AVHRR Data

Accuracy in the retrieval of ice sheet surface temperatures from satellite infrared data depends on successful cloud masking, which is challenging because of the low contrast in the albedo and emissivity of clouds and the surface. In *Comiso* (ref. 8), cloud masking was done by a combination of channel differencing and daily differencing, based on the change in observed radiances due to the movement of clouds. Here, we use an additional masking technique in which daily data that differ from the climatological mean by more than some threshold value are assumed to be cloud contaminated and are removed. We used a threshold value of 10°C, which produces the best validation statistics in the reconstruction procedure. A slightly different data set (based on a threshold value of 6°C) is reported in *Monaghan et al.* (ref. 7). Differences between the linear trends in the different versions of the AVHRR data are shown in Figure S1. The very strong trends in the earlier data set (ref. 8) are not supported by the ground-based observations. In particular, the very strong cooling trends (>1°C/decade) and the strong warming at ~30° to ~ 60°E in East Antarctica results in poor reconstruction statistics. Using any of the three slightly different data sets produces essentially the same results discussed in the text (statistically significant warming in West Antarctica throughout the 1957-2006 interval, and smaller but still generally significant warming in East Antarctica in the annual mean).

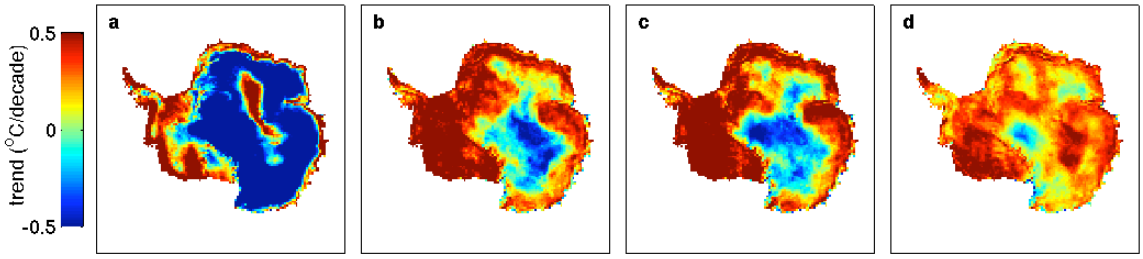

Figure S1. Comparison of linear trends for 1982-1999 from the AVHRR data processed with the cloud masking as used in **a**) Comiso (ref. 8), **b**) Monaghan et al. (ref. 7) and **c**) this paper. **d**) shows the same as (c) but for 1982-2006.

Verification Statistics

Likelihoods for given r^2, RE, and CE values arising by chance in the split calibration/verification experiments were calculated from 1000 Monte Carlo simulations for each time series at each grid point. RE and CE are defined as:

$$RE = 1 - \sum(x_i - \hat{x}_i)^2 / \sum(x_i - \bar{x}_C)^2$$

$$CE = 1 - \sum(x_i - \hat{x}_i)^2 / \sum(x_i - \bar{x}_V)^2$$

where x_i is the original data and $\hat{x}_i$ the reconstructed data for month i of the verification period, $\bar{x}_V$ is the mean of the original data in the verification period, and $\bar{x}_C$ is the mean of the original data in the calibration period. Each simulation consists of a red noise model for each grid point generated using normally distributed random numbers (Gaussian white noise) and the lag-1 autocorrelation for the time series at that grid point, calculated over the calibration interval. The 99th percentile value for each grid point is taken from the set of 1000 simulations. This value must be exceeded for the reconstruction to be less than 1% likely to have arisen by a chance realization of a red noise process. In all locations (Figure S2), the 99th percentile value is exceeded in all three tests (i.e. r^2, RE, and CE) for both early (1982-1994.5) and late (1994.5-2006) verification periods. For RE and CE, some locations yield verification scores <0. While there are sufficient degrees of freedom that these can still be considered statistically significant, an improved estimate over the climatological mean cannot be demonstrated at these locations. At all other locations, the results (RE, CE > 0) demonstrate improvement over both calibration period and verification period climatology.

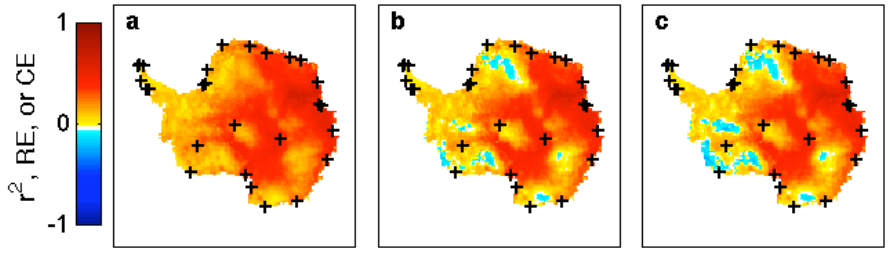

Figure S2. **a)** Difference between the minimum r^2 values obtained from the split calibration/verification experiments and the maximum 99th percentile values obtained from Monte Carlo red noise simulations. **b)** Minimum RE values obtained from the split calibration/verification calculations. **c)** Minimum CE values obtained from the split calibration/verification calculations. Crosses (+) show the locations of the weather station data from the READER database.

Reconstructed temperatures from automatic weather station data

Near surface (2-3 m) air temperature data are available from the READER database from 65 automatic weather stations (AWS) beginning in 1980. Like the T_{IR} satellite data, the AWS data have potential biases, including the possibility of variations in the height of the temperature sensors; they are also generally discontinuous, and data gaps are a particularly large problem in winter. Nevertheless, the AWS data provide a useful alternative reconstruction target to the T_{IR} data. We apply the same method as for the T_{IR}-based reconstruction to the AWS-data, using RegEM with $k = 3$ and using the READER occupied weather station temperature data to produce a reconstruction of temperature at each AWS site for the period 1957-2006. At many sites, the AWS data are insufficient in length to demonstrate reconstruction skill. Nevertheless verification tests indicate statistically skillful results at 26 sites (Table S1). As shown in Figure 2 in the main text, the results averaged over West and East Antarctica are nearly identical to those from the T_{IR}-based reconstruction. Figure S3 shows the spatial and seasonal pattern of the AWS-based reconstruction (cf. Figure 3 in the main text). Cooling at some sites in East Antarctica is more extensive and larger in magnitude than indicated by the T_{IR} data; however, trends are not significant, and verification skill is generally poor at those sites. Results for West Antarctica, however, agree very well with those for the T_{IR} reconstruction, showing warming over West Antarctica during all seasons, strongest in spring and winter.

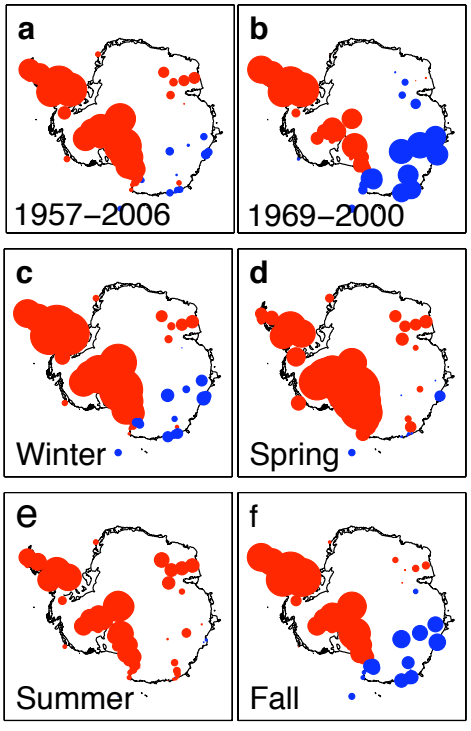

Figure S3. Spatial pattern of temperature trends (°C/decade) from reconstruction using AWS data. **a)** Mean annual trends for 1957-2006. **b**) Mean annual trends for 1969-2000, to facilitate comparison with ref. (2). **d-f**) Seasonal trends for 1957-2006. Winter = JJA, Spring = SON; Summer = DJF; Fall = MAM. Colored dots show negative (blue) and positive (red) trends at each AWS location scaled by magnitude. (To facilitate comparison with Figure 3, all non-zero trends are shown; in general only those sites with trends exceeding 0.15°C/decade are significant (2-tailed *t*-test)).

Additional Reconstructions

We obtained separate reconstructions of the T_{IR} satellite data using 1) conventional principal component analysis (PCA) rather than RegEM, and 2) using RegEM after linear detrending of the T_{IR} data. In the PCA, missing values in the weather station data were filled in with mean monthly climatology. We also repeated the calculations using a subset of the data. In the full reconstructions, we used the complete weather station data (1957-2006) from all 42 of the READER database locations listed in Table S2. In a subset reconstruction, we used only those 15 data sets (denoted by asterisks in Table S2) with the most available data coverage since 1957. The remaining 27 occupied stations, plus automatic weather station data, provide additional validation of our results; their correlations with the reconstruction based on monthly average values are shown in Table S2 (*r*, RE, and CE) and Figure S3 (*r* only). Strictly speaking RE is not defined because the excluded weather stations are incomplete over the calibration period. To calculate RE scores in this case, we used the calibration-period mean of the satellite data (or the reconstruction, if pre-1982) at the same location as the weather stations in the denominator of the RE calculation. Thus, the station data and the satellite data at the same location would have the same calibration period mean.

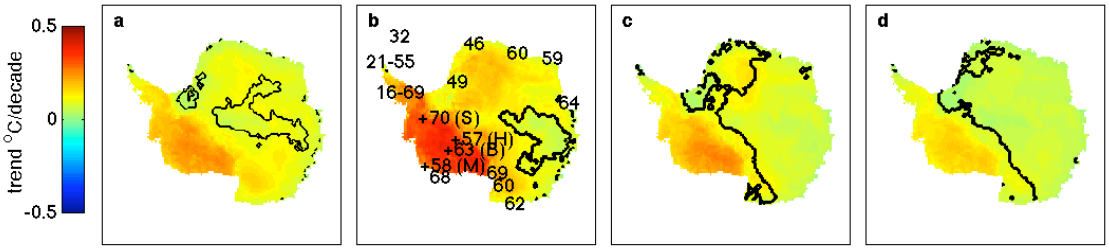

Figure S4. Comparison of linear trends in reconstructed surface temperatures for 1957-2006 obtained using **a**) RegEM as in Figure 3a. **b)** REGEM but using 15 rather than 42 weather station data from the READER data set as predictors; **c**) conventional PCA, using 3 principal components and 15 predictors, and **d**) RegEM, using 42 predictors but after linearly detrending the AVHRR data. Black lines separate areas of significant vs. insignificant trends (>95% confidence based on two-tailed *t*-test with number of degrees of freedom adjusted for temporal autocorrelation). The numbers shown in panel (b) are the correlation coefficients (x100) for the monthly temperatures from the 24 READER weather station data south of 60°S that were excluded from the 15-predictor RegEM calculation, and for Siple Station (S), Byrd (B), Mount Siple (M) and Harry (H) automatic weather stations, whose locations are denoted by crosses (+).

Modelling Results

In the main text, Figure 4 shows the mean trends in surface temperature from ensemble means of GISS Model E atmospheric model runs, with different forcing and boundary conditions[22]. Figure S5 shows additional results from GISS Model E. The left two panels show ensemble means using observed[23] SST and climatological sea ice boundary conditions, and including known radiative forcings. These results were subtracted from the ensemble means that used observed SST and observed sea ice and forcings (panels c and g in Figure 4, main text) to obtain the 'sea ice only' results shown in Figure 4 (panels d and h). This calculation provides an estimate of the influence of sea ice changes alone on the model results. Although it would be ideal to re-run GISS Model E with climatological SST, observed sea ice and no forcings, this would be prohibitively expensive and very unlikely to change the results significantly. The right two panels show the subtraction of ensemble means with the same boundary conditions (observed SST and sea ice), but with and without atmospheric forcings. This calculation provides an estimate of the influence of radiative forcings that take place solely via atmospheric response and not via changes in the ocean. As noted in the main text, forcing alone leads to cooling over both West and East Antarctica, especially during 1979-2003, due to the influence of stratospheric ozone and greenhouse gases on the SAM[17].

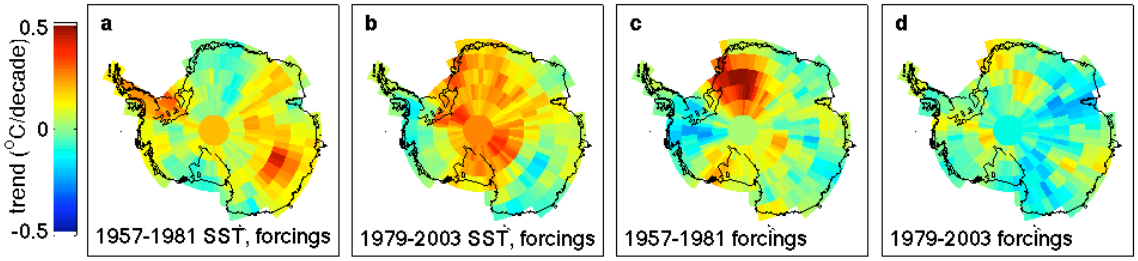

Figure S5. Modeled mean annual temperature trends (°C/decade) for the first and last twenty-five years of the period 1957-2003. **a,b)** Surface air temperature from 4-member GISS Model E ensemble simulations for 1957-1981 and 1979-2003, with observed SST and climatological sea ice boundary conditions, plus atmospheric forcings (changes in atmospheric concentrations of radiatively active species, including ozone). **c,d**) Difference between simulations with observed SST and sea ice boundary conditions but with and without atmospheric forcings, to isolate the effect of forcings alone.

Station Name	Latitude (°S)	Longitude (°E)	*r*	RE	CE	trend (°C/decade)
Butler Island	72.2	299.8	0.45	0.28	0.17	0.45
Cape King	73.6	166.6	0.80	0.50	0.54	0.13
Cape Philips	73.1	169.6	0.73	0.43	0.45	0.10
Cape Ross	76.7	163	0.85	0.69	0.69	0.22
Clean Air	90	360	0.54	0.26	0.22	0.28
D10	66.7	139.8	0.58	0.37	0.27	0.06
Elaine	83.1	174.2	0.78	0.60	0.59	0.37
Enigma Lake	74.7	164	0.76	0.60	0.52	0.15
Ferrel	77.9	170.8	0.88	0.77	0.75	0.25
GC41	71.6	111.3	0.64	0.37	0.37	0.08
Gill	80	181.4	0.77	0.50	0.51	0.08
LGB120	73.8	55.7	0.80	0.62	0.62	0.06
LGB35	76	65	0.85	0.69	0.70	0.05
Larsen Ice Shelf	66.9	299.1	0.59	0.22	0.28	0.36
Lettau	82.5	185.6	0.80	0.60	0.59	0.19
Manuela	74.9	163.7	0.75	0.44	0.48	0.07
Marble Point	77.4	163.7	0.89	0.72	0.74	0.22
Marily	80	165.1	0.77	0.59	0.57	0.26
Nansen Ice Sheet	74.8	163.3	0.74	0.41	0.46	0.08
Nico	89	89.7	0.59	0.06	0.22	0.20
Pegaus South	78	166.6	0.84	0.70	0.69	0.26
Priestley Gl	74.3	163.2	0.53	0.12	0.12	0.01
Schwerdtfeger	79.9	170	0.81	0.47	0.52	0.16
Terra Nova Bay	74.7	164.1	0.70	0.44	0.43	0.18
Tourm. Plateau	74.1	163.4	0.62	0.09	0.15	0.08
Whitlock	76.2	168.4	0.85	0.60	0.63	0.12

Table S1. Verification statistics (mean correlation coefficients (*r*), reduction of error (RE) and coefficient of efficiency (CE) statistics, from pre and post-1995 calibration/verification calculations) for the AWS-based reconstruction, for those 26 automatic weather stations where sufficient calibration data exist (data >40% complete) and verification skill is acceptable at >95% confidence. The last column shows the 1957-2006 trend calculated for each site in the full reconstruction (as shown in Figure S3).

Station Name	Latitude (°S)	Longitude (°E)	*r*	RE	CE
Adelaide	67.8	292.1	0.65	0.20	0.20
Amundsen_Scott*	90	0	--	--	--
Arturo_Prat	62.5	300.3	0.46	0.20	0.18
Asuka	71.5	24.1	0.60	0.34	0.34
Belgrano_I*	78	321.2	--	--	--
Belgrano_II	77.9	325.4	0.49	0.25	0.22
Bellingshausen	62.2	301.1	0.40	0.15	0.13
Byrd*	80	240	--	--	--
Campbell	52	169	*-0.22*	*-2.29*	*-2.16*
Casey*	66.3	110.5	--	--	--
Davis*	68.6	78	--	--	--
Deception	63	299.3	0.65	0.27	0.30
Dumont_Durville*	66.7	140	--	--	--
Esperanza	63.4	303	0.41	0.12	0.11
Faraday*	65.4	295.6	--	--	--
Ferraz	62.1	301.6	0.33	0.10	0.10
Great_Wall	62.2	301	0.27	0.08	0.07
Grytviken	54.3	323.5	*0.03*	*-0.53*	*-0.48*
Halley*	75.5	333.6	--	--	--
Jubany	62.2	301.4	0.21	0.05	0.04
King_Sejong	62.2	301.3	0.30	0.09	0.08
Leningradskaja	69.5	159.4	0.55	0.29	0.29
Macquarie	54.5	158.9	*-0.04*	*-1.68*	*-1.65*
Marambio	64.2	303.3	0.32	0.07	0.07
Marsh	62.4	301.1	0.45	0.16	0.16
Mawson*	67.6	62.9	--	--	--
McMurdo*	77.9	166.7	--	--	--
Mirny*	66.5	93	--	--	--
Molodeznaja	67.7	45.9	0.54	0.29	0.29
Neumayer	70.7	351.6	0.46	0.22	0.22
Novolazarevskaya*	70.8	11.8	--	--	--
O_Higgins	63.3	302.1	0.43	0.15	0.14
Orcadas	60.7	315.3	0.45	0.15	0.15
Rothera	67.5	291.9	0.35	0.13	0.11
Russkaya	74.8	223.1	0.53	0.25	0.25
San_Martin	68.1	292.9	*0.12*	0.00	-0.02
Scott_Base	77.9	166.7	0.65	0.38	0.39
Signy*	60.7	314.4	--	--	--
Syowa*	69	39.6	--	--	--
Terra_Nova_Bay	74.7	164.1	0.60	0.33	0.33
Vostok*	78.5	106.9	--	--	--
Zhongshan	69.4	76.4	0.64	0.40	0.40
Mt. Siple (AWS)	73.2	232.9	0.58	0.31	0.31
Siple (AWS)	75.9	276	0.70	0.44	0.44
Byrd (AWS)	80	240	0.64	0.39	0.39
Harry (AWS)	73.2	232.9	0.54	0.27	0.27

Table S2. List of the 42 occupied weather stations used as predictor variables, and four automatic weather stations (AWS) in West Antarctica. Weather stations used in the restricted 15-predictor reconstruction are marked with an asterix (*). Correlation coefficients (*r*), reduction of error (RE) and coefficient of efficiency (CE) statistics shown are for the monthly correlation between the weather station data withheld and the 15-predictor T_{IR} reconstruction at the corresponding location. Values that are not significant at better than 99% are italicized.