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

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 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; (b) improper determination of spatial structure during infilling; and (c) suboptimal determination of regularization parameters, particularly with respect to satellite principal component retention. We propose two methods to resolve these issues. One 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 that disagree with the previous method in several aspects. Rather than finding warming concentrated in West Antarctica, we find warming over the period of 1957-2006 to be concentrated in the Peninsula ($\approx 0.35^{o}$C decade^{-1}). We also show average trends for the continent, East Antarctica, and West Antarctica that are half or less than that found using the unimproved method. Notably, though we find warming in West Antarctica to be smaller in magnitude, we find that statistically significant warming extends at least as far as Marie Byrd Land. We also find differences in the seasonal patterns of temperature change, with winter and fall showing the largest differences and spring and summer showing negligible differences outside of the Peninsula.

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. 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} or S09 reconstruction[1]. 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 is dealt with implicitly, as our proposed modifications likewise separate the estimation of missing AVHRR PC and ground station information. The third

[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; however, due to the detrending, resulting trend magnitudes are not directly comparable.

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.

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 reader should note that to allow broader comparisons, the values 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. The shaded regions in Fig. 1 depict the definitions used for this study.

3. Calibration via infilling?

a. Sources of systematic error in the AVHRR data

The AVHRR instrument is carried aboard the NOAA 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 and some of which show latitudinal variations. Additionally, the AVHRR instrument measures skin temperature rather than 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 interchangeable[2]. 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 correlation[4] 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:

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

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

[2] An embedded trend in the satellite data is apparent from 1982 – 2001 (NOAA – 7 through NOAA – 14) that exceeds ground station measurements by 0.19 +/- 0.16^{o}C decade^{-1}.
[3] Since $\mathbf{A}$ and $\mathbf{b}$ are defined separately for each time step, the S09 definition is not strictly true (Section 3.c).
[4] Schneider (2001) and Mann et al. (2007) describe the algorithm as using the covariance matrix; however, RegEM scales to correlation prior to regularization.

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 $(1...k)$ and $(k+1...n)$, respectively:

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

The TTLS 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 † indicates the generalized inverse:

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

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

$$\tilde{\hat{\mathbf{b}}}_j = \tilde{\mathbf{A}}_j \mathbf{x}_{j,k} \tag{5}$$

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 $\mathbf{V}_{1...k}^{\mathrm{T}} = \begin{pmatrix} \mathbf{V}_{1,1} \\ \mathbf{V}_{2,1} \end{pmatrix}^{\mathrm{T}}$, yielding a full set of statistical weights to provide estimates for both missing and actual values:

$$\tilde{\hat{\mathbf{Y}}}_j = (\tilde{\hat{\mathbf{A}}}_j \;\; \tilde{\hat{\mathbf{b}}}_j) = \tilde{\mathbf{A}}_j (\mathbf{V}_{1,1}^{\mathrm{T}})^{\dagger} \mathbf{V}_{1...k}^{\mathrm{T}} \tag{6}$$

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):

$$\mathbf{C}_{\mathrm{res}} = \mathbf{V}_{2,2} \mathbf{\Lambda}_{k+1...n}^{2} \mathbf{V}_{2,2} \tag{7}$$

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 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 are not valid. In the special case that the response and explanatory variables are 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 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 temporally *incomplete*. 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. From equation (4), the missing data will be estimated using correlated – not

orthogonal – PCs. This means that the response variables are also partially calibrated to each other. In the case of PC 2 and PC 3, the correlation following regularization for the entire 1957 – 2006 period is a factor of 2.5 higher than that following the initial regularization (-0.2501 vs. -0.1001) after 30 iterations.

Next, RegEM 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 using a total least squares algorithm (which minimizes the errors in both the available and missing values) 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. 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 $\mathbf{x}_k$ have a spatial distribution identical to the corresponding AVHRR spatial eigenvector weights, recovering gridded estimates using the unaltered AVHRR spatial eigenvectors will result in a geographical translation of information and a change in the magnitude of the estimates.

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 $\tilde{\tilde{\mathbf{Y}}}$ of rank k, we may find the vector of coefficients $\mathbf{v}_i$ that describe the contribution of all variables to any i^{th} variable in $\tilde{\tilde{\mathbf{Y}}}$. If the i^{th} variable is an AVHRR PC, $\mathbf{v}_i$ then yields what the AVHRR spatial eigenvector weights *should be* in order to best reproduce the ground data used to predict the PC.

From equation (1), the SVD will yield the spatial eigenvectors and eigenvalues $\mathbf{\Lambda V}$ ("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 $\tilde{\tilde{\mathbf{Y}}}$. Since every series in $\tilde{\tilde{\mathbf{Y}}}$ 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 $\tilde{\tilde{\mathbf{Y}}}$ 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:

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

Fig. 2 (with metadata available in the Supporting Information) compares $\mathbf{v}_i$ to the AVHRR spatial eigenvector weights for the three PCs retained by S09. In order to use the AVHRR spatial weights (the stars in the figure) without geographically relocating temperature information, the TTLS weights (the circles) must be comparable. However, 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, 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. Additionally, the default output from RegEM does not yield

the calibrated PCs. 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). To perform these modifications, 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 since the ground station estimation is complete prior to the PCs being regressed. This also 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, appropriately scaled to reduce the influence of the PC on the decomposition. 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 calibration and spatial structure concerns is to avoid using the AVHRR PCs. 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 $\mathbf{L}_{AVHRR} = \sqrt{1/\tilde{n}}\ (\mathbf{\Lambda}_{AVHRR}\ \mathbf{V}_{AVHRR})$, where $\mathbf{\Lambda}_{AVHRR}$ contains the AVHRR eigenvalues $1...k$, $\mathbf{V}_{AVHRR}$ represents the weights of the spatial eigenvectors at the ground station locations, and $\tilde{n}$ represents the effective degrees of freedom. We then define a matrix of RegEM-infilled ground station observations $\mathbf{Y}$, unknown regression coefficients $\mathbf{a}$ and write:

$$\mathbf{L}_{AVHRR}\mathbf{a} = \mathbf{Y} \qquad (9)$$

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 $\mathbf{Y}$ (subscripts hereafter omitted for readability):

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

As we do not know the proper regularization parameter from any *a priori* physical arguments, we determine the parameter h^2 by minimizing the rms error between the reconstruction and all stations (including ones not used as predictors) via explicit leave-one-out cross-validation. The value of h^2 that minimizes error in withheld data can be interpreted as the maximum likelihood estimation of the true ratio of system measurement error and noise (Fitzpatrick 1991; Sima 2006), where noise refers to the amount of information on each predictor that is not usable for prediction.

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 regularization parameters

The final modification, which applies to both the E-W and RLS methods, is to determine the optimum regularization 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. 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, k_{RegEM} to denote the truncation parameters used by the infilling algorithm to regress the PCs against the ground stations for the E-W reconstructions, and h^2 to denote the regularization parameter for RLS.

We use several algorithms for the cross-validation experiments. The first is the TTLS algorithm of Schneider (2001) – which was the algorithm used by S09 – reprogrammed in the R Programming Language, and benchmarked against the original Matlab version. The second is a Truncated SVD (TSVD) algorithm similar to the DINEOF routine of Beckers and Rixen (2003). The third is the ridge regression (IRidge) algorithm of Schneider (2001) which utilizes Tikhonov regularization rather than truncation, also reprogrammed in R and benchmarked against the Matlab version.

Unlike IRidge where the regularization parameter is determined via minimization of the generalized cross validation (GCV) function of Golub et al. (1979), no analytical

cross-validation function is known for TTLS and TSVD. We therefore determine the optimal value of k_{gnd} by explicitly calculating reconstruction rms error to withheld stations. For this procedure, we withhold one station at a time from the ground station set and infill the remaining stations at different values of k_{gnd}. We then perform RLS and E-W reconstructions and calculate rms error to the withheld station. The value of k_{gnd} that results in the lowest overall rms error for the withheld stations is selected as optimal.

For the E-W reconstructions, we determine k_{RegEM} by regressing each AVHRR PC against the infilled ground station set, performing the reconstruction, and choosing the parameter that yields the lowest rms error between the reconstruction and all stations, including those not used as predictors. For RLS reconstructions, we determine the regularization parameter h^2 via explicit leave-one-out cross validation during the regression (section 6.d). The number of AVHRR PCs included in the reconstruction - k_{sat} - is likewise determined by minimizing the rms error on all stations.

Lastly, after the optimal parameters have been chosen, we evaluate how well the regression and reconstruction procedure may estimate temperatures at uninstrumented locations in a similar fashion as that used to determine k_{gnd}. One station at a time is entirely withheld from the ground station infilling and is also entirely withheld from the RLS and E-W reconstructions (including the cross-validation steps). We then calculate rms error (μ), coefficient of efficiency (CE), and correlation coefficient (r) to the withheld station. Because this is repeated for every station, we can obtain a complete set of verification statistics for all predictors. We also perform this procedure using the S09

reconstruction in order to have a like-to-like comparison of the effectiveness of the reconstruction methods.

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

1) PARAMETERS FOR GROUND STATION INFILLING

Because we find that including short-record stations results in degraded performance of the reconstruction, we limit our set of predictors to all READER stations

that are located within 100km of an AVHRR grid cell and have at least 96 months of data. This yields 63 stations (35 AWS and 28 manned ground stations), which are tabulated in the Supporting Information.

We find TTLS and TSVD to yield nearly identical results, with TTLS showing slightly lower (0 – 5%) reconstruction trends depending on the regularization parameter used. Due to the increased efficiency of TSVD (the algorithm is approximately 10 times faster than TTLS), we conduct most of the cross-validation using TSVD and spot-check with TTLS. Both algorithms yield an identical optimal k_{gnd} of 7.

However, we find the regularization method in TTLS and TSVD to be somewhat undesirable. While the optimal value of k_{gnd} does indeed yield comparable reconstructions to IRidge, small changes in the regularization parameter can cause large changes in the resulting reconstructions. Because the regularization parameter is fixed for the entire data set, the parameter that is ideal for the data set as a whole causes overfits during period with few predictors and underfits during periods with many predictors, yielding lower overall prediction effectiveness. In experiments where portions (5%) of the station data are withheld, we obtain maximum CEs to the withheld data of approximately 0.50 with TTLS and TSVD, but in excess of 0.63 with IRidge. This possibility – that the smooth regularization and ability to adapt the regularization to the number of predictors available that is provided by ridge regression would prove more effective in practice than TTLS – is pointed out in Schneider (2001). In addition to the problems resulting from coarse, fixed regularization, TTLS and TSVD lack a known analytical minimization function. This makes the computational cost advantage over

ridge regression noted as attractive by Mann et al. (2007) entirely evaporate due to the need for extensive, explicit cross-validation.

Given these issues with TTLS and TSVD, we choose to emphasize the reconstructions using IRidge-infilled ground stations in the main text. We do note, however, that the optimal parameters for TTLS and TSVD provide reconstructions that are quite comparable to IRidge. TTLS-infilled reconstructions are available in the Supporting Information.

2) E-W RECONSTRUCTION PARAMETERS

Unlike the ground station infilling, we found TTLS and TSVD to provide stable results for the PC regressions in the E-W reconstructions. In this case, the number of predictors is stable and there is only one predictand. This prevents having to select the truncation parameter based on best overall performance for multiple predictands and allows selection of the parameter based on the best performance for the single predictand. As a result, the underfitting / overfitting issues associated with the ground station infilling are avoided. For the E-W reconstructions – which are far more computationally intensive than the RLS reconstructions – we found the speed advantage of TTLS and TSVD to be particularly useful.

For the E-W reconstruction, imposing an upper limit of $k_{\mathrm{RegEM}} \leq 8$ during the ground station / AVHRR PC regression yields the best verification statistics for both TTLS and TSVD, with TSVD showing slightly higher trends (approximately 5%) and slightly better verification statistics. In contrast with the ground station infilling, the results are relatively insensitive to changes in the maximum allowable value for k_{RegEM}, with only a 0.01^{o}C decade^{-1} difference in the continental and West Antarctic trends as the

upper limit on k_{RegEM} is varied from 5 to 11. Below 5, trends in all areas are significantly reduced.

The optimal value for k_{sat} also shows little sensitivity to k_{RegEM}, yielding an ideal number of 150 retained satellite PCs with $k_{\mathrm{RegEM}} \leq 8$. Overfitting results in a rapid decrease in verification statistics and optimal number of retained satellite PCs when the upper limit on k_{RegEM} is allowed to exceed 11.

3) RLS RECONSTRUCTION PARAMETERS

For the RLS reconstructions, an optimal regularization parameter h^2 of 0.34 is obtained, with a resulting optimal value for k_{sat} of 126. Reconstruction trends show little sensitivity (decreasing by approximately 0.02^{o}C decade^{-1}) for values of h^2 up to 1.5, and are generally lower if fewer AVHRR eigenvectors are retained. An important observation is that a very small amount of regularization – approximately 99.3% of the predictor variance is retained – 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 (≈0.06 +/- 0.07 ^{o}C decade^{-1}) [5], nor is the warming in East Antarctica (≈0.03 +/- 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,

[5] 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).

we find the Peninsula regional average to demonstrate the most strongly positive trend, at 0.35 +/- 0.11 from 1957 – 2006. For West Antarctica, the RLS reconstruction demonstrates a statistically significant trend of 0.10 +/- 0.09 (approximately half that reported by S09) while the E-W reconstruction shows a trend of 0.06 +/- 0.07. Given that the E-W reconstruction displays some variance loss compared to both RLS and the raw ground station data, we consider the RLS result to be the more accurate of the two methods. This results in a total West Antarctic warming of 0.5^{o}C since 1957, which is roughly equivalent to the overall warming of the earth over the same period.

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, particularly in the Peninsula half of Marie Byrd Land and in the Pine Island and Thwaites glacier locations. 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 an area of mild cooling over the Ross Ice Shelf and adjacent land (not statistically significant) 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 5 provides spatial maps of statistically significant trends.

Our results – including the strong Peninsula warming, insignificant 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), our reconstructions show positive trends at Byrd AWS. For the entire 1957 – 2006 period, the reconstructed trends at the Byrd location are 0.27 +/- 0.10 (RLS) and 0.23 +/- 0.07 (E-W). For the satellite coverage period of 1982 – 2006, those become 0.44 +/- 0.30 (RLS) and 0.21 +/- 0.20 (E-W). By comparison, the S09 Byrd trends are 0.18 +/- 0.08 and 0.33 +/- 0.23, respectively.

Also unlike Monaghan et al. (2008), our reconstructions show little sensitivity to the removal of the manned Byrd station data, which decreases the West Antarctic regional trend by only 0.015. We do note, however, that the ground data from the interior of West Antarctica is sparse, with only the manned Byrd station providing any pre-1980 observations. The reconstructed trends in Marie Byrd Land, therefore, are essentially an interpolation between the long-record Peninsula and Ross area stations, using the AVHRR spatial data, the manned Byrd station, and infilled values for four West Antarctic stations (Byrd AWS, Erin, Henry, and Mount Siple) as constraints.

c. Seasonal patterns of temperature change

In comparing seasonal patterns of change, our reconstructions show minor to negligible differences compared to S09 in regional averages on the mainland for spring and summer (Table 2 and Figures 4, 5 and 6). Winter and fall show more substantial differences, especially in the Ross and Weddell regions. The Peninsula trends are substantially different for all seasons and all time periods.

In terms of seasonal maximums, S09 find the most 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 peaks primarily in spring and summer. We additionally show two separate patterns of change in West Antarctica, with the Ross region showing *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 again 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 are coupled to 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 winter trends at Amundsen-Scott (from the READER archive) of -0.34 +/- 0.07, compared to -0.33 +/- 0.07 (RLS) and -0.19 +/- 0.06 (E-W). 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 find cooling at the pole during all seasons, with a maximum cooling trend during winter in the 1958 – 2002 timeframe.

Statistical significance of the differences between our study and S09 were calculated via by a one sample t-test on the residual trend (or, equivalently, a paired t-test on the monthly estimates). Spatial maps of the differences between S09 and our reconstructions for all seasons and each season individually over the period of 1957 – 2006 appear in Fig. 6 and results for regional averages appear in Table 2.

d. Skill statistics

The top half of Table 3 summarizes the skill of the full reconstructions with respect to rms error (μ_{rms}), correlation coefficient (r), and average explained variance (R^2), which indicate how well the reconstruction performs at locations where ground information is available. Both the RLS and E-W reconstructions show a substantially better match to ground data than either the S09 reconstruction or the raw AVHRR data. This is not surprising for RLS, as the temporal information comes entirely from the ground stations. For the E-W reconstructions, which regress the AVHRR PCs against the ground data, the improvements validate our concerns with the S09 method.

Verification statistics (bottom half of Table 3) are calculated by comparing reconstructed temperatures to station data that is withheld from the reconstruction as described in Section 6.e. Statistics calculated are μ_{rms}, 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. A noteworthy result is that the verification statistics – wherein no station data is used at the locations being tested – *also* substantially exceed the performance of both the full S09

reconstruction *and* 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. 7 shows annual anomalies for the 7 most complete Peninsula ground stations vs. the corresponding grid cells from the RLS, E-W, and S09 reconstructions. A baseline period of 1970 – 1985 was chosen due to some stations being incomplete outside that period. The ground data trend is 0.46 +/- 0.18^{o}C decade^{-1}. The reconstruction trends are 0.44 +/- 0.16 (RLS), 0.39 +/- 0.14 (E-W), and 0.06 +/- 0.05 (S09).

e. 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_{sat} = 150 the overall pattern and magnitude of trends do not change appreciably. 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.

Mods 2 – 4, however, all exhibit a significant impact on the results. Each contribute approximately 33% of the difference in the magnitude of the continental average, which limits the first-order effect of errors in the AVHRR data on the S09 reconstruction to about 0.02 ^{o}C decade^{-1}. This indicates that although the detrended S09

reconstruction demonstrates similar magnitudes to our reconstructions, the similarity cannot be wholly due to trends in the AVHRR data. Loss of spatial covariance information (also recognized by S09) contributes to the reduction in magnitude by degrading the quality of the regression.

In terms of overall spatial pattern reorganization, Mod 4 (cross-validation) is the primary contributor, though Mods 2 and 3 also result in substantial reorganization, especially when examining reconstruction subperiods (Table 4). 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.

While the contribution of Mod 2 may be determined independently, Mod 3 requires both 1 and 2. We therefore estimate the influence of Mod 3 via deduction. For a qualitative examination, we may make use of equation (8) to evaluate whether Mod 3 results in a better match between the TTLS regression coefficients and the AVHRR spatial eigenvectors (Fig. 8). There is a noticeable improvement as compared to Fig. 2. 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, as compared to 0.93, 0.98, and 1.13 for S09. As expected, the E-W 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.

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 regularization 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 notable agreement with S09 (specifically, overall positive trends, statistically significant warming in West Antarctica, lesser warming in East Antarctica, and overall warming of the continent in summer and spring), 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 0.35 ^{o}C decade^{-1}, which is almost three times that of S09. Maximum warming in West Antarctica occurs in the area adjacent the Peninsula rather than on Ross. 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. All of these differences are statistically significant at the 5% level.

Though we find the general concept of regressing satellite principal components against ground information using an infilling algorithm 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.

Furthermore, we find the heuristic procedure of Mann et al. (2007) for determining truncation parameters to be suboptimal, at least in the case of Antarctica. This procedure warrants more investigation in order to establish the conditions under which it might provide optimal or near-optimal results. Additionally, we find strong evidence that regularization via truncation with a fixed parameter (as in TTLS and TSVD) provides near-optimal results only when the number of predictors is stable and the number of predictands is small (or, equivalently, highly spatially coherent).

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 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

Golub, G. H., M. T. Heath, and G. Wahba, 1979: Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21,** 215–223

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

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., 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 regions and station color codes. The Peninsula region is shaded in orange; West Antarctica in tan; and East Antarctica in gray. Circles: S09 station locations (with Campbell, Grytviken, and Macquarie too distant to be shown). Crosses: stations used as predictors for the improved reconstructions of this study. Triangles: stations used as additional verification targets for the improved reconstructions of this study, but were not used as predictors. The colors of the station symbols correspond to the colors used in Figs. 2 (S09) and 8 (this study).

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 TTLS weights for the PCs; stars represent the normalized AVHRR spatial eigenvector weights for the PCs. 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. 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. 5. Spatial patterns of change (1957 – 2006), with regions of statistically insignificant trends overlaid in gray ($p > 0.05$, two-tailed).

FIG. 6. Spatial patterns of change (1957 – 2006), with regions of statistically insignificant differences between S09 and RLS / E-W reconstructions overlaid in gray ($p > 0.05$, two-tailed).

FIG. 7. 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. 8. 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. Regional trend comparison between this study (RLS and E-W reconstructions) and the S09 reconstruction over the period of 1957 – 2006 with values in ^{o}C decade^{-1}.

Region	RLS[a]	E-W[a]	S09[a]
Continental Average	0.06 ± 0.08	0.04 ± 0.06	0.12 ± 0.09
East Antarctica	0.03 ± 0.09	0.02 ± 0.07	0.10 ± 0.10
West Antarctica	0.10 ± 0.09	0.06 ± 0.07	0.20 ± 0.09
Peninsula	0.35 ± 0.11	0.32 ± 0.09	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 2. Comparison of regional and seasonal trends with S09 (values in ^{o}C decade^{-1}) over the period of 1957 – 2006. Italics indicate trends in the residuals between S09 and either RLS or E-W reconstructions. Underline indicates where the S09 reconstruction is not significantly different from the RLS reconstruction. Bold indicates where the S09 reconstruction is not significantly different from the E-W reconstruction. All confidence intervals are 95% ($p > 0.05$, two-tailed) with degrees of freedom corrected for serial correlation of the residuals (Santer et al., 2000).

Season	Region	S09	RLS	S09 – RLS Residual Trend	E-W	S09 – EW Residual Trend
ALL	Cont.	0.12 ± 0.09	0.06 ± 0.08	*0.06 ± 0.05*	0.04 ± 0.06	*0.08 ± 0.05*
	East	0.10 ± 0.10	0.03 ± 0.09	*0.07 ± 0.06*	0.02 ± 0.07	*0.08 ± 0.06*
	West	0.20 ± 0.09	0.10 ± 0.09	*0.09 ± 0.06*	0.06 ± 0.07	*0.13 ± 0.06*
	Pen.	0.13 ± 0.05	0.35 ± 0.11	*-0.22 ± 0.10*	0.32 ± 0.09	*-0.20 ± 0.07*
Winter	Cont.	0.17 ± 0.10	0.08 ± 0.10	*0.09 ± 0.04*	0.03 ± 0.08	*0.14 ± 0.04*
	East	0.15 ± 0.11	0.06 ± 0.11	*0.09 ± 0.04*	-0.00 ± 0.09	*0.15 ± 0.04*
	West	0.27 ± 0.11	0.07 ± 0.11	*0.20 ± 0.07*	0.06 ± 0.08	*0.21 ± 0.07*
	Pen.	0.17 ± 0.07	0.51 ± 0.16	*-0.34 ± 0.13*	0.54 ± 0.13	*-0.36 ± 0.10*
Spring	Cont.	0.16 ± 0.07	0.11 ± 0.06	*0.06 ± 0.05*	0.09 ± 0.05	*0.07 ± 0.05*
	East	0.14 ± 0.08	0.07 ± 0.07	*0.07 ± 0.05*	0.08 ± 0.05	*0.07 ± 0.06*
	West	0.23 ± 0.08	0.21 ± 0.10	*0.03 ± 0.06*	0.13 ± 0.07	*0.10 ± 0.05*

	Pen.	0.14 ± 0.05	0.26 ± 0.09	*-0.12 ± 0.09*	0.24 ± 0.08	*-0.10 ± 0.07*
Summer	Cont.	**0.09 ± 0.10**	0.07 ± 0.06	*0.02 ± 0.07*	**0.05 ± 0.04**	***0.04 ± 0.08***
	East	**0.08 ± 0.11**	0.06 ± 0.07	*0.02 ± 0.07*	**0.04 ± 0.05**	***0.04 ± 0.09***
	West	**0.11 ± 0.09**	0.07 ± 0.06	*0.04 ± 0.07*	**0.04 ± 0.04**	***0.07 ± 0.07***
	Pen.	0.07 ± 0.05	0.20 ± 0.05	*-0.13 ± 0.05*	0.20 ± 0.04	*-0.12 ± 0.04*
Fall	Cont.	0.05 ± 0.08	-0.02 ± 0.09	*0.07 ± 0.06*	-0.02 ± 0.08	*0.07 ± 0.04*
	East	0.02 ± 0.09	-0.06 ± 0.11	*0.08 ± 0.06*	-0.05 ± 0.09	*0.07 ± 0.05*
	West	0.16 ± 0.07	0.05 ± 0.10	*0.10 ± 0.06*	0.02 ± 0.08	*0.14 ± 0.05*
	Pen.	0.12 ± 0.04	0.41 ± 0.09	*-0.30 ± 0.09*	0.32 ± 0.06	*-0.21 ± 0.05*

TABLE 3. Summary of full reconstruction performance and verification reconstruction statistics For RLS, E-W, and S09 using the 86 READER stations located within 120km of an AVHRR grid cell. 1982 – 2006 values are compared to the raw AVHRR data.

Full Reconstructions

	1957 – 2006			1957 – 1981			1982 – 2006		
	μ_{rms}	r	R^2	μ_{rms}	r	R^2	μ_{rms}	r	R^2
RLS	0.61	0.97	0.93	0.55	0.98	0.94	0.63	0.96	0.92
E-W	1.29	0.85	0.69	1.25	0.85	0.70	1.30	0.85	0.68
S09	1.92	0.56	0.31	1.89	0.61	0.32	1.91	0.55	0.30
AVHRR	-	-	-	-	-	-	1.65	0.70	0.48[a]

Verification Reconstructions

	1957 – 2006			1957 – 1981			1982 - 2006		
	μ_{rms}	r	CE	μ_{rms}	r	CE	μ_{rms}	r	CE
RLS	1.35	0.81	0.66	1.51	0.78	0.55	1.24	0.84	0.70
E-W	1.59	0.74	0.53	1.77	0.64	0.41	1.48	0.78	0.58
S09	1.94	0.54	0.29	1.96	0.54	0.27	1.91	0.55	0.30
AVHRR	-	-	-	-	-	-	1.65	0.70	0.48[a]

[a] R^2 and CE are equivalent calculations when both the verification period and calibration period means are zero, or if no calibration period exists (as is the case with the AVHRR / ground data comparison and the verification reconstructions).

TABLE 4. Effects of the methodological deficiencies in S09 on overall and regional trends (values in ^{o}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	0.04	0.06
East	0.10	0.08	0.08	0.06	0.02	0.03
West	0.20	0.18	0.14	0.09	0.06	0.10
Peninsula	0.13	0.11	0.29	0.28	0.32	0.35

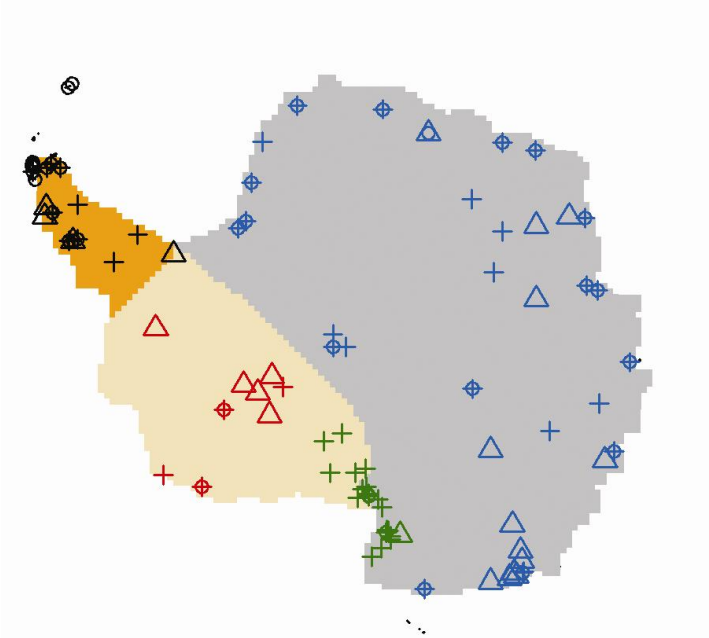

FIG. 1. Geographic regions and station color codes. The Peninsula region is shaded in orange; West Antarctica in tan; and East Antarctica in gray. Circles: S09 station locations (with Campbell, Grytviken, and Macquarie too distant to be shown). Crosses: stations used as predictors for the improved reconstructions of this study. Triangles: stations used as additional verification targets (not used as predictors) for the improved reconstructions of this study. The colors of the station symbols correspond to the colors used in Figs. 2 (S09) and 8 (this study).

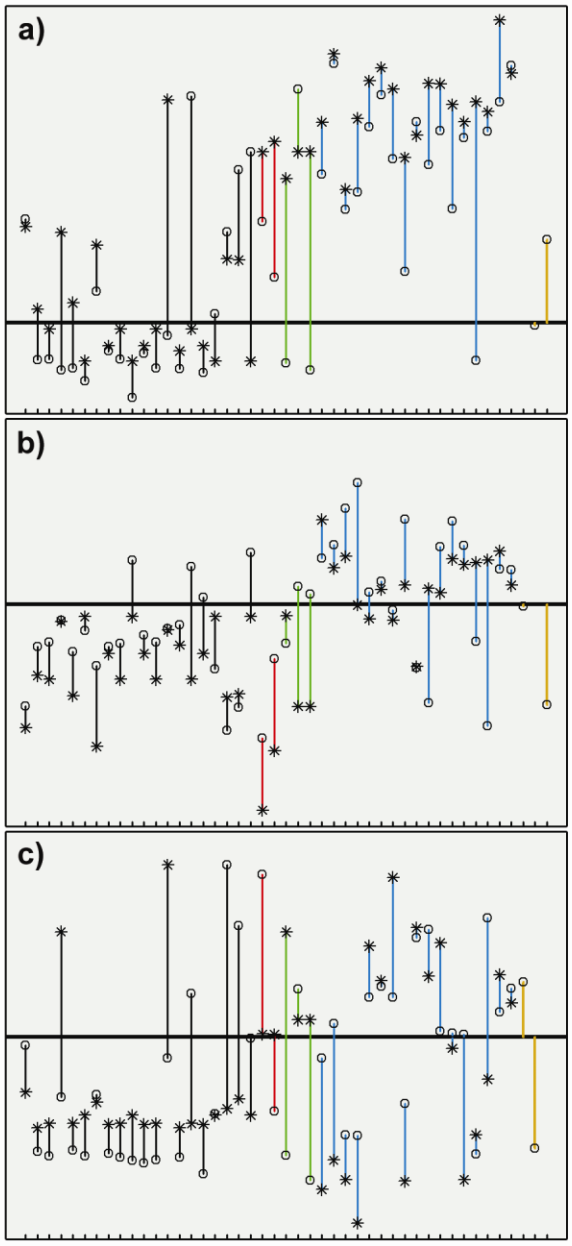

FIG. 2. Spatial structure used to estimate the satellite PCs in the 1957-1981 period by S09 vs. AVHRR spatial structure. Top: PC #1. Middle: PC #2. Bottom: PC #3. Circles represent TTLS weights for the PCs; stars represent the normalized AVHRR spatial eigenvector weights for the PCs. Geographic location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

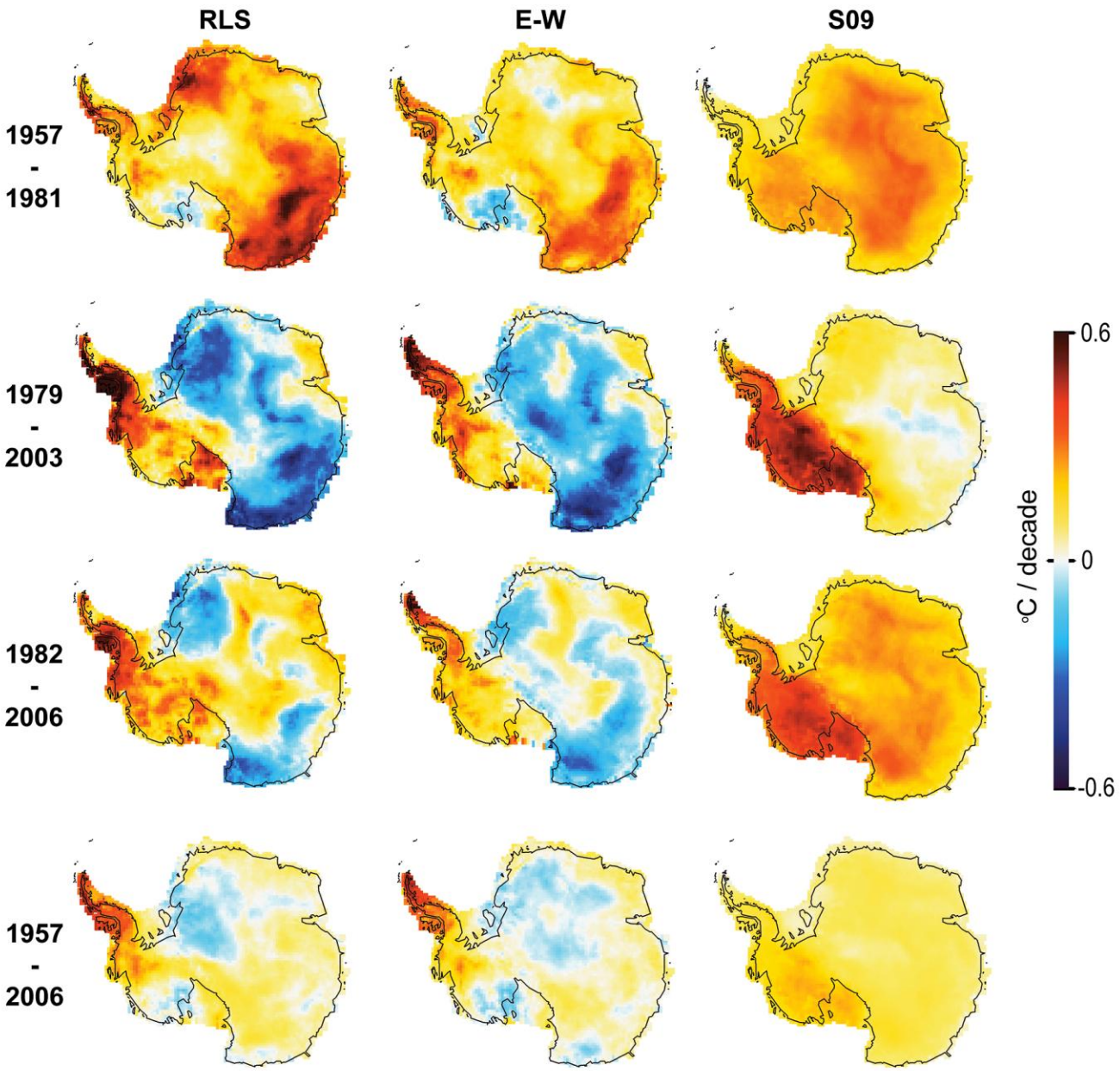

FIG. 3. Comparison of spatial patterns of change for RLS, E-W, and S09 reconstructions. Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.

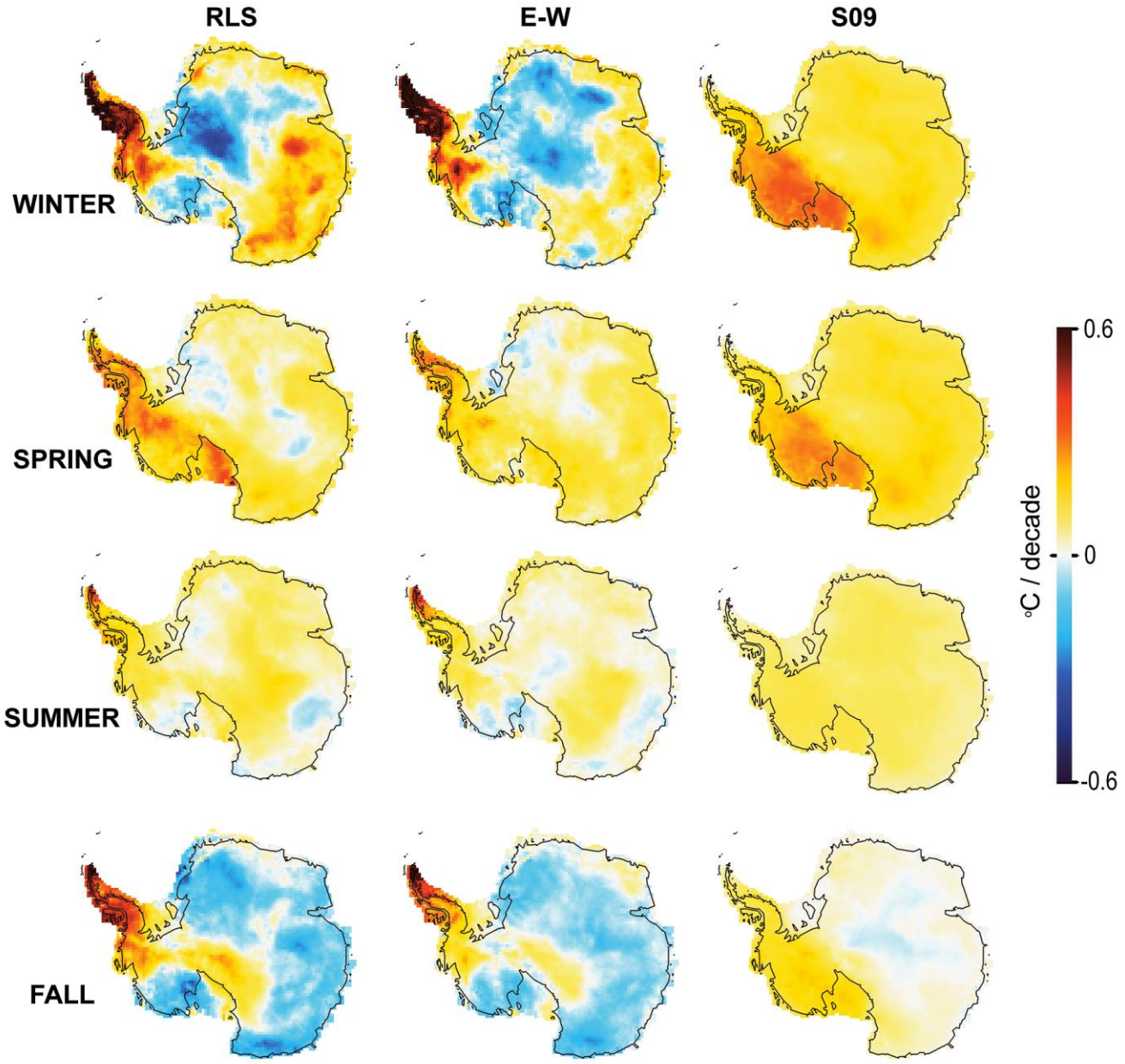

FIG. 4. Comparison of seasonal patterns of change for RLS, E-W, and S09 reconstructions for 1957 - 2006. Leftmost column is the RLS reconstruction; middle E-W; rightmost S09.

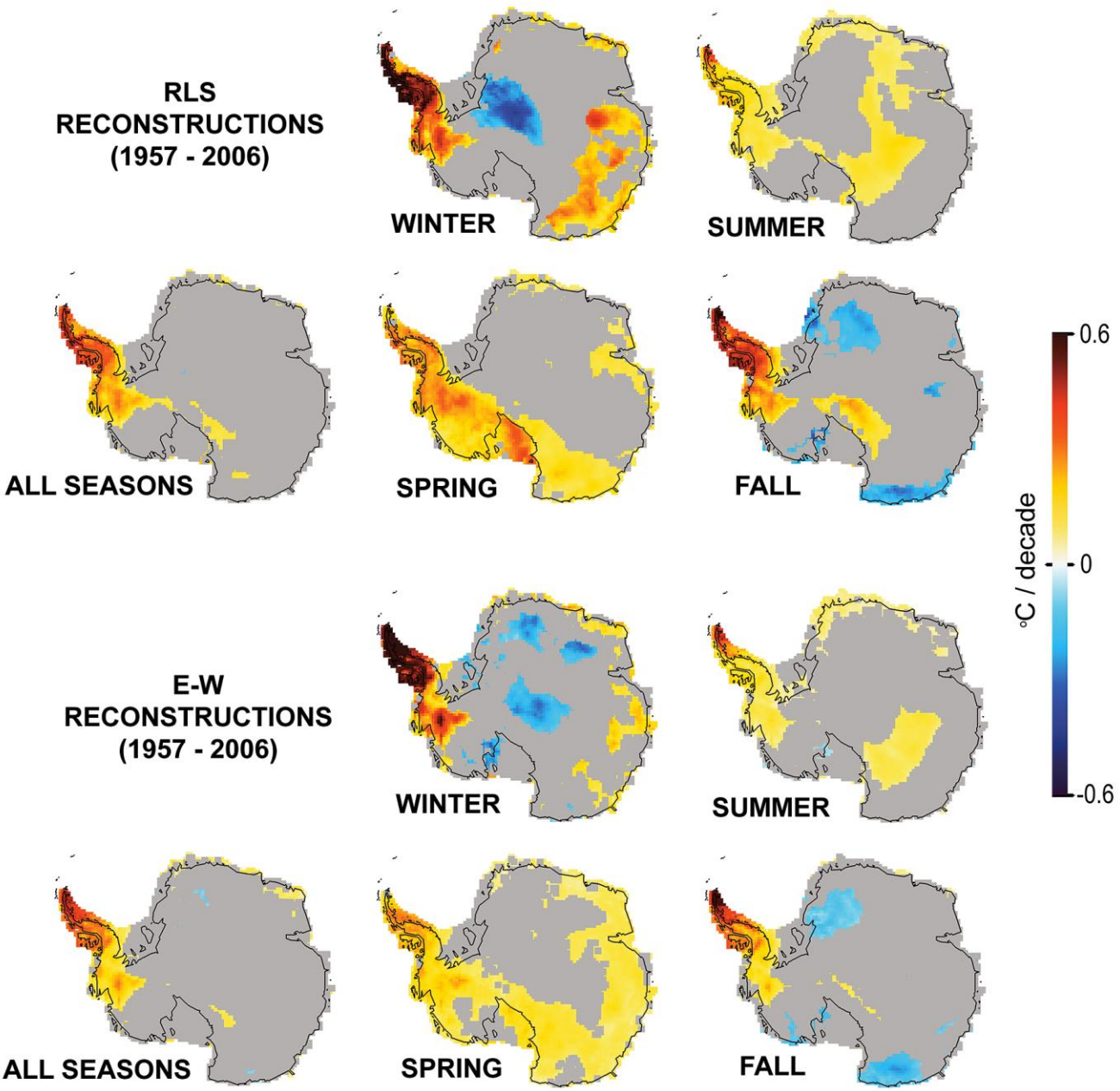

FIG. 5. Spatial patterns of change (1957 – 2006), with regions of statistically insignificant trends overlaid in gray ($p > 0.05$, two-tailed).

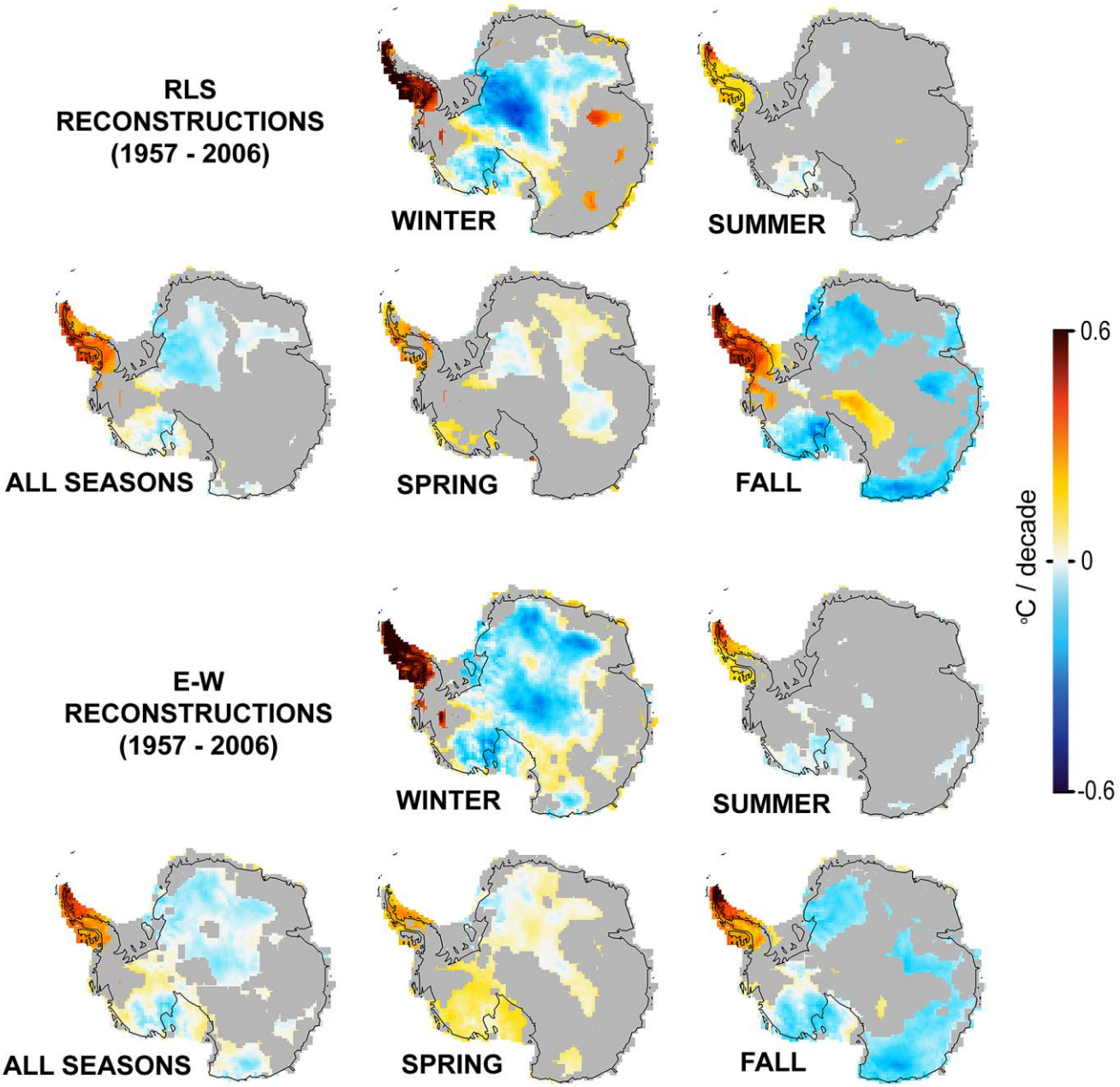

FIG. 6. Spatial patterns of change (1957 – 2006), with regions of statistically insignificant differences between S09 and RLS / E-W reconstructions overlaid in gray ($p > 0.05$, two-tailed).

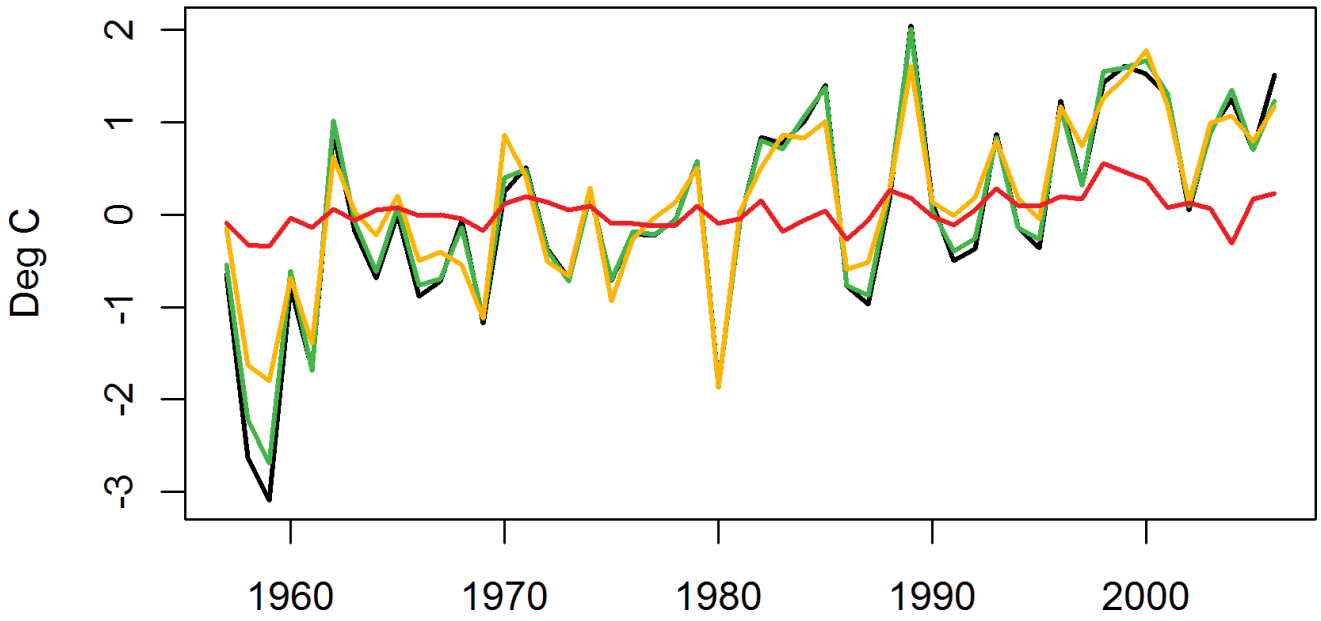

FIG. 7. Comparison of annual average anomalies in ^{o}C for the 7 most complete Peninsula stations (Arturo Prat, Bellingshausen, Esperanza, Faraday, Marambio, O'Higgins, and Rothera) versus reconstruction anomalies at the corresponding grid cells. Black: ground stations. Green: RLS. Gold: E-W. Red: S09.

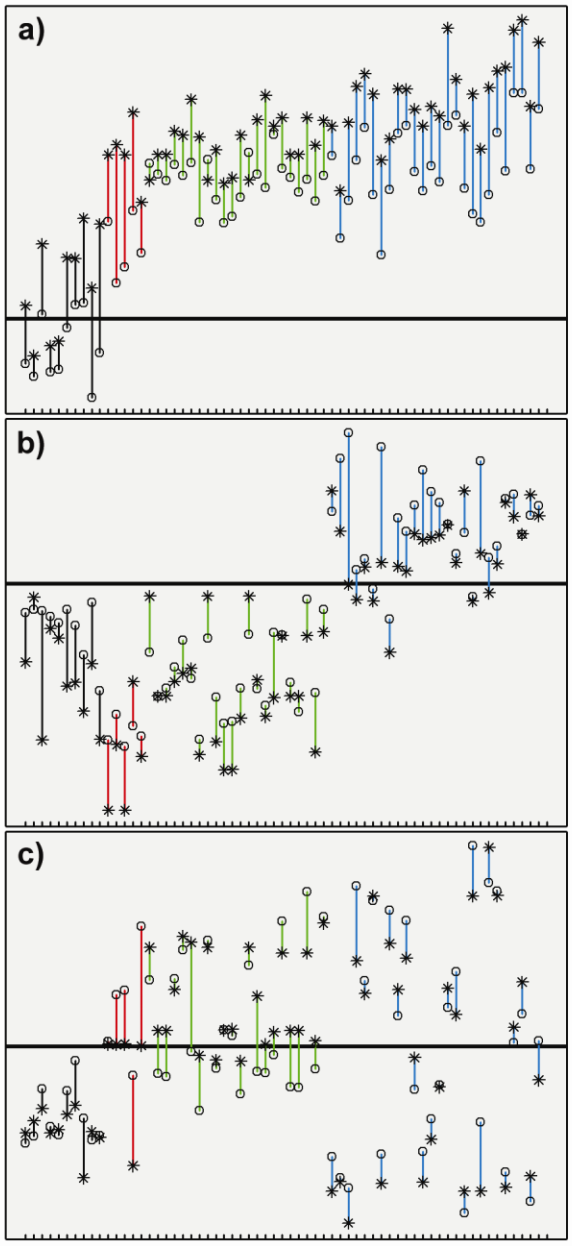

FIG. 8. Spatial structure used to estimate the satellite PCs when constrained by AVHRR eigenvector weighting vs. AVHRR spatial structure. Top: PC #1. Middle: PC #2. Bottom: PC #3. Colors, symbols, and normalization are identical to Fig. 2.