Supporting Information for "Arctic sea ice variability during the Instrumental Era"

M. Kathleen Brennan [1], Gregory J. Hakim[1], and Edward Blanchard-Wrigglesworth[1],

[1]Department of Atmospheric Sciences, University of Washington

Contents of this file

1. Data availability in Walsh et al 2017 (S1)
2. The early 20th century warming in reanalysis (S2)
3. Sensitivity of sea ice extent to temperature in our reconstructions (S3)
4. Sensitivity of results (S4 – S6)
5. Verification Statistics (Equation 1)

1. Data availability in Walsh et al 2017

Walsh et al. (2017) uses sea ice observations from a ranked list of 12 different sources, with priority given to the available highest-ranked source. When none of these sources are available at a given time, temporal interpolation (for a single month of missing data)

Corresponding author: M. Kathleen Brennan, Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA (mkb22@uw.edu)

or analog-based methods to fill in missing data are used (for periods with more than one month missing). In Figure S1 we plot the percentage of longitude ocean grid cells with an observation available for each month separated into two seasons from 1850–2013. Before March of 1953 (vertical green line) there is very little spatial coverage of sea ice observations in the winter months (September–March). Data coverage in the summer months (April–August) is also very low (<40% on average) before May of 1901 and with better coverage between 1902–1953 except for one time period in the late 1940s and early 1950s.

2. The early 20th century warming in reanalysis

Figure S2 shows a comparison between annually averaged Arctic (north of 70°N) mean temperature observations from HadCRUT and reanalysis data (NOAA-20C and ERA-20C) during the 20th century. Between 1953–2011 there is good agreement between HadCRUT and reanalysis datasets, with R^2-value of 0.85 for ERA-20C and 0.41 for NOAA-20C. In contrast, before 1953 the reanalysis products diverge from HadCRUT, R^2-values for ERA-20C and NOAA-20C of 0.27 and 0.01, respectively. In particular, we note from 1900–1953, ERA-20C temperature anomalies have little variability just below 0°C and NOAA-20C shows very cold anomalous temperatures of around −3°C, whereas HadCRUT increases from approximately −2°C to 1°C over the same time period. These discrepancies illustrate that neither of these reanalysis products capture the early 20th century warming.

3. Sensitivity of sea ice extent to temperature in our reconstructions

Here we add the relationship between our reconstructed Arctic mean 2 m air temperature and total Arctic sea ice extent to augment Figure 1 of the main paper. Results show that the sensitivity between temperature and sea ice in the reconstructions during 1850–2018 agree well with that found in satellite observations. In contrast, the Walsh et al. (2017) reconstruction shows a disconnect in this relationship before and after 1953, which we speculate is due to limited data availability (see Figure S1). When we calculate the linear least-squares regression between our reconstructed mean Arctic temperature and total Arctic sea ice extent pre-1953, we see slopes of -0.68 x 10^6 km^2/°C and -0.79 x 10^6 km^2/°C using CCSM4 and MPI priors and R^2-values of 0.76 and 0.89 respectively. For the same calculations on our reconstructions post-1953 we see slopes of of -0.66 x 10^6 km^2/°C and -0.75 x 10^6 km^2/°C using CCSM4 and MPI priors and R^2-values of 0.87 and 0.93 respectively. Therefore, the linear relationship does not shift in the mid 1900s in our reconstructions in the same way it does in the Walsh et al. (2017) reconstructions. Furthermore, a large shift in sea ice and temperature sensitivity is not supported by the literature (Armour et al., 2011; Mahlstein & Knutti, 2012).

4. Sensitivity of data assimilation results

Here, we quantify the sensitivity of our reconstructions to the choice of gridded temperature product assimilated (and their errors), climate model prior, and sample-error mediation in the DA (localization length scale and ensemble variance inflation factor).

4.1. Sensitivity to the observations

To test the sensitivity to the choice of assimilated observations, we assimilate three temperature products: HadCRUT, GISTEMP and Berkeley Earth (BE). The original temperature measurements used to create these products are mostly the same, and the main difference is the amount of interpolation (or infill) from grid cells with observations to grid cells with no observations for GISTEMP and BE. Reconstructions using all three temperature products are shown in Figure S4. The skill of the reconstructions during the satellite period is slightly higher when HadCRUT is assimilated, as measured by the R^2 values and coefficient of efficiency. Overall, the source of the temperature observations has little effect on the overall variability of the reconstructions, with R^2 values with satellite data ranging from 0.82–0.89 for the MPI prior and 0.79–0.89 for the CCSM4 prior (described below). This is expected given the overall agreement among temperature products (e.g. (Rohde et al., 2013)).

For all of these experiments, an observed uncertainty of ($\mathbf{R}$ in Equation 2) 0.4 K^2 is used for all three products as explained in the main manuscript. Other uncertainty estimates tested include: (1) using the annually averaged diagonal elements of the error covariance matrix provided with HadCRUT, and (2) using the variance across all three datasets at each point.

Method (1) is ideal, but can only be applied to HadCRUT which has fewer data points than GISTEMP and BE because it does not use interpolation. For method (2) the variance across these datasets is very small, given that they often use the same original temperature

observations. This led to an over-weighting of observations in the Kalman gain and a SIE reconstruction with an inter-annual variability much larger than the satellite record.

4.2. Sensitivity to the prior

We use the MPI and CCSM4 Last Millennium simulations to test the sensitivity of the results to the choice of model prior. Figure S4 shows Arctic SIE from these two experiments (note that we use different inflation factors of 1.8 and 2.6 for the MPI and CCSM4 priors respectively, see below). Results show differences in inter-annual variability, but overall the decadal variability and the timing and magnitude of the ETCW agree well (Figure S4). MPI-based reconstructions show slightly higher correlation with satellite data, with R^2-values of 0.82–0.89, as compared to CCSM4-based reconstruction, with R^2-values of 0.79–0.89.

4.3. Sensitivity to sample error: prior inflation and localization

The prior ensemble-perturbation inflation factor and prior spatial localization length scale are both determined empirically based on correlations with the trend in Arctic SIE in satellite observations, and correlation and coefficient of efficiency with satellite observations between 1979–2017. A series of experiments are performed with inflation factors ranging from 1.6–2.0 (incremented by 0.1) for the MPI prior and 2.3–2.7 (incremented by 0.1) for the CCSM4 prior. For each inflation factor, reconstructions are performed for localization radii of 5,000, 7,500, 10,000, 15,000, 20,000, and 25,000 km. As the basis of comparison, the trend, detrended variance, correlation and coefficient of efficiency with respect to satellite observations between 1979–2017 are determined across all iterations and ensemble members for each of the 30 parameter combinations (see Figure S5 and S6).

Increasing the localization length scale and the ensemble inflation of sea ice relative to temperature, both increase the temporal variance and trend in the reconstructions of SIE (Figures S5, S6). The results indicate that there is not only a trade-off between capturing the trend versus inter-annual variability, but that there are various parameter combinations that show similar performance. Overall, all experiments described above using HadCRUT observations resulted in R^2 values greater than 0.86 and CE greater than 0.77 for MPI and R^2 values greater than 0.76 and CE greater than 0.62 for CCSM4.

5. Verification Statistics

To test the performance of our reconstructions we use both R^2 value (correlation coefficient squared) and coefficient of efficiency against satellite observations. The coefficient of efficiency (CE, Nash and Sutcliffe (1970)), like the correlation coefficient, measures the synchronicity in the variability of two datasets, but also quantifies mean bias and the difference in variance between the two datasets. This is a much stricter skill metric. Its maximum value is 1.0 and it is unbounded in the negative direction. A CE value of zero occurs when the sum of squared errors is equal to the variance in the verification data. Generally, positive CE values represent skill. It is defined as:

$$CE = 1 - \frac{\sum_i^n (v_i - x_i)^2}{\sum_i^n (v_i - \bar{v})^2}. \quad (1)$$

Here v is the verification value, with mean value $\bar{v}$, and x is the value being evaluated (the reconstructed value).

References

Armour, K. C., Eisenman, I., Blanchard-Wrigglesworth, E., McCusker, K. E., & Bitz, C. M. (2011). The reversibility of sea ice loss in a state-of-the-art climate model. *Geophysical Research Letters*, *38*. doi: 10.1029/2011GL048739

Fetterer, F., Knowles, K., Meier, W. N., Savoie, M., & Windnagel, A. K. (2017). Sea Ice Index, Version 3. *NSIDC: National Snow and Ice Data Center, Boulder, Colorado USA*. doi: 10.7265/N5K072F8

Mahlstein, I., & Knutti, R. (2012). September Arctic sea ice predicted to disappear near 2ÂC global warming above present. *Journal of Geophysical Research: Atmospheres*, *117*(D6). doi: 10.1029/2011JD016709

Nash, J. E., & Sutcliffe, J. V. (1970). Riverflow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282–290.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., & Mosher, S. (2013). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, *01*(02). doi: 10.4172/2327-4581.1000103

Walsh, J. E., Fetterer, F., Scott Stewart, J., & Chapman, W. L. (2017). A database for depicting Arctic sea ice variations back to 1850. *Geographical Review*, *107*(1), 89–107. doi: 10.1111/j.1931-0846.2016.12195.x

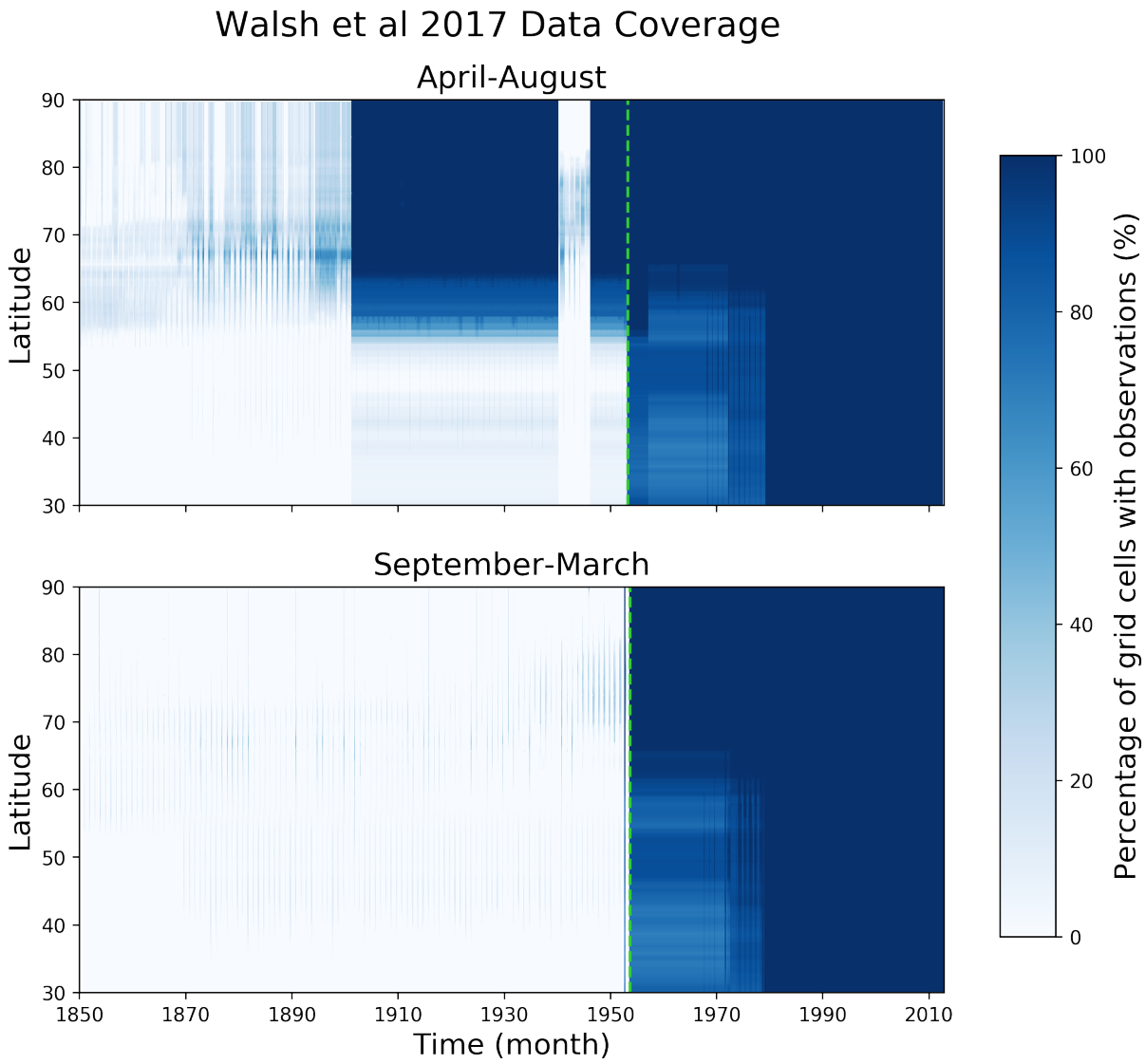

Figure S1. Shown is the data availability incorporated into the Walsh et al. (2017) Arctic sea ice record separated by two seasons over time. The color indicates the percentage of ocean longitude grid cells with an observation available at each latitude for each month. The vertical green lines indicate April 1953 and September 1953 respectively.

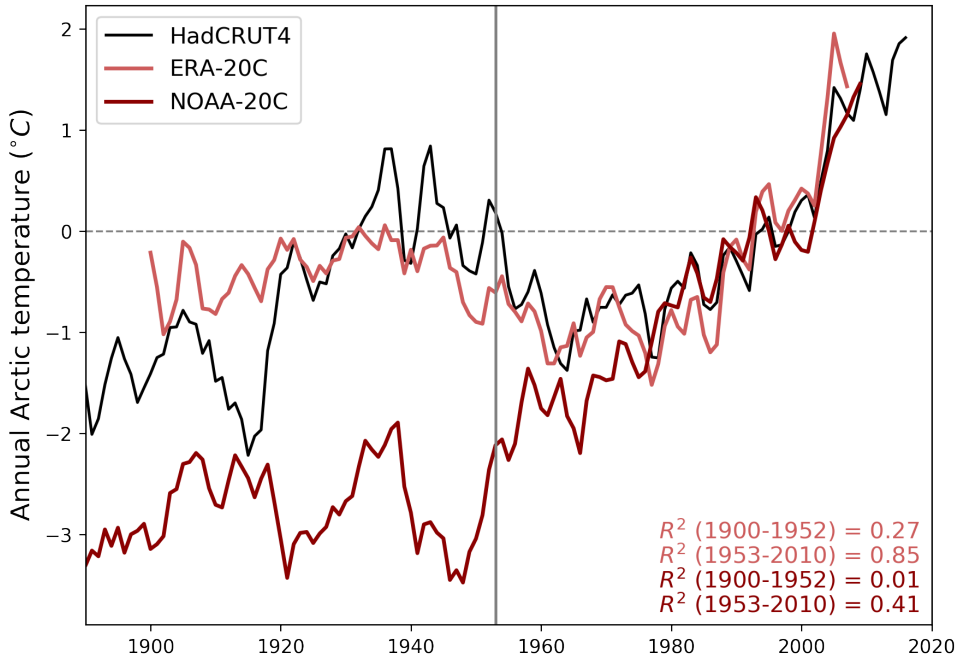

Figure S2. Arctic (north of 70°N) mean surface air temperatures anomalies from HadCRUT, NOAA-20C, and ERA-20C. The vertical gray line indicates the year 1953, when availability of observations of sea ice in the Arctic increase substantially in the Walsh et al. (2017) record. Anomalies are centered about 1979–2011.

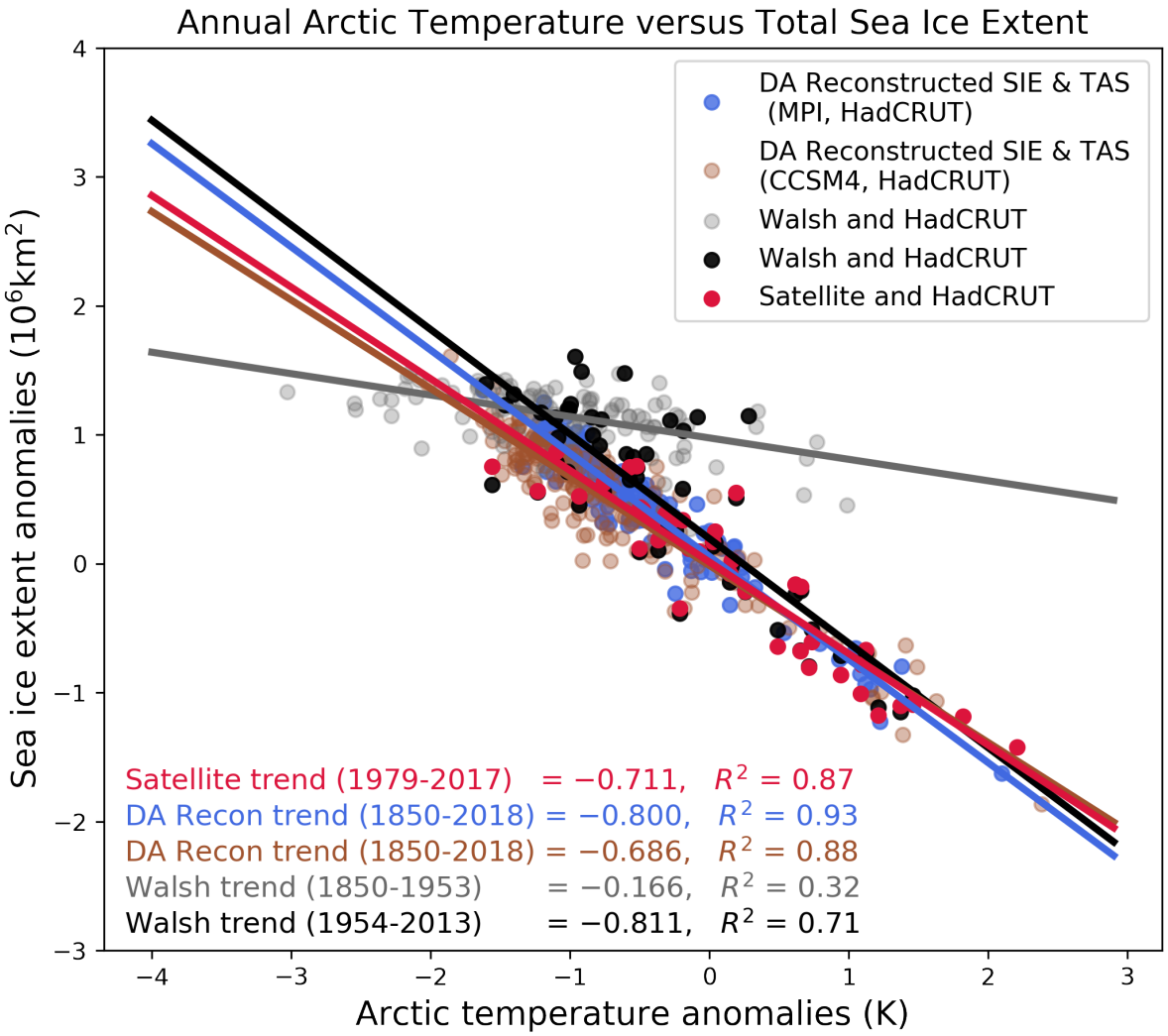

Figure S3. Arctic 2 m air temperature (averaged north of 65°N, derived from HadCRUT) and total Arctic sea ice extent in both the satellite data between 1979–2017 (in red) and Walsh et al. (2017) data set between 1850–1952 (in gray) and 1953–2013 (in black). The same relationship from reconstructed SAT and SIE between 1850–2018 are also shown for two reconstructions: one using MPI as the prior and HadCRUT temperature observations (in blue) and one using CCSM4 as the prior and HadCRUT temperature observations (in brown). Anomalies are relative to 1979–2013.

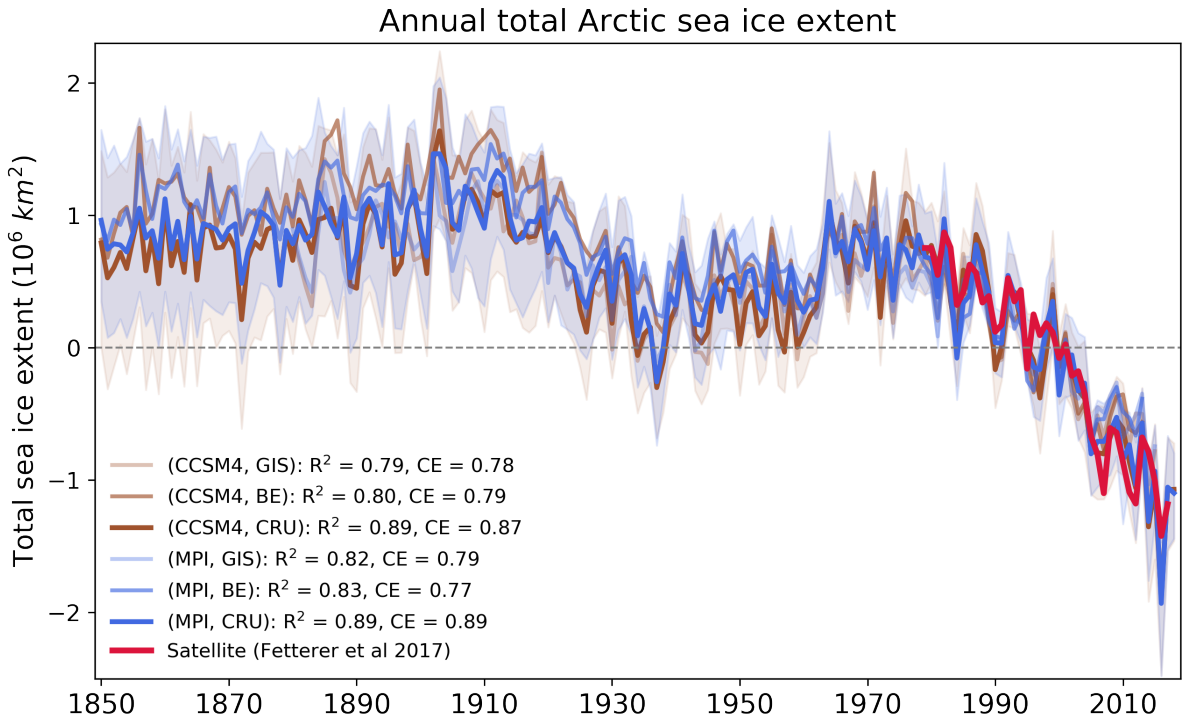

Figure S4. Total Arctic SIE reconstructed using priors drawn from two models (MPI and CCSM4 Last Millennium simulations) and three temperature datasets (HadCRUT, GISTEMP, and BE). For all experiments a localization length scale of 15,000 km is used and an inflation factor of 1.8 for MPI and 2.6 for CCSM4. The 97.5 and 2.5 percentiles of the ensemble spread are shown in blue and brown shading.

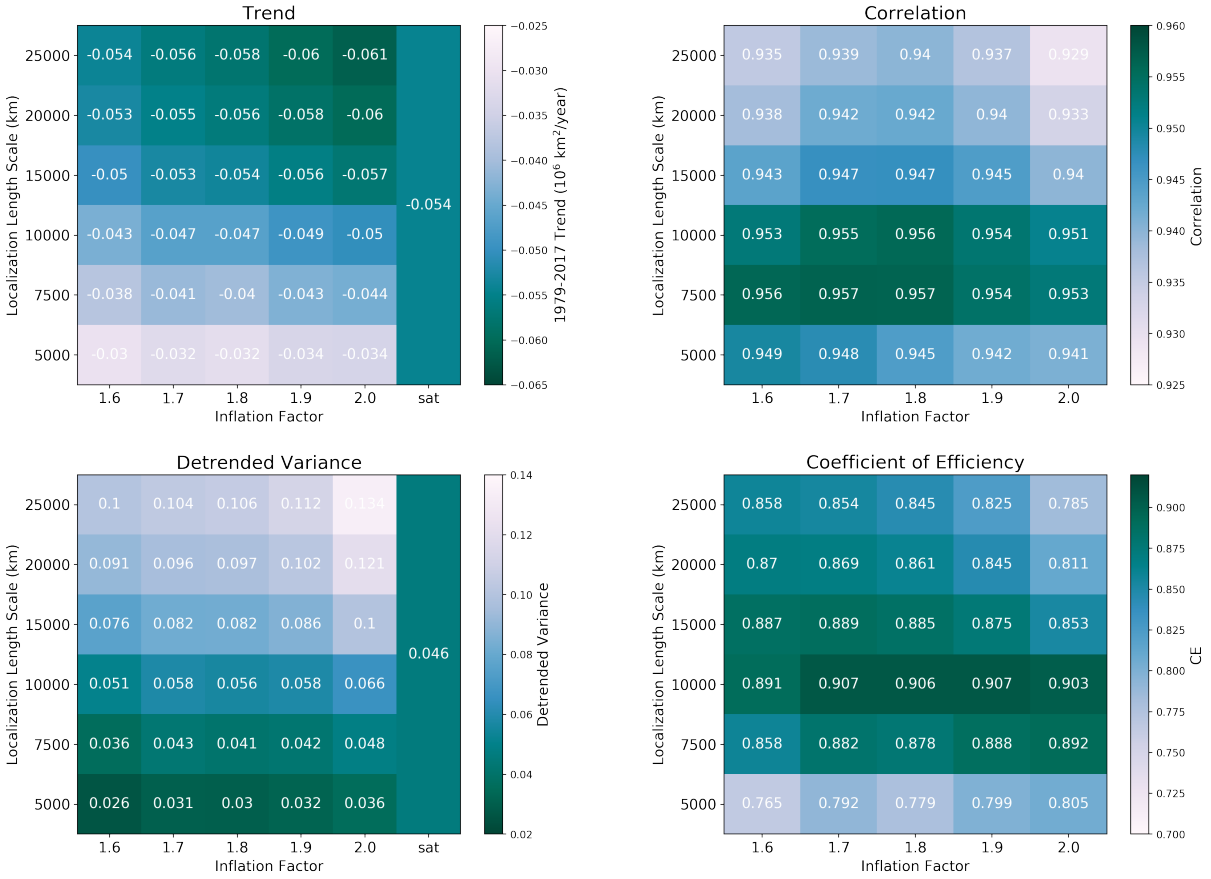

Figure S5. Verification statistics for 30 reconstructions performed using MPI as a model prior, HadCRUT observations, and different combinations of localization length scales (y-axis) and inflation factors (x-axis) are shown. Trends and detrended variances during the satellite era are shown in the two boxes on the left and the values observed in the satellite record (Fetterer et al., 2017) are shown in the column on the right. The correlation and coefficient of efficiency of these reconstructions when compared with (Fetterer et al., 2017) are shown in the two boxes on the right.

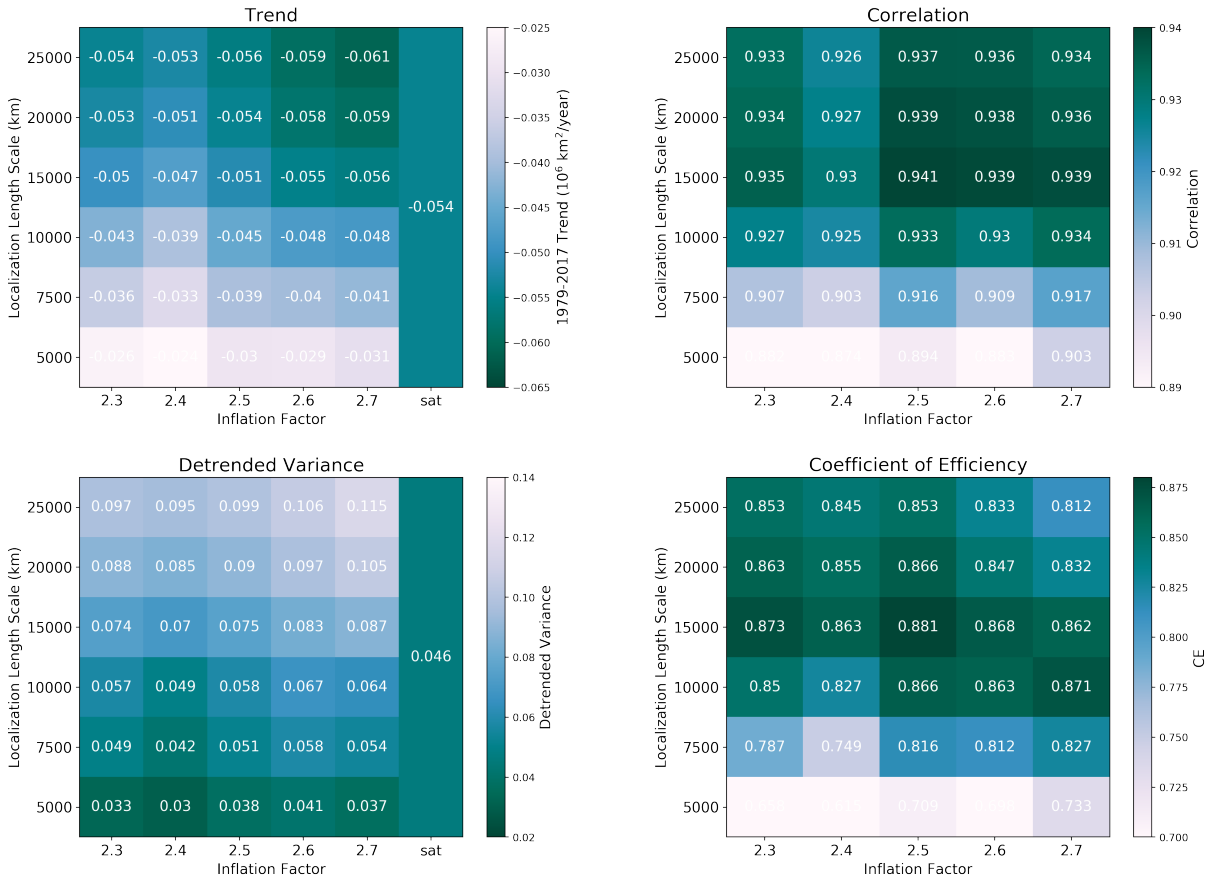

Figure S6. Verification statistics for 30 reconstructions performed using CCSM4 as a model prior, HadCRUT observations, and different combinations of localization length scales (y-axis) and inflation factors (x-axis) are shown. Trends and detrended variances during the satellite era are shown in the two boxes on the left and the values observed in the satellite record (Fetterer et al., 2017) are shown in the column on the right. The correlation and coefficient of efficiency of these reconstructions when compared with (Fetterer et al., 2017) are shown in the two boxes on the right.