

Improved methods for PCA-based reconstructions: case study using 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

Abstract

A detailed analysis is presented of a recently published Antarctic temperature reconstruction that combines satellite and ground information using a regularized expectation-maximization algorithm. Though the general reconstruction concept has merit, it is susceptible to spurious results for both temperature trends and patterns. The deficiencies include: (a) improper calibration of satellite data to ground data; (b) improper determination of spatial structure during infilling; and (c) suboptimal determination of truncation parameters, particularly with respect to satellite-derived principal component retention. This study proposes two different methods 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, we find warming over the period of 1957-2006 to be concentrated in the Peninsula ($\approx 0.3^\circ\text{C decade}^{-1}$). We also show average trends for the continent and East Antarctica that are approximately half that found using the unimproved method. The trends for the continent and East Antarctica, while positive, are not significant at the 5% level. Though we conclude the West Antarctic regional average is statistically significant, the trend is again approximately half of that found using the unimproved method. Finally, we find similar seasonal patterns for the Peninsula, but substantially different seasonal 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 & Walsh 2007; Monaghan et al. 2008) relied on interpolation or kriging methods to estimate temperatures at non-instrumented points. In Chapman & 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.

S09 present three separate reconstructions. The primary reconstruction is the focus of this paper and will be referred to as the T_{IR} reconstruction¹. They also present a reconstruction that does not combine AVHRR data with ground data, which will be referred to as the AWS (Automatic Weather Station) reconstruction. This reconstruction

¹ S09 additionally present a detrended variant which we will not directly address. This variant retains the linearly detrended AVHRR data as-is, and is justifiably de-emphasized in the S09 text. All criticisms apply equally to the detrended reconstruction, though, due to the effects of the detrending, trend magnitudes are not directly comparable.

is dealt with implicitly, as our proposed modifications likewise separate the estimation of missing AVHRR PC and ground station information. The third reconstruction, utilizing standard principal component analysis, appears in S09's Supplementary Information and is not accompanied by sufficient information for a quantitative comparison. However, as this version also utilized the same number of retained AVHRR PCs as the T_{IR} reconstruction, our criticisms apply to the major parameter choice for this reconstruction as well.

The primary S09 method involves the following major steps: a) cloud masking and regridding of the raw AVHRR data; b) decomposition of the cloud masked AVHRR data into principal components (PCs) and spatial eigenvectors; c) augmentation of a matrix of station data starting in 1957 with the first three AVHRR PCs; d) estimation of missing data in the augmented matrix with an infilling algorithm; e) extraction of the completely infilled PCs; and f) estimation of temperatures at all grid points by reconstituting the PCs with their corresponding spatial eigenvectors (Steig et al. 2009; Steig, personal communication). The last step provides a time series of maps containing the temperature contribution from each PC / spatial eigenvector pair, which are then summed together to provide the gridded temperature estimates for all months.

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 and magnitudes 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 cloud masking of the Advanced Very High Resolution Radiometer (AVHRR) data. We do not attempt to replicate the cloud masking operations by S09, as these are similar to previously published studies (e.g., Comiso 2000), and instead utilize the archived set provided by Steig on his university website. For ground data, we utilized S09's archived READER data set (Turner et al. 2003), also published on the same website. Each series in the READER and AVHRR data are then centered, converted to anomalies and scaled to unit variance prior to further processing.

For the period of 1957-2006, our replication yields linear trends in $^{\circ}\text{C decade}^{-1}$ 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

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 our conclusions and 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 AVHRR data used by S09 is 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. 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, some of which may change from satellite to satellite. Additionally,

the AVHRR instrument measures skin temperature in grid cell averages rather than point measurements of near-surface air temperature. These factors highlight the need to calibrate the AVHRR data to ground data, as the measurements cannot *a priori* be expected to be interchangeable². The mathematical description provided by S09 establishes the ground data as the explanatory 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 $\mathbf{Y} = (\mathbf{A} \ \mathbf{b})$, where \mathbf{A} is said to represent the ground station data (predictors, or explanatory variables) and \mathbf{b} is said to represent the AVHRR principal components to be estimated (predictands, or response variables)³. Regularization is accomplished by performing a singular value decomposition of the correlation⁴ matrix \mathbf{C} with k retained eigenvectors (Mann et al. 2007). From Schneider (2001), this yields the spatial eigenvectors and squared eigenvalues of the $n \times p$ matrix of observations $\tilde{\mathbf{Y}} = \mathbf{Y} / \tilde{\mathbf{s}}$ (where n consists of the time steps, p consists of the variables, and $\tilde{\mathbf{s}}$ is a vector of unbiased standard deviation estimators), since:

² For example, a statistically significant embedded trend in the satellite data is apparent in the 1982 – 2001 period (NOAA – 7 through NOAA – 14) that exceeds ground station measurements by 0.19 +/- 0.16°C decade⁻¹ (Supporting Information).

³ Since \mathbf{A} and \mathbf{b} are defined separately for each time step, the S09 definition is not strictly true.

⁴ 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.

$$\tilde{\mathbf{Y}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \quad (1)$$

$$\mathbf{C} = \mathbf{V}\mathbf{\Lambda}^2\mathbf{V}^T \quad (2)$$

In this notation, \mathbf{U} represents the temporal eigenvectors, $\mathbf{\Lambda}$ the eigenvalues, and \mathbf{V} the spatial eigenvectors. We may then partition \mathbf{V} into subspaces, where rows 1 and 2 indicate available observations and missing observations, and columns 1 and 2 indicate eigenvectors $1 \dots k$ and $k+1 \dots n$, respectively:

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

The TLS solution yielding the set of statistical weights \mathbf{x} for prediction of \mathbf{b} from \mathbf{A} is given by Fierro et al. (1997), where symbol \dagger indicates the generalized inverse:

$$\mathbf{x}_k = (\mathbf{V}_{1,1}^T)^\dagger \mathbf{V}_{2,1}^T \quad (4)$$

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

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

Alternatively, rather than limit the estimation to subspace $\mathbf{V}_{2,1}$, we can replace

$\mathbf{V}_{2,1}^T$ in equation (4) with $\mathbf{V}_{1 \dots k}^T = \begin{pmatrix} \mathbf{V}_{1,1} \\ \mathbf{V}_{2,1} \end{pmatrix}$, yielding a full set of statistical weights to provide

estimates for both missing and actual values:

$$\tilde{\mathbf{Y}}_j = (\tilde{\mathbf{A}}_j \tilde{\mathbf{b}}_j) = \tilde{\mathbf{A}}_j (\mathbf{V}_{1,1}^T)^\dagger \mathbf{V}_{1 \dots k}^T \quad (6)$$

As all eigenvectors greater than k are discarded, subspace $\mathbf{V}_{2,2}$ provides an estimate of the covariance matrix of the scaled predictand residuals (Schneider, 2001):

$$\mathbf{C}_{\text{res}} = \mathbf{V}_{2,2} \mathbf{\Lambda}_{k+1 \dots n}^2 \mathbf{V}_{2,2} \quad (7)$$

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

c. Theoretical and practical difficulties with the S09 approach

Fundamentally, calibration is simply the process by which the relationship between a set of response and explanatory variables is established. A valid calibration places the response variables in terms of the explanatory variables to allow estimation of a response from a given observation of the explanatory variable. If this is not performed, then the relationship between the variables is undefined and subsequent predictions cannot be shown to be valid. In the special case that the response and explanatory variables are already known to be equivalent quantities and are interchangeable, no formal calibration is necessary. Formal calibration is required if this does not hold. As already discussed, the ground and AVHRR data are not interchangeable, and the latter consideration therefore applies.

A critical aspect of the S09 methodology is that *both* the satellite PCs (which exist only from 1982 – 2006) and the station data matrix are *incomplete* when supplied to RegEM. During 1982 – 2006, the PCs appear in \mathbf{A} (not \mathbf{b}) and are directly used to predict missing ground station values. Missing ground station information, then, is estimated by linear combinations of the ground stations *and* the PCs. Since the correlation matrix \mathbf{C} is computed using both actual and estimated values, the response

variables (the PCs) are partially calibrated to the ground station data and partially calibrated to linear combinations of themselves.

Secondly, the process of regularization in RegEM destroys the orthogonality of the PCs (Table 1). From equation (4), the missing data will be estimated using correlated – not orthogonal – PCs. This means that not only are the response variables partially calibrated to themselves, they are also partially calibrated to each other. In the case of PC 2 and PC 3, the correlation for the entire 1957 – 2006 period 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). If one examines just the estimated values (i.e., the 1957 – 1981 period), the correlation coefficient between PCs 2 and 3 is -0.6, despite the fact that the original values are orthogonal.

A third concern is that the RegEM version as used by S09 does not extract modeled (calibrated) data. Rather, it extracts the *original data* with estimates in place of the missing values. Because the algorithm implicitly assumes that all variables are already equivalent and interchangeable, the original response variables (the AVHRR PCs) are never expressed in terms of the explanatory variables (the ground data). Even assuming the previous considerations to be of negligible importance, this means, at best, the PCs used by S09 are properly calibrated only in the 1957 – 1981 period, and have 25 years of uncalibrated data spliced on the end.

The final concern is that the concept of using a *total least squares* algorithm (which minimizes the errors in both the available and missing values) to regress PCs against temperature observations 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*). Since systemic errors are more likely to affect the PCs than the ground data (Section 3.a), the filtering effect of the truncation parameter k will be less effective for the PCs. This is because random errors in the AVHRR data have already been relegated to the truncated modes, while the systemic errors will be interpreted as signal. Moreover, the assumption that the relative variance of errors in the data set are homogenous (Schneider, 2001) is violated by the AVHRR PCs, because the relative variance of errors increases (i.e., signal to noise ratio decreases) as one proceeds from the low-order to high-order modes. This translates into additional estimation error for the ground stations when the PCs appear in \mathbf{A} .

4. Spatial structure considerations

Another concern with S09 is the difference in spatial structure used to infill the PCs and the corresponding AVHRR spatial eigenvectors. The assumption that the spatial structure is similar is implicit in the S09 method, which recovers the gridded temperature estimates without altering the AVHRR spatial eigenvectors. From equations (4) and (6), the estimated portion of the PCs (1957 – 1981) is comprised of linear combinations of station data *and themselves*, with coefficients given by matrix \mathbf{x}_k . Unless the coefficients in \mathbf{x}_k yield a spatial distribution identical to the corresponding AVHRR spatial eigenvector weights, recovering gridded estimates using the unaltered spatial eigenvectors will result in error. Except in the unique case where the regression coefficients in \mathbf{x}_k yield a spatial distribution that differs from the eigenvector weights by

a scalar multiplier (which will only affect magnitudes) this will result in the gridded estimates showing a different magnitude *and* spatial pattern than the ground station matrix.

We can investigate whether the AVHRR eigenvector weights are directly compatible with the TTLS regression coefficients. Since equation (6) yields a fully populated matrix of estimates $\tilde{\mathbf{Y}}$ of rank k , we may find the vector of coefficients \mathbf{v}_i that describe the remainder of the series in $\tilde{\mathbf{Y}}$ in terms of any given series i . In the case of an estimated PC, \mathbf{v}_i describes how much variation the PC explains in $\tilde{\mathbf{Y}}$. In other words, \mathbf{v}_i is what the AVHRR spatial eigenvector weights *should be* in order to reproduce the ground data used to predict the PC.

From equation (1), the SVD will yield the spatial eigenvectors and eigenvalues $\mathbf{\Lambda V}$ (denoted “imputation EOFs”) and the temporal eigenvectors \mathbf{U} . The j^{th} column in $\mathbf{\Lambda V}$ contains the coefficients that yield the contribution of the j^{th} column in \mathbf{U} to each series in $\tilde{\mathbf{Y}}$. Since every series in $\tilde{\mathbf{Y}}$ is comprised of linear combinations of \mathbf{U} using coefficients given by the corresponding i^{th} row in $\mathbf{\Lambda V}$, we can calculate the contribution of series i to the rest of the data in $\tilde{\mathbf{Y}}$ by taking the vector sum of the imputation EOFs with series i removed ($\mathbf{\Lambda}_j \mathbf{V}_{-i,j}$), scaled by the appropriate element $v_{i,j}$ in \mathbf{V} , and a normalization constant c :

$$\mathbf{v}_i = c \sum_{j=1}^k v_{i,j} \mathbf{\Lambda}_j \mathbf{V}_{-i,j} \quad (8)$$

Figure 2 (with Figure 1 as the geographic key) compares \mathbf{v}_i to the AVHRR spatial eigenvector weights for the three PCs retained by S09. There are some obvious

and important differences. While PC 1 and PC 3 demonstrate similar spatial makeups for the Peninsula and East Antarctica, the TTLS algorithm predicts PC 2 using five East Antarctic stations in the opposite orientation. In 9 out of 17 cases for PC 2, the coefficients assigned to East Antarctica differ from the AVHRR weights by a factor of 2 or more. In West Antarctica, the differences are more significant, and the overall match between the TTLS weights and the AVHRR eigenvectors is poor. For PCs 1 and 2, TTLS uses two out of five stations in the opposite orientation, and four of the stations (80%) differ from the AVHRR eigenvector weight by a factor of 1.5 or more. For PC 3, all of the West Antarctic stations differ by a factor of at least three, and three of the four most heavily weighted stations in TTLS are used in the opposite orientation. 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 lower set of weights for East Antarctica between TTLS and the AVHRR eigenvectors. This necessarily results in a redistribution of the Peninsula trend across the entire continent.

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 / PC matrix. In their study, S09 state that they use the procedure described in Mann et al. (2007) to determine their truncation parameters. This procedure involves inspection of the log eigenvalue spectrum and calculation of eigenvalue sampling error. The modes

that correspond to separable eigenvalues (i.e., the error bars on the eigenvalues do not fall into the continuum of overlapping error estimates) are selected for use. The remainder are discarded.

This calculation is identical to that performed by North et al. (1982), where it was shown that overlapping eigenvalue sampling error estimates indicate mixing (degeneracy) of the underlying modes. Importantly, North et al. (1982) note that this procedure provides *no guidance* on determining the truncation parameter. The effect of splitting degenerate modes during truncation depends on the analysis being performed. If one is trying to find a smaller basis for representing a data set, the effect is limited to the amount of variance the multiplet explains in the original data. If the explained variance of the multiplet is small, the error due to splitting it is correspondingly small. In the case of S09, the latter concern applies. The procedure used by S09, which selects the truncation parameter based solely on sampling error, is incomplete because it omits any investigation of the error caused by early truncation.

Furthermore, the procedure described in Mann et al. (2007) was evaluated during the course of pseudoproxy experiments under different conditions than exist in Antarctica. It was tested using separate noise realizations for each pseudoproxy, which does not admit the possibility of nonlocal correlation between predictor and predictand residuals. As discussed earlier, this assumption is violated in Antarctica. Mann et al. (2007) also note that the procedure is too conservative at high signal-to-noise ratios (SNRs). The average coefficient of determination (r^2) between the AVHRR and ground data is approximately 0.45, which roughly corresponds to an SNR of 1.0. This was the highest SNR tested by Mann et al. (2007), with the exception of perfect pseudoproxies.

Finally, they note that the procedure is heuristic, describing it as a “conservative tool that works well in practice,” and suggest cross-validation as a possible alternative and more objective method. We will show via extensive cross-validation testing that this procedure led S09 to select suboptimal truncation parameters.

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. 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 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 more accurate temporal information.

This 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. (As noted in Section 3.c, this modification is *required* for a valid calibration.) The estimates 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 discussed, RegEM is not capable of providing a calibration function if the explanatory and response variables are both incomplete, or if multiple response variables are simultaneously included. Given that the explanatory variables (station data) display wide gaps in coverage and the response variables (AVHRR data) are entirely missing from 1957 – 1981, the method used by S09 cannot perform the necessary calibration. Additionally, as S09 do not use the modeled PCs in the 1982 – 2006 period, the second half of their reconstruction is entirely uncalibrated.

We address these issues first by ensuring the explanatory variables are temporally complete (which we denote Mod 1), and second by taking advantage of equation (6) to extract the modeled PCs (which we denote Mod 2). The latter is explained in the prior section. To perform the former, we utilize the RegEM algorithm to infill a matrix comprised solely of ground station data (analogous to the AWS reconstruction in S09). The completed matrix is then augmented by the AVHRR PCs and the entire 1957 – 2006 period for the PCs is predicted via equation (6). 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.

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 eigenvectors is to constrain the ground station regression by the corresponding eigenvector weights. Because each set of weights is unique to a particular AVHRR eigenvector, this requires that each PC be infilled separately and has the added benefit of entirely resolving the issues of mutual reinforcement and cross-contamination of the PCs noted in Section 3. We denote this modification as Mod 3. Additionally, we denote the combination of Mods 1 – 3 for direct AVHRR PC regression as the *eigenvector-weighted* (E-W) method.

d. Eliminating use of principal components

A more elegant means to resolve the concerns with calibration and potential differences in spatial structure is to avoid using the AVHRR PCs at all. Since we assume the AVHRR 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 eigenvectors. To do this, we first define our spatial EOFs as $\mathbf{L} = \sqrt{1/\tilde{n}} (\mathbf{\Lambda} \mathbf{V})$, where $\mathbf{\Lambda}$ contains the AVHRR eigenvalues $1...k$, \mathbf{V} represents the weights of the corresponding spatial eigenvectors at the ground station (predictor) locations, and \tilde{n} represents the effective degrees of freedom. We may then define a matrix of ground station observations \mathbf{Y} and unknown set of statistical weights \mathbf{a} and write:

$$\mathbf{L}\mathbf{a} = \mathbf{Y} \quad (9)$$

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} (subscripts omitted for readability):

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

As we do not know the proper regularization parameter from any *a priori* physical arguments, we determine the 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 $\mathbf{L}\mathbf{a}$ 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. This drives determining μ^2 via iterative estimation (Fitzpatrick 1991) and yields the reasonable physical constraint that the error is approximately constant over the analysis timeframe. The value of μ^2 that minimizes prediction error in withheld data may then be interpreted as the maximum likelihood estimation of the true ratio of system measurement error and noise (Fitzpatrick 1991; Sima 2006).

We denote this method – which inseparably combines Mods 1 – 3 without use of the AVHRR PCs – as the *regularized least squares* (RLS) method.

e. Determining truncation parameters

The final modification, which applies to both the E-W and RLS methods, is to determine the optimum truncation parameters through a series of cross validation

experiments. This provides an objective criterion for determining important modes without resorting to heuristic tools (inspecting the log eigenvalue spectrum, bootstrapped eigenvalue/eigenvector, broken stick, scree plots, etc.) that can give vastly different answers for the same 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 PCs, and k_{RegEM} to denote the truncation parameter used by the infilling algorithm to regress the PCs against the ground stations (applicable only to the E-W reconstructions). The adjustable arguments for the cross-validation then consist of:

- Correlation and covariance settings
- Two different infilling algorithms for the ground station infilling⁵
- 14 different ground station sets
- 10 different values for k_{gnd}
- 99 values for k_{sat} (2 to 100 retained eigenvectors)
- Two different infilling algorithms for the PC estimation (E-W only)
- 12 values for k_{RegEM} (E-W) or 15 values of c (RLS)
- Two different levels of withholding for verification

The total number of permutations for RLS (1,663,200) and E-W (2,661,120) based on this set of adjustable arguments is computationally prohibitive. To reduce the total number of permutations to a reasonable amount, we first screen for the infilling

⁵ We evaluated the performance of TTLS and an iterative truncated SVD (TSVD) approach similar to the DINEOF routing for infilling cloud-masked data sets (Alvera-Azcárate et al. 2009; Beckers and Rixen 2003; Beckers et al. 2006; Beckers, personal communication). Additional details concerning the TSVD algorithm are available in the Supporting Information.

algorithm, station sets and values of k_{gnd} that provide reasonable verification statistics in a split early/late calibration verification test during the ground station infilling step. We also limit the E-W tests (due to the high computational cost of the E-W method) to sets of 3, 13, 28, 50, and 100 retained AVHRR eigenvectors. This reduces the number of cross-validation permutations to a manageable 23,760 for RLS and 1,920 for E-W (flowcharts are available in the Supporting Information). After performing the required number of RLS and E-W reconstructions, we select the combination of adjustable arguments that yield the highest verification statistics for the gridded reconstruction. Note that stations used as verification targets for the gridded reconstructions are simply *omitted* in their entirety from the ground station matrix and subsequent E-W and RLS regressions. This allows us to evaluate how well the reconstruction might perform at uninstrumented grid locations. We denote this modification – wherein all truncation parameters are determined via cross-validation – as Mod 4.

f. Summary of modifications

To address the three primary issues noted in our abstract with the S09 method, we propose the following modifications:

- Mod 1: Infill ground stations separately from the AVHRR PCs
- Mod 2: Use the calibrated, modeled PCs at all times
- Mod 3: Constrain the prediction of the PCs by the AVHRR spatial eigenvectors
- Mod 4: Determine all adjustable parameters via cross-validation testing

These are implemented in two ways. In E-W, the PCs are regressed against the station data. In RLS, the ground station data is directly regressed against the retained AVHRR spatial eigenvectors. Optimal parameters for both are selected via Mod 4.

7. Results

a. Optimal parameters

The ground station screening experiments yield an optimal station set consisting of all READER stations that are located within 150km of an AVHRR grid cell and have 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 2. 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. We briefly discuss the source of this instability in Section 7.f and the Supporting Information.

Another important observation is that the smallest amount of regularization for the RLS reconstructions yields the best predictions, indicating that the ground data is largely free from systemic error. This provides tangible evidence that the AVHRR data contains larger errors than the ground data.

b. Overall 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 ($\approx 0.06 \pm 0.07$ °C decade⁻¹)⁶, nor is the warming in East Antarctica ($\approx 0.05 \pm 0.09$). This is similar to S09, wherein the trends for the continent and East Antarctica are less positive than West Antarctica and the Peninsula. Unlike S09, we find only the Peninsula regional average to consistently demonstrate a statistically significant trend. Though the regional average for West Antarctica displays mixed results for trend significance based on the optimal parameters determined from our cross-validation testing, based on additional testing (Section 7.d) the most likely estimate for the West Antarctic regional average is a statistically significant 0.10 ± 0.07 °C decade⁻¹, which is approximately ½ of that found by S09. We also find that the land area adjacent to the Peninsula displays statistically significant warming that slightly exceeds the warming found by S09 in the same area. While cooling in S09 is restricted primarily to East Antarctica in the 1969 – 2000 period, the RLS and E-W reconstructions provide evidence of cooling in various areas of the continent for all periods analyzed – including in the Ross area of West Antarctica during 1957 – 1981. Table 3 compares trend magnitude by region for the RLS, E-W and S09 reconstructions.

⁶ 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).

Figure 3 compares the spatial patterns of temperature change using major subperiods that appear in the S09 text. One feature that is similar to S09 is a strong indication of warming in portions of West Antarctica. However, the pattern of West Antarctic warming is substantially different. In S09, warming is concentrated in the Ross region and the warming over all of West Antarctica is statistically significant. In our reconstructions, statistically significant warming is concentrated in the area adjacent to the Peninsula and qualitatively appears to be an extension of the Peninsula warming. Additionally, we show a large area of cooling over the Ross Ice Shelf and adjacent land (statistically significant in the E-W reconstruction, but not in the RLS reconstruction), which is distinctly at odds with S09. A similar cooling feature – corroborated by ground station records – extending from the South Pole to the Weddell Sea is also absent from the S09 reconstructions. Figure 4 depicts areas of statistically significant trends.

Our reconstructions result in a significantly poorer match to the 1957-1981 and 1979-2003 patterns in the atmosphere-only GISS ModelE simulation 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. Comparison of our results to the 11-model ensemble reported in Chapman & Walsh (2007) and the 19-model winter ensemble reported in Connolley & Bracegirdle (2009) is more positive; the ensembles show significant warming in the Peninsula region and lesser warming in East Antarctica. Both ensembles, however, exhibit warming in the Weddell region and enhanced warming on Ross that is at odds with our results.

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 & Walsh (2007) and Monaghan et al. (2008) than S09. One notable difference between the Monaghan et al. (2008) reconstruction and the present work is in West Antarctica. Unlike the Monaghan et al. (2008) reconstruction, our reconstructions show positive trends in recent years at Byrd AWS. For the entire satellite coverage period of 1982 – 2006, we show trends of 0.15 ± 0.32 (RLS) and 0.02 ± 0.23 (E-W). For the latter half of the satellite period, those become 0.16 ± 0.87 (RLS) and 0.15 ± 0.66 (E-W).

c. Seasonal patterns of temperature change

In comparing seasonal patterns of change, substantial differences between our results and S09 are apparent (Figure 5 and Table 4). S09 find maximum warming in winter and spring for all areas. In our reconstructions, the Peninsula shows maximum warming in winter and fall, and the remainder of the continent shows maximum warming in spring and summer. We additionally show two separate patterns of change in West Antarctica, with the Ross region showing maximum *cooling* in winter rather than maximum warming, while the area adjacent to the Peninsula follows the Peninsula pattern. Overall, the seasonal behavior for the Peninsula and West Antarctica is closer to Chapman & Walsh (2007) and Monaghan et al. (2008) than S09. Those studies find 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). The observation that the seasonal

patterns in S09 for all regions follow the Peninsula is additional evidence that the S09 method allows the Peninsula to unduly influence the reconstruction.

Our reconstructions, like S09, show cooling over large portions of East Antarctica in the fall. In addition to this, however, we also find significant cooling in the region stretching from the South Pole to the Weddell Sea 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 °C decade⁻¹ during 1957 – 2006 for winter, spring, summer and fall, respectively. The winter cooling is absent in S09, who show the greatest *warming* occurring at the pole during this time. 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 & Walsh (2007), who also find cooling at the pole during all seasons, with a maximum cooling trend during winter in the 1958 – 2002 timeframe.

d. Uncertainties in West Antarctica

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 $\approx 0.05 \pm 0.07$ °C decade⁻¹. Both the magnitude and statistical significance are dependent on the truncation parameter for the ground station infilling. Table 5 summarizes this behavior for the values of k_{gnd} that were used in the full-grid cross validation experiments, and shows verification statistics for West Antarctica only. Like the rest of the continent, the highest verification statistics in West Antarctica are achieved for $k_{\text{gnd}} = 7$.

For the full 63-station reconstructions, only $k_{\text{gnd}} = 7$ yields an insignificant trend in West Antarctica, with a magnitude about half of what is computed for neighboring values of k_{gnd} ⁷. In order to determine if this trend is artificially low, we conduct two separate sets of experiments. In one experiment, we perform the ground station infilling using the individual ridge regression option in RegEM (Schneider, 2001). We optimize the ridge parameter for each time step using a generalized cross validation function. This experiment yields similar patterns of change as shown in Fig. 3, with less intense cooling on Ross, comparable verification statistics and a statistically significant average West Antarctic trend of 0.11 ± 0.08 °C decade⁻¹.

For the second experiment, we avoid infilling the ground stations altogether. Because not all of the stations are complete for any given period, the appropriate offsets are determined based on periods of mutual overlap. RLS reconstructions are then performed, with separate regression coefficients calculated for each unique combination of missing and available values. The resulting patterns and verification statistics in West Antarctica are quite comparable to the ridge regression test. In this test, West Antarctic trends were 0.08, 0.12, 0.11, and 0.10 (± 0.07 °C decade⁻¹ for all) depending on whether 20, 40, 60, or 80 overlapping months were used to determine the offsets.

Based on these results, we conclude that the West Antarctic regional average is likely to be ≈ 0.10 °C decade⁻¹, with a low estimate of 0.05 and a high estimate of 0.12. Trend uncertainty based on unexplained variance in the linear model is approximately ± 0.07 for all estimates.

⁷ Testing indicates the dependence on k_{gnd} when using TTLS or TSVD may be the result of the fixed truncation parameter providing insufficient filtering when the number of predictors is low, and is the subject of ongoing work by the authors (Supporting Information).

e. 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 verification targets are completely omitted from the reconstructions; hence, no target 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 withhold the additional 35 withheld stations for 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 statistics exceed the 99th percentile for Monte Carlo experiments using the mean, variance and lag-1 autoregression coefficients for all stations. Verification skill to ground data exceeds the explained variance and r^2 values calculated using either the S09 reconstruction or the raw AVHRR data. Full statistics are available in the Supporting Information.

Variance loss in our reconstructions is small, with RLS slightly outperforming E-W. Since the Peninsula displays the highest trends, variance loss in that region will be most noticeable. Fig. 6 shows annual anomalies for the 7 most complete Peninsula ground stations vs. the corresponding grid cells from the RLS, E-W, and S09

reconstructions. The ground data trend is $0.41 \pm 0.16^{\circ}\text{C decade}^{-1}$. The reconstruction trends are 0.40 ± 0.14 (RLS), 0.34 ± 0.12 (E-W), and 0.06 ± 0.04 (S09).

f. Relative importance of Mods 1 – 4

Based on performing S09-style replications with and without Mod 1 (which separates the ground station and PC estimation), we find that this modification has a negligible impact on the gridded results. Though the impact becomes greater as additional AVHRR eigenvectors are included, even at $k_{\text{sat}} = 100$ the overall pattern and magnitude of temperature estimates does not change appreciably with this step. Therefore, while the combined infilling of PCs and ground stations technically invalidates the calibration (and should be avoided), it has no material impact on the S09 results. We note, however, that this is not a general statement, as other applications may not be as forgiving.

Mods 2 – 4, however, all exhibit a significant impact on the results. In terms of overall trend magnitude, Mods 2 – 4 each contribute approximately 33% of the difference between S09 and our results. In terms of overall spatial pattern reorganization, Mod 4 (truncation parameter) is the primary contributor. Mods 2 and 3, however, also result in substantial reorganization, especially when examining reconstruction subperiods. Since the patterns in our reconstructions cannot be obtained simply by increasing the number of retained AVHRR eigenvectors, it follows that the calibration and spatial structure concerns with S09 are significant. These results are summarized in Table 7. Note that Mod 3 is not broken out separately, as proper implementation requires both Mods 1 and 2. We therefore arrive at an estimate for the influence of Mod 3 via deduction.

We can, however, qualitatively examine the improvement due to Mod 3. Again making use of equation (8), we can evaluate the difference in spatial structure between the AVHRR data and the spatial structure used to infill the PCs (Fig. 7). Because our reconstructions use more stations than S09, some stations not present in Fig. 2 are included. There is a noticeable improvement in the match between the TTLS regression coefficients and the AVHRR spatial eigenvectors as compared to the S09 method. The rms error estimates between the normalized spatial weights for this method are 0.91, 0.56, and 0.44 for PCs 1-3, respectively. For S09, the corresponding estimates are 0.93, 0.98, and 1.13. 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 error in the reconstruction. We also note that PC 1 shows a very similar pattern with a nearly constant offset, which reflects the lower trends in the ground data than in the AVHRR data.

Fig. 7 also provides some insight into the differences between the RLS and E-W reconstructions. In the E-W method, constraining the regression by the AVHRR spatial eigenvectors does not result in an exact match. This may be due to the effects of sampling error on the correlation matrix during the regression, errors in the AVHRR spatial structure, errors in the ground data, and / or other, as-yet-unidentified reasons. In the case of the RLS reconstructions, however, the AVHRR spatial eigenvectors are *directly* regressed against the ground station information, and therefore match perfectly. This will result in slightly different gridded estimates. More complete understanding of these reasons is the subject of ongoing work.

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: a) improper calibration; b) spatial structure differences between the infilling operation and recovery of gridded estimates; and, c) suboptimal determination of truncation parameters. We propose four modifications to correct these issues.

We demonstrate that our concerns have a material impact on the results. When resolved, the results obtained differ from S09 in several key aspects. While we find some agreement with S09 (specifically, overall positive trends, lesser warming in East Antarctica, and similar seasonal patterns in the Peninsula and in parts of West and East Antarctica), we also find substantial differences. Average 1957 – 2006 temperature trends for the continent, East Antarctica and West Antarctica are halved. We find an average Peninsula trend of approximately $0.3\text{ }^{\circ}\text{C decade}^{-1}$, which is almost three times that of S09. Statistically significant warming in West Antarctica is limited to the area adjacent the Peninsula, and the Ross region shows cooling, not maximum warming. East Antarctica displays a persistent cooling feature extending from the South Pole to the Weddell Sea, and large portions of West and East Antarctica display substantially different seasonal behavior.

Though we find the general concept of regressing satellite principal components against ground information using an infilling algorithm appears to have merit, care must be taken to ensure a proper calibration. We observe that in cases where the temporal

component of a data set may be suspect, a method using only spatial information from the satellite data may provide more accurate results. The method also presents itself as a diagnostic tool; one could easily compare results between temporal and spatial methods. This diagnostic may be performed 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 differences between the AVHRR data set used by S09 and temperatures measured at ground station locations. The small regularization parameters required for optimal verification statistics using only ground station temporal information indicate the error is likely to be with the AVHRR data. 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 suggest similar problems may exist in other AVHRR temperature products.

Acknowledgments.

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

References

- 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
- 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
- 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
- Connolley, W. M., and T. J. Bracegirdle, 2007: An Antarctic assessment of AR4 coupled models. *Geophys. Res. Lett.*, **34**, L22505, doi:10.1029/2007GL031648
- 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
- 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
- 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. 1. Geographic color codes used in Figs. 2 and 5.

FIG. 2. Spatial structure used to estimate the satellite PCs in the 1957-1981 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, with station colors per Fig. 1. Gold indicates AVHRR PCs. Weights are normalized such that the variance is unity.

FIG. 3. 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. 4. Spatial patterns of change, with regions of statistically insignificant trends overlaid in gray.

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.

FIG. 6. Comparison of annual average anomalies for the 7 most complete Peninsula stations (Arturo Prat, Bellinghousen, Esperanza, Faraday, Marambio, O'Higgins, and Rothera) versus reconstruction anomalies at the corresponding grid cells. Black: ground stations. Green: RLS. Gold: E-W. Red: S09.

FIG. 7. Spatial structure used to estimate the satellite PCs when constrained by 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.

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. 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	<u>7</u>	<u>80</u>	<u>0.1</u>	<u>Cor</u>
		6	87	0.1	Cov
	TTLS	7	100	9	Cor
		6	100	9	Cov
E-W	TSVD	<u>7</u>	<u>100</u>	<u>9</u>	<u>Cor</u>
		6	100	9	Cov

TABLE 3. Regional trend comparison between this study (RLS and E-W reconstructions) and the S09 reconstruction with values in $^{\circ}\text{C decade}^{-1}$.

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 4. Comparison of seasonal trends with S09 (values in $^{\circ}\text{C decade}^{-1}$). The two most rapidly warming seasons are bolded and italicized (3 in the case of a tie).

Method	Region	Winter	Spring	Summer	Fall
RLS	Continent	0.07 +/- 0.04	<i>0.09 +/- 0.03</i>	<i>0.08 +/- 0.03</i>	0.01 +/- 0.04
	East	<i>0.09 +/- 0.05</i>	<i>0.07 +/- 0.03</i>	<i>0.07 +/- 0.03</i>	-0.02 +/- 0.01
	West	-0.08 +/- 0.05	<i>0.15 +/- 0.04</i>	<i>0.08 +/- 0.03</i>	0.05 +/- 0.04
	Peninsula	<i>0.40 +/- 0.07</i>	0.17 +/- 0.04	0.18 +/- 0.02	<i>0.40 +/- 0.04</i>
E-W	Continent	0.04 +/- 0.04	<i>0.09 +/- 0.04</i>	<i>0.07 +/- 0.02</i>	0.01 +/- 0.04
	East	0.04 +/- 0.05	<i>0.08 +/- 0.03</i>	<i>0.06 +/- 0.03</i>	-0.01 +/- 0.01
	West	-0.01 +/- 0.03	<i>0.11 +/- 0.03</i>	<i>0.06 +/- 0.02</i>	0.02 +/- 0.03
	Peninsula	<i>0.41 +/- 0.05</i>	0.19 +/- 0.03	0.20 +/- 0.02	<i>0.34 +/- 0.03</i>
S09	Continent	<i>0.17 +/- 0.05</i>	<i>0.16 +/- 0.04</i>	0.09 +/- 0.05	0.05 +/- 0.04
	East	<i>0.15 +/- 0.05</i>	<i>0.14 +/- 0.04</i>	0.08 +/- 0.05	0.02 +/- 0.04
	West	<i>0.27 +/- 0.05</i>	<i>0.23 +/- 0.04</i>	0.11 +/- 0.04	0.16 +/- 0.04
	Peninsula	<i>0.17 +/- 0.03</i>	<i>0.14 +/- 0.02</i>	0.07 +/- 0.02	0.12 +/- 0.02

TABLE 5. 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. For values of k_{gnd} other than 7, the remaining parameters were not optimized in order to isolate the effects of changing k_{gnd} alone.

Reconstruction	k_{gnd}	Trend ^a		CE ^b	
		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
	<u>7</u>	<u>0.05 ± 0.08</u>	<u>0.11 ± 0.07</u>	<u>0.65</u>	<u>0.51</u>
	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
	<u>7</u>	<u>0.04 ± 0.06</u>	<u>0.06 ± 0.06</u>	<u>0.51</u>	<u>0.42</u>
	8	0.09 ± 0.06	0.09 ± 0.06	0.49	0.41
S09	3 ^c	$0.20 \pm 0.09^{\text{d}}$		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 6. Summary verification statistics and comparison to values calculated from the published T_{IR} reconstruction.

	Full recon ^a			Verification recon ^b		
	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

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

^b For the additional 35 stations withheld from the verification reconstructions

TABLE 7. Effects of the methodological deficiencies in S09 on overall and regional trends (values in $^{\circ}\text{C decade}^{-1}$).

	S09	Mod 2 Only	Mod 4 Only	Mods 2 & 4	E-W	RLS
Continent	0.12	0.10	0.10	0.07	0.05	0.06
East	0.10	0.08	0.08	0.07	0.04	0.05
West	0.20	0.16	0.13	0.08	0.04	0.05
Peninsula	0.13	0.11	0.28	0.26	0.29	0.29

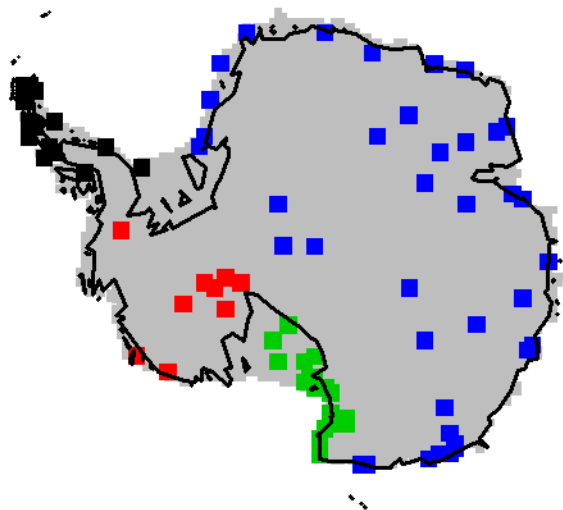


FIG. 1. Geographic color codes used in Figs. 2 and 5.

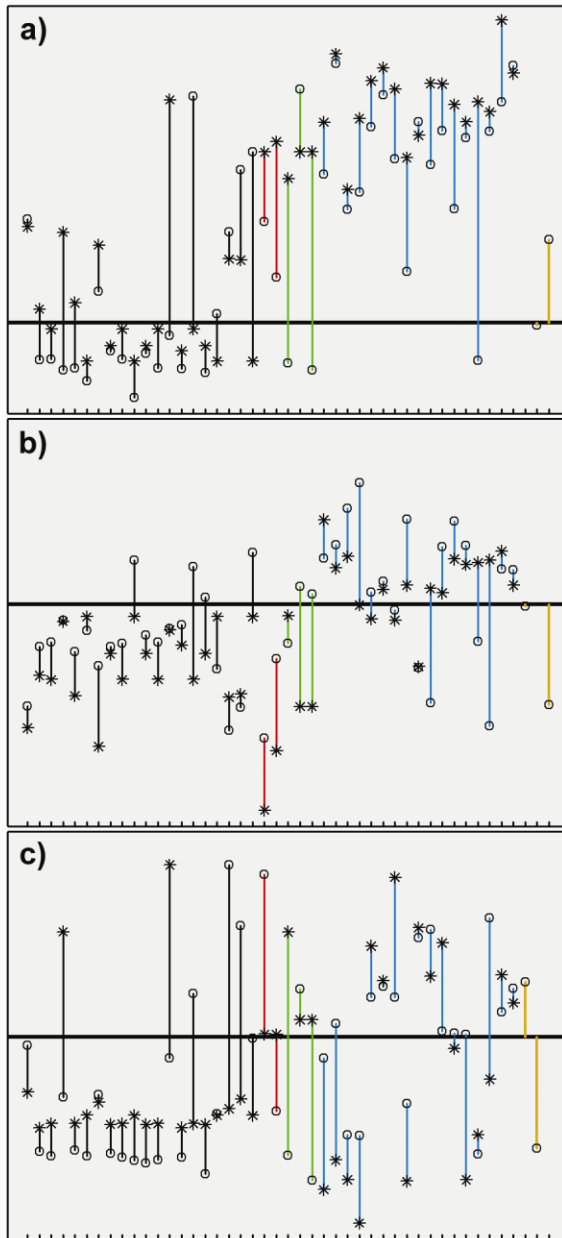


FIG. 2. Spatial structure used to estimate the satellite PCs in the 1957-1981 period by S09 vs. AVHRR spatial structure. Top: PC #1. Middle: PC #2. Bottom: PC #3. Circles represent the normalized weight in the imputation EOFs; stars represent the normalized AVHRR spatial eigenvector weights. Geographic location indicated by color, with station colors per Fig. 1. Gold indicates AVHRR PCs. Weights are normalized such that the variance is unity.

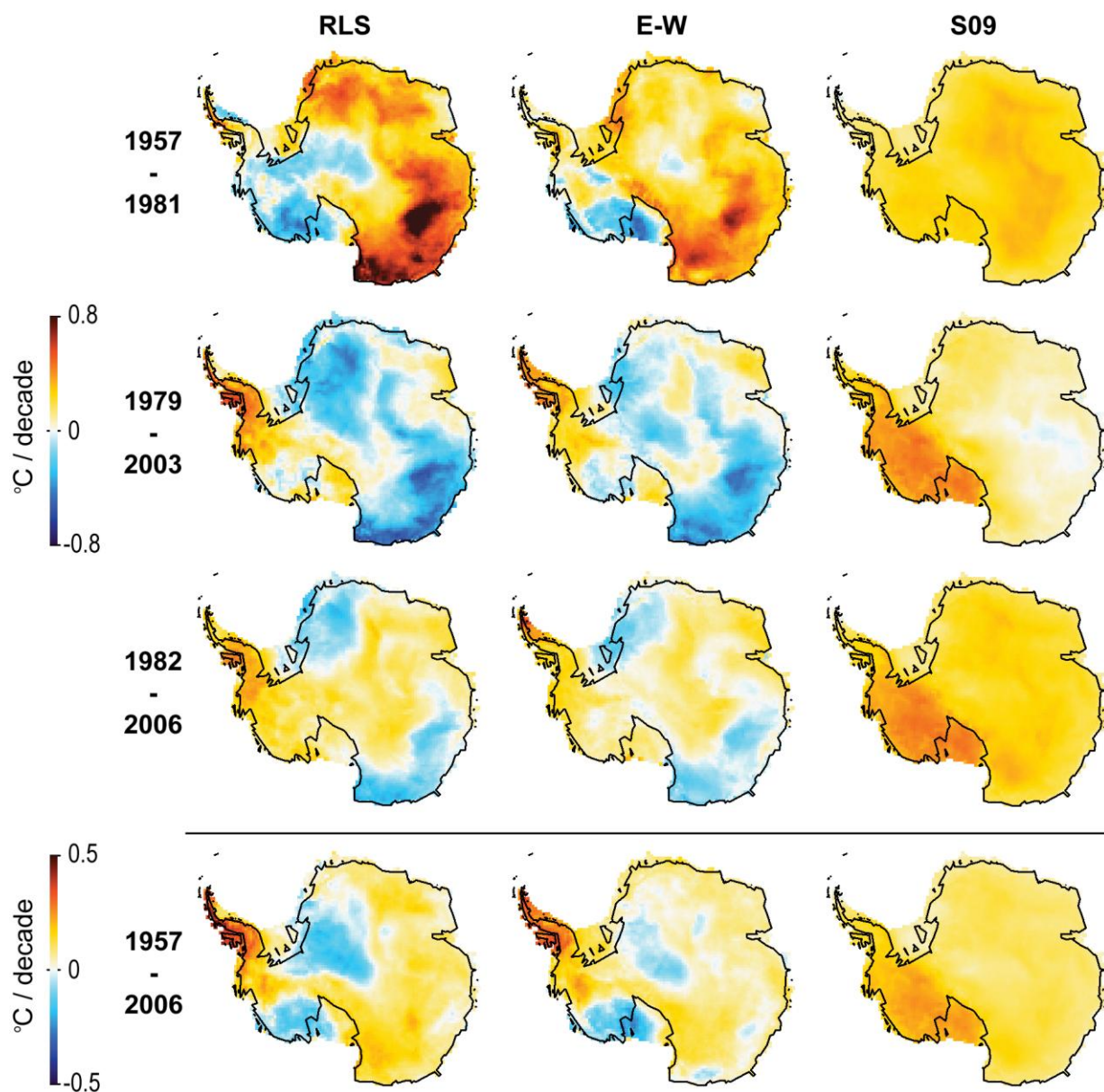


FIG. 3. Comparison of spatial patterns of change for RLS, E-W, and S09 reconstructions.

Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.

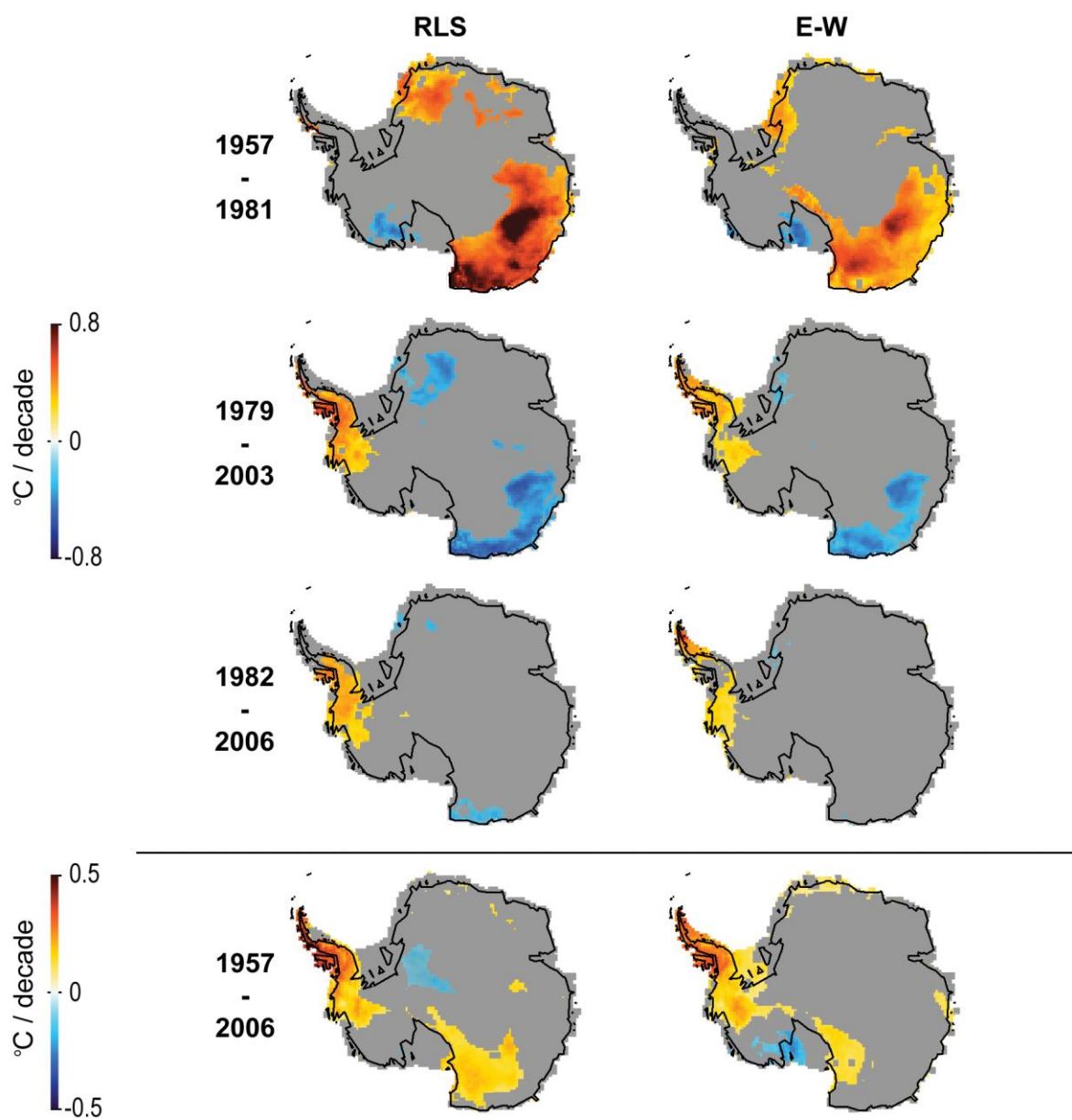


FIG. 4. Spatial patterns of change, with regions of statistically insignificant trends overlaid in gray.

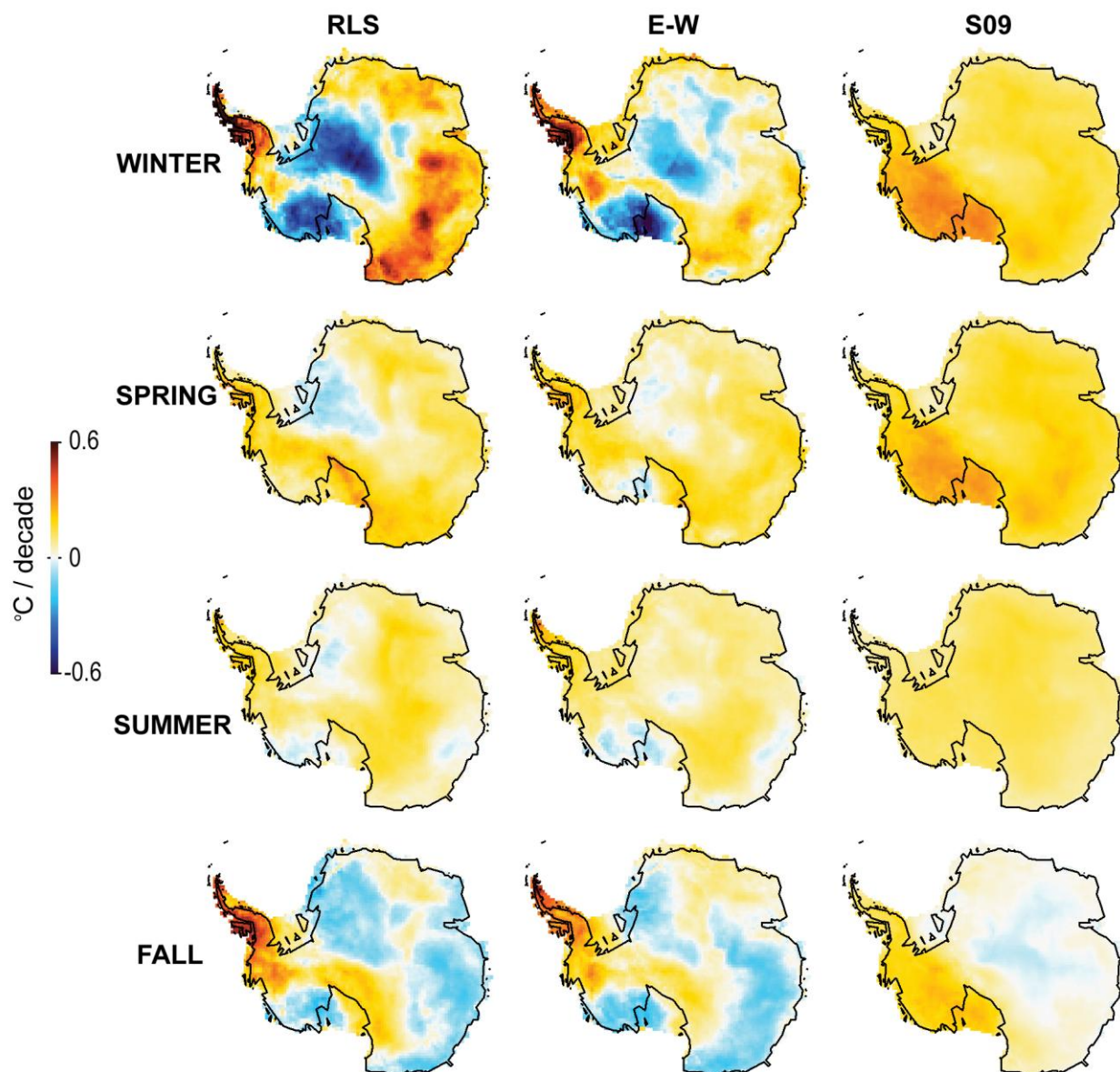


FIG. 5. Comparison of seasonal patterns of change for RLS, E-W, and S09 reconstructions for 1957 - 2006. Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.

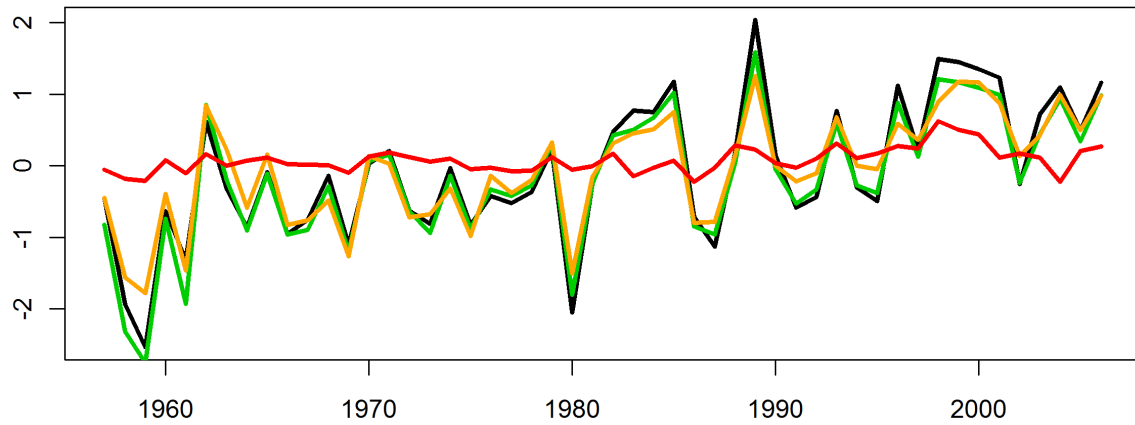


FIG. 6. Comparison of annual average anomalies in $^{\circ}\text{C}$ for the 7 most complete Peninsula stations (Arturo Prat, Bellinghousen, Esperanza, Faraday, Marambio, O'Higgins, and Rothera) versus reconstruction anomalies at the corresponding grid cells. Black: ground stations. Green: RLS. Gold: E-W. Red: S09.

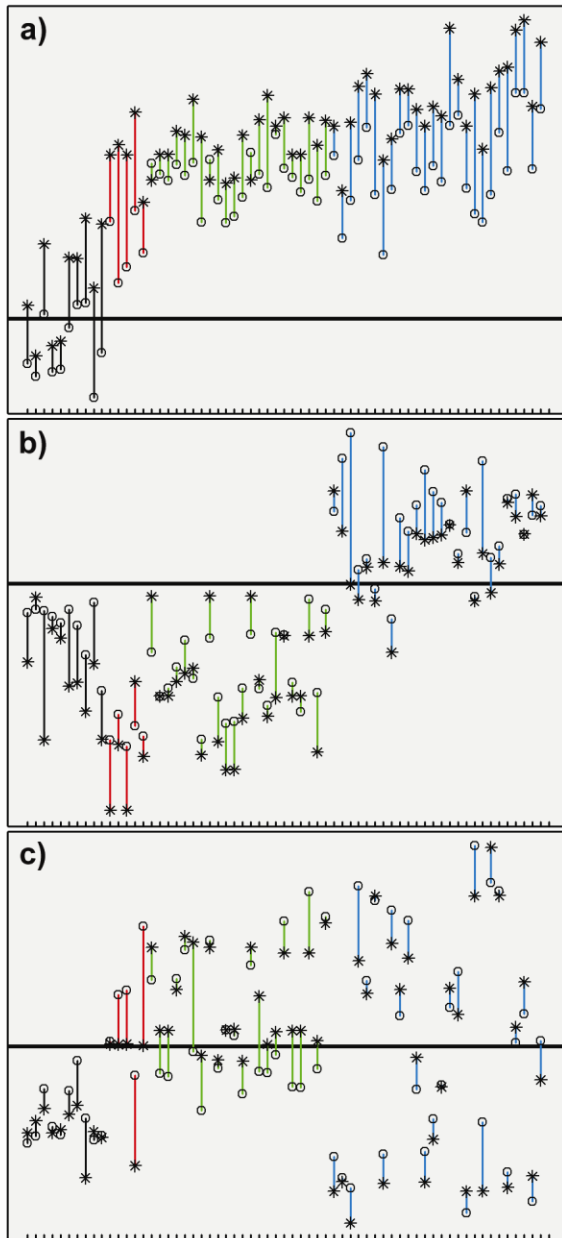


FIG. 7. Spatial structure used to estimate the satellite PCs when constrained by AVHRR eigenvector weighting vs. AVHRR spatial structure. Top: PC #1. Middle: PC #2. Bottom: PC #3. Colors, symbols, and normalization are identical to Fig. 2.