution and presented as monthly means.

The AVHRR data is not a continuous set of measurements from a single instrument. Like other satellite imaging products, measurements from different satellites must be combined to produce a continuous record, which admits the possibility of splicing errors. Sensor degradation, calibration errors, time-of-observation drifts, atmospheric conditions, and cloud opacity at infrared wavelengths (Comiso 2000; Fowler et al. 2009; Gleason et al. 2002; Jiménez-Muñoz and Sobrino 2006; Jin and Treadon 2003; Sobrino et al. 2008; Trishchenko and Li 2001; Trishchenko 2002; Trishchenko et al. 2002) all contribute non-negligible measurement error. When combining with other sources (such as *in situ* measurements), an additional consideration is that the AVHRR instrument measures skin temperature rather than near-surface air temperature. These factors all highlight a need to calibrate the AVHRR data to ground data (or vice-versa) when either set is used to predict the other, as these types of systemic errors imply that the noise structure may be different between the sets. The mathematical description provided by S09 establishes the ground data as the independent variables and indicates that the infilling algorithm provides the calibration.

b. Description of the total least squares algorithm

The infilling method utilized by S09 is an implementation of truncated total least squares (TTLS) in a regularized expectation-maximization algorithm (RegEM) developed by Schneider (2001). The TTLS algorithm provides a solution to the linear model $\mathbf{Ax} = \mathbf{b}$, where both $\mathbf{A}$ and $\mathbf{b}$ are assumed to contain errors. S09 define an augmented matrix $\mathbf{X} = (\mathbf{A} \mid \mathbf{b})$, where $\mathbf{A}$ is said to represent the ground station data (predictor) and $\mathbf{b}$

is said to represent the AVHRR principal components to be estimated (predictand)[1]. Regularization is accomplished by performing a singular value decomposition of the correlation[2] matrix $\mathbf{C}$ with k retained eigenvectors (Mann et al. 2007). From Schneider (2001), this yields the right singular vectors and squared eigenvalues of the $n \times p$ matrix of observations $\tilde{\mathbf{X}} = \mathbf{X} / \tilde{\mathbf{s}}$ (where $\tilde{\mathbf{s}}$ is a vector of unbiased standard deviation estimators), since:

$$\left(\tilde{\mathbf{X}}\right) = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}} \qquad (1)$$

$$\left(\mathbf{C}\right) = \mathbf{V}\mathbf{\Lambda}^{2}\mathbf{V}^{\mathrm{T}} \qquad (2)$$

We may then partition $\mathbf{V}$ into subspaces, where rows 1 and 2 indicate predictor and predictand, and columns 1 and 2 indicate eigenvectors $\left(1...k\right)$ and $\left(k+1...n\right)$, respectively:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{1,1} & \mathbf{V}_{1,2} \\ \mathbf{V}_{2,1} & \mathbf{V}_{2,2} \end{pmatrix} \qquad (3)$$

The TLS solution is given by Fierro et al. (1997), where symbol † indicates the generalized inverse:

$$\mathbf{x}_{k} = \left(\mathbf{V}_{1,1}^{\mathrm{T}}\right)^{\dagger} \mathbf{V}_{2,1}^{\mathrm{T}} \qquad (4)$$

We can now estimate the missing values in $\tilde{\mathbf{X}}$ for any moment in time j:

$$\tilde{\hat{\mathbf{b}}}_{j} = \tilde{\mathbf{A}}_{j}\mathbf{x}_{j,k} \qquad (5)$$

[1] The use of the words "said to represent" is deliberate, as we will show that the S09 method does not mathematically follow the definitions presented in their study.
[2] Both Schneider (2001) and Mann et al. (2007) describe the algorithm as using the covariance matrix. However, RegEM as used by S09 scales the covariance matrix to correlation prior to regularization.

where the original data $\mathbf{A}$ and estimated data $\hat{\mathbf{b}}$ can be recovered via rescaling by the factor $\tilde{\mathbf{s}}$.

Alternatively, rather than limit the estimation to subspace $\mathbf{V}_{2,1}$, we can replace $\mathbf{V}_{2,1}^{\mathrm{T}}$ in equation (4) with $\mathbf{V}_{1\ldots k}^{\mathrm{T}} = \begin{pmatrix} \mathbf{V}_{1,1} \\ \mathbf{V}_{2,1} \end{pmatrix}$, yielding a full set of statistical weights to provide estimates for both missing and actual values:

$$\tilde{\tilde{\mathbf{X}}} = \left(\tilde{\tilde{\mathbf{A}}} \;\; \tilde{\tilde{\mathbf{b}}} \right) = \tilde{\mathbf{A}}_j (\mathbf{V}_{1,1}^{\mathrm{T}})^{\dagger} \mathbf{V}_{1\ldots k}^{\mathrm{T}} \qquad (6)$$

As all eigenvectors greater than k are discarded, subspace $\mathbf{V}_{2,2}$ provides an estimate of the predictand residuals, and thus an estimate of the residual correlation matrix is given by:

$$\mathbf{C}_{\mathrm{res}} = \mathbf{V}_{2,2} \mathbf{\Lambda}_{k+1\ldots n}^{2} \mathbf{V}_{2,2} \qquad (7)$$

RegEM defines a new correlation matrix using the original data $\mathbf{A}$, the newly estimated data $\hat{\mathbf{b}}$ and the estimated correlation matrix of the residuals $\mathbf{C}_{\mathrm{res}}$. A new solution is then computed. The algorithm iterates until the rms change in $\hat{\mathbf{b}}$ is less than a pre-defined tolerance.

c. Practical and theoretical difficulties with the S09 approach

A critical aspect of the S09 methodology is that *both* the satellite PCs and the station data matrix are *incomplete* when supplied to RegEM. Therefore, in the satellite period (1982-2006), the PCs appear in matrix $\mathbf{A}$ (not $\mathbf{b}$) and are directly used to predict missing ground station values. As $\mathbf{C}$ is computed using $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{C}_{\mathrm{res}}$, this influence

affects the prediction for $\mathbf{b}$ at *all* times; hence, the effect of the PCs on estimating ground station data during 1982-2006 influences the prediction of the PCs during 1957-1981.

A second source of mutual influence is regularization itself. Regularization destroys the orthogonality of the PCs (Table 1). From equation (4), the ground data will be estimated using correlated – not orthogonal – PCs, which influences the estimation of the PCs on the subsequent iteration and suggests the possibility of mutual reinforcement. In the case of PC 2 and PC 3, the correlation following regularization at the final iteration (using a convergence tolerance of 0.005) is a factor of 2.5 higher than that following the initial regularization (-0.2501 vs. -0.1001). The effect is *a priori* unpredictable with respect to changes in k and becomes more dramatic if additional PCs are included. This places an undesirable practical limit on the amount of information that may be used in the S09 method.

These observations present a major difficulty in ascribing a calibration function to RegEM. One cannot claim that a regression that allows mutual interaction between multiple dependent variables (PCs) constitutes a calibration to the independent variables (ground data). One is forced to conclude that when the independent variables have missing observations, or when the dependent variables each have differing temporal completeness, RegEM cannot provide a valid calibration. 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.

A final point is that the concept of using a *total least squares* algorithm (which minimizes the errors in both the predictors and predictands) to regress PCs against

temperature observations when both are incomplete presents a theoretical difficulty. The error in a PC (which represents the temporal component of a temperature *field*) does not mean the same thing to the reconstruction as an error in an observation (which represents temperature at a *point*). Additionally, errors in the principal components are more likely to be systemic in nature (section 3.a), which implies a greater likelihood of non-climatic signals affecting components with larger eigenvalues than in the ground data. In that case, the filtering effect of the truncation parameter k will be less effective for the principal components, which translates into additional estimation error for the ground stations when the PCs appear in $\mathbf{A}$.

4. Spatial structure considerations

Another source of error is the difference in spatial structure used to infill the PCs and the corresponding original AVHRR eigenvectors. The assumption that the spatial structure be similar is implicit in the S09 method, which recovers anomaly estimates for all grid cells by combining the infilled PCs with the unaltered AVHRR spatial eigenvectors. Differences in spatial structure result in a geographic transfer of information from the dominant spatial features in the EOFs used to estimate missing values (imputation EOFs) to the dominant spatial features in the AVHRR eigenvectors.

As equation (6) yields a fully populated $n \times p$ matrix of estimates $\tilde{\tilde{\mathbf{X}}}$, we may then obtain the imputation EOFs via SVD. For a given series of estimates i, the contribution of a particular imputation EOF to the series is given by the spatial weight $v_{i,j}$. Neglecting the contribution of a series to itself, we can then approximate the overall

spatial structure used to estimate a series by taking the vector sum of the imputation spatial EOFs with the series removed ($\mathbf{L}_{-i,j}$), scaled by the series spatial weights $v_{i,j}$ and a normalization constant c:

$$v_i = c\sum_{j=1}^{k} v_{i,j}\mathbf{L}_{-i,j} \qquad (8)$$

Figure 2 shows the comparison of the imputation weights vs. AVHRR weights at the grid cells nearest the station locations for PCs 1-3. While PC 1 and PC 3 demonstrate a similar spatial arrangement between AVHRR and RegEM for the Peninsula and East Antarctica, five East Antarctic stations in PC 2 are used in the opposite orientation and less than half are weighted similarly. None of the PCs demonstrate agreement between RegEM and AVHRR in West Antarctica, with the largest deviations occurring in the Ross Ice Shelf region. It is also clear that the PCs have a non-negligible contribution to each other, especially in the case of PC 3 (where the magnitude of the contribution of PC 2 is greater than over half of the actual ground stations). Lastly, PC 1 – which primarily determines the average temperature trend for the continent – displays a noticeably higher set of weights for the Peninsula stations and a noticeably lower set of weights for East Antarctica in RegEM than is present in the AVHRR data. This necessarily results in a redistribution of the Peninsula trend across the entire continent by the first AVHRR spatial eigenvector.

5. Significant principal components

A critical aspect of the reconstruction method employed by S09 – which is essentially principal components regression – is the choice of the truncation parameter k_{sat} for the satellite data and k_{RegEM} for the infilling of the augmented ground station/satellite PC matrix. In their study, S09 state that they use a combination of physical meaningfulness and "statistical separability" to define k_{sat}, which they determined to be 3. The latter was later clarified by Steig on the blog *RealClimate* to mean breaks in the eigenvalue spectrum that exceed the sampling uncertainty calculation of North et al. (1982). They determine k_{RegEM} by stating that a decomposition of the ground data produces similar results to the satellite data and again mention the concept of separable components.

The direct physical interpretation of the first 3 satellite components as meaningful (and no others) is suspect. Since it is unlikely that physically orthogonal modes will also be mathematically orthogonal, the well-known result is that physical modes are mixed among principal components and that the boundary conditions can have a greater impact on the spatial structure than the physical modes themselves (Aires, Rossow, and Chedin 2002; Buell 1979; North et al. 1982; North 1984). Indeed, the statistical authority cited by S09 as their source for determining k discusses this mixing in detail, and, based on the high degree of overlap in the error estimates for eigenvalues 2 and 3, it is almost certain that they represent a mix of physical modes and are not each a single physical mode.

While S09 state that the first 3 spatial eigenvectors visually appear similar to physical modes – specifically the SAM index and zonal 3-wave pattern – we argue that they appear just as similar to sets of standing waves (such as Chladni patterns) that would

be expected from decomposing a spatially autocorrelated data set with the shape of Antarctica. It is relevant to note that the SAM and NAM indices themselves are the low-order spatial eigenvectors from decompositions of atmospheric pressure time series that have hemispherical (annular) shapes (e.g., Thompson and Wallace 1998; Thompson and Wallace 2000, Thompson, Wallace, and Hegerl 2000; Baldwin 2001). Given the dependence of spatial EOF patterns on boundary conditions, it is not surprising that a decomposition of spatially autocorrelated temperatures yields the same shapes when using the roughly disk-shaped Antarctic continent as the spatial constraint. This is not necessarily indicative of direct physical meaningfulness; rather, it can be indicative of using spatial boundaries with similar shapes during decomposition. Examination of nodal patterns supports this supposition (Supporting Information).

Extensive testing on synthetic data by Compagnucci and Richman (2007) demonstrate that many of the familiar single-pole and dipole patterns that appear in orthogonal decompositions of atmospheric data sets are quite possibly mathematical artifacts, and that even survival of the feature after rotation of the extracted components does not necessarily indicate physical meaningfulness. They also note cases where significant mixing occurs even when the North et al. (1982) criterion is met. Since eigenvectors #2 and #3 plainly do not meet the North et al. (1982) criterion regardless and S09 do not show robustness of the spatial patterns to rotations (both of which would simply *indicate* – not prove – *possible* physical meaningfulness), the validity of this truncation criterion is highly questionable. Neither our analysis nor established statistical theory supports the notion that the choice of truncation parameter should be influenced based on visual similarity with known or theorized physical modes – especially when the

point of comparison is to mathematical objects that themselves may be partially or wholly artifacts of EOF analysis.

Secondly, use of the sampling error calculation to determine the truncation point is discussed in North (1982), wherein the sampling error calculation is specifically stated as *not* providing any guidance for determining the truncation point. Rather, the sampling error calculation is proposed as a rule-of-thumb for identifying and retaining all eigenvectors of a degenerate multiplet rather than retaining some and discarding others with an arbitrary choice of k. North (1982) also discusses the case where the purpose of the decomposition is to obtain a smaller basis for representing the raw data (as is the case for S09). As long as the purpose of the decomposition is not to uncover the true underlying physical modes, the impact of splitting a multiplet is limited to the amount of variance that multiplet explains in the original data. While this is undesirable because it adds to the error associated with the representation, it is certainly no more or less undesirable than adding an equivalent amount of error to the representation through early truncation. The rationale advanced by S09 for selecting 3 eigenvectors based solely on sampling error is incomplete because it ignores any comparison to the error in representation due to early truncation. We will demonstrate that this led S09 to choose suboptimal truncation parameters for the both the satellite data and for RegEM.

6. Corrections to methodology

a. Spatial and temporal assumptions

Performing a reconstruction of this type necessarily requires assumptions that if not met potentially invalidate the results. It is important, then, that the assumptions be ones that are *likely* to hold, at least approximately. A stated assumption of S09 is that the AVHRR data provides a reasonably accurate spatial representation of temperatures. However, by retaining the 1982-2006 portion of the AVHRR PCs unchanged, S09 implicitly make the additional assumption that the AVHRR data provides a reasonably accurate *temporal* representation of temperatures. We find that, for reasons discussed in Section 3 and further detailed in the Supporting Information, the latter assumption is not likely to hold. To correct this, our approach shares the spatial assumption of S09 and assumes that the *ground data* provides a more accurate temporal representation of temperature.

The improved temporal assumption may be mathematically expressed in one of two ways. If an infilling algorithm is used, one may make use of equation (6) to extract regression estimates for the AVHRR PCs at all times rather than only times where the original PCs are incomplete. The estimates obtained using equation (6) may then be reconstituted with the corresponding spatial eigenvectors to obtain the reconstruction. In this way, the reconstruction contains no direct AVHRR temporal information. An alternative means of expressing the revised temporal assumption is to perform the regression by using *only* spatial information and exclude the PCs altogether (Section 6.d).

b. Calibration

As shown in Section 3, RegEM is not capable of providing a valid calibration function if either the independent variables have missing observations or the dependent

variables have differing temporal completeness. Given that the dependent variables are simply the first k_{sat} principal components derived from the AVHRR data and all have the same temporal coverage, the latter concern does not apply. The former, however, requires resolution. We address this issue by simply infilling a matrix composed solely of ground station data (analogous to the AWS reconstruction in S09). The completed matrix is then augmented with the AVHRR PCs and the PCs are infilled. This prevents the estimation of the PCs from influencing each other via their influence on the estimation of ground data, as the estimation of ground data has already been completed. It additionally helps resolve the theoretical difficulty of errors in the PCs meaning something different than errors in the ground stations, as the PCs are never used to estimate ground temperatures.

To make this modification, we must consider the choice of truncation parameter and which ground stations to use. For clarity, we will denote the truncation parameter associated with infilling *only* ground data (i.e., excluding any satellite data) as k_{gnd}. Both k_{gnd} and the optimal station selection are determined by running a series of cross-validation experiments (Table 2). The experiments compare the performance of 14 different station sets with permutations including covariance and correlation networks, values of k_{gnd} ranging from 1 to 10, and two different infilling algorithms: (a) an implementation of RegEM TTLS in the R programming language; and, (b) an iterative truncated SVD (TSVD) approach similar to the DINEOF routine (Alvera-Azcárate et al. 2009; Beckers and Rixen 2003; Beckers et al. 2006; Beckers, personal communication) developed for infilling cloud-masked data sets (Supporting Information).

c. Spatial structure considerations when regressing principal components

One way to resolve the issue of differing spatial structures between the ground station infilling and the AVHRR decomposition is to scale the ground stations in RegEM by the corresponding AVHRR spatial eigenvector weights when estimating the PCs. This requires that each PC be infilled separately and has the added benefit of entirely resolving the issue of mutual reinforcement noted in Section 3. Following the infilling, we can again make use of Equation (8) to evaluate the difference in spatial structure between the AVHRR data and the spatial structure used to infill the PCs. Results for the first 3 PCs (to provide a direct comparison to Fig. 2) are provided in Fig. 3.

There is a noticeable improvement in the match between the structure for infilling and that of the AVHRR data as compared to the S09 method. The rms differences in the normalized spatial weights for this method are 0.91, 0.56, and 0.44 for PCs 1-3, respectively. For S09, the corresponding rms differences are 0.93, 0.98, and 1.13. We note that the spatial structure in PC 1 is very similar between the AVHRR data and RegEM for our method with the exception of a nearly constant offset. We also note that, as expected, our method greatly reduces the percentage of stations used in the opposite orientation. When this does occur, it is with lower-weight stations than in the S09 method, which lessens the impact on the reconstruction. We denote this method as the *eigenvector-weighted* (E-W) method.

d. Eliminating use of principal components

A more elegant means to resolve the difference in spatial structure is to avoid using the AVHRR PCs at all. This is the simplest solution, as it simultaneously

eliminates all of the concerns discussed in 6.a through 6.c above. Since we assume the spatial structure to be accurate, the most efficient way to perform the reconstruction is to directly regress the ground station data against the AVHRR spatial structure. To do this, we first define our spatial EOFs as $\mathbf{L}_k = \sqrt{1/\tilde{n}}\left(\mathbf{\Lambda}_k \mathbf{V}_k\right)$, where $\mathbf{\Lambda}_k$ represents the AVHRR eigenvalues $1...k$, $\mathbf{V}_k$ represents the corresponding spatial eigenvectors, and $\tilde{n}$ represents the effective degrees of freedom. We may then define a matrix of ground station observations $\mathbf{Y}$ and write:

$$\mathbf{La} = \mathbf{Y} \qquad (9)$$

The regularized least squares solution can be found in Lawson and Hanson (1974), where a vector solution is computed separately for each time j in matrices $\mathbf{a}$ and $\mathbf{Y}$:

$$\mathbf{a} = \left(\mathbf{L}^{\mathrm{T}}\mathbf{L} + \mathbf{I}\mu^2\right)^{-1}\mathbf{L}^{\mathrm{T}}\mathbf{Y} \qquad (10)$$

As we do not know the proper regularization parameter from any *a priori* physical arguments, we determine the regularization parameter $\mu^2 = c\mu_0^2$ through a series of cross validation experiments. Parameter μ_0^2 represents the rms error between $\mathbf{Y}$ and $\mathbf{La}$ at the station locations, and the scaling constant c is the parameter that is varied. For simplicity, we assume that the noise on the system is Gaussian in nature. We then impose an additional constraint that μ^2 should produce a reconstruction where the same value for μ_0^2 is obtained in both the calibration and verification periods, which drives determining μ^2 via iterative estimation (Fitzpatrick 1991). Given that the regularization parameter μ^2 can be interpreted as an assumed ratio of system mean squared measurement error to our prior assumption of Gaussian noise (Fitzpatrick 1991; Sima

2006), this yields the reasonable physical constraint the error and noise be approximately constant over the analysis timeframe. We choose the combination $c\mu_0^2$ that minimizes the error in the withheld data and note that this combination may be interpreted as the maximum likelihood estimation of the true ratio of system measurement error and noise (Fitzpatrick 1991). We denote this method as the *regularized least squares* (RLS) method.

e. Determining truncation parameters

The final correction, which applies to both the E-W and RLS methods, is to determine the optimum truncation point for both the ground data and full-grid reconstruction through a series of cross validation experiments. This provides an objective criterion for determining important modes without resorting to heuristical tools (bootstrapped eigenvalue/eigenvector, broken stick, scree plots, etc.) or subjective arguments (visual similarity to known/theorized physical modes) that can give vastly different answers for a given set of data. The cross-validation criterion is simple and objective: those modes which improve the prediction of withheld data are retained.

To avoid confusion, we will use k_{gnd} to refer to the truncation parameter for the initial ground data infilling, k_{sat} to denote the number of retained AVHRR components, and k_{RegEM} to denote the truncation parameter used by the infilling algorithm to estimate each AVHRR component (applicable only to the E-W reconstructions). As before, we use cross validation testing to determine the optimum settings for all truncation parameters and, for the RLS reconstructions, the scaling parameter c. Table 3 contains

the permutations used. Flowcharts depicting the methods and cross validation experiments are available in the Supporting Information.

7. Results

a. Optimal parameters

The ground station cross validation experiments yield an optimal station set consisting of all on-grid READER stations with at least 96 months of data. A total of 63 stations (35 AWS and 28 manned ground stations) are included. The full-grid cross validation experiments, utilizing this set, yield the optimal settings summarized in Table 4. All four reconstruction variants have similar magnitudes and spatial distributions of temperature trends, and all have similar cross validation statistics. Although the covariance network for RLS (Supporting Information) demonstrates slightly higher verification statistics, we use the correlation version for the main text for three reasons. One is that S09 is calculated in a correlation setting; the second is that we found both the ground-only and E-W reconstructions demonstrate superior verification performance in a correlation setting. The final and most compelling reason is that the covariance reconstructions for all cases (E-W, RLS and attempts to perform the S09 procedure in a covariance setting) are significantly more unstable and subject to overfitting than the correlation counterparts. The primary source of the instability is the ground station infilling using a covariance network (Supporting Information).

b. Temporal and spatial patterns of temperature change

While we do find overall warming of the continent, the continental average is not significant at the 5% level (≈0.06 +/- 0.07 ^{o}C decade^{-1}) [3], nor is the warming in East Antarctica (≈0.05 +/- 0.09). The Peninsula is the only region that consistently demonstrates a statistically significant trend. The reconstructions also provide evidence of cooling in various parts of Antarctica for all time frames analyzed. Table 5 summarizes the differences in trend magnitude by region between the RLS, E-W and S09 reconstructions. Figure 4 displays the comparison between the spatial patterns of temperature change for the RLS, E-W and S09 reconstructions using the major subperiods that appear in the S09 text.

One feature that is qualitatively similar to S09 is a strong indication that the Peninsula warming does in fact extend into continental West Antarctica (albeit with a lesser magnitude than what S09 report). However, we find that the warming is concentrated on the Peninsula side, not in the Ross region of West Antarctica. Because of this, the match between reconstructions and the 1957-1981 and 1979-2003 patterns of temperature change from the GISS ModelE simulations is poorer than what is presented in the S09 text. In particular, S09 note that both the T_{IR} reconstruction and the ModelE results show stronger and more persistent warming in continental West Antarctica than on the Peninsula, which is not supported by our results or the historical ground station information. Our results – including the strong Peninsula warming, mild cooling to neutral trend in the Ross region, and generally insignificant trends elsewhere on the continent – compare more favorably to Chapman and Walsh (2007) and Monaghan et al. (2008) than S09.

[3] All uncertainty intervals in this study are 95% confidence intervals, with degrees of freedom corrected for AR(1) serial correlation of the residuals (Santer et al. 2000).

While we find that West Antarctica displays significantly less warming than the Peninsula, there are important uncertainties in our result. For the correlation RLS and E-W reconstructions pictured in the main text, the West Antarctic trend is ≈0.05 +/- 0.07 ^{o}C decade^{-1}. Both the magnitude and statistical significance of the trend are highly dependent on the truncation parameter for the ground station infilling. Table 6 summarizes this behavior for the values of k_{gnd} that were used in the full-grid cross validation experiments. For the full 63-station reconstructions, only $k_{gnd} = 7$ yields an insignificant trend in West Antarctica, with a magnitude about half of what is computed using other values for k_{gnd}. While the verification statistics for West Antarctic stations (like the remainder of the continent) are still superior with $k_{gnd} = 7$, the differences are minor. This indicates that the West Antarctic trend for $k_{gnd} = 7$ may be artificially low. We cautiously conclude that the trend in West Antarctica is likely significant and may be closer to 0.10 ^{o}C decade^{-1} than 0.05 ^{o}C decade^{-1}. While this general result (stronger trends in West Antarctica than East Antarctica) is similar to S09, the magnitudes of the trends in both regions are approximately ¼ to ½ of that reported by S09.

In comparing seasonal patterns of change, substantial differences between our results and S09 are apparent (Figure 5). While S09 report that both the Peninsula and West Antarctica show the greatest warming in winter and spring, our results yield the greatest warming in winter and fall for the Peninsula and the portion of West Antarctica adjacent to the Peninsula. In contrast, the Ross region of West Antarctica shows the greatest *cooling* during these same seasons. Chapman and Walsh (2007) and Monaghan et al. (2008) find similar patterns, with maximum Peninsula warming during winter and

fall (both studies) and maximum Ross region cooling during winter and fall (Monaghan et al. 2008) over slightly different periods (1958 – 2002 and 1960 – 2005, respectively).

In addition to resolving general cooling of East Antarctica during the fall, we also find significant cooling at the Antarctic pole during winter. This corresponds well to seasonal trends at Amundsen-Scott (from the READER archive) of -0.34, +0.03, -0.00, and -0.01 ^{o}C decade^{-1} during 1957 – 2006 for winter, spring, summer and fall, respectively. The result contrasts sharply with S09, who show the greatest *warming* occurring at the pole in winter. It differs to a lesser extent from the Monaghan et al. (2008) result of approximately neutral winter trends at the pole from 1960 – 2005, and matches well with Chapman and Walsh (2007), who also find cooling at the pole during all seasons, with a maximum cooling trend during winter in the 1958 – 2002 timeframe. We note that the seasonal patterns are not unique to the $k_{\mathrm{gnd}} = 7$ solution; the patterns are common to all four analyzed choices for k_{gnd}.

c. Verification statistics

Verification statistics are calculated by comparing reconstructed temperatures to station data that is withheld from the reconstruction. Statistics calculated are rms error (μ_{rms}), correlation coefficient (r), and coefficient-of-efficiency (CE). Reduction-of-error (RE) statistics are undefined as these stations are entirely withheld; hence, no calibration period exists. For the primary 63-station set selected for use in the RLS and E-W reconstructions, there are 24 unused on-grid stations available for verification. We also conduct reconstructions using the 28 stations with the longest records, and use the

additional 35 withheld stations as verification targets. This provides two independent sets of stations for verification.

Table 6 shows the mean station μ_{rms}, r and CE values for the RLS and E-W reconstructions using optimum settings and compares them to values obtained using the T_{IR} reconstruction from S09. Our methods demonstrate significantly improved verification skill to ground data over the S09 reconstruction. We additionally conduct Monte Carlo experiments using the mean, variance and lag-1 autocorrelation coefficients for each station and find that our results exceed the 99th percentile for all stations. Full statistics are available in the Supporting Information.

8. Conclusions and recommendations

S09 present a novel means of using an infilling algorithm to produce a high-resolution gridded reconstruction of Antarctic temperatures using ground and satellite data. We have shown that the method has three primary areas of concern. The first concern is that S09 do not perform an acceptable calibration prior to using the AVHRR data to estimate missing ground temperature values. This can result in unnecessary estimation error and mutual reinforcement of error in the satellite data. S09 also do not use the same spatial structure to infill the PCs as they use to recover the gridded estimates, which transfers trend information from the Peninsula to the remainder of the continent. The final major concern is that S09 use non-standard, subjective criteria to determine truncation parameters. This causes S09 to discard an excessive amount of

satellite information, with the result that the Peninsula stations unduly influence seasonal patterns and trend magnitudes in regions outside the Peninsula.

We demonstrate that these issues have a material impact on the results. When resolved, the results obtained differ from S09 in several key aspects. Average 1957 – 2006 temperature trends for the continent, East Antarctica and West Antarctica are halved. Warming is concentrated in the Peninsula, not in West Antarctica. The seasonal patterns of temperature change are more complex, with portions of West Antarctica, the pole, and the Weddel region demonstrating behavior opposite of that reported by S09. While we do find support for the S09 conclusion that the Peninsula warming extends into continental West Antarctica, we do not find support for the notion of greater and more consistent warming in West Antarctica than in the Peninsula. Instead we show the Peninsula warming to be greater by a factor of 3 (or more). As a corollary, the discrepancy between the coupled GISS ModelE runs cited by S09 and Antarctic temperatures are greater than indicated by S09.

Though we find the general concept of regressing the satellite principal components against ground information using an infilling algorithm to have merit, we note that in cases where the temporal component of a data set may be suspect, a simpler method using only the spatial information may provide more accurate results. This method also presents itself as a diagnostic tool; one could easily compare results between temporal and spatial methods. Discrepancies in the results could then be investigated in detail. This has applicability beyond temperature reconstructions. The concept may be used for any problem that requires both temporal and spatial analysis of incomplete data sets where the temporal and spatial information are derived from different sources.

Finally, we recommend that more study be undertaken to resolve the significant discrepancy in temporal evolution between the AVHRR data set used by S09 and temperatures measured at ground station locations. Though the scope of this work limits our analysis to a single data set, the potential sources for error outlined in the main text and Supporting Information are of a nature that suggest similar problems may exist in other AVHRR temperature products, Antarctic and otherwise.

Acknowledgments.

We gratefully acknowledge assistance from Hu McCullough, Roman Mureika, Jean-Marie Beckers, and Eric Steig.

References

Aires, F., W. B. Rossow, and A. Chedin, 2002: Rotation of EOFs by the Independent Component Analysis: Toward a Solution of the Mixing Problem in the Decomposition of Geophysical Time Series. *J. Atmos. Sci.*, **59**, 111–123, doi:10.1175/1520-0469(2002)059<0111:ROEBTI>2.0.CO;2

Alvera-Azcárate, A. Barth, D. Sirjacobs, J. –M. Beckers, 2009: Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. *Ocean Science*, **5**, 475–485

Baldwin, M. P., 2001: Annular modes in global daily surface pressure. *Geophys. Res. Lett.*, **28**, 4115 – 4118, doi:10.1029/2001GL013564

Beckers, J. –M., and M. Rixen, 2003: EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *J. Atmos. Oceanic Technol.*, **20**, 1839–1956, doi:10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2

Beckers, J. –M., A. Barth, and A. Alvera-Azcárate, 2006: DINEOF reconstruction of clouded images including error maps – application to the Sea Surface Temperature around Corsican Island. *Ocean Science*, **2**, 183–199

Buell, C. E., 1979: On the physical interpretation of empirical orthogonal functions. *6th Conference on Probability and Statistics in Atmospheric Science*, Banff, Alberta, Amer. Meteor. Soc., 112–177

Chapman, W. L., and J. E. Walsh, 2007: A synthesis of Antarctic temperatures. *J. Climate*, **20**, 4096–4117, doi:10.1175/JCLI4236.1

Comiso, J. C., 2000: Variability and trends in Antarctic surface temperatures from in situ and satellite infrared measurements. *J. Climate*, **13**, 1674–1696, doi:10.1175/1520-0442(2000)013<1674:VATIAS>2.0.CO;2

Compangnucci, R. H., and M. B. Richman, 2007: Can principal component analysis provide atmospheric circulation or teleconnection patterns? *Int. J. Climatol.*, **28**, 703–726, doi:10.1002/joc.1574

Fierro, R. D., G. H. Golub, P. C. Hansen, and D. P. O'Leary, 1997: Regularization by Truncated Total Least Squares. *SIAM Journal on Scientific Computing*, **18**, 1223–1241

Fitzpatrick, B. G., 1991: Bayesian analysis in inverse problems. *Inverse Problems*, **7**, 675–702, doi:10.1088/0266-5611/7/5/003

Fowler, C., J. Maslanik, T. Haran, T. Scambos, J. Key, and W. Emery, cited 2009: AVHRR Polar Pathfinder Twice-daily 5 km EASE-Grid Composites V003. [http://nsidc.org/data/docs/daac/nsidc0066_avhrr_5km.gd.html]

Gleason, A. C., S. D. Prince, S. J. Goetz, and J. Small, 2002: Effects of orbital drift on land surface temperature measured by AVHRR thermal sensors. *Remote Sens. Environ.*, **79**, 147–165, doi:10.1016/S0034-4257(01)00269-3

Jiménez-Muñoz, J. C., and J. A. Sobrino, 2006: Error sources on the land surface temperature retrieved from thermal infrared single channel remote sensing data. *Int. J. Remote Sens.*, **27**, 999–1014, doi:10.1080/01431160500075907

Jin, M., and R. E. Treadon, 2003: Correcting the orbit drift effect on AVHRR land surface skin temperature measurements. *Int. J. Remote Sens.*, **24**, 4543–4558, doi:10.1080/0143116031000095943

Lathauwer, L. D., B. D. Moor, and J. Vandewalle, 2000: An introduction to independent component analysis. *Journal of Chemometrics*, **14**, 123–149, doi:10.1002/1099-128X(200005/06)14:3<123::AID-CEM589>3.0.CO;2-1

Lawson, C. L., and R. J. Hanson, 1974: *Solving least squares problems*. Prentice-Hall, 340 pp.

Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2007: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, doi:10.1029/2006JD008272

Monaghan, A.J., D. H. Bromwich, W. Chapman, and J. C. Comiso, 2008: Recent variability and trends of Antarctic near-surface temperature. *J. Geophys. Res.*, **113**, doi:1029/2007JD009094

North, G. R., T. L. Bell, R. F. Cahalan, and F. T. Moeng, 1982: Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Mon. Wea. Rev.*, **110**, 699–706, doi:10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2

North, G. R., 1984: Empirical Orthogonal Functions and Normal Modes, *J. Atmos. Sci.*, **41**, 879–887, doi:10.1175/1520-0469(1984)041<0879:EOFANM>2.0.CO;2

Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor, 2000: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.*, **105**, 7337–7356, doi:10.1029/1999JD901105

Schneider, T., 2001: Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. *J. Climate*, **14**, 853–871, doi:10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2

Sima, D. M., 2006: Regularization techniques in model fitting and parameter estimation. Ph.D dissertation, Departement Elektrotechniek, Katholieke Universiteit Leuven, 195 pp. [Available from Katholieke Universiteit Leuven, Electrical Engineering Department, ESAT-SCD/SISTA, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, and on-line at ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/dsima/reports/thesisDianaSima.html]

Sobrino, J. A., Y. Julien, M. Atitar, and F. Nerry, 2008: NOAA-AVHRR Orbital Drift Correction From Solar Zenithal Angle Data. *Geoscience and Remote Sensing, IEEE Transactions on*, **46**, 4014–4019, doi:10.1109/TGRS.2008.2000798

Steig, E. J., D. P. Schneider, S. D. Rutherford, M. E. Mann, J. C. Comiso, and D. T. Shindell, 2009: Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature*, **457**, 459–463, doi:10.1038/nature07669

Trishchenko, A. P., and Z. Li, 2001: A method for the correction of AVHRR onboard IR calibration in the event of short-term radiative contamination. *Int. J. Remote Sens.*, **22**, 3619–3624, doi:10.1080/01431160110069935

Thompson, D. J. W., and J. M. Wallace, 1998: The Arctic Oscillation signature in wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300, doi:10.1029/98GL00950

Thompson, D. J. W., and J. M. Wallace, 2000: Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability. *J. Climate*, **13**, 1000–1016, doi:10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2

Thompson, D. J. W., J. M. Wallace, and G. Hegerl, 2000: Annular Modes in the Extratropical Circulation. Part II: Trends. *J. Climate*, **13**, 1018–1036, doi:10.1175/1520-0442(2000)013<1018:AMITEC>2.0.CO;2

Trishchenko, A. P., and Z. Li, 2001: A method for the correction of AVHRR onboard IR calibration in the event of short-term radiative contamination. *Int. J. Remote Sens.*, **22**, 3619–3624, doi:10.1080/01431160110069935

Trishchenko, A. P., 2002: Removing unwanted fluctuations in the AVHRR thermal calibration data using robust techniques. *J. Atmos. Oceanic Technol.*, **19**, 1939–1954, doi:10.1175/1520-0426(2002)019<1939:RUFITA>2.0.CO;2

Trishchenko, A. P., G. Fedosejevs, Z. Li, and J. Cihlar, 2002: Trends and uncertainties in thermal calibration of AVHRR radiometers onboard NOAA-9 to -16. *J. Geophys. Res.*, **107**, doi:10.1029/2002JD002353

Turner, J., S. R. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M. Carleton, P. D. Jones, V. Lagun, P. A. Reid, and S. Iagovkina, 2003: The SCAR READER Project: Toward a High-Quality Database of Mean Antarctic Meteorological Observations, *J. Climate*, **17**, 2890–2898, doi:10.1175/1520-0442(2004)017<2890:TSRPTA>2.0.CO;2

List of Figures

FIG. 5. Comparison of seasonal patterns of change for RLS, E-W, and S09 reconstructions for 1957 - 2006. Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.

TABLE 1. Correlation coefficients for AVHRR PCs following regularization in RegEM TTLS, using S09 station selection and $k = 3$.

Iteration 1

r	PC 1	PC 2	PC 3
PC 1	0.4505	0.0600	-0.0743
PC 2	0.0600	0.2727	-0.1001
PC 3	-0.0743	-0.1001	0.1311

Iteration 35

r	PC 1	PC 2	PC 3
PC 1	0.8365	-0.0550	-0.0708
PC 2	-0.0550	0.8311	-0.2501
PC 3	-0.0708	-0.2501	0.2507

TABLE 2. Permutation matrix for the 1,496 ground station only cross validation experiments. "Cor" and "Cov" indicate correlation and covariance, respectively.

k_{gnd}	With-holding	Algo-rithm	Network
1 … 10	Random, Early, Late	TTLS	Cor
			Cov
		TSVD	Cor
			Cov

TABLE 3. Permutation matrix for the 23,040 RLS and 1,920 E-W reconstruction cross-validation experiments.

Type	k_{gnd}	k_{sat}	c or k_{RegEM}	Algorithm	Network
RLS	5 … 8	2 … 100	0.1, 0.2, … 1.1, 1.3, 1.5, 1.75, 2.0	TSVD	Correlation
					Covariance
				TTLS	Correlation
					Covariance
E-W	5 … 8	3, 13, 28, 50, 100	1 … 12	TSVD	Correlation
					Covariance
				TTLS	Correlation
					Covariance

TABLE 4. Optimal parameters determined from the RLS and E-W cross-validation experiments. "Cor" and "Cov" indicate correlation and covariance, respectively. Parameters used for reconstructions presented in the text are bolded and underlined.

Type	Algo-rithm	k_{gnd}	k_{sat}	c / k_{RegEM}	Cor/ Cov
RLS	TTLS	7	80	0.1	Cor
		6	87	0.1	Cov
	TSVD	**7**	**80**	**0.1**	**Cor**
		6	87	0.1	Cov
E-W	TTLS	7	100	9	Cor
		6	100	9	Cov
	TSVD	**7**	**100**	**9**	**Cor**
		6	100	9	Cov

TABLE 5. Regional trend comparison between this study (RLS and E-W reconstructions) and the S09 reconstruction.

Region	RLS[a]	E-W[a]	S09[a]
Continental Average	0.06 ± 0.07	0.05 ± 0.07	0.12 ± 0.09
East Antarctica	0.05 ± 0.09	0.04 ± 0.06	0.10 ± 0.10
West Antarctica	0.05 ± 0.08	0.04 ± 0.08	0.20 ± 0.09
Peninsula	0.29 ± 0.10	0.29 ± 0.08	0.13 ± 0.05

[a] Confidence intervals are 95%, with degrees of freedom corrected for serial correlation of the residuals (Santer et al. 2000).

TABLE 6. West Antarctic trend sensitivity for varying k_{gnd}. Bold and underline indicates optimal parameters determined by the cross-validation experiments. Full reconstructions use the optimal 63-station set; verification reconstructions use a 28-station subset that is comprised of the on-grid stations used by S09.

		Trend[a]		CE[b]	
Reconstruction	k_{gnd}	Full recon	Verification recon	Full recon	Verification recon
RLS Correlation	5	0.17 ± 0.09	0.14 ± 0.08	0.60	0.45
	6	0.16 ± 0.09	0.14 ± 0.08	0.60	0.46
	7	**0.05 ± 0.08**	**0.11 ± 0.07**	**0.65**	**0.51**
	8	0.10 ± 0.07	0.12 ± 0.07	0.63	0.47
E-W Correlation	5	0.10 ± 0.07	0.06 ± 0.07	0.51	0.41
	6	0.11 ± 0.07	0.07 ± 0.06	0.51	0.39
	7	**0.04 ± 0.06**	**0.06 ± 0.06**	**0.51**	**0.42**
	8	0.09 ± 0.06	0.09 ± 0.06	0.49	0.41
S09	3[c]	0.20 ± 0.09[d]		0.47[d]	0.42[d]

[a] Trends in deg C / decade. Confidence intervals are 95%, with degrees of freedom corrected for serial correlation of the residuals (Santer et al. 2000).

[b] CEs are for withheld West Antarctic stations only: Doug, Elizabeth, Harry, Siple, and Theresa for full reconstructions; Byrd, Erin, and Mount Siple for verification reconstructions.

[c] S09 combine ground stations and PCs using a truncation parameter of 3.

[d] Values calculated using the published T_{IR} reconstruction.

TABLE 7. Summary verification statistics and comparison to values calculated from the published T_{IR} reconstruction.

	Full recon*			Verification recon**		
	μ_{rms}	r	CE	μ_{rms}	r	CE
RLS	1.01	0.87	0.73	1.35	0.85	0.68
E-W	1.24	0.83	0.60	1.47	0.82	0.62
S09	1.52	0.61	0.37	1.91	0.60	0.36

* For the 24 stations not used in the full 63-station reconstructions

** For the additional 35 stations withheld from the verification reconstructions

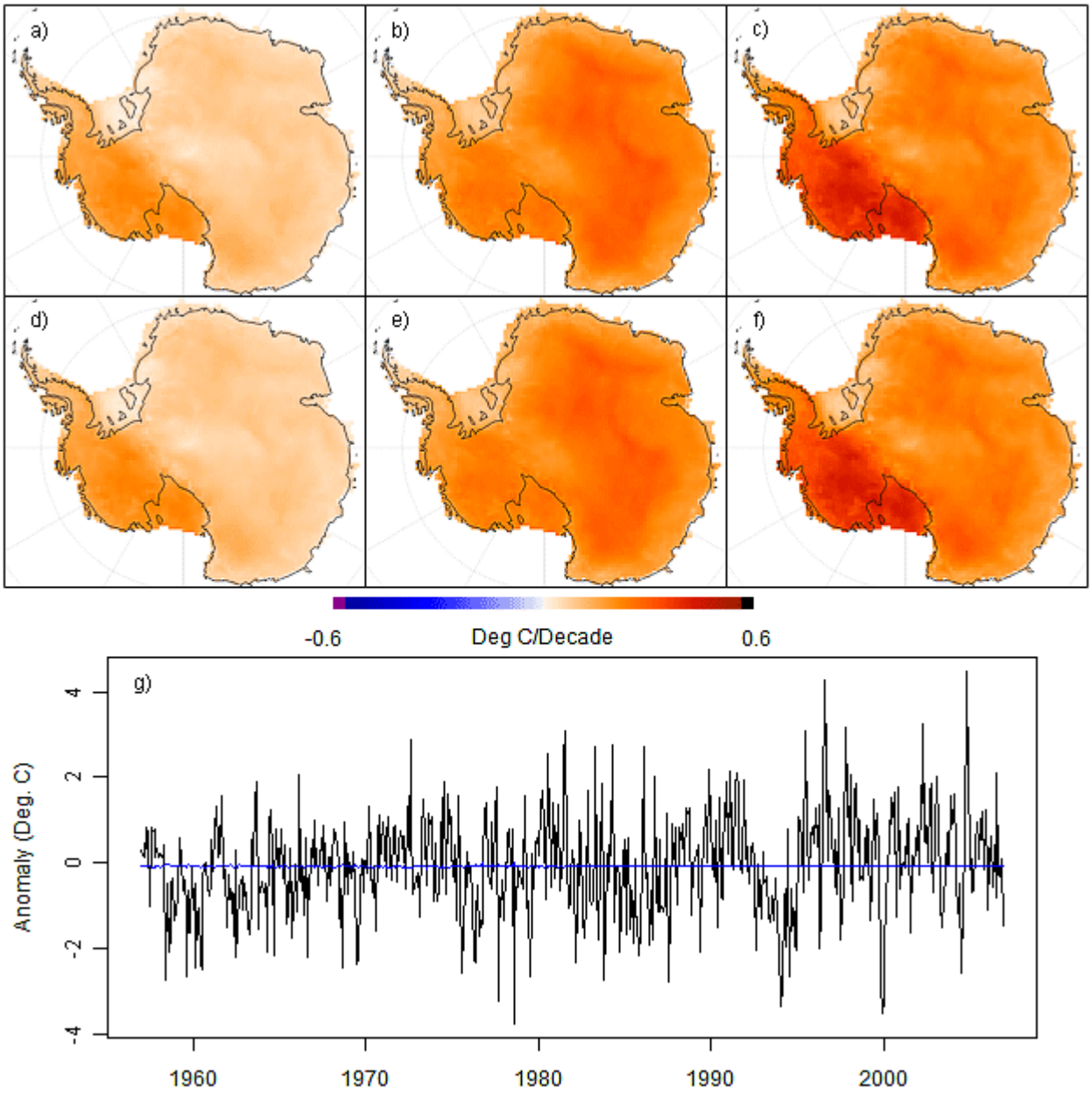

FIG. 1. Comparison of spatial distribution of temperature trends between the S09 T_{IR} reconstruction and the replication effort of this paper. a-c, S09 reconstruction: 1957-2006 (a), 1957-1981 (b), 1982-2006 (c). d-f, Replication effort: 1957-2006 (d), 1957-1981 (e), 1982-2006 (f). g, Monthly replication reconstruction means with the blue line indicating the difference between the S09 and replication reconstructions.

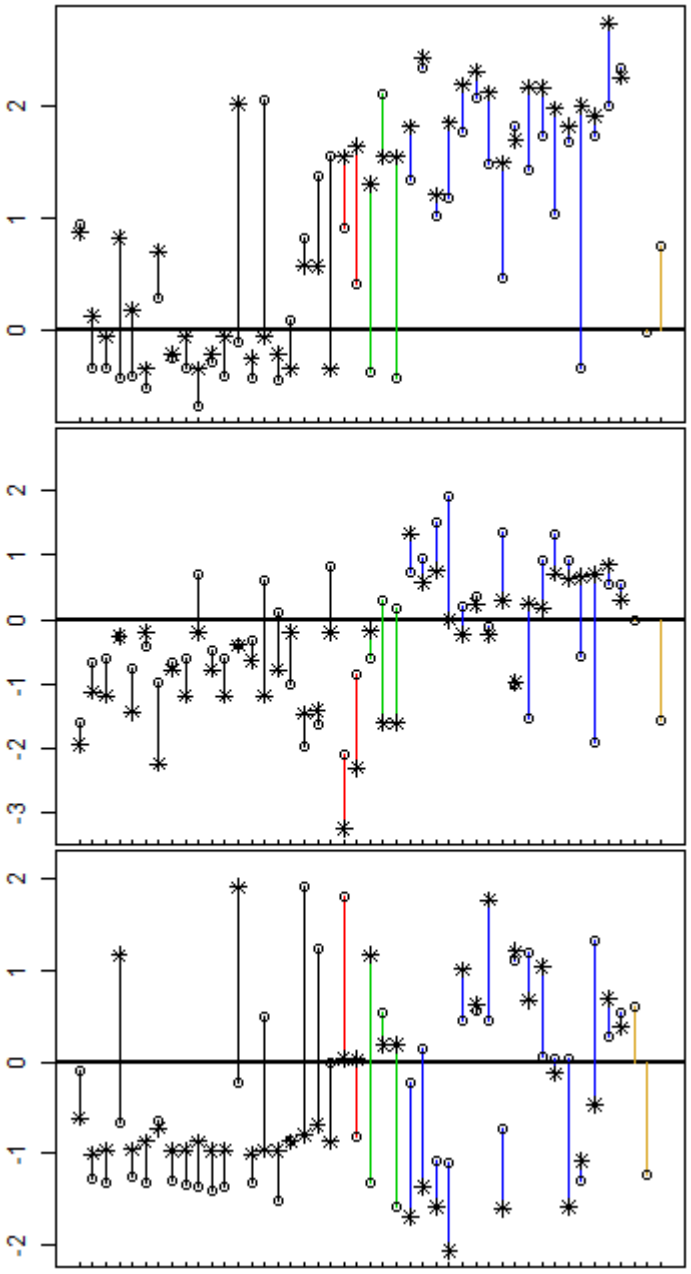

FIG. 2. Spatial structure used to estimate the satellite PCs in the 1957-1982 period by S09 vs. AVHRR spatial structure. Top: PC #1. Middle: PC #2. Bottom: PC #3. Circles represent the normalized weight in the imputation EOFs; stars represent the normalized AVHRR spatial eigenvector weights. Geographic location indicated by color. Black (Peninsula), Red (West Antarctica, excluding Ross), Green (Ross), Blue (East Antarctica), Gold (AVHRR PC). Weights are normalized such that the variance is unity.

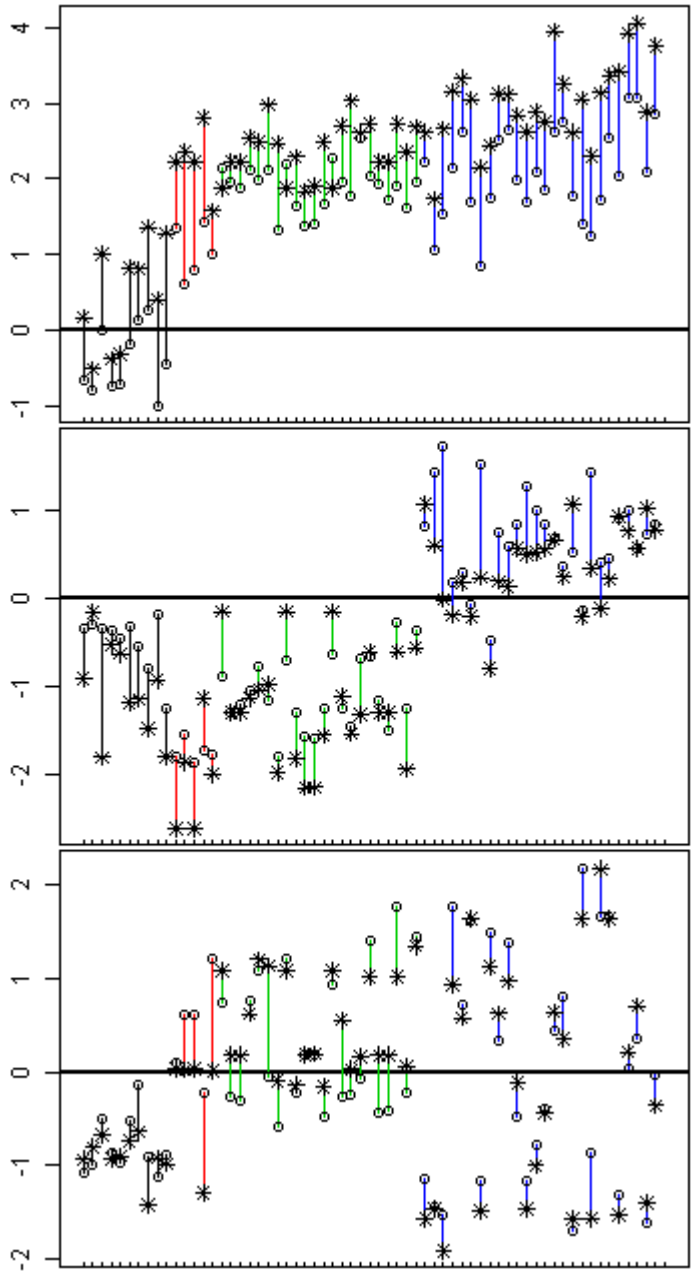

FIG. 3. Spatial structure used to estimate the satellite PCs in the 1957-1982 period with AVHRR eigenvector weighting vs. AVHRR spatial structure. Top: PC #1. Middle: PC #2. Bottom: PC #3. Colors, symbols, and normalization are identical to Fig. 2.

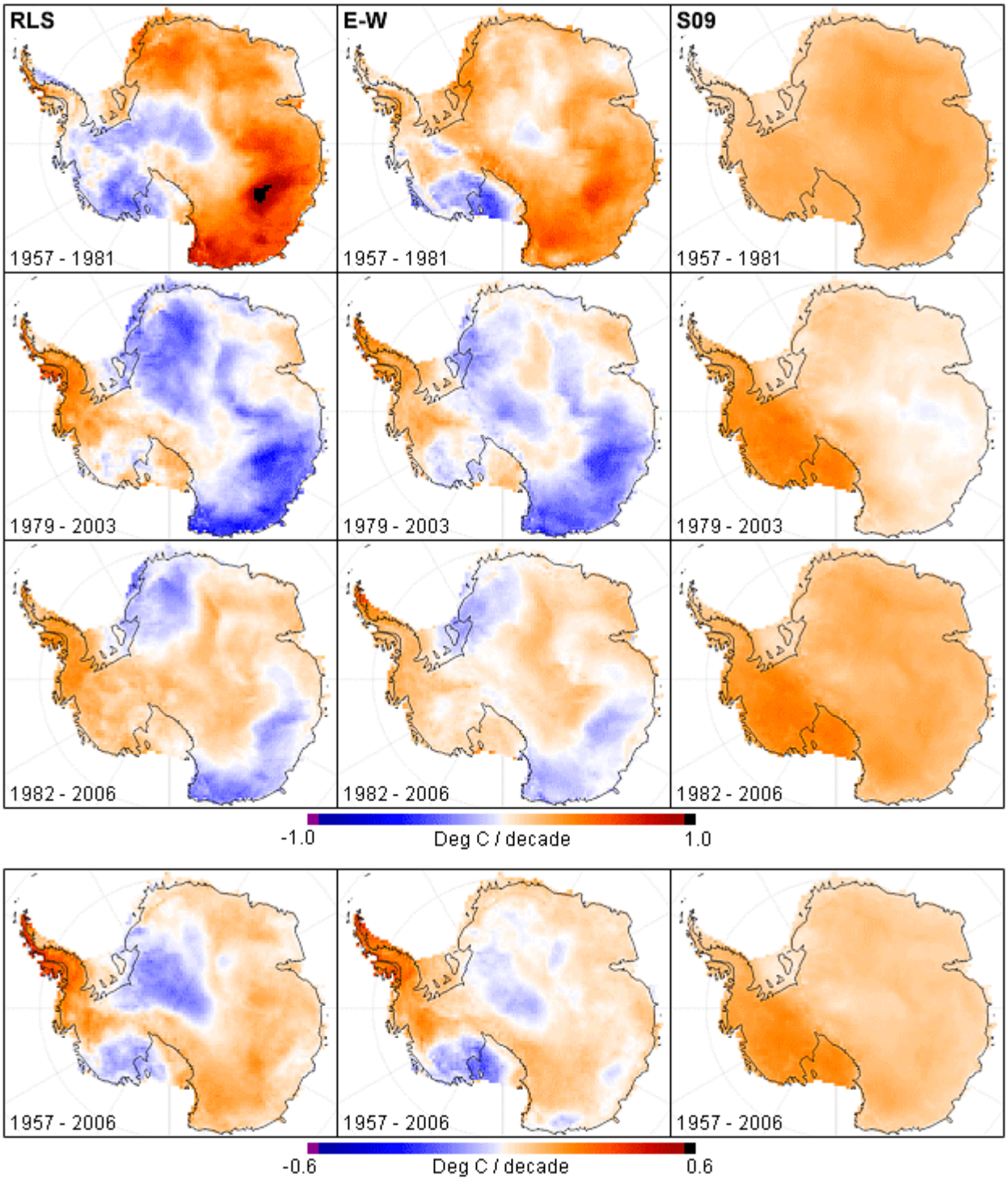

FIG. 4. Comparison of spatial patterns of change for RLS, E-W, and S09 reconstructions. Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.

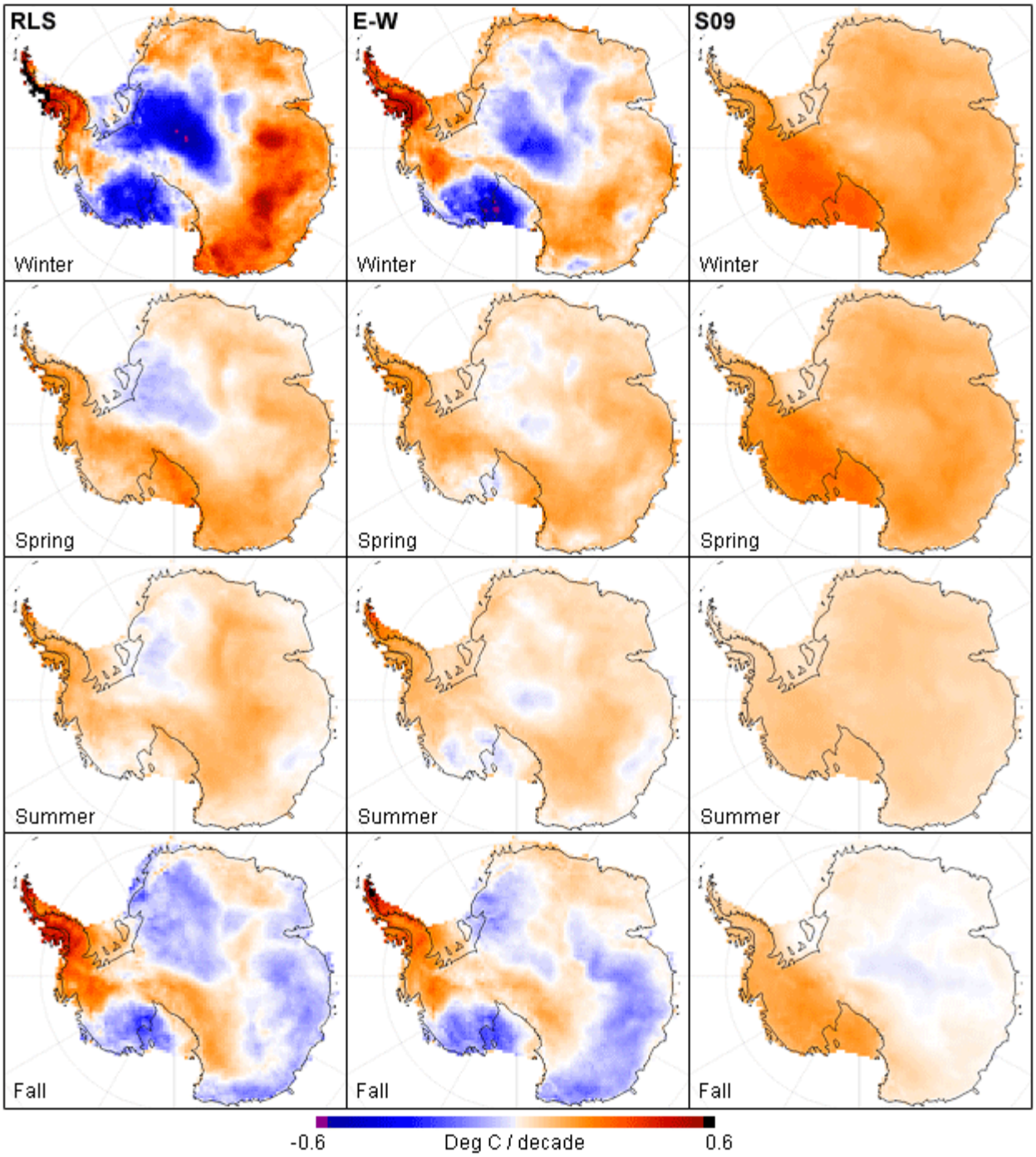

FIG. 5. Comparison of seasonal patterns of change for RLS, E-W, and S09 reconstructions for 1957 - 2006. Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.