Improved methods for PCA-based reconstructions: case study using the Steig *et al.* (2009) Antarctic temperature reconstruction

Supporting Information

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

S1. Geographic definitions

For our study, we define the following regions in Antarctica:

- Peninsula: The portion of West Antarctica that lies north of a line between Cape Adams and the mainland south of the Eklund Islands.
- West Antarctica: The portion of the continent to the west of the Transantarctic Mountains, including the Ross Sea and excluding the Peninsula.
- East Antarctica: The remainder of the continent.

All trend calculations are performed using the regional masks in the left panel of Figure S1. The conclusions in the main text do not depend on the regional definitions. The results of the replication effort are depicted in Figure S2.

S2. Sources of uncertainty in AVHRR trends

The AVHRR instrument is a multi-channel sensor. Surface skin temperatures are developed using channel 4 and 5 information. Since clouds are opaque at channel 4 and 5 wavelengths, channel 3 is provided for assistance in cloud detection. Additionally, the AVHRR/3 instrument carried aboard NOAA-15 and later has a split channel 3 to enhance cloud detection capabilities. Whether this enhanced capability was used for the final two satellites in the S09 set has not been published, and whether this would materially affect the homogeneity of the S09 set – also used for this study – remains an open question.

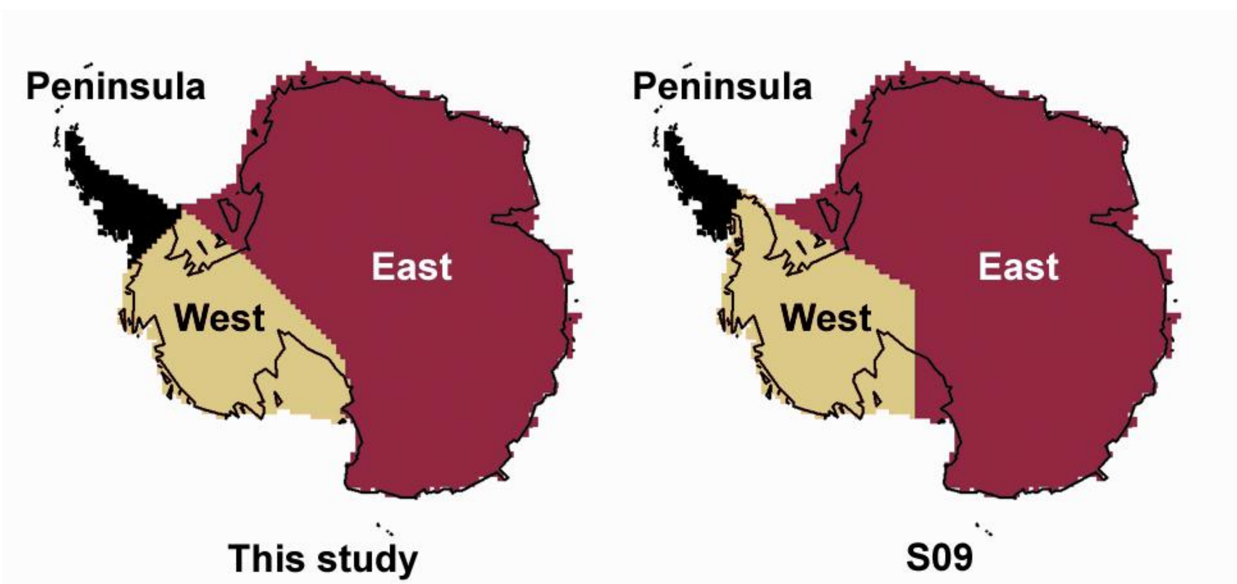

FIG. S1. Geographic masks for this study (left) vs. S09 (right).

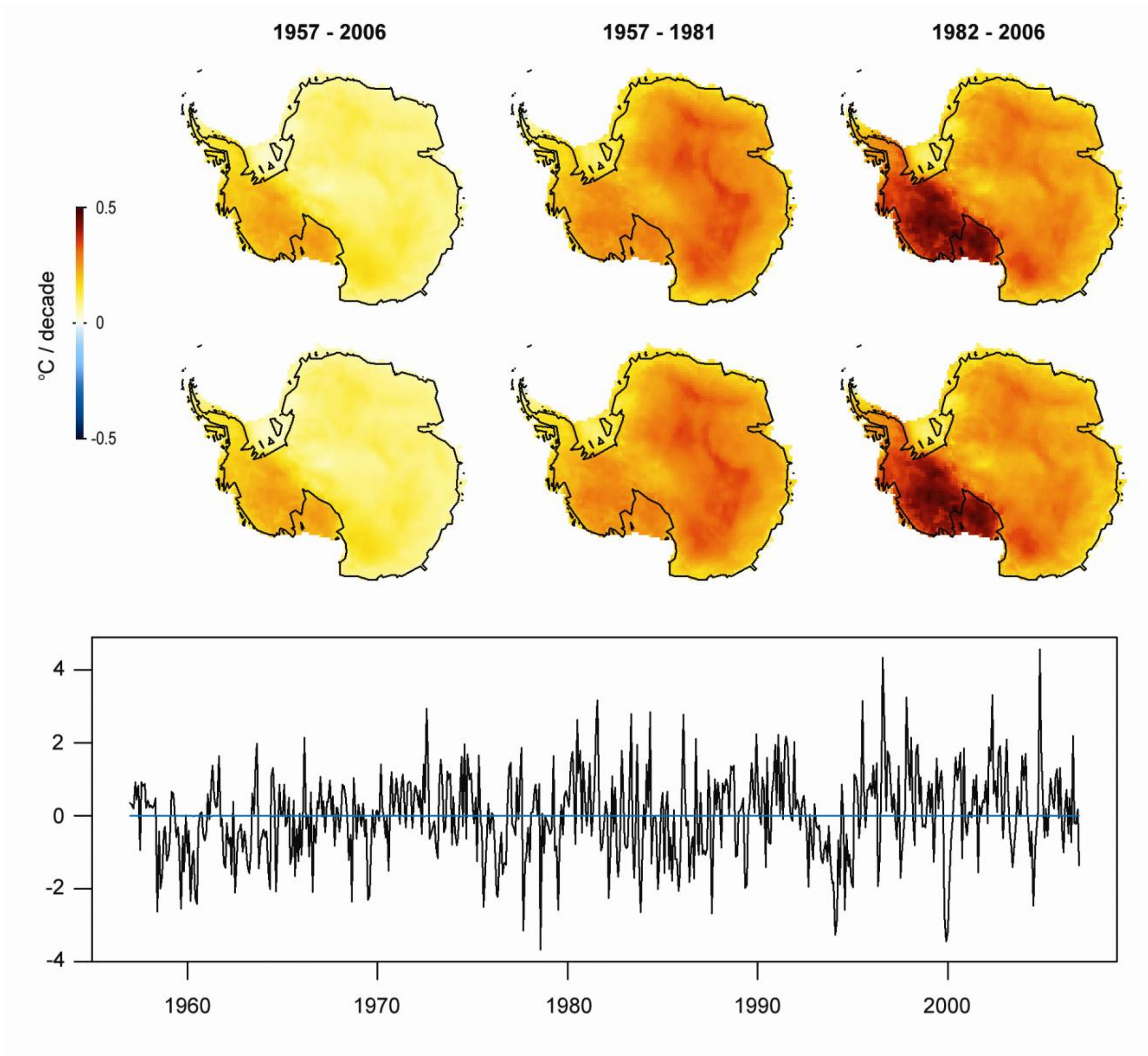

FIG. S2. Results of S09 replication. Top three panels: S09. Middle 3 panels: Replication effort. Bottom panel: Monthly means for the S09 reconstruction, with the blue line indicating the difference between the replication and S09.

a. Cloud masking

S09 utilize an AVHRR data set processed via a procedure similar to the cloud masking method described in Comiso (2000), but impose an additional constraint that daily values differing from local climatology by more than 10°C are assumed to be cloud contaminated and were removed. In general, the procedure relies on detecting clouds using channel differencing, with slightly different methods for night and day. Only clear-sky values are retained for use in computing monthly averages. Comiso reported a mean bias due to using only clear-sky values of 0.3 °C during summer months and 0.5 °C during winter months, and found an rms error of 3°C when comparing clear-sky AVHRR to ground measurements. The error for any particular measurement was dependent on time of observation, surface type (i.e., water vs. ice), and time of year. The S09 data set, when compared to ground observations from the READER database, produces similar values, with mean biases of -0.6, +0.5, +0.9, and +0.5 ^{o}C for fall, winter, spring and summer, respectively, and an rms error of 4.4 ^{o}C for all data.

b. Instrument calibration error

While an important source of error, cloud masking is not the only source. Calibration uncertainties can contribute significant error. The thermal channels for the AVHRR instruments are calibrated in-flight by observing an internal blackbody and deep space. Solar contamination of the internal calibration target (ICT) and subsequent thermal inertia effects on the platinum resistance thermocouples (PRTs) that measure ICT temperature were investigated by Trishchenko and Li (2001). PRT thermal inertia can result in up to 0.6 K of error in channels 4 and 5 for NOAA-12 and is highly dependent on latitude. NOAA-14 and -15 demonstrate smaller errors of ~0.4 K, and show a different latitudinal dependence. Trishchenko and Li do not explicitly conclude that the latitudinal profile changes with time; however, based on the physical description proposed, such a time dependence is possible. As only one satellite used for the S09 study[1] (NOAA-14) is investigated, this contribution to measurement error for the remainder of the satellites in the S09 data set is unknown.

Further investigation by Trishchenko et al. (2002) and Trishchenko (2002) on the overall uncertainty budget of the thermal channels includes two additional satellites used by S09 (NOAA-11 and NOAA-16). They find the noise equivalent error (NEΔT) to vary from ~0.1 K at 300 K to ~0.2 K at 200 K for channels 4 and 5, with a wide spread between the satellites. Extending the previous work by Trishchenko and Li, the authors find that orbital effects on ICT temperature are not constant over the life of the satellites. The ICT temperature variability for NOAA-9 increases fairly monotonically throughout life, while the ICT for NOAA-11 shows peculiar intra-orbit variability starting in 1991 and worsening significantly in 1993. NOAA-14 shows step changes in variability in mid-1998, mid-1999, and mid-2000, with the last being the largest jump. As the ICT is used for one of two points on the calibration curve for in-flight calibration and the PRTs display significant thermal inertia, fluctuations in ICT temperatures can translate into errors in measured temperatures exceeding +/- 0.5 K. The amount of error changes from satellite to satellite.

c. Other sources of error

Several other studies indicate temperature drifts for a given satellite and discontinuities between satellites. Gleason et al. (2002), Jin and Treadon (2003), and Sobrino et al. (2008) find correlation between time-of-observation changes and temperature drifts, with this effect being largest at low latitudes. Gleason et al. (2002) additionally find a larger jump in temperature between NOAA-9 and NOAA-11 than between NOAA-7 and NOAA-9 (which also appears in our comparison of AVHRR and surface temperatures; Figure S3), with each subsequent satellite showing warmer temperatures than the previous. More generally, Jiménez-Muñoz and Sobrino (2006) place a minimum bound on errors

[1] NOAA-12 was also studied by Trishchenko and Li (2001). Though the scan motor aboard NOAA-11 failed in September of 1994 resulting in brief use of NOAA-12, this use was too short for NOAA-12 to have a significant impact on the results.

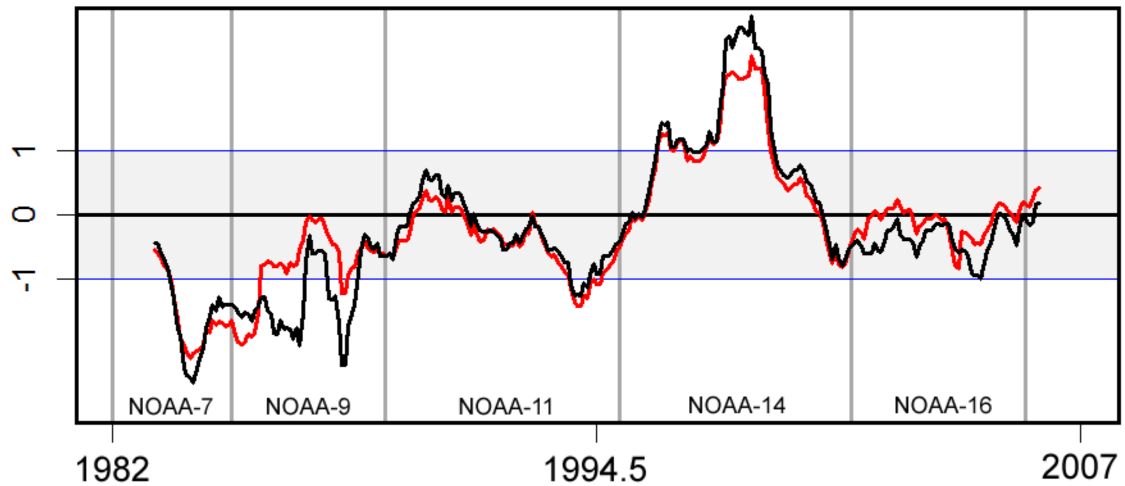

FIG. S3. Estimates of difference in means between ground station data and corresponding AVHRR anomalies, normalized to the 95% confidence intervals. Black: paired Wilcoxon test; red: paired t-test.

associated with remote thermal sensing of approximately 0.8 K when sufficient *in situ* data is not available to monitor surface emissivity and atmospheric aerosols, as is the case in Antarctica.

d. Trends in AVHRR data vs. ground data

Figure S3 presents a comparison of ground measurements against the cloud masked AVHRR data set provided by S09. All available on-grid READER sites are used for the comparison. Anomalies are calculated for both the corresponding AVHRR grid locations and ground data using only months where ground data exists. Statistical significance is determined by a paired Wilcoxon test on a running 24-month sample, with the resulting median estimate scaled to the 95% confidence intervals (two-tailed). Due to the possibility that the distribution of the residuals can change between satellites and there is no reason to believe that the residuals are Gaussian, a non-parametric Wilcoxon test is used as the primary metric (the results of a t-test are also shown). Both find statistically significant differences between the AVHRR data and the ground stations, which results in a difference in trend of approximately 0.08 +/- 0.14°C decade^{-1} over the period of 1982 – 2006. During the period of 1982 – 2001, however, where each satellite showed increasingly warmer temperatures, the difference in trend is a statistically significant 0.19 +/- 0.16.

In particular, we find that NOAA-14 manifests a statistically significant warm offset, with the larger-than-anticipated NEΔT error noted by Trischenko suggesting a possible explanation. Both NOAA-7 and NOAA-9 demonstrate statistically significant cool offsets. NOAA-9 is the only satellite analyzed by Trischenko that shows a substantial deviation in the NEΔT error between channels 4 and 5. We also note that from NOAA-7 to NOAA-14, each successive satellite shows warmer temperatures relative to ground data than the previous.

From the results of the Wilcoxon test, we determine the offset amount for each satellite that minimizes the difference in trend. The offsets range from 0.2 ^{o}C for NOAA-9 to -0.3 ^{o}C for NOAA-14. However, as the grid is sparsely sampled by ground stations (87 locations out of 5509) and the locations are typically in areas of the most cloud cover (Comiso 2000), it is not possible to separate errors due to cloud masking, satellite calibration, or other measurement errors (some of which are dependent on latitude). This means it is not known if the calculated offsets are valid throughout the grid. Additionally, the station data is unlikely to be free from error. For these reasons, these offsets are not used for any reconstruction presented in the main text or in the Supporting information. A discussion on possible sources of error in the ground data is contained in Section S4.

S3. Significant AVHRR eigenvectors

A close examination of the first three spatial eigenvectors (Figure S4) – which were the only eigenvectors retained by S09 – reveals insufficient geographic resolution to isolate West Antarctica from the Peninsula for analysis. Neither eigenvector #1 nor #3 has significant weights in West Antarctica. Eigenvector #1 is dominated by East Antarctica, while eigenvector #3 contrasts the Weddel Sea with the opposite side of East Antarctica.

The eigenvector describing West Antarctica is primarily eigenvector #2. It appears prominently in all of the plots of reconstructed temperatures for S09. For this eigenvector, there are 23 predictor stations in the high-weight (blue) region, where S09 would expect a good correlation between the station temperature data and PC #2. Of these, only 5 are located in West Antarctica (Byrd, Russkaya, Scott Base, McMurdo, and Mario Zuchelli/Terra Nova Bay). 18 of the stations are located in the Peninsula. Due to the larger number of data points in the Peninsula, the regression results necessarily will be determined primarily by Peninsula stations. When the reconstruction estimates are recovered by reconstituting PC #2 with eigenvector #2, the Peninsula trend is therefore transferred throughout West Antarctica.

Because no eigenvectors are included that display high weights in either the Peninsula or West Antarctica independently, it is *not mathematically possible* for the S09 results to show any differentiation between Peninsula trends and West Antarctica. This also applies to seasonal results. Including only eigen-vectors 1 – 3 firmly couples West Antarctica to the Peninsula (eigenvector #2) and the Weddel region/South Pole to the Peninsula (eigenvector #3). It is not surprising, then, that S09 find that West Antarctica and the Peninsula share the same seasonal pattern, nor is it surprising that S09 find the greatest *warming* at the South Pole and land adjacent to the Weddel Sea in winter despite the fact that the ground data shows the greatest *cooling* during this season. S09 do not have sufficient spatial degrees of freedom to obtain any other result.

S4. Ground data

The ground data chosen by S09 for the main T_{IR} reconstruction consists of monthly averages for 42 of the 46 manned surface stations in the British Antarctic Survey READER database (Turner et al. 2003). Coverage starts as early as 1903 for Southern Ocean stations (Orcadas); however, coverage of the continent itself generally begins in 1956 or 1957. Two stations (Almirante Brown and Primavera) have no archived temperature data. S09 additionally exclude two Southern Ocean stations (Gough and Marion), yet make the questionable choice of including two similar

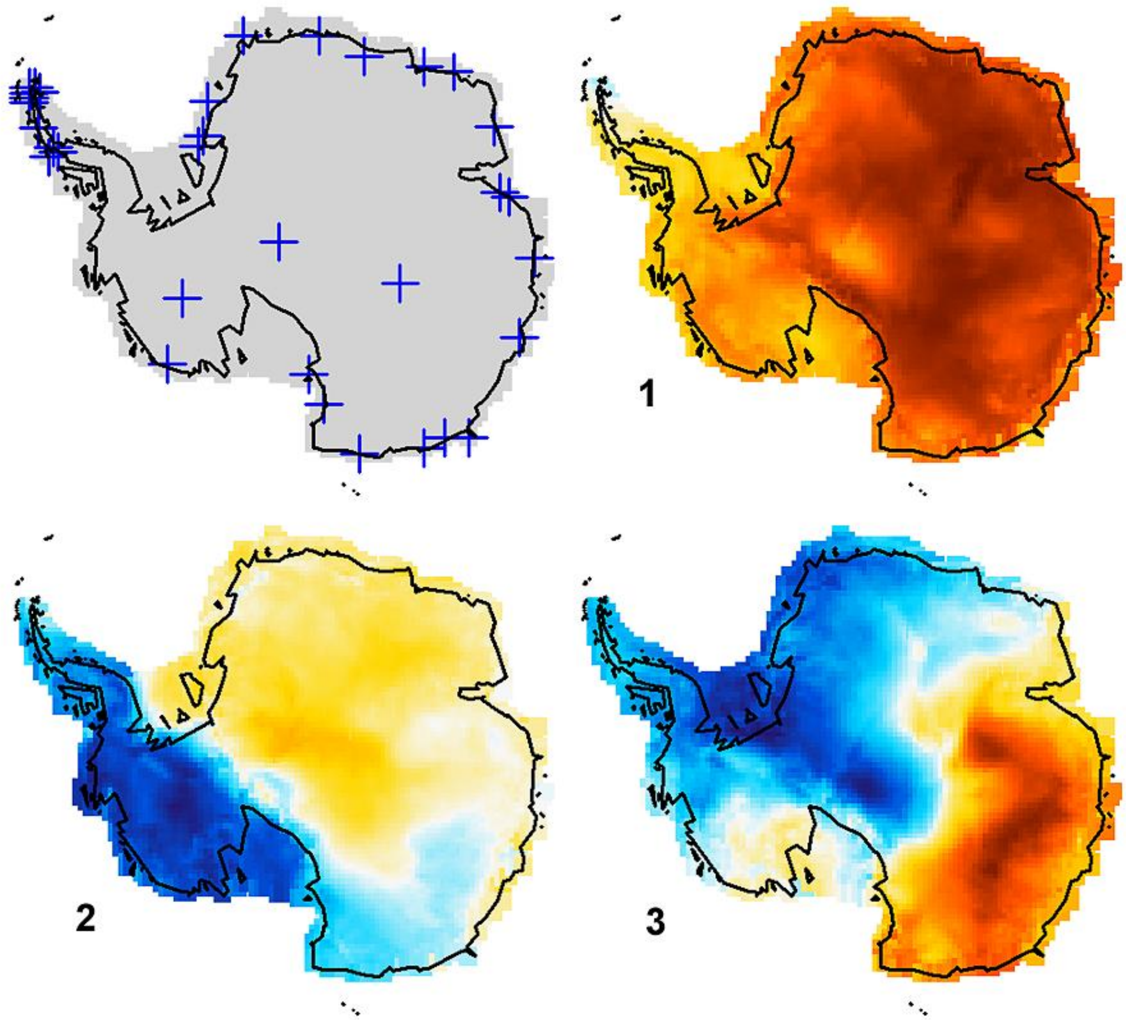

FIG. S4. Top left: S09 station locations for those stations within 150km of an AVHRR grid cell. Other panels: AVHRR spatial eigenvectors, with eigenvector order listed.

stations (Orcadas and Signy) that are both off the AVHRR grid by more than 700 km. Of the 42 stations used, 34 have data outside of the satellite coverage period. 30 sites are at least 40% complete (240 values out of 600 possible).

In a separate reconstruction that did not utilize satellite information, S09 used Automatic Weather Station (AWS) data from the READER database for all 65 AWS sites. AWS data is more sparse, with only 28 sites being more than 40% complete (120 values out of 300 possible) during the satellite period of 1982-2006. The earliest coverage for AWS begins in 1980.

Figure S5 shows the number of stations, by month, with sufficient data to compute a monthly average. Daily data less than 90% complete are flagged by BAS (http://www.antarctica.ac.uk/met/READER/data.html) and are neither used in this paper nor included in Fig. S5. During the pre-satellite era, approximately 20 individual values per month are available. This

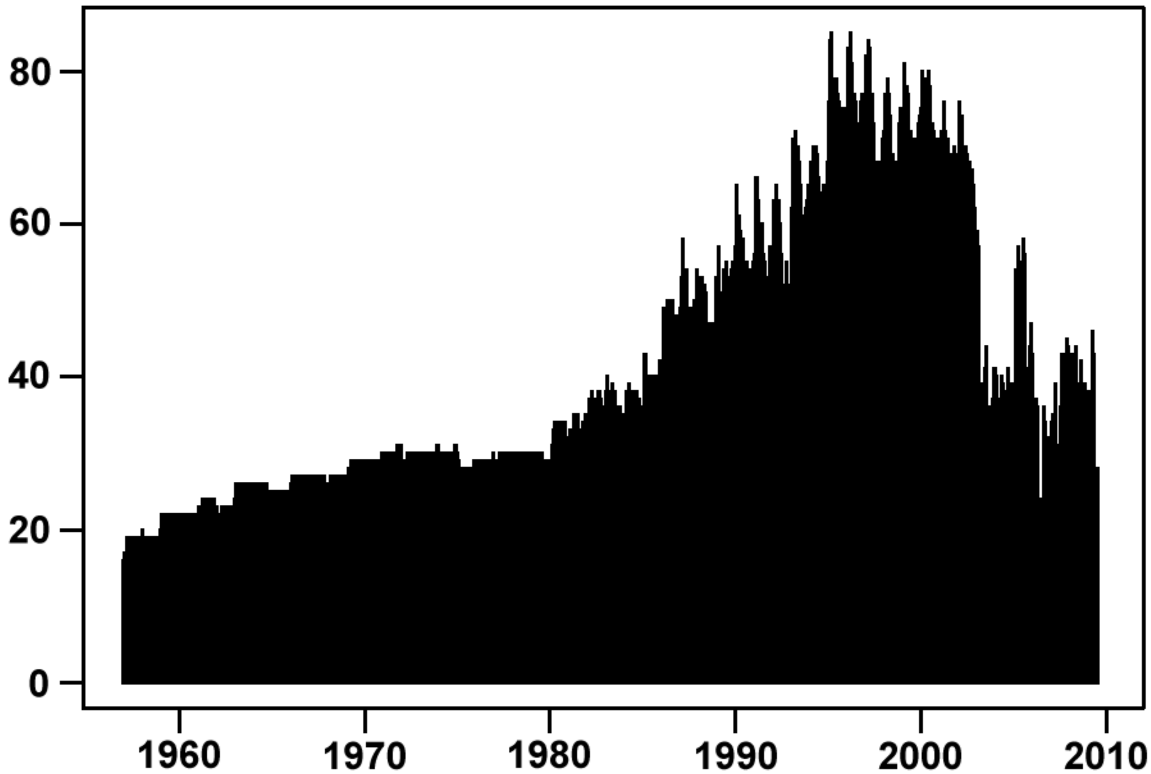

FIG. S5. Number of available monthly observations for all stations in the READER database.

increases to ~80 by 1995 and drops precipitously in 2002 to approximately 35 values per month.

Like the satellite data, the ground data is subject to error. Turner et al. (2003) describe the potential sources of error in detail, along with the quality control methods used to create an end product usable for climatology. Many of the sources of error – such as undocumented station moves, instrumentation changes, and differing methods to calculate daily means – are shared with other historical surface air temperature data sets, like GHCN (Peterson et al. 1998), though the paucity of stations and extreme climate in Antarctica exacerbate these issues. Clerical errors are also a concern:

- http://www.antarctica.ac.uk/met/READER/aws_corrections.html
- http://www.antarctica.ac.uk/met/READER/surface_corrections.html

While the ground data is certainly not without error, the potential impact is different than errors in the satellite data. The impact of ground errors on reconstruction trends is minimized by the fact that the observations come from multiple instruments. It is implausible that station moves, instrumentation changes, burial in snow, and/or biases due to missing observations would result in offsets in the same direction simultaneously at multiple stations and that these would occur coincident with satellite transitions. Despite a significant number of clerical corrections to the READER database since the publication of S09, reconstruction trends using the corrected and uncorrected sets are nearly identical.

S5. Ground station cross-validation screening

As described in Section 6 of the main text, a screening test is performed to reduce the total number of permutations to a reasonable amount. For clarity, we provide a flowchart for the cross validation tests in Figure S7 and show the results of the experiments in Figure S8. Of the 109 stations with temperature data in the READER database, the following stations are not used for any station set:

- Casey New Airstrip (insufficient data for anomaly calculations)
- Dome F (insufficient data for anomaly calculations)
- Terra Nova Bay (duplicate of Mario Zuchelli)
- High Priestley Glacier (unexplained 2°C drop in temperature in 1992 following a data collection frequency change; Shepherd 1999)

Two additional stations display suspect observations, and permutations were run including and excluding these stations:

- Adelaide (unexplained 5°C difference in temperature pre- and post-1972)
- Deception (unexplained drop in pre-1962 temperatures)

The station set we use for the full-grid reconstructions is set Grid.1C. Tables S1 and S2 provide a descriptions and verification CEs for various settings of k_{gnd}. These statistics do *not* include the use of satellite data, and therefore differ from the full-grid reconstruction statistics (presented later). The purpose of calculating verification statistics at this stage of the process is to identify the station set and preliminary values of k_{gnd} that will be used as a starting point for the full-grid reconstructions. The reason k_{gnd} is preliminary is due to two primary reasons:

1. Because RegEM cannot predict values when a series is entirely withheld, the verification testing must be performed as a set of early/late withholding experiments. Especially in cases where the station record length is short, truncating half of the record greatly increases the effects of sampling error on the correlation matrix and subsequent eigendecomposition. In the later gridded reconstruction verification tests, the AVHRR spatial eigenvectors naturally provide interpolation between points, allowing stations to be completely omitted without greatly affecting their prediction (Tables S5 – S6 at the end of this document).

2. The geographic distribution of stations is quite uneven. The Peninsula and eastern edge of the Ross Ice Shelf contain a disproportionate number of stations relative to the land area they represent, while West Antarctica is significantly under-represented. If one examines the amount of *actual data* available by region, the deficit in West Antarctic coverage is even more striking (Table S3). Since the regression coefficients determined by RegEM in the ground-station only infilling are entirely unconstrained by the AVHRR spatial eigenvectors, this results in a different prediction – and different verification statistics – as compared to the gridded reconstructions. Figure S6 depicts this difference in spatial structure.

TABLE S1. Description of sets used for cross validation experiments. Grid.1C and No.Ocean.1C are the two sets with the highest (and nearly identical) performance in the cross validation experiments.

Set Name	Adelaide (On-Grid)	Deception (Off-Grid)	Manned On-Grid, 96+ Months	AWS On-Grid, 96+ Months*	Manned On-Grid, 12 – 96 Months	AWS On-Grid, 12 – 96 Months	Manned, Off-Grid	AWS, Off-Grid	Southern Ocean	Total Stations
Full	X	X	X	X	X	X	X	X	X	105
S09	X	X	X		X				(2)	42
No.Ocean.1A	X	X	X	X	X	X	X	X		97
No.Ocean.1B			X	X	X	X	X	X		95
No.Ocean.1C			X	X			X	X		71
No.Ocean.2A	X	X	X		X		X			37
No.Ocean.2B			X		X		X			35
No.Ocean.2C			X				X			34
Grid.1A	X		X	X	X	X				87
Grid.1B			X	X	X	X				86
Grid.1C			X	X						63
Grid.2A	X		X		X					30
Grid.2B			X		X					29
Grid.2C			X							28

* Also includes Erin, with 62 months of data, to provide an additional verification target for West Antarctica.

TABLE S2. Grid.1C station composition and selected ground-only cross validation experiment results. Minimum CE values obtained from early and late experiments were recorded. Pointwise mean CE results for 5% random withholding are also shown. This provides an indication of the deterioration of the quality of the regression due to the amount of data withheld in early/late experiments.

Station Metadata				Minimum CE Early/Late Cross Validation Experiments (SATELLITE INFORMATION EXCLUDED)							
				Correlation k_{gnd}				Covariance k_{gnd}			
Name	Lat	Lon	Record Length	4	5	6	7	4	5	6	7
Amundsen Scott	-90.0	292.1	600	-	-	-	-	-	-	-	-
Arturo Prat	-62.5	300.3	423	-	-	-	-	-	-	-	-
Belgrano I	-78.0	321.2	273	-	-	-	-	-	-	-	-
Belgrano II	-77.9	325.4	128	-	-	-	-	-	-	-	-
Byrd (manned)	-80.0	240.0	177	-	-	-	-	-	-	-	-
Casey	-66.3	110.5	575	-	-	-	-	-	-	-	-
Davis	-68.6	78.0	547	-	-	-	-	-	-	-	-
Dumont Durville	-66.7	140.0	584	-	-	-	-	-	-	-	-
Esperanza	-63.4	303.0	475	-	-	-	-	-	-	-	-
Faraday	-65.4	295.6	600	-	-	-	-	-	-	-	-
Halley	-75.5	333.6	600	-	-	-	-	-	-	-	-
Leningradskaja	-69.5	159.4	240	-	-	-	-	-	-	-	-
Marambio	-64.2	303.3	415	-	-	-	-	-	-	-	-
Mario Zuchelli	-74.7	164.1	192	-	-	-	-	-	-	-	-
Mawson	-67.6	62.9	600	-	-	-	-	-	-	-	-
McMurdo	-77.9	166.7	577	-	-	-	-	-	-	-	-
Mirny	-66.5	93.0	600	-	-	-	-	-	-	-	-
Molodeznaja	-67.7	45.9	437	-	-	-	-	-	-	-	-
Neumayer	-70.7	351.6	308	-	-	-	-	-	-	-	-
Novolazarevskaya	-70.8	11.8	549	-	-	-	-	-	-	-	-
O'Higgins	-63.3	302.1	492	-	-	-	-	-	-	-	-
Rothera	-67.5	291.9	356	-	-	-	-	-	-	-	-
Russkaya	-74.8	223.1	119	-	-	-	-	-	-	-	-
San Martin	-68.1	292.9	203	-	-	-	-	-	-	-	-
Scott Base	-77.9	166.7	596	-	-	-	-	-	-	-	-
Syowa	-69.0	39.6	535	-	-	-	-	-	-	-	-
Vostok	-78.5	106.9	540	-	-	-	-	-	-	-	-
Zhongshan	-69.4	76.4	167	-	-	-	-	-	-	-	-
Butler Island	-72.2	299.8	176	0.12	0.17	*-0.15*	*0.00*	0.26	0.25	*-0.61*	*-0.38*
Byrd (AWS)	-80.0	240.6	187	0.07	0.06	*-0.03*	*-0.00*	0.04	*-0.46*	*-0.45*	*-0.83*
Cape King	-73.6	166.6	201	0.64	0.66	0.66	0.72	0.50	0.50	0.43	0.63
Cape Phillips	-73.1	169.6	151	0.53	0.55	0.51	0.49	0.43	0.47	0.37	0.43
Cape Ross	-76.7	163.0	169	0.77	0.77	0.76	0.78	0.71	0.72	0.77	0.74
Clean Air	-90.0	0.0	192	0.16	0.21	*-0.10*	0.31	0.06	0.29	*-0.28*	0.37
D10	-66.7	139.8	162	0.50	0.39	0.55	0.54	0.40	0.36	0.37	0.38
Drescher	-72.9	341.0	108	0.25	0.39	*-0.07*	0.07	0.16	0.22	*-0.27*	0.19
Elaine	-83.1	174.2	151	0.60	0.57	0.53	0.38	0.61	0.52	0.17	0.39
Enigma Lake	-74.7	164.0	126	0.68	0.67	0.59	0.58	0.55	0.66	0.57	0.55
Erin	-84.9	231.2	62	0.22	0.18	0.14	0.04	0.22	0.21	0.04	*-0.19*
Ferrell	-77.9	170.8	204	0.71	0.73	0.77	0.77	0.80	0.78	0.73	0.73
GC41	-71.6	111.3	177	*-0.05*	0.02	0.05	0.03	*-0.54*	*-1.22*	*-2.88*	*-24.0*
GF08	-68.5	102.1	133	0.39	0.53	0.52	0.41	0.32	0.53	0.49	0.39
Gill	-80.0	181.4	193	0.48	0.52	0.45	0.38	0.47	0.45	0.32	0.32
Henry	-89.0	359.0	109	0.54	0.54	0.58	0.70	0.40	0.63	0.71	0.73
LGB20	-73.8	55.7	136	0.53	0.60	0.60	0.55	0.49	0.61	0.57	0.58
LGB35	-76.0	65.0	151	0.51	0.53	0.56	0.51	0.55	0.54	0.49	0.50
Larsen Ice Shelf	-66.9	299.1	129	0.30	0.26	0.29	0.26	0.35	0.33	0.45	0.30
Lettau	-82.5	185.6	149	0.54	0.51	0.47	0.45	0.47	0.50	0.37	0.19

TABLE S2. (continued)

Station Metadata				Minimum CE Early/Late Cross Validation Experiments (SATELLITE INFORMATION EXCLUDED)							
				Correlation k_{gnd}				Covariance k_{gnd}			
Name	Lat	Lon	Record Length	4	5	6	7	4	5	6	7
Linda	-78.5	168.4	112	0.73	0.74	0.66	0.62	0.74	0.73	0.73	0.70
Manuela	-74.9	163.7	222	0.55	0.57	0.64	0.69	0.49	0.43	0.37	0.63
Marble Point	-77.4	163.7	266	0.80	0.81	0.82	0.82	0.80	0.82	0.84	0.82
Marilyn	-80.0	165.1	152	0.67	0.67	0.55	0.39	0.65	0.66	0.47	0.33
Minna Bluff	-78.6	166.7	110	0.52	0.57	0.49	0.33	0.55	0.54	0.39	0.48
Mount Siple	-73.2	232.9	140	0.17	0.09	*-0.14*	-0.20	0.14	0.16	*-0.12*	*-0.17*
Nansen Ice Sheet	-74.8	163.3	163	0.58	0.38	0.54	0.61	0.45	0.04	*-0.47*	0.54
Nico	-89.0	89.7	120	0.32	0.37	0.46	0.47	0.16	0.50	0.56	0.52
Pegasus North	-77.9	166.5	115	0.75	0.72	0.76	0.75	0.72	0.72	0.77	0.70
Pegasus South	-78.0	166.6	136	0.77	0.76	0.80	0.81	0.80	0.77	0.78	0.80
Priestley Glacier	-74.3	163.2	176	0.18	0.16	0.31	0.29	0.12	0.11	0.04	0.21
Relay Station	-74.0	43.1	103	0.29	0.52	0.51	0.51	0.12	0.48	0.54	0.43
Schwerdtfeger	-79.9	170.0	201	0.54	0.57	0.51	0.49	0.55	0.55	0.45	0.45
Tourmaline Plateau	-74.1	163.4	166	0.75	0.04	0.21	0.20	0.09	0.25	0.05	0.11
Uranus Glacier	-71.4	291.1	119	0.21	0.20	0.15	0.20	0.23	0.27	0.25	0.20
			MEAN:	**0.44**	**0.46**	**0.43**	**0.43**	**0.40**	**0.40**	**0.23**	**-0.32**
			MEAN: (random withholding)	**0.55**	**0.59**	**0.60**	**0.55**	**0.55**	**0.58**	**0.50**	**0.49**

Note: The dramatic drop-off in mean CE for covariance, $k_{gnd} = 7$, is due entirely to one station, GC41. Excluding GC41 (which demonstrates a -24.0 CE) results in a mean CE of 0.38.

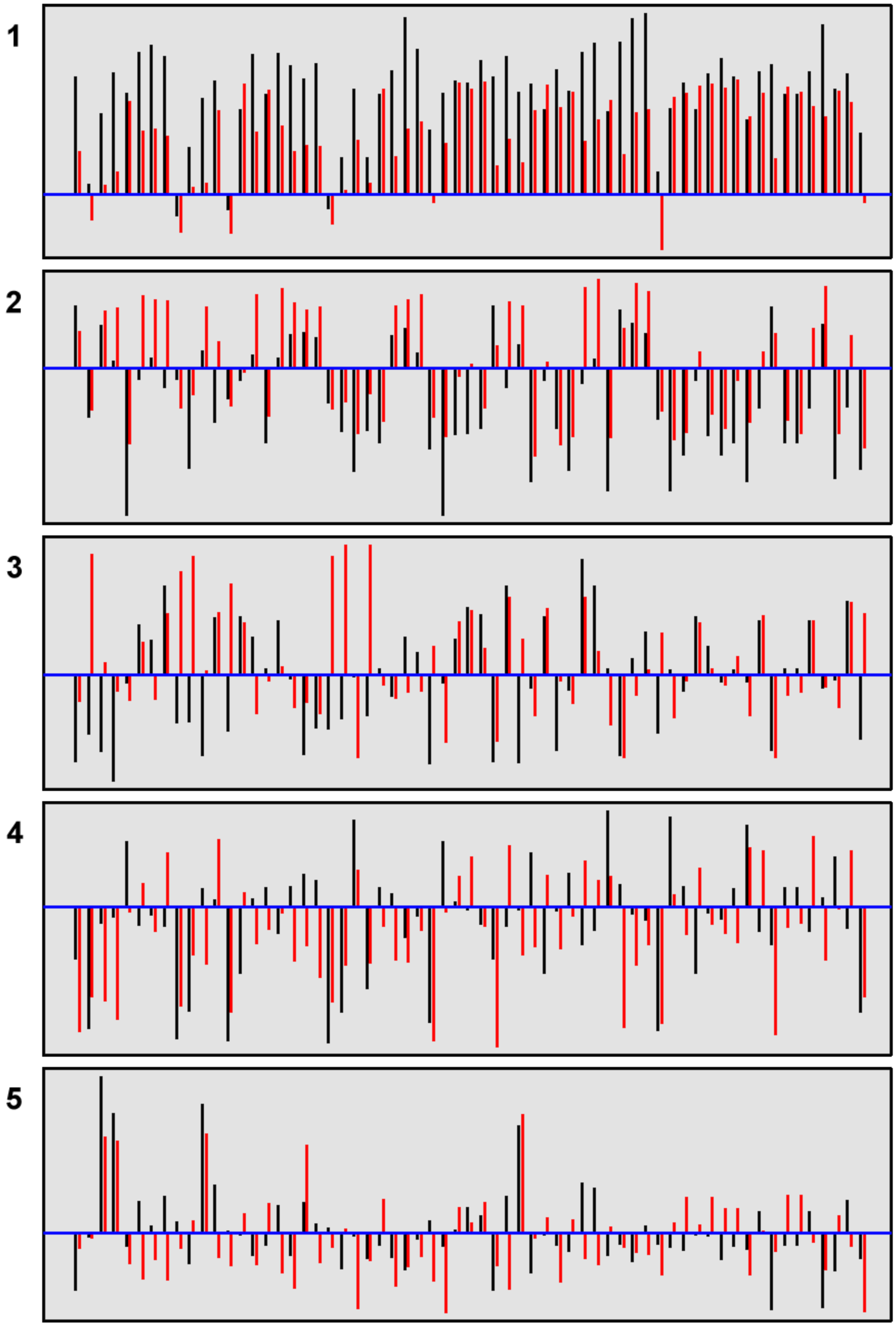

FIG. S6. Comparison of spatial weights between the AVHRR data at station locations (black) vs. spatial weights for a matrix of infilled ground stations (red). Numbers indicate eigenvector order.

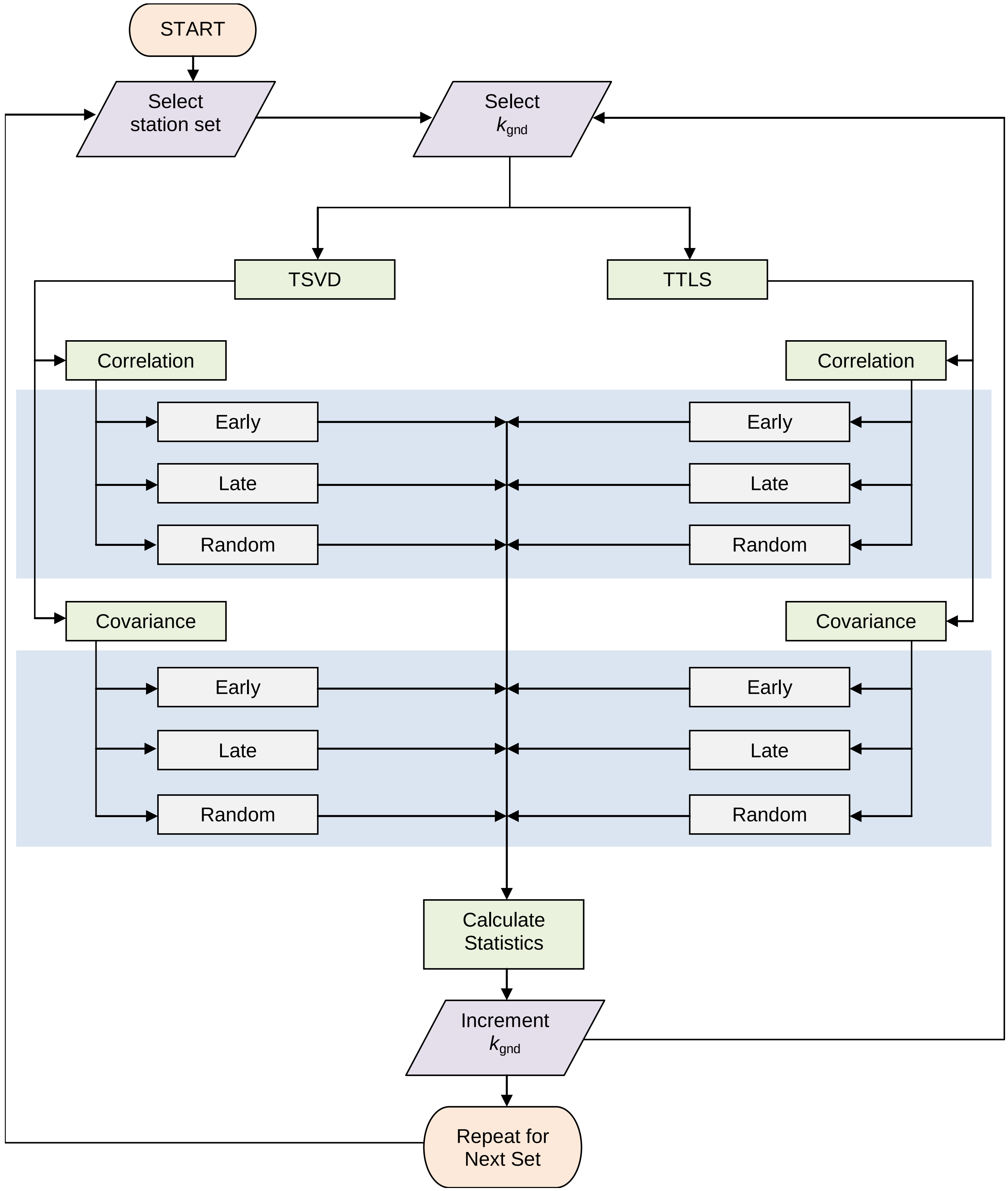

FIG. S7. Flowchart depicting ground station only cross validation experiments. The processes in the blue shaded regions are infilling processes with the method of withholding listed. For early/late withholding, the first ½ and second ½ of the verification station data are withheld, respectively. For random withholding, 5% of the data in the station set is withheld via a randomization function. This process is repeated for all 14 station sets.

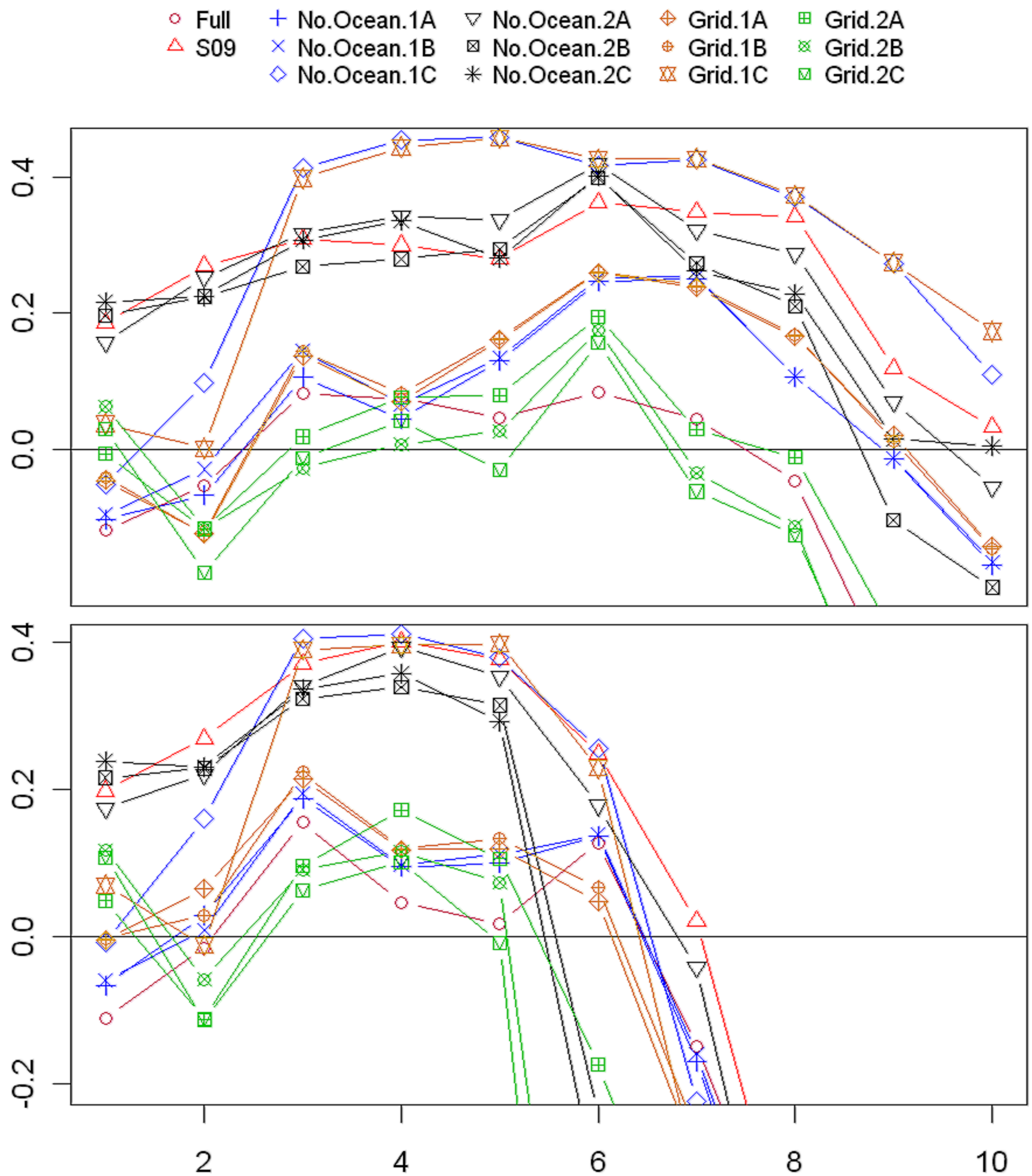

FIG. S8. Cross validation CE results for 14 station sets and early/late withholding versus truncation parameter k_{gnd}, which was varied from 1 to 10. Top panel: correlation network. Bottom panel: covariance network. Results shown are mean verification station CEs for experiments using the TSVD algorithm. Slightly lower CE numbers with identical patterns are obtained using TTLS. Station sets "No.Ocean.1C" and "Grid.1C" show superior cross validation performance, and set "Grid.1C" is the set used for full-grid reconstructions.

TABLE S3. Ground data distribution by area. Land area percentage calculated by dividing the grid cells in the corresponding mask from Fig. S1 by 5,509 (total number of grid cells). Relative contribution is calculated as the ratio of monthly means percentage to land area percentage.

Station Description	Quantity Measured	Total	Peninsula	West Antarctic Land	Ross Ice Shelf	East Antarctica
	Land Area	**100%**	**4.1%**	**17.4%**	**3.9%**	**74.5%**
On-Grid Stations 1957 – 2006	Number of Stations	87	14 (16.1%)	10 (11.5%)	23 (26.4%)	40 (46.0%)
	Number of Monthly Means	18,931	3,741 (19.8%)	1,031 (5.4%)	4,611 (24.4%)	9,548 (50.4%)
	Relative Contribution	**1.00**	**4.88**	**0.31**	**6.24**	**0.68**
On-Grid Stations 1957 – 1981	Number of Stations	28	8 (28.6%)	2 (7.1%)	3 (10.7%)	17 (60.7%)
	Number of Monthly Means	5,663	1,375 (24.3%)	219 (3.9%)	630 (11.1%)	3,439 (60.7%)
	Relative Contribution	**1.00**	**5.92**	**0.22**	**2.85**	**0.82**

S6. Infilling algorithms

Two infilling algorithms are used in this study. For the TTLS algorithm (described in the main text), in order to maintain the open-source nature of the code for this study, the Matlab code from Schneider (2001) was transliterated into the R Programming Language. Both the Matlab and R versions include options for inputting a pre-defined correlation or covariance matrix and assigning parameters for truncation, convergence tolerance, maximum number of iterations, and variance inflation. Additionally, the R version incorporates equation (6) from the main text, which is absent from the Matlab version, along with options for centering the data matrix (column or global), selecting covariance/correlation networks, and supplying weighting vectors. The R version of the TTLS algorithm was benchmarked against the Matlab version for equivalent settings and yielded equivalent results.

The second algorithm is a truncated singular value decomposition (TSVD) algorithm, similar to the DINEOF routine (Alvera-Azcárate et al. 2009; Beckers and Rixen 2003; Beckers et al. 2006; Beckers, personal communication). Like the TTLS algorithm, the input consists of the $n \times p$ matrix of station data such that individual series are arranged in columns and time is represented by the rows. Each series is centered and missing values infilled by zeros. If desired, the algorithm scales to correlation using a vector of unbiased standard deviation estimators $\tilde{\mathbf{s}}$ and can apply user-defined spatial weights.

The TSVD algorithm operates directly on the input matrix rather than the correlation/covariance matrix. Like the DINEOF routine, rather than immediately regularize using the final desired truncation parameter (which can result in wildly divergent results based on the truncation parameter chosen), initial estimates are obtained via truncated SVD with $k = 1$.

$$\left(\hat{\mathbf{X}}\right)_{k=1} = \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^{\mathrm{T}} \qquad \text{(S1)}$$

This provides a matrix of estimates $\hat{\mathbf{X}}$, which are substituted for the missing values. The expectation-maximization process continues until the rms change in estimates is less than a predefined tolerance. The truncation parameter k is then incremented by 1 and the process is repeated. This continues until the algorithm converges at the final desired value for k.

This process (rather than immediately proceeding to the final truncation value) provides vast improvement in solution stability by reducing the effect of sampling error. In the case of sparse data sets – such as the S09 ground station matrix – this sampling error can be quite large. Not only are the number of observations limited, but the initial infill of zeros contributes observational error as the infilled zeros are treated as actual data during regularization. Since higher-order eigenvalues tend to be more closely spaced than lower-order ones (as is the case for the READER data), directly proceeding to large values for k increases the chance that sampling error in the higher-order modes will materially affect the estimation. Iteratively approaching the final value provides increasingly accurate estimations of the missing values – reducing sampling error – prior to higher-order modes being included. This stabilizes the eigenvalue spectrum and prevents wild swings in the results for regressions using different values of k. For our study, this method provides significant increases in verification skill versus a standard truncated SVD method for all values of k greater than ≈3. At low values of k where sampling errors are less of a concern, it yields approximately equal results.

Figure S9 compares the results of verification tests using identical settings for TTLS and TSVD for the primary 63-station data set used in this study. In general, the TTLS and TSVD solutions are very close, though TSVD typically demonstrates slightly better verification statistics, especially at higher values of k. This result is not limited to this particular station choice. We find that the iterative TSVD slightly outperforms TTLS for most station sets and most values of k.

TSVD has some additional benefits over TTLS. One particularly attractive benefit is speed. For a given value of k and convergence tolerance, TSVD reaches a solution approximately 5 to 10 times faster. A second benefit is that TSVD provides solutions for all values of k from 1 to the full rank of the input matrix. TTLS, however, limits k to the lowest number of actual observations present at any time in the matrix. This is because a separate total least squares solution must be computed for *each* time based on the available observations for *that particular* time. If k exceeds the number of actual observations, at least one row/column set in matrix $\mathbf{V}_{1,1}^{\mathrm{T}}$ in Equation (4) will be zero, making the

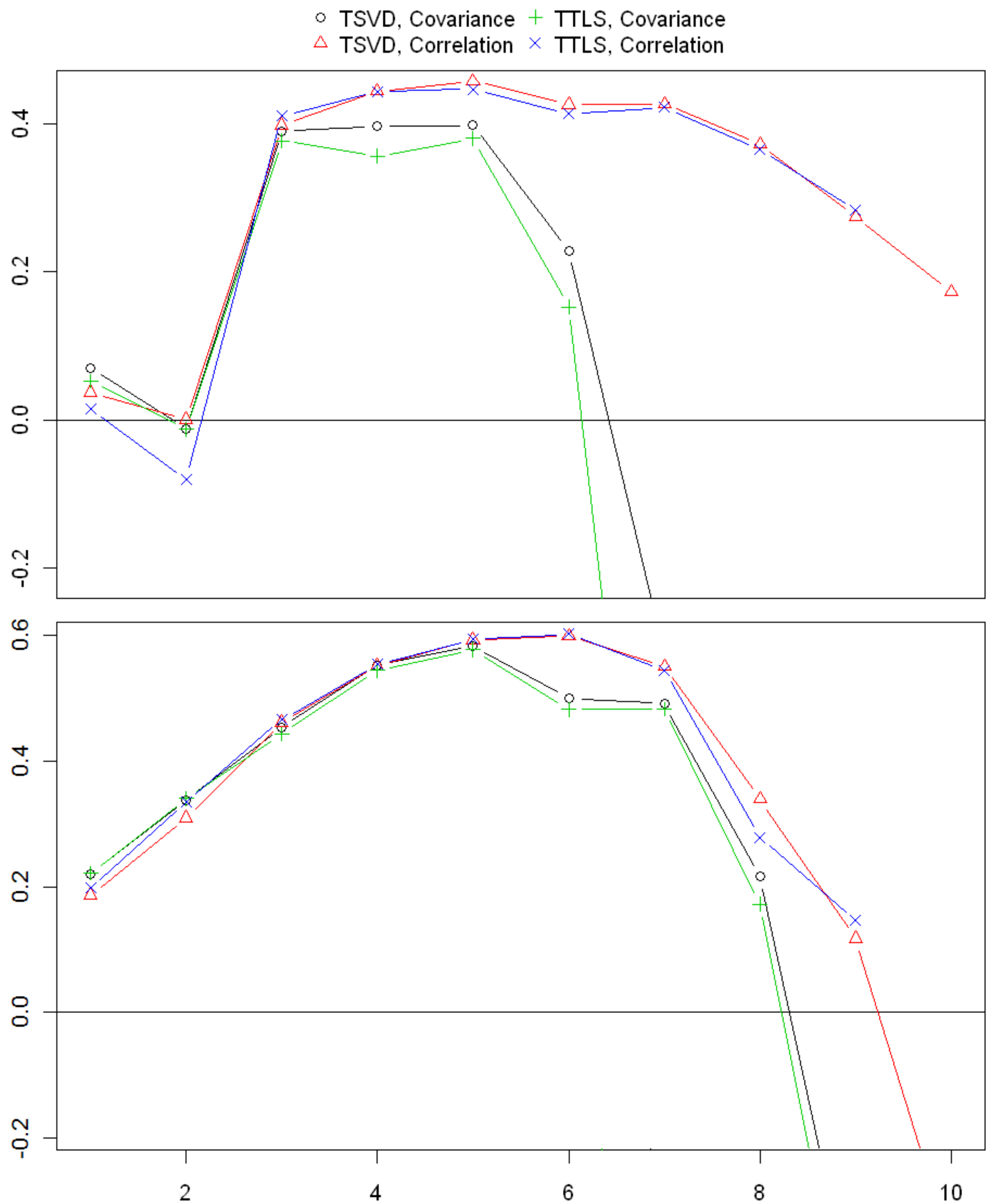

FIG. S9. Comparison of TSVD vs. TTLS algorithms for station set Grid.1C. Top panel: minimum CEs from early/late withholding of verification stations. Bottom panel: CE results for random withholding of 5% of the data.

pseudoinverse undefined. This early termination of k in TTLS can be seen in Fig. S9, where TTLS values are not available for $k = 10$ as there are only 9 actual observations for some early months in the Grid.1C set.

S7. Additional information concerning E-W reconstruct-tions

The *eigenvector-weighted* (E-W) reconstructions utilize the infilling algorithm to perform multiple linear regression of each retained satellite principal component against the ground data. The estimates for missing values for the PCs are linear combinations of the ground station data with coefficients determined from the spatial structure, either by solving Equation (4) for TTLS, or direct decomposition of the data matrix for TSVD. A key consideration is that for both algorithms, the spatial structure of the data is determined *empirically* from the augmented matrix[2].

From Fig. S6, it is clear that the spatial structure of the AVHRR data set differs from the spatial structure determined via SVD of the unaugmented, infilled ground station matrix. This is due to the spatial boundary conditions being different (Aires, Rossow, and Chedin, 2002) between the ground data and AVHRR data. In some cases, the spatial weights for a given station have the opposite orientation. This leads to the perverse situation where a predictor can be used in one orientation for estimating the principal component, but the temperature estimate for that grid cell is recovered by multiplying the principal component by a spatial weight with the opposite sign (Figs. 2 and 7 in the main text). In order to avoid this problem, spatial constraints must be applied to the ground data to ensure that the principal component is being predicted using the magnitude and orientation of the associated AVHRR spatial eigenvector.

For the E-W reconstructions, we define an $m \times n$ matrix $\underset{\sim}{\mathbf{V}}$ containing AVHRR spatial weights with columns corresponding to the eigenvector number and rows corresponding to the ground station locations. As the relative magnitudes of the weights are important, each column is normalized to have unit range. For the *n*th principal component, we can then define a vector of weights $\mathbf{w}_n$:

$$\mathbf{w}_n = \alpha \underset{\sim}{\mathbf{V}}_n \qquad \text{(S2)}$$

The scaling factor α provides a means of emphasizing or de-emphasizing the importance of the principal component relative to the set of ground stations. As $\alpha \to 0$, the ratio of principal component variance to ground station variance approaches infinity, and the SVD preferentially selects the PC. As $\alpha \to \infty$, the SVD preferentially selects ground information. This provides a means to choose whether the algorithm minimizes the PC residuals, ground station residuals, or some combination of both. Therefore, in the limit of either $\alpha \to 0$ or $\alpha \to \infty$, the regression approaches an OLS regression. If the error in the data set violates the homogeneity assumption of TTLS – such as PC regression where the SNR changes from low- to high-order modes – the OLS cal-culation is more appropriate. Since we have assumed that the temporal information in the AVHRR data is suspect, we therefore assign a large value to α. For $\alpha \geq 10$ the difference in regression results is negligible.

---

[2] While it is possible to supply a covariance / correlation matrix for TTLS, that supplied matrix is used only as a starting point for the infilling and estimates obtained using it are progressively overwritten. If the convergence tolerance is set appropriately, TTLS converges to a similar answer whether this information is supplied or omitted.

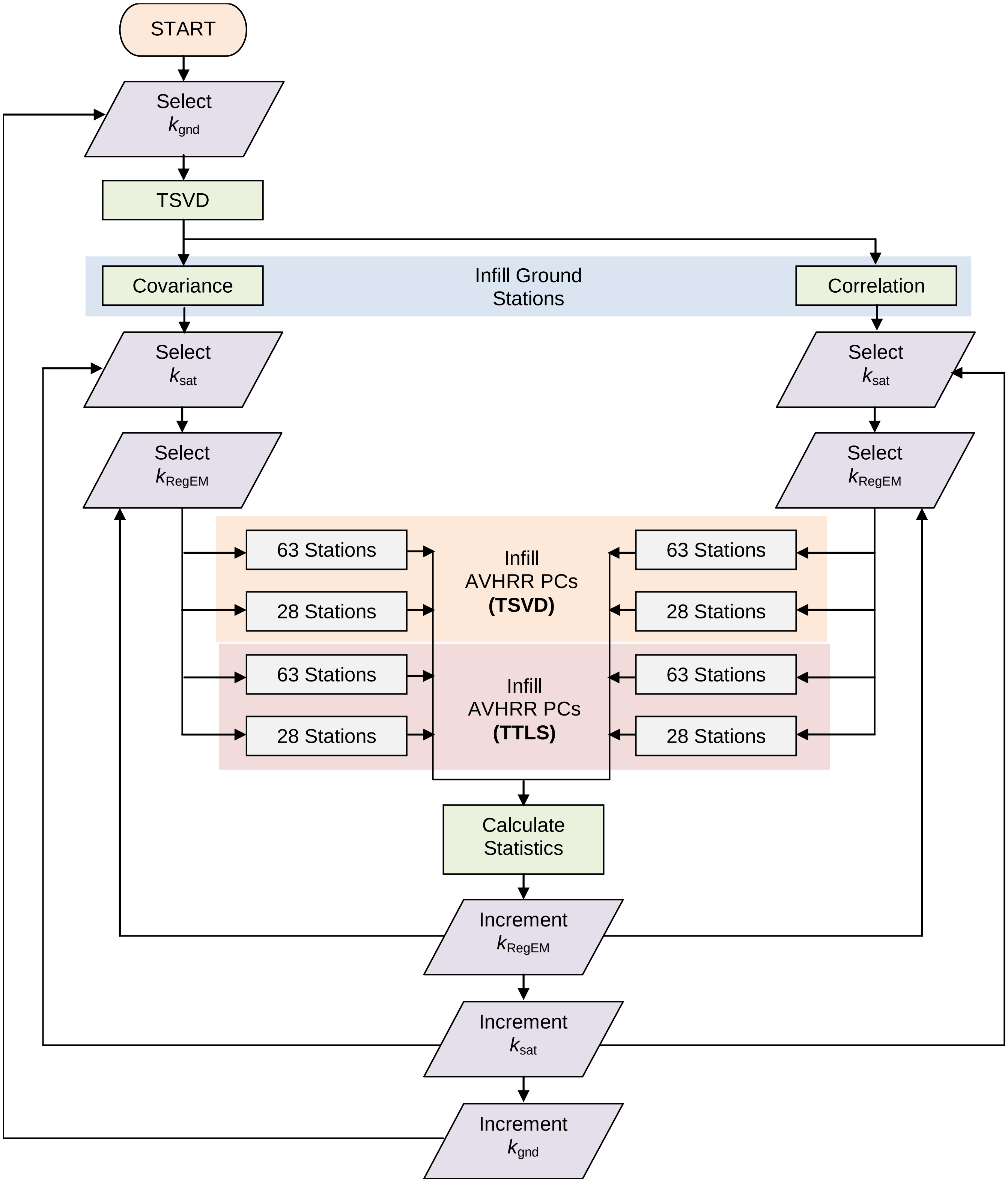

FIG. S10. Flowchart depicting E-W reconstruction cross validation experiments. The range for k_{RegEM} is 1 – 12; k_{sat} is 3, 13, 28, 50, and 100; k_{gnd} is 5 – 8. Full reconstructions use the 63-station Grid.1C set and use the 24 unused, on-grid stations as verification targets. The 28-station verification reconstructions withhold 35 stations from set Grid.1C for use as verification targets.

Significant overfitting of the PCs and dramatic decrease in verification statistics occurs as α drops below ≈5, indicating that our assumption of larger errors on the PCs than the ground stations was accurate.

Figure S10 provides a graphical demonstration for the E-W procedure and cross-validation tests. The process starts by infilling the selected ground station set using truncation parameter k_{gnd}. Infilling is performed in both a covariance and correlation setting. These completely infilled ground stations are augmented by the AVHRR PCs (one at a time) and the PCs are then infilled using both TSVD and TTLS. This process is repeated after withholding 35 stations from the infilled ground matrix to provide additional verification targets. The range of settings shown in Fig. S10 results in a total of 1,920 reconstructions being performed. We then select the optimal reconstructions, based on verification statistics.

For correlation networks, scaling is performed in the following manner:

- Scale the AVHRR data by dividing each series by a vector of unbiased standard deviation estimators $\tilde{\mathbf{s}}_{\text{sat}}$
- Perform SVD on the AVHRR data and extract k_{sat} principal components
- Augment the infilled ground station matrix with each principal component, one at a time, and infill using the correlation setting of the infilling algorithm

Following infilling, the estimated AVHRR principal components are reconstituted with their associated spatial eigenvectors, providing a 5509 x 600 matrix of scaled estimates. Estimated gridded temperature ($\hat{\mathbf{T}}$) is recovered across the grid by unscaling by $\tilde{\mathbf{s}}_{\text{sat}}$:

$$\hat{\mathbf{T}} = \tilde{\mathbf{s}}_{\text{sat}} \left(\tilde{\hat{\mathbf{U}}} \mathbf{\Lambda} \mathbf{V}^{\text{T}} \right), \qquad \text{(S3)}$$

where $\tilde{\hat{\mathbf{U}}}$ indicates the estimated PCs, scaled to correlation, and the vector of unbiased standard deviation estimators $\tilde{\mathbf{s}}_{\text{sat}}$ is given by:

$$\tilde{\mathbf{s}}_{\text{sat}} = \left\{ \frac{\text{diag}\left(\mathbf{T}^{\text{T}}\mathbf{T}\right)}{\tilde{n}} \right\}^{1/2}, \qquad \text{(S4)}$$

where $\tilde{n}$ represents the effective degrees of freedom. For an $n \times p$ matrix of anomalies, with individual series organized in columns, $\tilde{n} = n - 12$, as converting to anomalies removes 12 separate means. We note that the default setting in the Matlab version of RegEM has a hard-coded $\tilde{n} = n - 1$, and it is unclear if S09 changed this to compute their reconstructions. For correctness, however, we use $\tilde{n} = n - 12$ for our reconstructions.

S8. Additional information concerning RLS reconstructions

The *regularized least squares* (RLS) reconstructions directly utilize the AVHRR spatial information. Like the E-W reconstructions, the RLS reconstructions start with a completely infilled ground station matrix. However, unlike S09 and the E-W reconstructions, the AVHRR principal components are not used.

We implement two versions of the algorithm. One version computes the spatial structure simply via SVD of the AVHRR data (scaled to correlation, if desired). The second version preprocesses the AVHRR data by row centering prior to the SVD. The resulting time series of row means is then used to compute a spatial EOF with uniform weights on the grid. This uniform

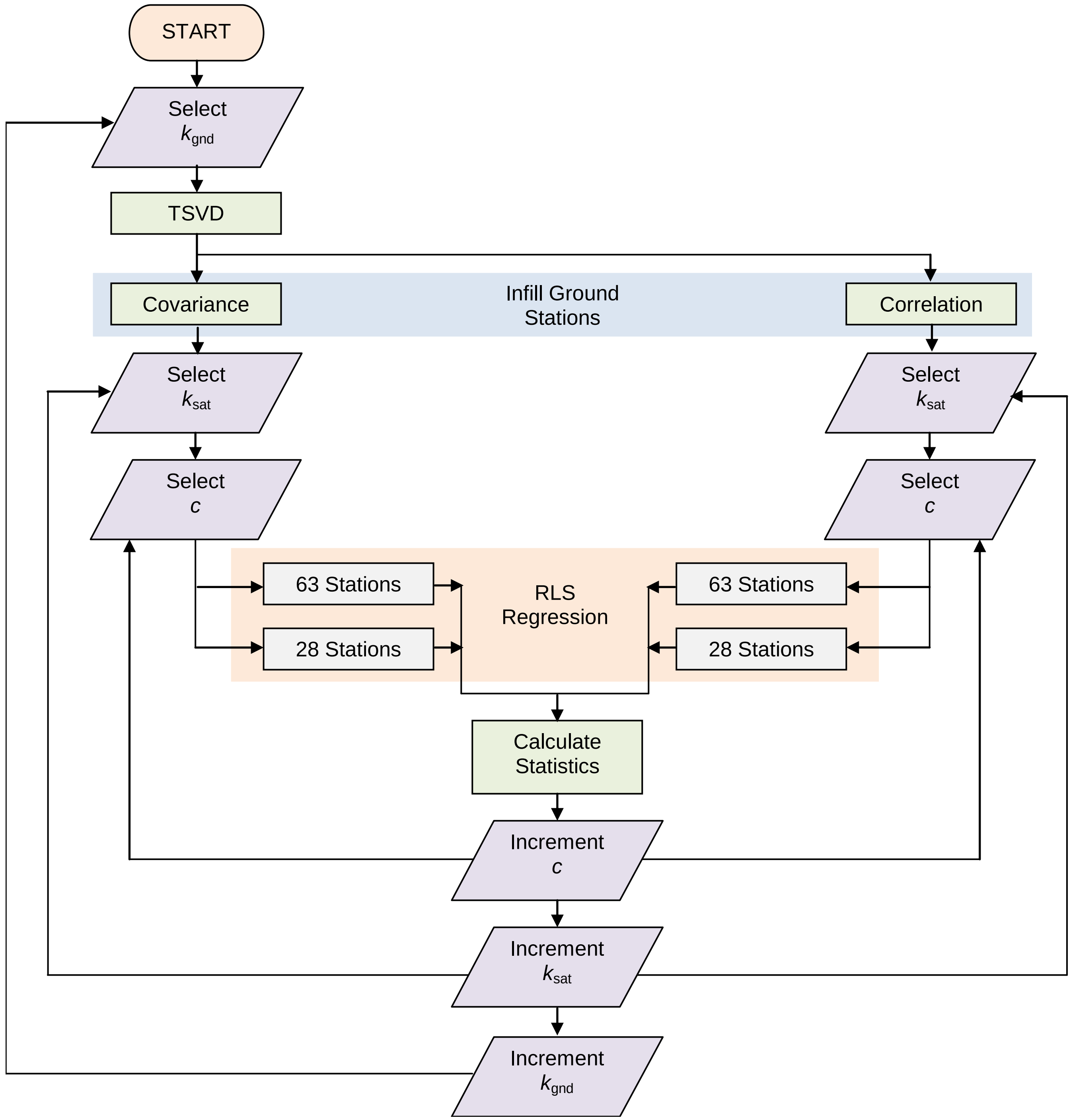

FIG. S11. Flowchart depicting RLS reconstruction cross validation experiments. The range for k_{RegEM} is 1 – 12; k_{sat} is 2 – 100; c is 0.1, 0.2, …, 1.1, 1.3, 1.5, 1.75, 2.0. Full reconstructions use the 63-station Grid.1C set and use the 24 unused, on-grid stations as verification targets. The 28-station verification reconstructions withhold 35 stations from set Grid.1C for use as verification targets.

EOF is included in Equation (10) as $\mathbf{L}_1$, and is used in the RLS regression to determine the continental trend. The reason for developing the second option was to prevent satellite measurement errors that are the same magnitude everywhere – such as a splicing error – from affecting the determination of the spatial eigenvectors.

When using the full set of AVHRR data from 1982 – 2006, both methods give nearly identical results (within ± 0.005 ^{o}C decade^{-1} for 1957 – 2006 continent-wide and regional trends). However, when using subsets of the AVHRR data, the row centered version provides significant stability in trend magnitude and geographic distribution regardless of the AVHRR time period used. We therefore use the row centered version throughout this study (though the option for the standard version is preserved in the code).

A flowchart of the cross-validation experiments for determining the optimal settings for the RLS reconstructions appears as Figure S11. The process is analogous to the E-W process, with scaling factor c replacing k_{RegEM} and k_{sat} varying smoothly from 2 to 100. Using the ranges listed in Fig. S11, the experiments result in performing 23,040 RLS reconstructions.

For correlation networks, scaling is performed in the following manner:

- Scale the AVHRR data by dividing each series by a vector of unbiased standard deviation estimators $\tilde{\mathbf{s}}_{\text{sat}}$
- Perform SVD on the AVHRR data and extract the first k_{sat} spatial EOFs
- Scale the infilled ground station matrix by dividing each series by a vector of unbiased standard deviation estimators $\tilde{\mathbf{s}}_{\text{gnd}}$

The least squares solution is then found by solving Eqn. [10] using the scaled spatial EOFs $\tilde{\mathbf{L}}$ and scaled ground station data $\tilde{\mathbf{Y}}$ for each time j. From the definition of $\mathbf{L}$, this provides a solution vector $\mathbf{a}_j$, which is the estimate of the correlation AVHRR temporal eigenvectors at time j divided by the square root of the effective degrees of freedom:

$$\mathbf{a}_j = \sqrt{\tilde{n}}(\hat{\tilde{\mathbf{U}}}) \qquad \text{(S5)}$$

The 5509 x 600 matrix of temperature estimates is then given by:

$$\hat{\mathbf{T}} = \frac{\tilde{\mathbf{s}}_{\text{sat}}}{\sqrt{\tilde{n}}}\left(\mathbf{a}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}\right) \qquad \text{(S6)}$$

S9. Additional results and full verification statistics

a. E-W parameter sensitivity

For the E-W reconstructions, there are 3 major adjustable parameters: k_{gnd} (truncation parameter for infilling the ground stations), k_{sat} (number of AVHRR modes to regress), and k_{RegEM} (truncation parameter for regressing the AVHRR principal components against the ground station data). Figure S12 shows the variation in the CE statistic based on changes to k_{sat} and k_{RegEM}. Continental and regional trends show some minor dependence on these parameters; using $k_{\text{sat}} = 28$ and $k_{\text{RegEM}} = 3$, for example, yields trends of 0.05, 0.01, and 0.15 ^{o}C decade^{-1} for the continent, West Antarctica, and the Peninsula, respectively. Increasing k_{sat} and k_{RegEM} generally result in larger trends for West Antarctica and the Peninsula. We find large values for k_{sat} (100) and k_{RegEM} (9) to provide optimal verification statistics.

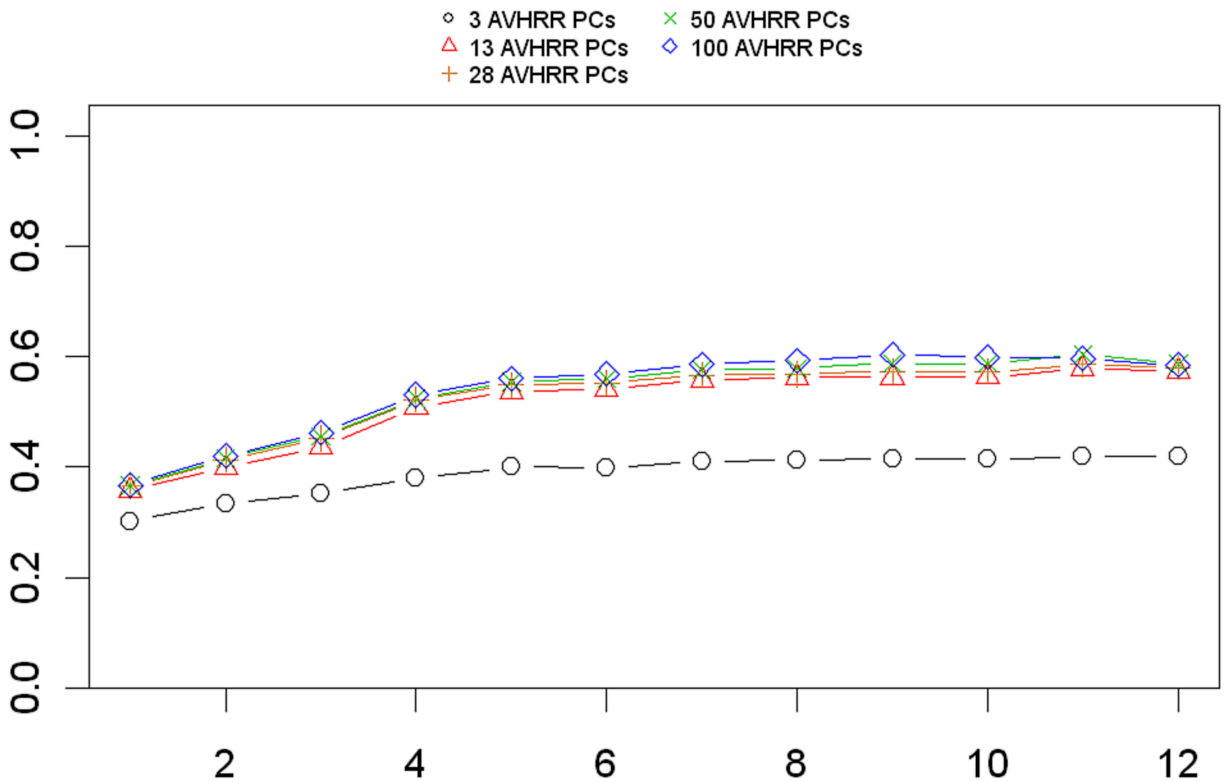

FIG. S12. CE statistic vs. k_{RegEM} for various values of k_{sat}. Statistics computed based on reconstructions using the TSVD algorithm in a correlation setting, with $k_{\text{gnd}} = 7$. Patterns for TTLS, covariance, and/or other settings for k_{gnd} are similar. The reconstructions presented in the text utilize values of $k_{\text{sat}} = 100$ and $k_{\text{RegEM}} = 9$.

b. RLS parameter sensitivity

The RLS reconstruction parameters include k_{gnd}, k_{sat}, and c (regularization parameter scaling factor). These reconstructions demonstrate an even smaller dependence on k_{sat} than the E-W reconstructions, and very little dependence on c (for small values of c). An RLS reconstruction using settings of $k_{\text{sat}} = 13$ and $c = 0.5$, for example, yields trends of 0.06, 0.07, and 0.36 ^{o}C decade^{-1} for the continent, West Antarctica, and the Peninsula, respectively. These are almost identical to the trends calculated using the optimal settings of $k_{\text{sat}} = 80$ and $c = 0.1$. Including too few AVHRR eigenvectors results in an inability to reproduce high frequency variability, but has little effect on trends. Varying c produces a similar effect until c exceeds ≈1.0, where excessive regularization begins to result in a suppression of trend magnitudes. Values smaller than ≈0.1 result in computational singularities due to insufficient regularization when k_{sat} exceeds the number of stations. When fewer eigenvectors than stations are used (which allows c = 0), the change in average trends is in the third significant digit and spatial

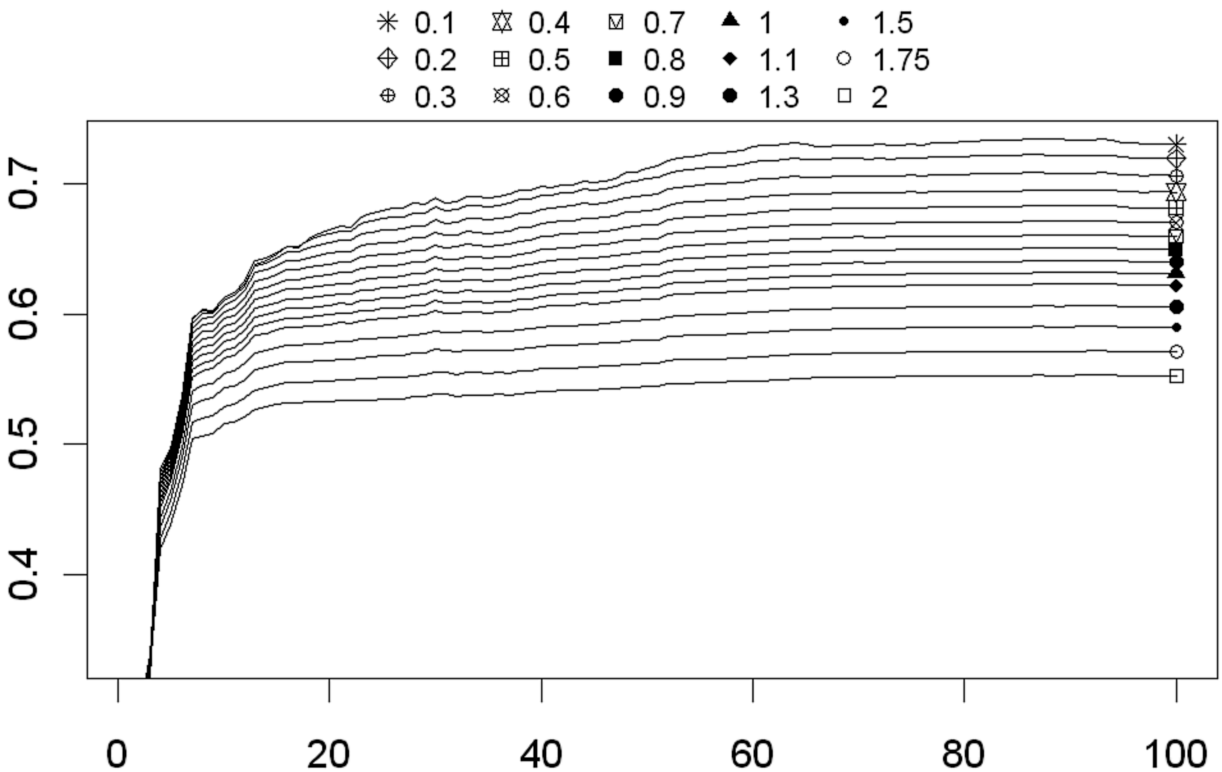

FIG. S13. CE statistic vs. k_{sat} for various values of c. Statistics computed based on reconstructions in a correlation setting, with $k_{gnd} = 7$. Patterns for covariance and/or other settings for k_{gnd} are similar. The reconstructions presented in the main text utilize values of $k_{sat} = 80$ and $c = 0.1$. Values smaller than $c = 0.1$ result in computational singularities due to insufficient regularization when the AVHRR eigenvectors outnumber the included stations.

patterns were indistinguishable. We therefore determined it unnecessary to find the smallest allowable value for c with a resolution better than 0.1. Figure S13 shows the dependence of CE on k_{sat} and c.

c. Sensitivity to k_{gnd}

The parameter for which the E-W and RLS (and, incidentally, S09) reconstructions display the most sensitivity is k_{gnd}, which is the truncation parameter used for the infilling of the ground stations. As the ground stations are used to determine the time series of loadings to be applied to each AVHRR spatial eigenvector, they contain the temporal information in the reconstructions. The AVHRR data is simply used to fill in the empty space between stations.

Figures S14 – S17 and Table S4 show reconstructions holding k_{sat}, k_{RegEM}, and c constant, but with k_{gnd} varied from 5 to 8. Reconstructions using both correlation and covariance networks are presented. For the correlation networks, $k_{gnd} = 7$ yields superior verification statistics. For covariance networks, $k_{gnd} = 6$ is optimal. Verification statistics for West Antarctica alone were also superior with those settings of k_{gnd}.

As discussed in Section 7 of the main text, the area most sensitive to the choice of k_{gnd} is West Antarctica. While a small area of statistically insignificant cooling to neutral trend in the Ross region of West Antarctica is a feature common to almost all of the reconstructions, the more prominent cooling that results when using optimal parameters is unique to those parameter settings (though only when k_{sat} is large). However, if fewer AVHRR eigenvectors are included, the Ross cooling becomes a robust feature of the reconstruction for all values of k_{gnd} investigated. Due to the paucity of ground data in the region, it is not possible to determine whether the cooling is an artifact of overfitting during the E-W or RLS regressions, or whether including the optimal number of AVHRRs for elsewhere on the continent results in masking a robust West Antarctic cooling feature with noise, since few predictors are available in the area of concern. Regardless, based on the ridge regression and RLS-without-infilling tests, the magnitude of West Antarctic trends in the S09 reconstruction are larger than would be supported by either explanation.

Significantly decreased solution stability – especially for West Antarctica – is found for covariance networks. The primary reason for this effect is not the RLS or E-W regressions themselves; it is the initial ground station infilling that feeds the regressions (Fig. S8 and Table S3). Covariance networks are far more sensitive to k_{gnd} than are correlation networks. As an example, Nansen Ice Sheet, Cape King, Cape Phillips, and Engima Lake are all co-located (~ 74^{o}S, 165^{o}E) and have similar record lengths (~ 160 months), yet display drastically different behavior as k_{gnd} is varied from 5 to 7. In a correlation network, on the other hand, all four of those stations display very similar behavior for the same variation in k_{gnd}, which is what we would expect for a closely-positioned group of stations for which the raw temperature data shows a high correlation of ~0.85 for all stations to each other.

The precise reason for this behavior is the subject of an ongoing work by the present authors. Initial results suggest that the reason may be due to the fixed truncation parameter in TTLS and TSVD resulting in overfitting when the number of available predictors for that particular time step is low. Since high variance indicates more local microclimate variability or “weather noise”, stations displaying increased variance may provide less accurate predictions. This problem is exacerbated in a covariance setting, as their larger variance results in preferential selection of those stations. When the number of predictors is small, the fixed truncation parameter does not provide adequate filtering. Scaling to correlation helps decrease the influence of the high variance stations, reducing the amount of overfitting during periods of sparse predictors. Experiments using ridge regression with the ridge parameter optimized for *each* time step based on the number of available predictors demonstrate enhanced stability (with solutions similar to the optimal values of k_{gnd}) and provide some evidence that this reasoning is accurate.

Regardless of the precise cause, the impact of instability is naturally more important for the short record length stations, which have a higher percentage of points that require infilling. This would indicate that we should expect more stable results in the RLS and E-W regressions as the average record length of the included ground stations increases. We find this is, indeed, the case. Table S4 shows continental and regional trends for various settings of k_{gnd}. In a covariance network, reconstruction trends are much more consistent for the 28-station verification reconstructions (which have an average record length of 435 months for included stations) than the 63-station reconstructions (which have an average record length of 274 months for included stations).

For correlation networks, however, the difference between the 63- and 28-station reconstructions is much less striking, especially in the Peninsula and West Antarctica, with the E-W method being the most stable. Almost all correlation E-W settings for k_{gnd} (for both the 63- and 28-station reconstructions) yield continental and regional trends that are similar to the optimal settings in correlation RLS. In particular, the trends computed for West Antarctica in correlation E-W all provide point estimates that are well within each other's 95% confidence intervals and well within the confidence intervals for the optimal settings in correlation RLS. The point estimate for S09 of 0.20^{o}C decade^{-1} lies *outside* the 95% confidence intervals of every correlation E-W result with the exception of the 63-station, $k_{\text{gnd}} = 5$ solution, where it lies on the 95% CI. Most importantly, the magnitudes and spatial patterns of temperature change (including covariance) are similar when the optimal settings are used, and the point estimates for the Peninsula and West Antarctic trends in S09 are outside the 95% confidence intervals for all reconstructions using optimal settings. These observations, coupled with the ridge regression and RLS-without-infilling experiments give us a reasonable degree of confidence in the results using the optimal settings.

d. Spatial similarity between ground station and AVHRR data

As noted above and in the main text, for the reconstruction to be valid, the ground data and AVHRR data must have similar spatial structures. In order to provide a basic check on our results, we additionally conduct a reconstruction using no satellite data and no interpolation. For this reconstruction, each of the 5,509 grid cells are infilled using the actual monthly anomaly from the nearest ground station, plus an offset determined by station overlaps (where the chronologically-nearest 60 months of data to the point being infilled is used to compute the offset). This reduces the resolution across Antarctica to a series of polygons, which change in number, size, and shape based on the available stations. It provides a gross estimate of the underlying spatial structure for the ground stations alone. If the spatial structures between the ground data and AVHRR data are compatible, this method should produce a spatial distribution of temperature change that is similar to that produced by the RLS and E-W reconstructions.

Fig. S24 at the end of this document presents this reconstruction alongside the optimal RLS and E-W correlation reconstruction. We note good agreement between the three reconstructions, which indicates that fundamentally the ground and AVHRR data share similar spatial makeups. The overall trends are also similar (0.06, 0.03, 0.14, and 0.36 ^{o}C decade^{-1} for the continent, East Antarctica, West Antarctica, and the Peninsula, respectively), with West Antarctica being higher than the RLS and E-W reconstructions due to the lack of any satellite spatial information for the West Antarctic interior. With no satellite data to constrain the infilling, the nearest-station approach copies the higher-magnitude Transantarctic Mountain trends throughout the Ross region. We also note that the nearest-station reconstructed variances are much higher during the subperiods (factor of ~ 1.5) than the RLS or E-W reconstructions. This is because the nearest-station reconstruction lacks the filtering provided by the truncation parameters in the RLS and E-W reconstructions. However, overall, the match is quite satisfactory. We also provide Figure S25, which shows that the resulting continental trend is not strongly dependent on the maximum number of overlapping months for determining offsets, except when few months are chosen. The choice of 60 months corresponds to the middle of the flat region of the plot.

e. Notes on the code

The entirety of the code (AntarcticaFinal.txt) for this project was written in the R programming language, version 2.9.1. While most features operate well on prior versions, the code was not explicitly tested for backwards compatibility. We recommend that interested readers obtain version 2.9.1 or higher from CRAN before attempting to run the code. Readers will also need to obtain the following packages: mapproj, maps, and waveslim.

The code is organized into the following major sections:

- PLOTTING FUNCTIONS, which defines the commands used to generate the plots in both the main text and this SI.
- GENERAL FUNCTIONS, which defines data loading and parsing commands.
- VERIFICATION FUNCTIONS, which defines commands for computing verification statistics, extracting spatial weights, performing Monte Carlo analysis and generating the Chladni patterns.
- EM ALGORITHMS, which defines the TTLS and TSVD algorithms.
- IMPUTATION EXPERIMENTS, which defines the commands for performing the cross-validation experiments.
- RECONSTRUCTION FUNCTIONS, which defines the commands for regressing the satellite PCs and spatial eigenvectors against the ground data.
- VARIABLE ASSIGNMENTS, which contain the station sets, standard plot labels, grid cell assignments for ground stations and geographic masks.
- LOAD AND PARSE DATA, which obtains all data from on-line sources and parses it for use.
- SATELLITE OFFSETS, which calculates differences in ground and satellite data.
- RECONSTRUCTION, which performs the E-W, RLS and S09 replication reconstructions.
- VERIFICATION, which computes full verification statistics.

During the initial run-through of the code, all necessary data will be downloaded from public sources and saved into the current working directory automatically. Following the initial run, we recommend users comment-out the “down.data()” command at the beginning of the LOAD AND PARSE DATA segment and utilize the saved data rather than download the data every time.

The code as provided will produce the E-W and RLS correlation reconstructions presented in the main text and takes approximately 20 minutes to run on a 1.3 GHz Centrino processor. At least 2 GB of RAM is required. For readers interested in performing their own sensitivity analyses, each function is thoroughly commented to describe the purpose and nomenclature used to change the settings and perform different reconstructions, including using different station sets or establishing new station sets.

Complete verification statistics for reader-defined settings can also be computed.

An important note is that the reconstructions are quite memory intensive. As R is restricted in the amount of RAM it may use, readers interested in performing a large number of reconstructions in a row are encouraged to modify the imputation experiment commands to use the parameters they choose. The imputation experiments will save the reconstructions in independent files and clear the volatile memory before proceeding to the next reconstruction, which will prevent out-of-memory errors. The saved reconstructions can then be loaded and analyzed.

Additionally, while the EM algorithms were written using the notation in the main text and SI, the reconstruction functions were written utilizing the *transpose* of the data matrices, which causes the spatial information to appear in the left-hand eigenvector (u) instead of the right-hand eigenvector (v). In the event of confusion, the comments provided in the code clearly indicate which eigenvector contains the spatial information for each function.

Lastly, any questions about the operation of the code should be referred to the primary author. Email is the preferred method of communication.

TABLE S4. Continental and regional trend sensitivity as a function of reconstruction method and k_{gnd}. Full reconstructions use the optimal 63-station set; verification reconstructions use the 28 on-grid stations for S09. Settings which correspond to maximum verification statistics are bolded. The correlation variants are featured in the main text.

			Trend (°C decade^{-1} +/- 95% CI)	
Type	k_{gnd}	Region	Full Recon	Ver. Recon
RLS Correlation	5	Continent	0.09 ± 0.08	0.06 ± 0.07
		Peninsula	0.41 ± 0.11	0.36 ± 0.11
		West	0.17 ± 0.09	0.14 ± 0.08
		East Antarctica	0.05 ± 0.09	0.02 ± 0.08
	6	Continent	0.08 ± 0.07	0.06 ± 0.07
		Peninsula	0.39 ± 0.11	0.36 ± 0.11
		West	0.16 ± 0.08	0.14 ± 0.08
		East Antarctica	0.04 ± 0.09	0.02 ± 0.08
	7	**Continent**	**0.06 ± 0.07**	**0.06 ± 0.07**
		Peninsula	**0.29 ± 0.10**	**0.35 ± 0.11**
		West	**0.05 ± 0.08**	**0.11 ± 0.07**
		East	**0.05 ± 0.09**	**0.03 ± 0.08**
	8	Continent	0.06 ± 0.07	0.06 ± 0.07
		Peninsula	0.29 ± 0.09	0.34 ± 0.11
		West	0.10 ± 0.07	0.12 ± 0.07
		East Antarctica	0.03 ± 0.09	0.02 ± 0.08
RLS Covariance	5	Continent	0.10 ± 0.08	0.06 ± 0.08
		Peninsula	0.52 ± 0.14	0.47 ± 0.13
		West	0.24 ± 0.12	0.18 ± 0.10
		East Antarctica	0.04 ± 0.10	-0.00 ± 0.09
	6	**Continent**	**0.06 ± 0.08**	**0.05 ± 0.08**
		Peninsula	**0.41 ± 0.12**	**0.43 ± 0.12**
		West	**0.13 ± 0.11**	**0.14 ± 0.11**
		East	**0.02 ± 0.09**	**0.00 ± 0.09**
	7	Continent	0.03 ± 0.08	0.03 ± 0.08
		Peninsula	0.27 ± 0.12	0.41 ± 0.12
		West	-0.05 ± 0.13	0.04 ± 0.13
		East Antarctica	0.04 ± 0.09	0.00 ± 0.09
	8	Continent	0.11 ± 0.08	0.14 ± 0.09
		Peninsula	0.42 ± 0.13	0.50 ± 0.13
		West	0.24 ± 0.16	0.50 ± 0.18
		East Antarctica	0.06 ± 0.09	0.02 ± 0.09

TABLE S4. (continued)

			Trend (°C decade^{-1} +/- 95% CI)	
Type	k_{gnd}	Region	Full Recon	Ver. Recon
E-W Correlation	5	Continent	0.09 ± 0.07	0.05 ± 0.07
		Peninsula	0.35 ± 0.09	0.31 ± 0.08
		West	0.13 ± 0.07	0.06 ± 0.07
		East Antarctica	0.07 ± 0.08	0.03 ± 0.08
	6	Continent	0.08 ± 0.07	0.05 ± 0.07
		Peninsula	0.34 ± 0.09	0.28 ± 0.07
		West	0.12 ± 0.07	0.07 ± 0.06
		East Antarctica	0.06 ± 0.08	0.03 ± 0.08
	7	**Continent**	**0.05 ± 0.07**	**0.05 ± 0.07**
		Peninsula	**0.29 ± 0.08**	**0.32 ± 0.08**
		West	**0.04 ± 0.06**	**0.06 ± 0.08**
		East	**0.04 ± 0.08**	**0.03 ± 0.08**
	8	Continent	0.07 ± 0.07	0.05 ± 0.07
		Peninsula	0.28 ± 0.08	0.31 ± 0.08
		West	0.09 ± 0.06	0.08 ± 0.06
		East Antarctica	0.05 ± 0.08	0.03 ± 0.08
E-W Covariance	5	Continent	0.07 ± 0.07	0.05 ± 0.07
		Peninsula	0.34 ± 0.09	0.29 ± 0.08
		West	0.12 ± 0.07	0.08 ± 0.07
		East Antarctica	0.04 ± 0.08	0.03 ± 0.08
	6	**Continent**	**0.06 ± 0.07**	**0.05 ± 0.07**
		Peninsula	**0.27 ± 0.08**	**0.27 ± 0.08**
		West	**0.07 ± 0.06**	**0.07 ± 0.06**
		East	**0.04 ± 0.08**	**0.04 ± 0.08**
	7	Continent	0.06 ± 0.07	0.04 ± 0.07
		Peninsula	0.19 ± 0.07	0.22 ± 0.07
		West	-0.02 ± 0.06	-0.01 ± 0.06
		East Antarctica	0.07 ± 0.08	0.05 ± 0.08
	8	Continent	0.09 ± 0.07	0.08 ± 0.07
		Peninsula	0.28 ± 0.07	0.28 ± 0.07
		West	0.08 ± 0.07	0.07 ± 0.07
		East Antarctica	0.08 ± 0.08	0.07 ± 0.08

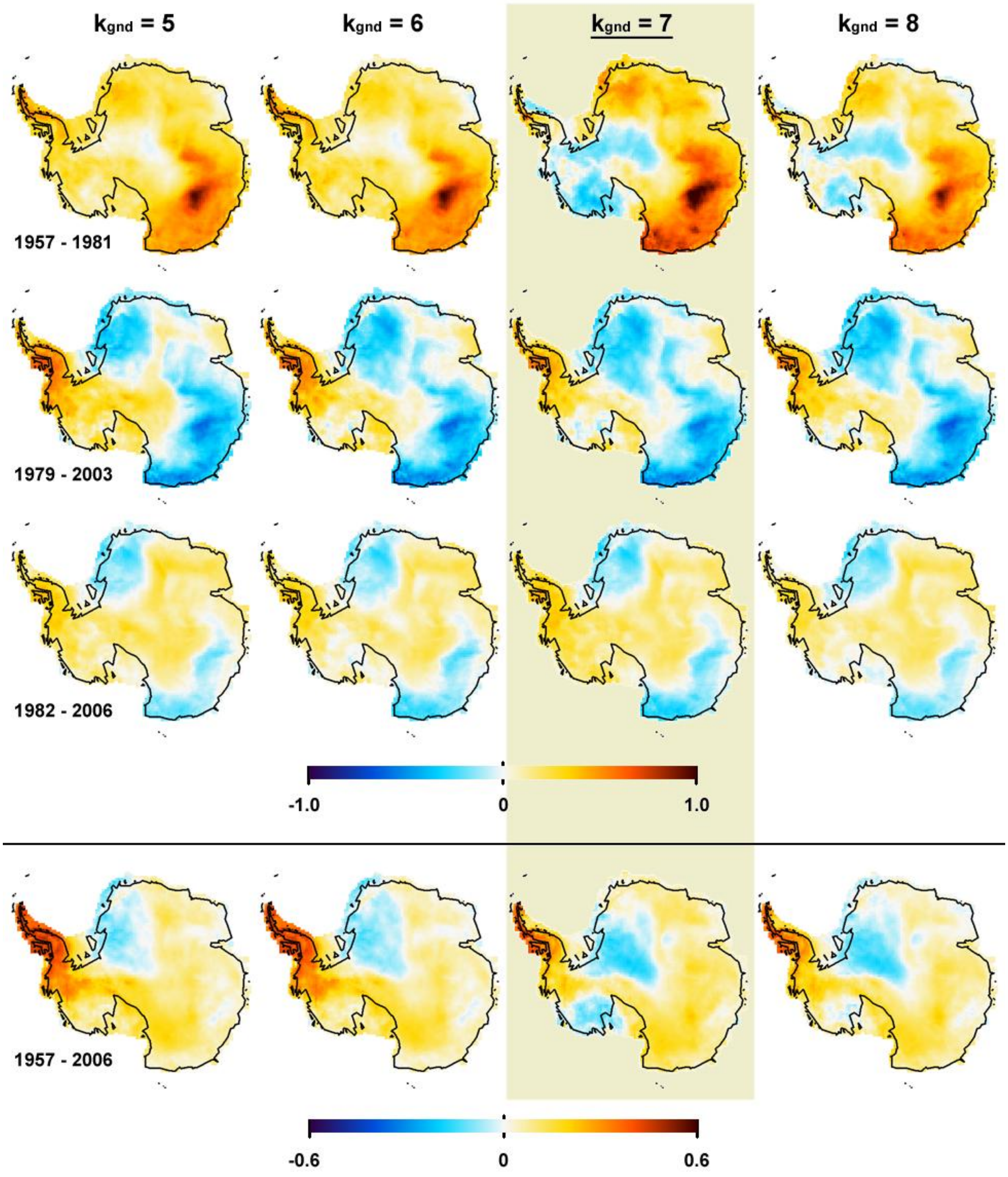

FIG. S14. Spatial distribution of trends for various timeframes for RLS correlation network reconstructions, with varying k_{gnd}. Maximum verification statistics for the entire continent and West Antarctica specifically are achieved with $k_{\text{gnd}} = 7$.

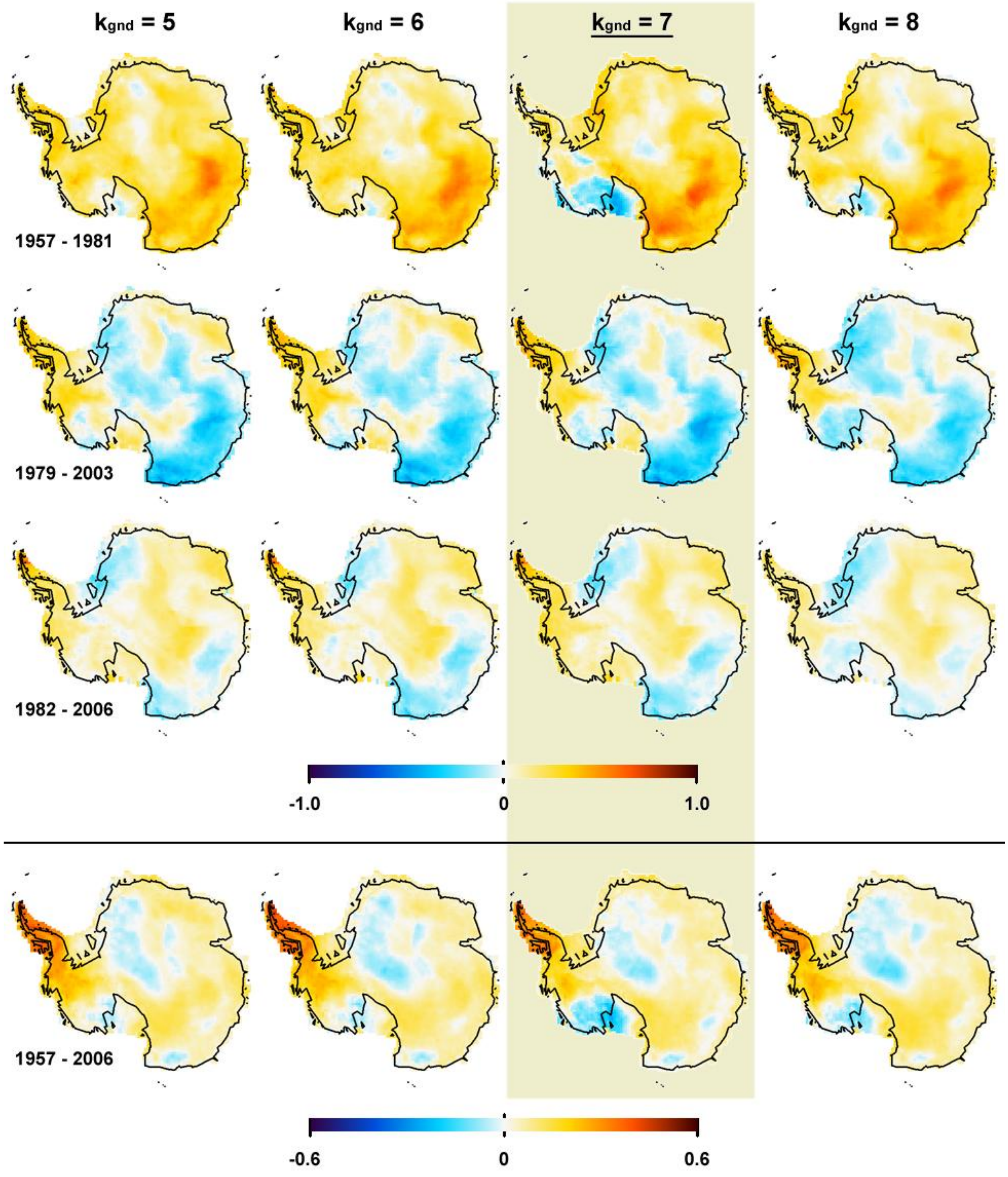

FIG. S15. Spatial distribution of trends for various timeframes for E-W correlation network reconstructions, with varying k_{gnd}. Maximum verification statistics for the entire continent and West Antarctica specifically are achieved with $k_{\text{gnd}} = 7$.

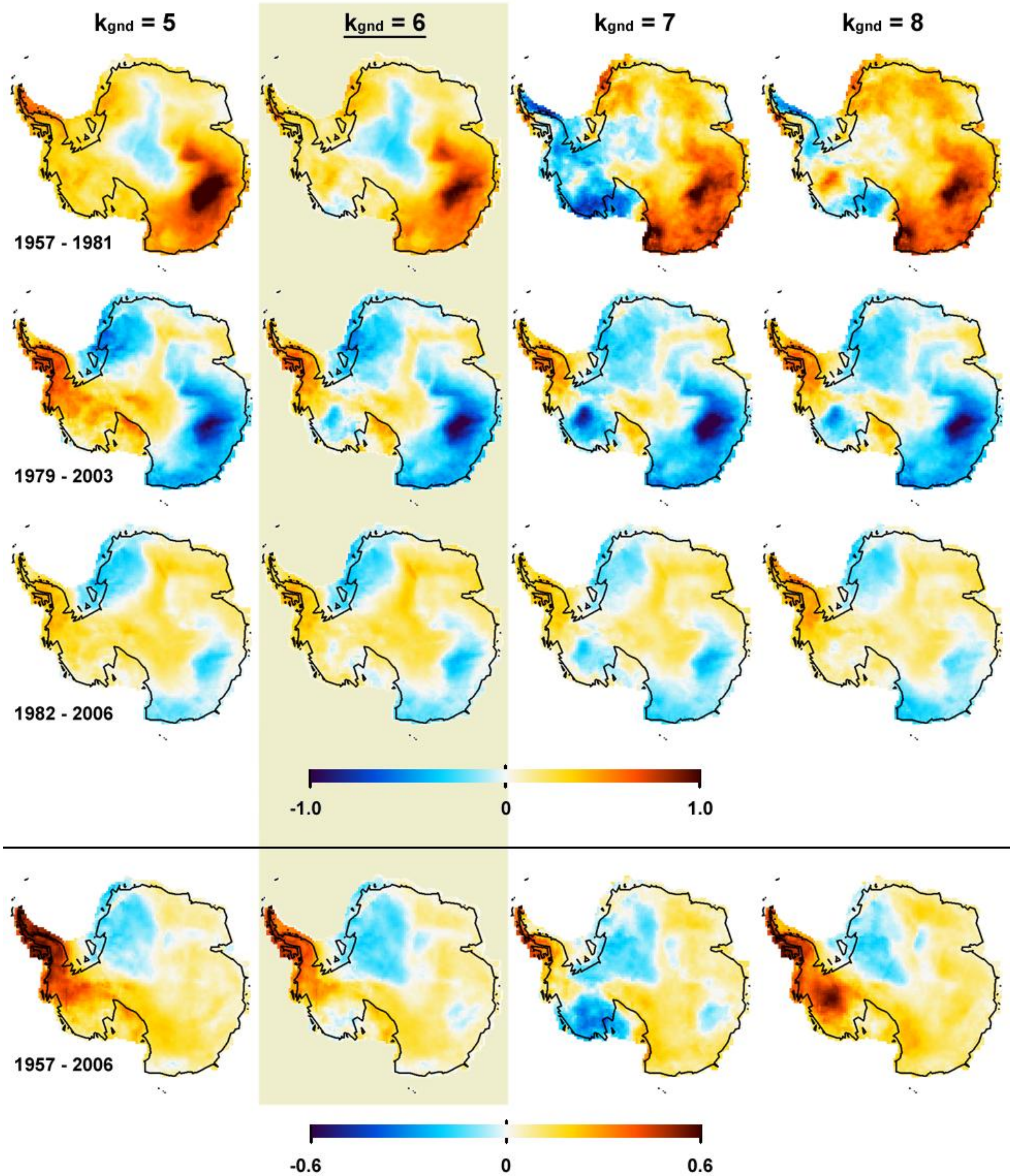

FIG. S16. Spatial distribution of trends for various timeframes for RLS covariance network reconstructions, with varying k_{gnd}. Maximum verification statistics for the entire continent and West Antarctica specifically are achieved with $k_{gnd} = 6$.

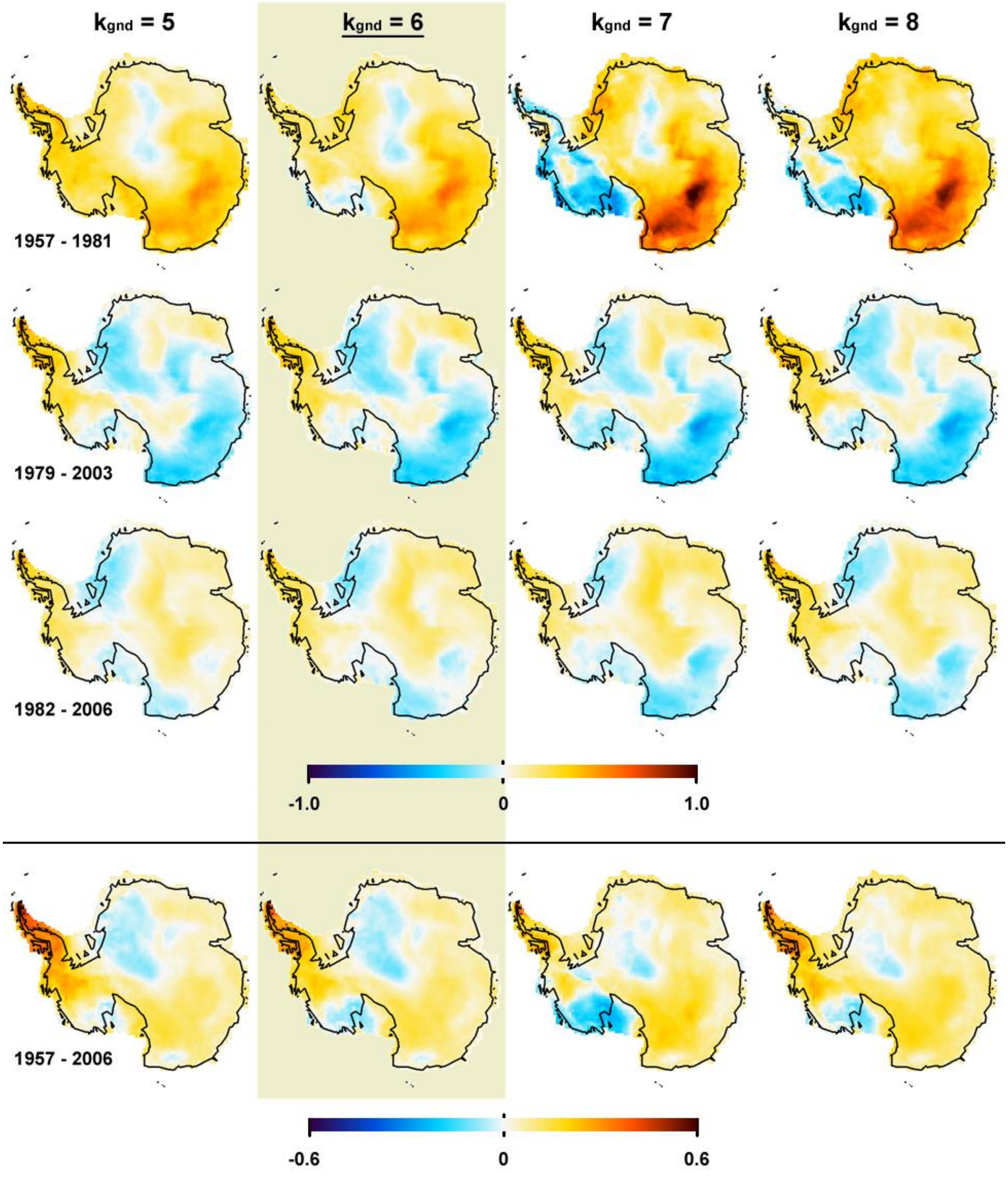

FIG. S17. Spatial distribution of trends for various timeframes for E-W covariance network reconstructions, with varying k_{gnd}. Maximum verification statistics for the entire continent and West Antarctica specifically are achieved with $k_{\text{gnd}} = 6$.

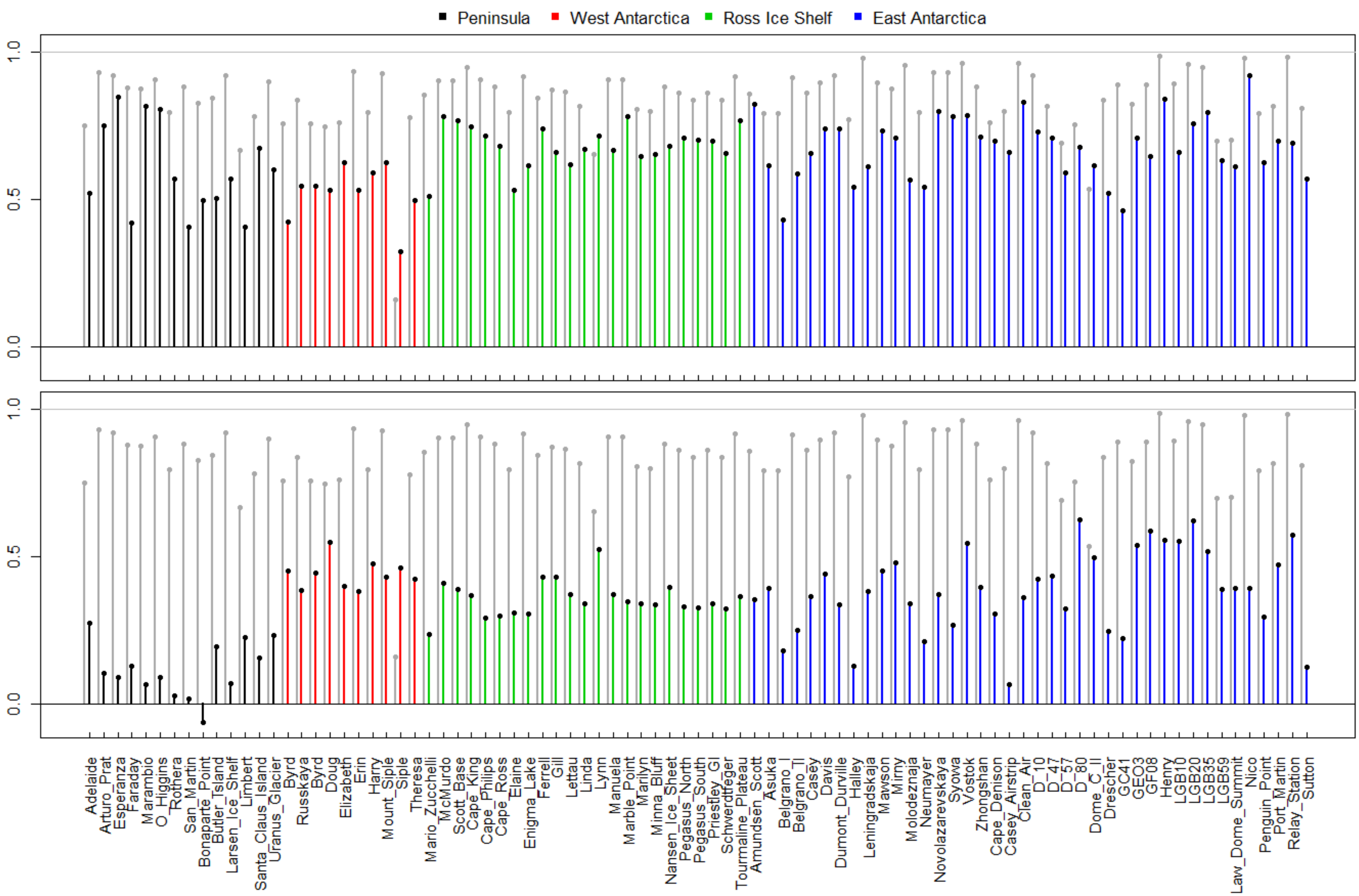

FIG. S18. Average explained variance / CE for the RLS **correlation** reconstruction vs. the E-W correlation reconstruction (top panel) and the S09 reconstruction (bottom panel). RLS values are in gray; E-W and S09 values are color coded by geographic location.

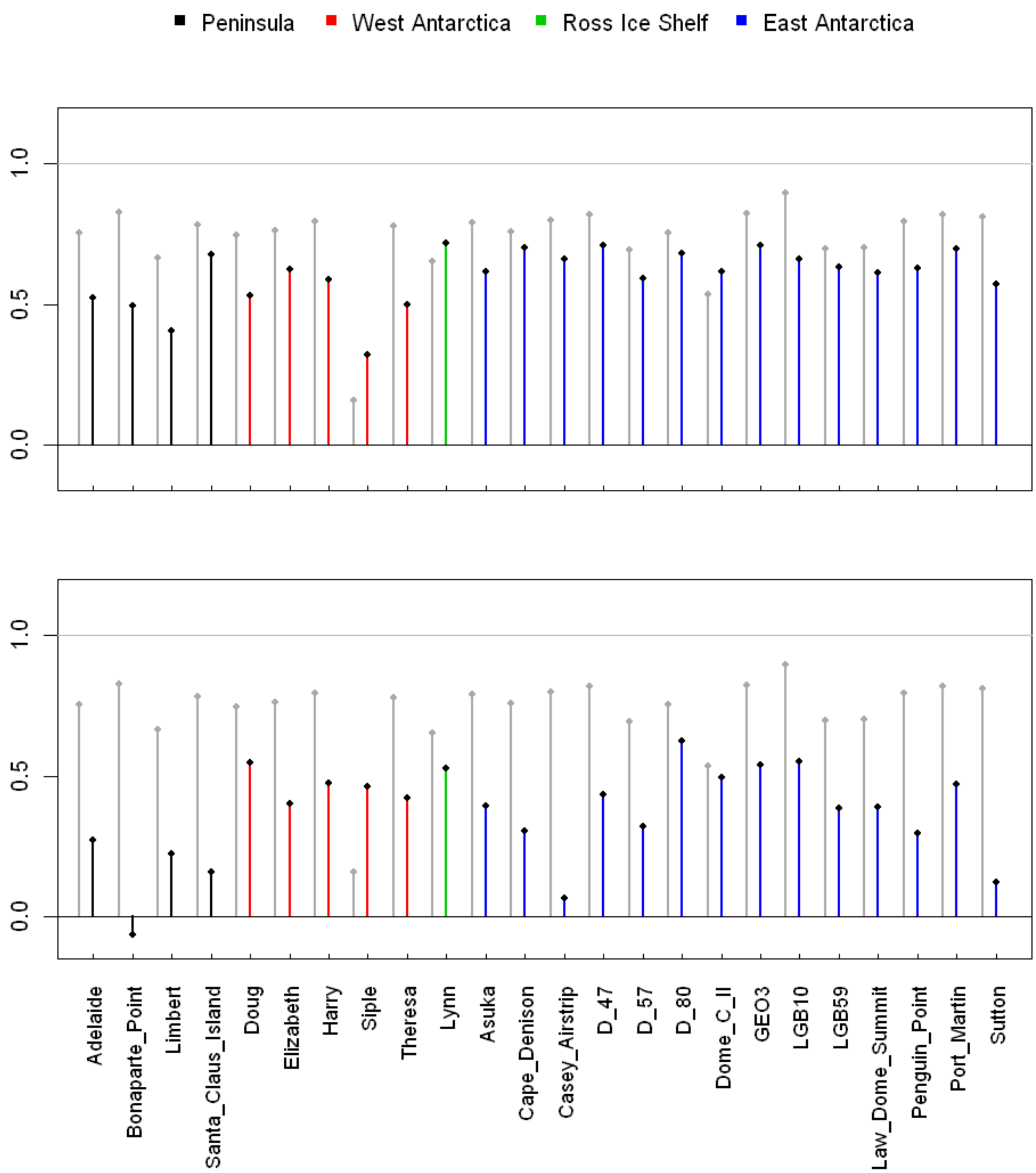

FIG. S19. CE for stations entirely withheld from the RLS **correlation** reconstruction versus the E-W correlation reconstruction (top panel) and the S09 reconstruction (bottom panel). RLS values are in gray; E-W and S09 values are color coded by geographic location.

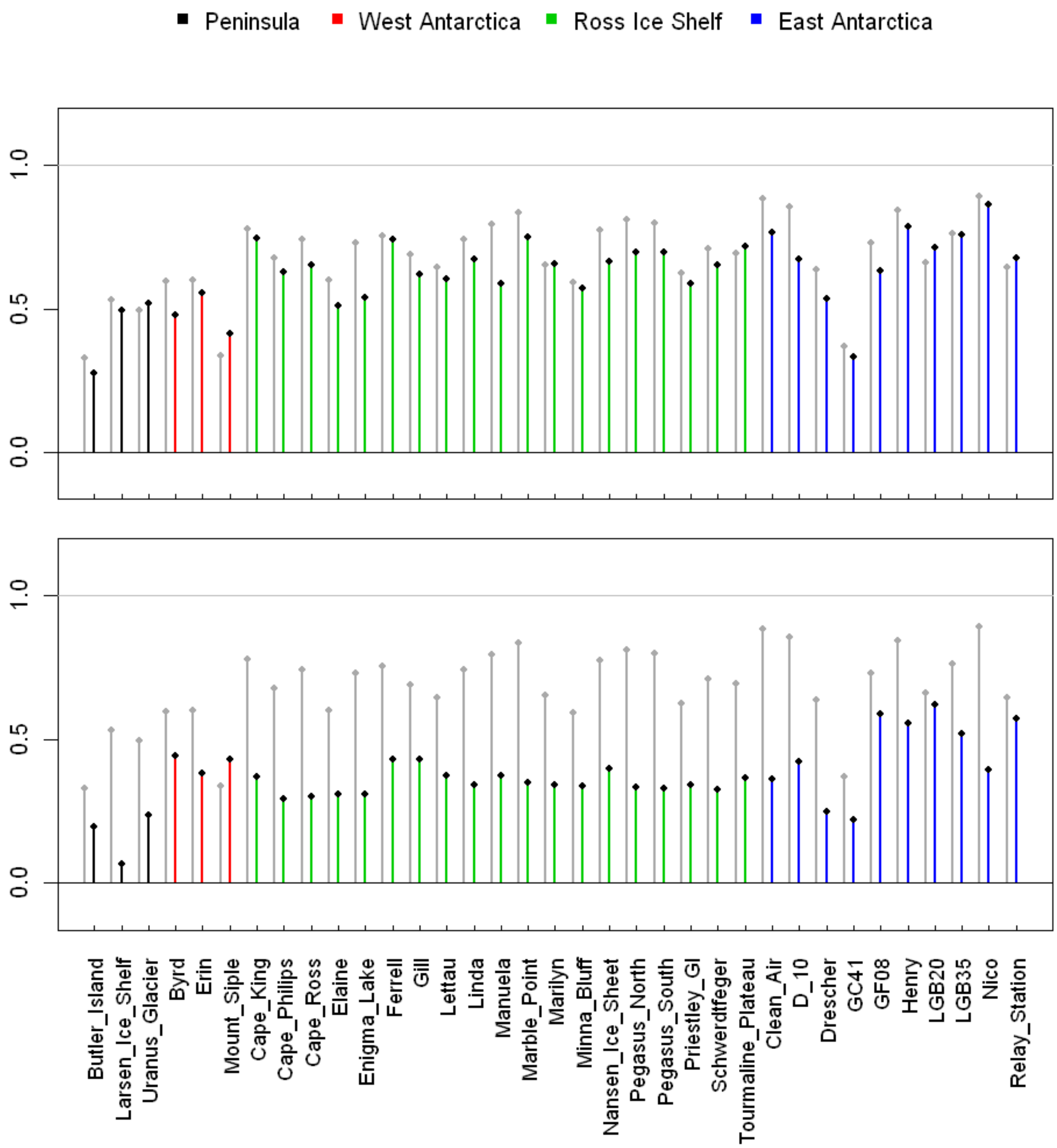

FIG. S20. CE for the 35 additional stations entirely withheld from the 28-station RLS **correlation** ***verification*** reconstruction versus the E-W correlation verification reconstruction (top panel) and the S09 reconstruction (bottom panel). RLS values are in gray; E-W and S09 values are color coded by geographic location.

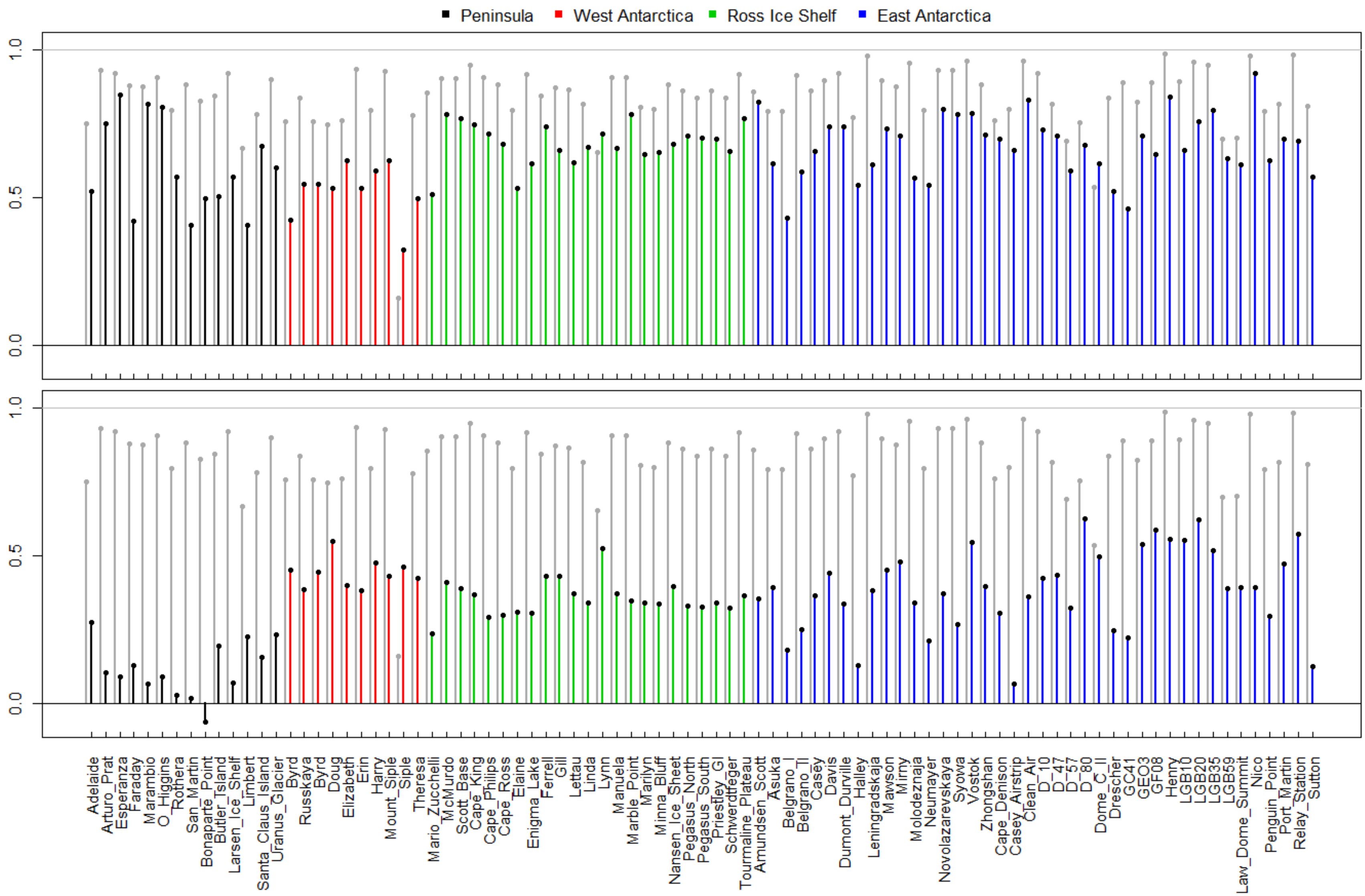

FIG. S21. Average explained variance / CE for the RLS **covariance** reconstruction vs. the E-W covariance reconstruction (top panel) and the S09 reconstruction (bottom panel). RLS values are in gray; E-W and S09 values are color coded by geographic location.

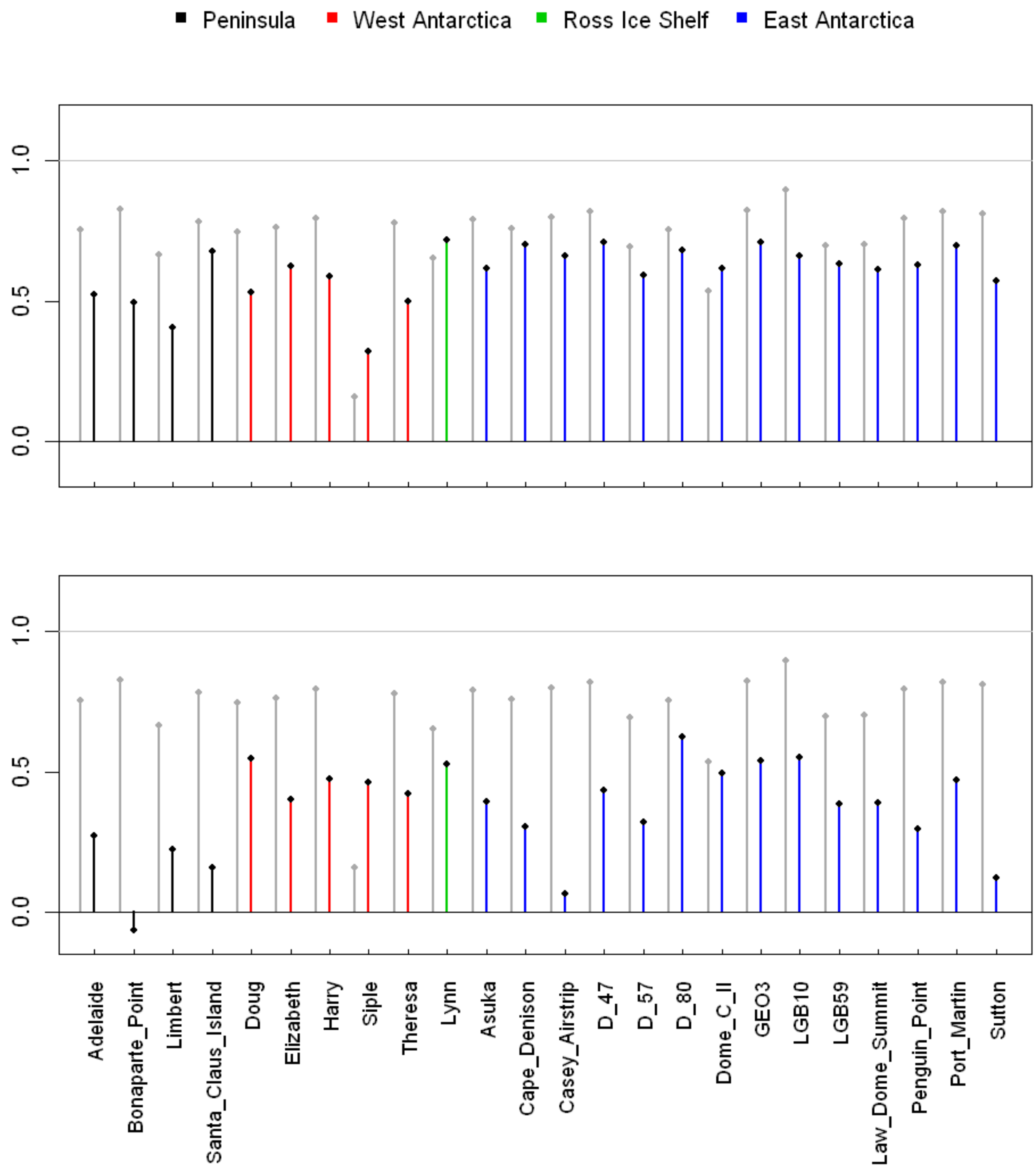

FIG. S22. CE for stations entirely withheld from the RLS **covariance** reconstruction versus the E-W covariance reconstruction (top panel) and the S09 reconstruction (bottom panel). RLS values are in gray; E-W and S09 values are color coded by geographic location.

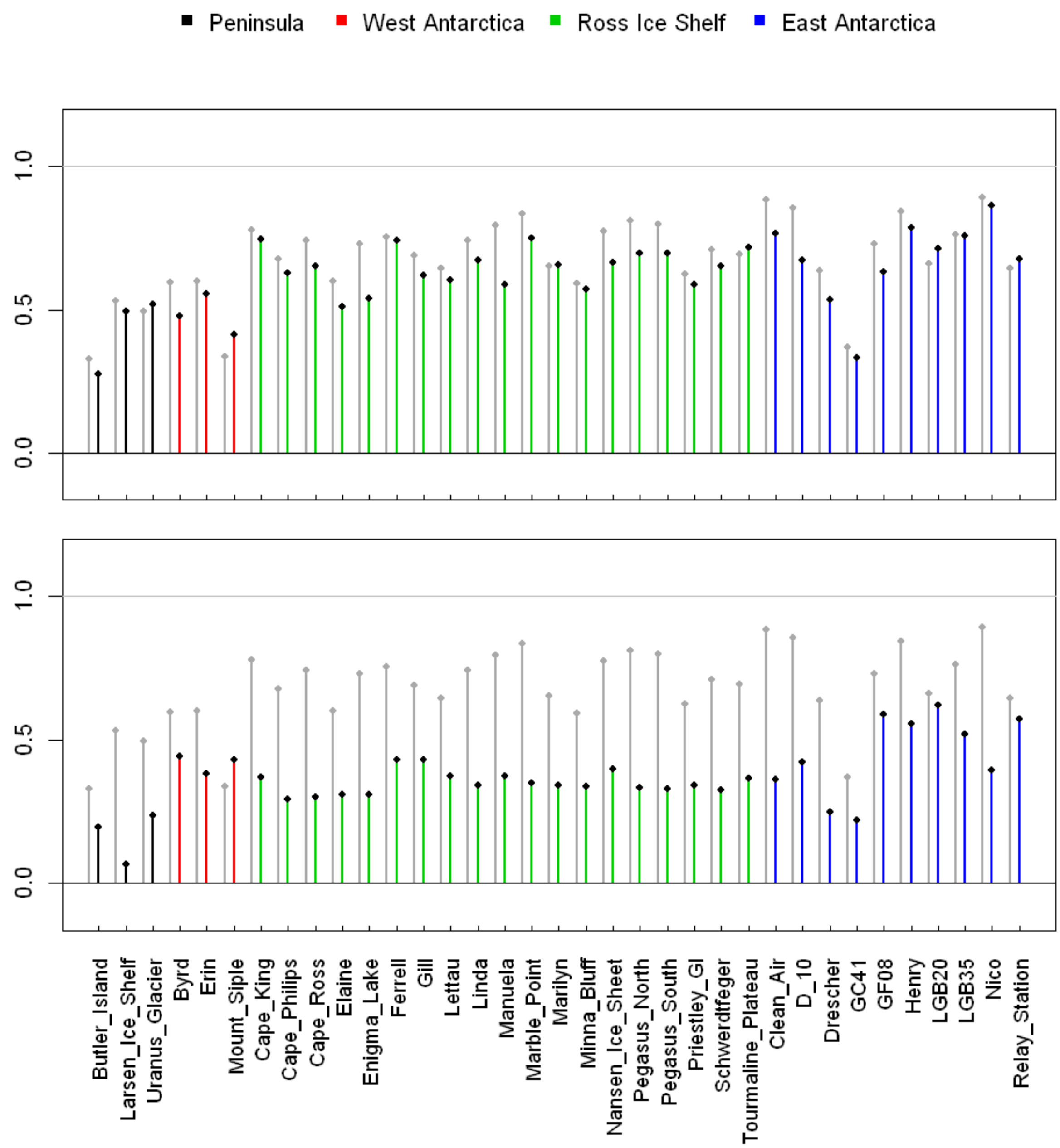

FIG. S23. CE for the 35 additional stations entirely withheld from the 28-station RLS **covariance** ***verification*** reconstruction versus the E-W covariance verification reconstruction (top panel) and the S09 reconstruction (bottom panel). RLS values are in gray; E-W and S09 values are color coded by geographic location.

TABLE S5. Full statistics for the 63-station reconstructions.

	RLS		E-W			
Trend Summary	Correlation[a]	Covariance[a]	Correlation[a]	Covariance[a]	S09[b]	Monte Carlo[c]
Continent	**0.06 ± 0.07**	0.06 ± 0.08	**0.05 ± 0.07**	0.06 ± 0.07	0.12 ± 0.09	-
Peninsula	**0.29 ± 0.10**	0.41 ± 0.12	**0.29 ± 0.08**	0.27 ± 0.08	0.13 ± 0.05	-
West Antarctica	**0.05 ± 0.08**	0.13 ± 0.11	**0.04 ± 0.06**	0.07 ± 0.06	0.20 ± 0.09	-
East Antarctica	**0.05 ± 0.09**	0.02 ± 0.09	**0.04 ± 0.08**	0.04 ± 0.08	0.10 ± 0.10	-

Station Name	R^2 (Station Data Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Amundsen Scott	0.86	0.87	0.82	0.84	*0.36*	-0.77
Arturo Prat	0.93	0.89	0.75	0.67	*0.11*	-0.82
Belgrano I	0.79	0.97	0.43	0.44	*0.18*	-0.63
Belgrano II	0.91	0.79	0.59	0.45	*0.25*	-0.48
Byrd (manned)	0.76	0.84	0.43	0.47	*0.45*	-0.56
Casey	0.86	0.92	0.66	0.61	*0.37*	-0.78
Davis	0.90	0.92	0.74	0.75	*0.44*	-0.79
Dumont Durville	0.92	0.86	0.74	0.64	*0.34*	-0.74
Esperanza	0.92	0.94	0.85	0.85	*0.09*	-0.78
Faraday	0.88	0.84	0.42	0.44	*0.12*	-1.02
Halley	0.77	0.91	0.54	0.50	*0.12*	-0.76
Leningradskaja	0.98	0.87	0.61	0.70	*0.38*	-0.63
Marambio	0.88	0.95	0.82	0.79	*0.06*	-0.75
Mario Zuchelli	0.85	0.84	0.51	0.54	*0.23*	-0.56
Mawson	0.90	0.92	0.73	0.73	*0.45*	-0.79
McMurdo	0.90	0.90	0.78	0.80	*0.41*	-0.76
Mirny	0.88	0.94	0.71	0.70	*0.48*	-0.79
Molodeznaja	0.96	0.86	0.57	0.59	*0.34*	-0.75
Neumayer	0.80	0.88	0.54	0.46	*0.21*	-0.68
Novolazarevskaya	0.93	0.92	0.80	0.70	*0.37*	-0.76
O’Higgins	0.91	0.90	0.81	0.75	*0.09*	-0.80
Rothera	0.80	0.87	0.57	0.52	*0.03*	-0.87
Russkaya	0.84	0.91	0.55	0.43	*0.39*	-0.48
San Martin	0.88	0.88	0.41	0.44	*0.01*	-0.69
Scott Base	0.90	0.96	0.77	0.78	*0.39*	-0.79
Syowa	0.93	0.91	0.78	0.69	*0.27*	-0.78
Vostok	0.96	0.97	0.79	0.84	*0.55*	-0.78
Zhongshan	0.88	0.95	0.71	0.71	*0.39*	-0.58
Butler Island	0.85	0.95	0.50	0.52	0.20	-0.60
Byrd (AWS)	0.76	0.79	0.55	0.51	0.44	-0.61
Cape King	0.95	0.95	0.75	0.76	0.37	-0.57
Cape Phillips	0.91	0.89	0.72	0.66	0.29	-0.51
Cape Ross	0.88	0.89	0.68	0.67	0.30	-0.56
Clean Air	0.96	0.97	0.83	0.82	0.36	-0.58
D10	0.92	0.93	0.73	0.69	0.42	-0.55
Drescher	0.84	0.93	0.52	0.50	0.25	-0.41
Elaine	0.80	0.96	0.53	0.53	0.31	-0.50
Enigma Lake	0.92	0.93	0.61	0.64	0.31	-0.44
Erin	0.94	0.95	0.53	0.58	0.38	-0.15

TABLE S5. (continued)

	RLS		E-W			
	Correlation	Covariance	Correlation	Covariance	S09[b]	Monte Carlo[c]
Station Name	R^2 (Station Data Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Ferrell	0.84	0.95	0.74	0.73	0.43	-0.62
GC41	0.89	0.98	0.46	0.50	0.22	-0.65
GF08	0.89	0.95	0.71	0.65	0.59	-0.59
Gill	0.87	0.98	0.66	0.71	0.43	-0.57
Henry	0.99	0.98	0.84	0.85	0.56	-0.43
LGB20	0.96	0.95	0.76	0.78	0.63	-0.45
LGB35	0.95	0.95	0.80	0.79	0.52	-0.56
Larsen Ice Shelf	0.92	0.95	0.57	0.56	0.07	-0.54
Lettau	0.87	0.73	0.62	0.63	0.37	-0.46
Linda	0.82	0.92	0.67	0.68	0.22	-0.42
Manuela	0.91	0.99	0.67	0.66	0.34	-0.62
Marble Point	0.91	0.95	0.78	0.77	0.35	-0.65
Marilyn	0.81	0.90	0.65	0.65	0.34	-0.52
Minna Bluff	0.80	0.67	0.65	0.58	0.34	-0.37
Mount Siple	0.93	0.88	0.63	0.42	0.43	-0.47
Nansen Ice Sheet	0.88	0.88	0.68	0.67	0.40	-0.51
Nico	0.98	0.99	0.92	0.91	0.39	-0.44
Pegasus North	0.86	0.95	0.71	0.71	0.33	-0.39
Pegasus South	0.84	0.94	0.70	0.69	0.33	-0.48
Priestley Glacier	0.86	0.83	0.70	0.66	0.34	-0.55
Relay Station	0.99	0.98	0.69	0.73	0.57	-0.42
Schwerdtfeger	0.84	0.95	0.66	0.66	0.32	-0.58
Tourmaline Plateau	0.92	0.87	0.77	0.77	0.36	-0.53
Uranus Glacier	0.90	0.95	0.60	0.59	0.23	-0.43
Station Name	CE (Station Data Not Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Adelaide	0.75	0.79	0.52	0.50	0.27	-0.53
Asuka	0.79	0.77	0.61	0.66	0.39	-0.14
Bonaparte Point	0.83	0.76	0.50	0.46	**-0.06**	-0.40
Cape Denison	0.76	0.72	0.70	0.61	0.31	-0.25
Casey Airstrip	0.80	0.88	0.66	0.62	**0.06**	0.61
D47	0.82	0.80	0.71	0.64	0.43	-0.17
D57	0.69	0.71	0.59	0.54	0.32	0.09
D80	0.76	0.79	0.68	0.65	0.63	0.16
Dome C II	0.54	0.54	0.62	0.50	0.50	-0.46
Doug	0.75	0.82	0.53	0.54	0.55	-0.10
Elizabeth	0.76	0.74	0.52	0.62	0.40	-0.22
GEO3	0.82	0.80	0.71	0.71	0.54	-0.36
Harry	0.80	0.77	0.59	0.59	0.48	-0.21
LGB10	0.89	0.88	0.66	0.68	0.55	-0.28
LGB59	0.70	0.68	0.63	0.65	0.39	-0.56

TABLE S5. (continued)

	RLS		E-W			
	Correlation	Covariance	Correlation	Covariance	S09[b]	Monte Carlo[c]
Station Name	CE (Station Data Not Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
LGB10	0.89	0.88	0.66	0.68	0.55	-0.28
LGB59	0.70	0.68	0.63	0.65	0.39	-0.56
Law Dome Summit	0.70	0.73	0.61	0.54	0.39	-0.36
Limbert	0.67	0.71	0.41	0.38	0.22	-0.35
Lynn	0.65	0.72	0.72	0.68	0.53	-0.30
Penguin Point	0.79	0.69	0.63	0.49	0.30	-0.23
Port Martin	0.82	0.83	0.70	0.56	0.47	-0.30
Santa Claus Island	0.78	0.81	0.68	0.63	0.16	0.07
Siple	0.16	0.17	0.32	0.27	0.46	-0.31
Sutton	0.81	0.77	0.57	0.35	<u>**0.12**</u>	0.30
Theresa	0.78	0.76	0.50	0.53	0.42	-0.27
MEAN R^2	**0.89**	0.91	**0.67**	0.65	0.33	-0.60
MEAN CE	**0.73**	0.74	**0.60**	0.57	0.38	-0.21
MEAN COMBINED	**0.84**	0.86	**0.65**	0.63	0.34	-0.49
MEAN r	**0.95**	0.93	**0.85**	0.85	0.59	-
MEAN r^2	**0.90**	0.88	**0.73**	0.73	0.37	-

[a] Trends and summary statistics bolded correspond to the reconstructions used in the main text.

[b] Values in *italics* indicate stations that were used in the S09 reconstructions, which designates those statistics shown as R^2 (average explained variance). All other statistics for S09 are CE values.

[c] 1,000 Monte Carlo simulations using the mean, variance, and lag-1 autocorrelation coefficient from the station data were conducted for each station. Numbers listed correspond to the 99th percentile. Values <u>bolded and underlined</u> designate stations for which CE / R^2 values are negative or do not exceed the results of the Monte Carlo simulations.

[d] R^2 denotes *average explained variance* (Cook et al. 1999) and is computed as $1-\sum(x_i-\hat{x}_i)^2/\sum(x_i-\bar{x}_c)^2$, where x is the original data and $\hat{x}$ is the estimated data in the calibration period. CE is *coefficient of efficiency* (Cook et al. 1999) as is computed identically, except that all values are taken from the verification period. Thus, stations that were used in the reconstructions utilize R^2 as the measure of explained variance, and stations that were withheld from the reconstruction utilize CE.

TABLE S6. Full statistics for the 28-station verification reconstructions.

	RLS		E-W			
Trend Summary[a]	Correlation[a]	Covariance[a]	Correlation[a]	Covariance[a]	S09[b]	Monte Carlo[c]
Continent	**0.06 ± 0.07**	0.05 ± 0.08	**0.05 ± 0.07**	0.05 ± 0.07	0.12 ± 0.09	-
Peninsula	**0.35 ± 0.11**	0.43 ± 0.12	**0.32 ± 0.08**	0.27 ± 0.08	0.13 ± 0.05	-
West Antarctica	**0.11 ± 0.08**	0.14 ± 0.11	**0.06 ± 0.06**	0.07 ± 0.06	0.20 ± 0.09	-
East Antarctica	**0.03 ± 0.08**	0.01 ± 0.09	**0.03 ± 0.08**	0.04 ± 0.08	0.10 ± 0.10	-

Station Name	R^2 (Station Data Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Amundsen Scott	0.97	0.99	0.80	0.84	*0.36*	-0.77
Arturo Prat	0.94	0.90	0.64	0.68	*0.11*	-0.82
Belgrano I	0.79	0.97	0.41	0.46	*0.18*	-0.63
Belgrano II	0.94	0.83	0.65	0.56	*0.25*	-0.48
Byrd (manned)	0.86	0.98	0.61	0.47	*0.45*	-0.56
Casey	0.87	0.93	0.66	0.63	*0.37*	-0.78
Davis	0.90	0.94	0.69	0.76	*0.44*	-0.79
Dumont Durville	0.95	0.91	0.67	0.73	*0.34*	-0.74
Esperanza	0.95	0.95	0.81	0.84	*0.09*	-0.78
Faraday	0.91	0.89	0.31	0.44	*0.12*	-1.02
Halley	0.78	0.93	0.52	0.52	*0.12*	-0.76
Leningradskaja	0.99	0.93	0.60	0.69	*0.38*	-0.63
Marambio	0.88	0.96	0.76	0.82	*0.06*	-0.75
Mario Zuchelli	0.96	0.94	0.66	0.55	*0.23*	-0.56
Mawson	0.90	0.94	0.76	0.74	*0.45*	-0.79
McMurdo	0.93	0.96	0.79	0.78	*0.41*	-0.76
Mirny	0.88	0.95	0.73	0.71	*0.48*	-0.79
Molodeznaja	0.97	0.92	0.50	0.61	*0.34*	-0.75
Neumayer	0.80	0.90	0.55	0.49	*0.21*	-0.68
Novolazarevskaya	0.94	0.93	0.77	0.71	*0.37*	-0.76
O'Higgins	0.92	0.91	0.75	0.75	*0.09*	-0.80
Rothera	0.81	0.89	0.54	0.53	*0.03*	-0.87
Russkaya	0.86	0.95	0.52	0.53	*0.39*	-0.48
San Martin	0.90	0.90	0.33	0.46	*0.01*	-0.69
Scott Base	0.90	0.96	0.79	0.79	*0.39*	-0.79
Syowa	0.94	0.93	0.71	0.72	*0.27*	-0.78
Vostok	0.98	0.98	0.90	0.88	*0.55*	-0.78
Zhongshan	0.89	0.97	0.66	0.72	*0.39*	-0.58

Station Name	CE (Station Data Not Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Adelaide	0.77	0.81	0.52	0.53	0.27	-0.53
Asuka	0.73	0.73	0.64	0.66	0.39	-0.14
Bonaparte Point	0.86	0.83	0.31	0.52	**<u>-0.06</u>**	-0.40
Butler Island	0.33	0.41	0.28	0.35	0.20	-0.60
Byrd (AWS)	0.60	0.31	0.48	0.51	0.44	-0.61
Cape Denison	0.77	0.72	0.69	0.66	0.31	-0.25

TABLE S6. (continued)

Station Name	RLS Correlation	RLS Covariance	E-W Correlation	E-W Covariance	S09[b]	Monte Carlo[c]
	CE (Station Data Not Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Cape King	0.78	0.77	0.75	0.68	0.37	-0.57
Cape Phillips	0.68	0.65	0.63	0.55	0.29	-0.51
Cape Ross	0.74	0.75	0.65	0.63	0.30	-0.56
Casey Airstrip	0.78	0.86	**0.56**	**0.57**	**0.06**	0.61
Clean Air	0.88	0.89	0.77	0.79	0.36	-0.58
D10	0.86	0.85	0.67	0.74	0.42	-0.55
D47	0.79	0.77	0.71	0.71	0.43	-0.17
D57	0.63	0.61	0.61	0.58	0.32	0.09
D80	0.73	0.73	0.67	0.71	0.63	0.16
Dome C II	0.57	0.55	0.62	0.57	0.50	-0.46
Doug	0.55	0.65	0.56	0.51	0.55	-0.10
Drescher	0.63	0.76	0.53	0.51	0.25	-0.41
Elaine	0.60	0.68	0.51	0.52	0.31	-0.50
Elizabeth	0.57	0.34	0.59	0.52	0.40	-0.22
Enigma Lake	0.73	0.72	0.54	0.53	0.31	-0.44
Erin	0.60	0.69	0.55	0.57	0.38	-0.15
Ferrell	0.76	0.81	0.74	0.71	0.43	-0.62
GC41	0.37	0.42	0.33	0.43	0.22	-0.65
GEO3	0.79	0.78	0.69	0.71	0.54	-0.36
GF08	0.73	0.79	0.63	0.64	0.59	-0.59
Gill	0.69	0.68	0.62	0.60	0.43	-0.57
Harry	0.61	0.58	0.57	0.56	0.48	-0.21
Henry	0.84	0.86	0.79	0.82	0.56	-0.43
LGB10	0.66	0.68	0.67	0.63	0.55	-0.28
LGB20	0.66	0.71	0.71	0.70	0.63	-0.45
LGB35	0.76	0.78	0.76	0.73	0.52	-0.56
LGB59	0.63	0.64	0.58	0.64	0.39	-0.56
Larsen Ice Shelf	0.53	0.57	0.49	0.52	0.07	-0.54
Law Dome Summit	0.69	0.73	0.62	0.60	0.39	-0.36
Lettau	0.65	0.68	0.60	0.54	0.37	-0.46
Limbert	0.37	0.37	0.30	0.31	0.22	-0.35
Linda	0.74	0.80	0.67	0.64	0.22	-0.42
Lynn	0.70	0.70	0.70	0.61	0.53	-0.30
Manuela	0.79	0.76	0.59	0.60	0.34	-0.62
Marble Point	0.84	0.88	0.75	0.74	0.35	-0.65
Marilyn	0.65	0.72	0.66	0.63	0.34	-0.52
Minna Bluff	0.59	0.49	0.57	0.61	0.34	-0.37
Mount Siple	0.33	0.18	0.41	0.29	0.43	-0.47
Nansen Ice Sheet	0.77	0.76	0.66	0.61	0.40	-0.51
Nico	0.89	0.90	0.86	0.87	0.39	-0.44
Pegasus North	0.81	0.88	0.70	0.70	0.33	-0.39
Pegasus South	0.80	0.87	0.70	0.69	0.33	-0.48

TABLE S6. (continued)

	RLS		E-W			
	Correlation	Covariance	Correlation	Covariance	S09[b]	Monte Carlo[c]
Station Name	CE (Station Data Not Used in Reconstruction)[c,d]				(R^2 or CE)[c,d]	(R^2 or CE)[c,d]
Penguin Point	0.75	0.61	0.63	0.54	0.30	-0.23
Port Martin	0.82	0.78	0.56	0.66	0.47	-0.30
Priestley Glacier	0.62	0.54	0.59	0.56	0.34	-0.55
Relay Station	0.64	0.66	0.68	0.62	0.57	-0.42
Santa Claus Island	0.76	0.77	0.61	0.63	0.16	0.07
Schwerdtfeger	0.71	0.75	0.65	0.63	0.32	-0.58
Siple	0.05	0.22	0.27	0.27	0.46	-0.31
Sutton	0.80	0.71	0.48	0.51	**0.12**	0.30
Theresa	0.54	0.63	0.52	0.56	0.42	-0.27
Tourmaline Plateau	0.69	0.64	0.72	0.67	0.36	-0.53
Uranus Glacier	0.50	0.64	0.52	0.56	0.23	-0.43
MEAN R^2	**0.90**	0.93	**0.65**	0.66	0.28	-0.72
MEAN CE	**0.67**	0.68	**0.61**	0.60	0.37	-0.38
MEAN COMBINED	**0.75**	0.76	**0.62**	0.62	0.34	-0.49
MEAN r	**0.89**	0.88	**0.82**	0.84	0.59	-
MEAN r^2	**0.80**	0.78	**0.68**	0.71	0.37	-

[a] Trends and summary statistics bolded correspond to the reconstructions used in the main text.

[b] Values in *italics* indicate stations that were used in the S09 reconstructions, which designates those statistics shown as R^2 (average explained variance). All other statistics for S09 are CE values.

[c] 1,000 Monte Carlo simulations using the mean, variance, and lag-1 autocorrelation coefficient from the station data were conducted for each station. Numbers listed correspond to the 99th percentile. Values **bolded and underlined** designate stations for which CE / R^2 values are negative or do not exceed the results of the Monte Carlo simulations.

[d] R^2 denotes *average explained variance* (Cook et al. 1999) and is computed as $1 - \sum (x_i - \hat{x}_i)^2 / \sum (x_i - \bar{x}_c)^2$, where x is the original data and $\hat{x}$ is the estimated data in the calibration period. CE is *coefficient of efficiency* (Cook et al. 1999) as is computed identically, except that all values are taken from the verification period. Thus, stations that were used in the reconstructions utilize R^2 as the measure of explained variance, and stations that were withheld from the reconstruction utilize CE.

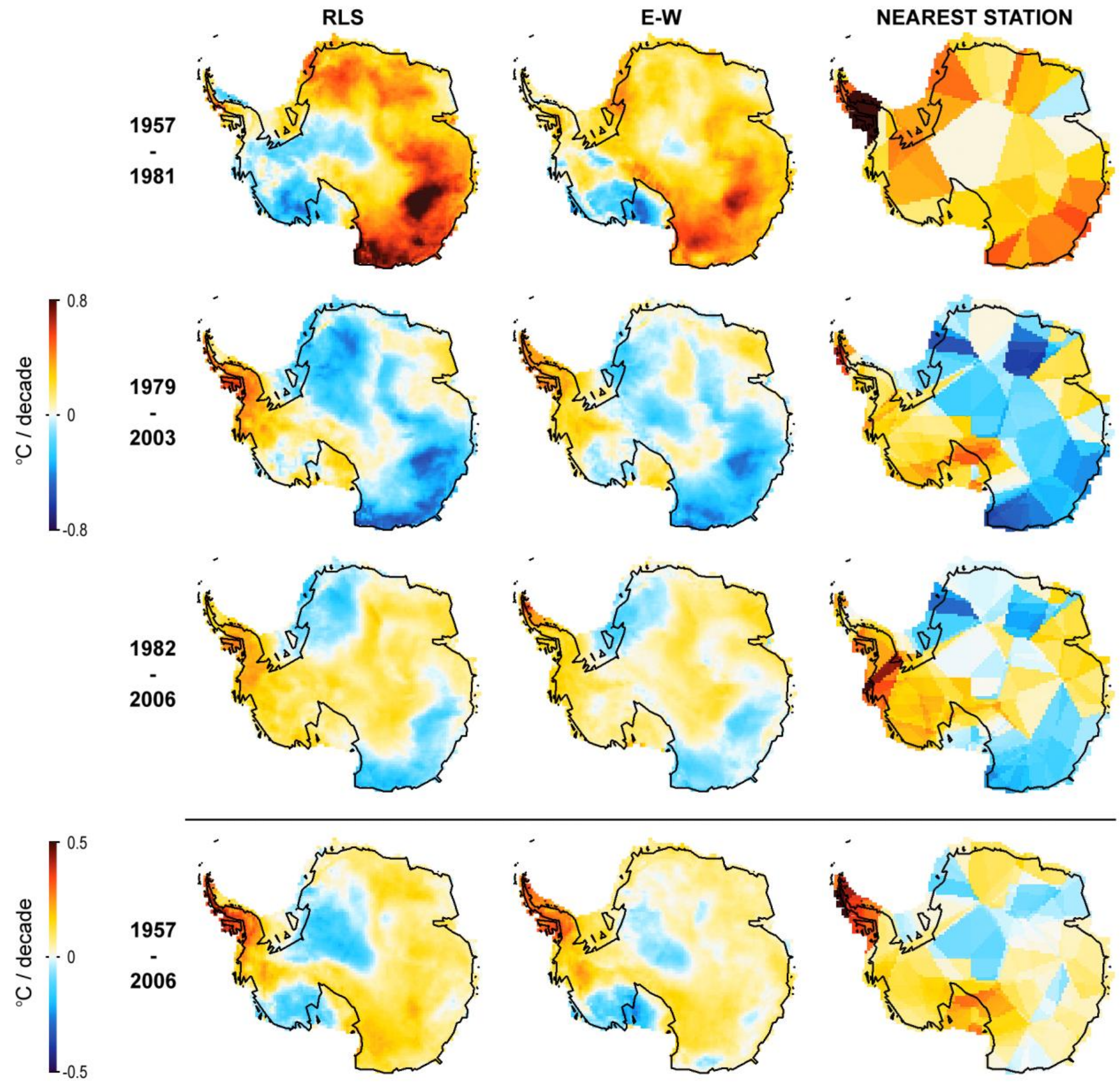

FIG. S24. Comparison of spatial distribution of trends for RLS correlation, E-W correlation, and nearest-station reconstructions.

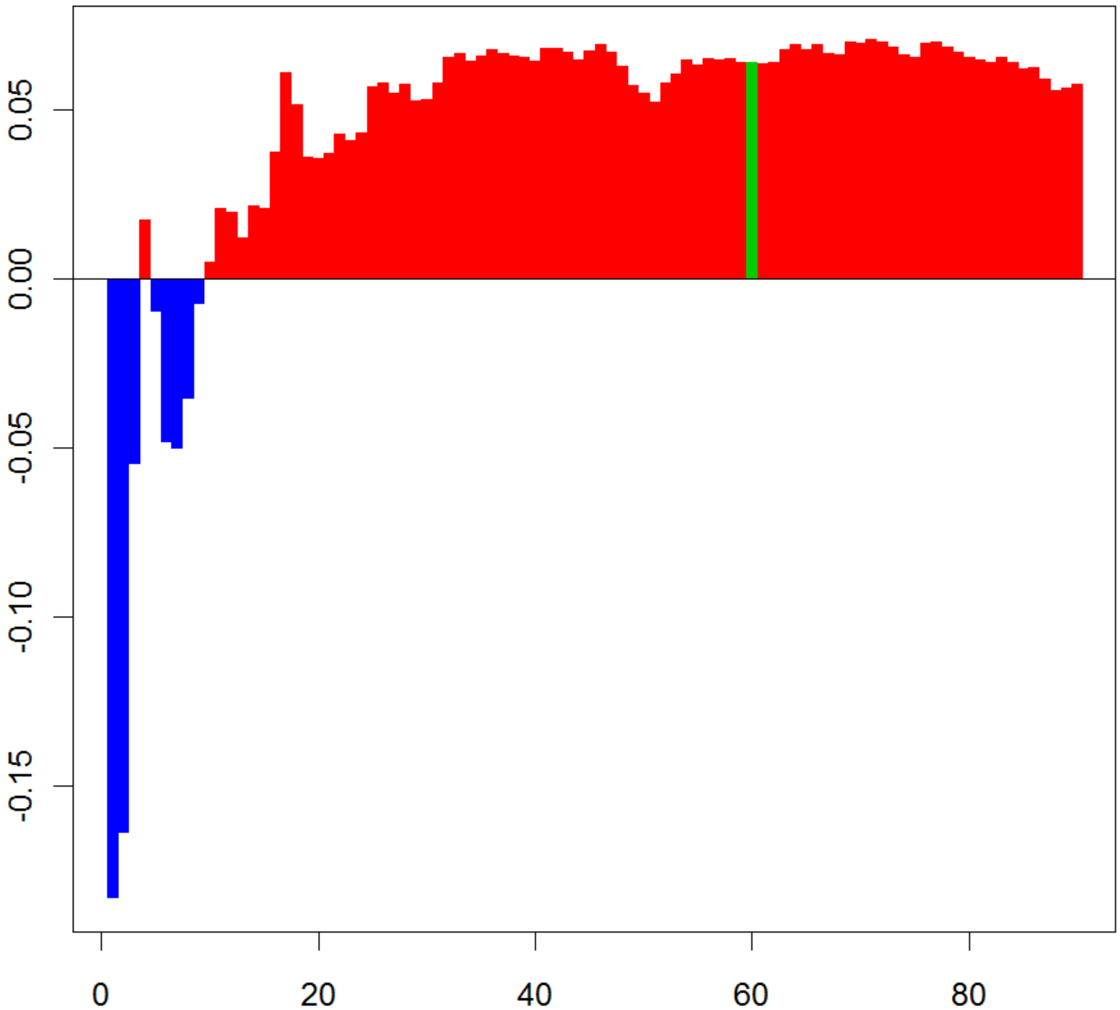

FIG. S25. Continental trend vs. maximum months-of-overlap for determining station offsets for the nearest-station reconstructions. The green bar indicates the trend when using 60 months of overlap.

References

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

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

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

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

Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleaveland, 1999: Drought Reconstructions for the Continental United States. *J. Climate*, **12**, 1145–1162, doi:10.1175/1520-0442(1999)012<1145:DRFTCU>2.0.CO;2

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

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

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

Peterson, T. C., R. Vose, R. Schmoyer, and V. Razuvaëv, 1998: Global historical climatology network (GHCN) quality control of monthly temperature data, *Int. J. Clim.*, **11**, 1169–1179, doi:10.1002/(SICI)1097-0088(199809)18:11<1169::AID-JOC309>3.0.CO;2-U

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

Shepherd, D., 1999: Catalogue of Antarctic Climate Data. *Executive Council Working Group on Antarctic Meteorology, World Meteorological Organization*, (Preliminary Report from the Australian Bureau of Meteorology). [Available on-line at http://www.wmo.int/pages/prog/www/Antarctica/Catalogue-Antarctic.doc]

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

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