The Benefits of Continuous Local Regression for Quantifying Global Warming

David C Clarke[1,1,1], Mark Richardson[2,2,2,2], and David C Clarke[1]

[1]None
[2]Jet Propulsion Laboratory, California Institute of Technology

December 1, 2022

Abstract

Change in global mean surface temperature (?GMST), based on a blend of land air and ocean water temperatures, is a widely cited climate change indicator that informs the Paris Agreement goal to limit global warming since preindustrial to "well below" 2°C. Assessment of current ?GMST enables determination of remaining target-consistent warming and therefore a relevant remaining carbon budget. In recent IPCC reports, ?GMST was estimated via linear regression or differences between decade-plus period means. We propose non-linear continuous local regression (LOESS) using +-20 year windows to derive ?GMST across all periods of interest. Using the three observational GMST datasets with almost complete interpolated spatial coverage since the 1950s, we evaluate 1850—1900 to 2019 ?GMST as 1.14degC with a likely (17—83 %) range of 1.05—1.25degC, based on combined statistical and observational uncertainty, compared with linear regression of 1.05degC over 1880—2019. Performance tests in observational datasets and two model large ensembles demonstrate that LOESS, like period mean differences, is unbiased. However, LOESS also provides a statistical uncertainty estimate and gives warming through 2019, rather than the 1850—1900 to 2010—2019 period mean difference centered at the end of 2014. We derive historical global near-surface air temperature change (?GSAT), using a subset of CMIP6 climate models to estimate the adjustment required to account for the difference between ocean water and ocean air temperatures. We find ?GSAT of 1.21degC (1.11—1.32degC) and calculate remaining carbon budgets. We argue that continuous non-linear trend estimation offers substantial advantages for assessment of long-term observational ?GMST.

The Benefits of Continuous Local Regression for Quantifying Global Warming

David C. Clarke[1], Mark Richardson[2,3]

[1]Independent Researcher, Montreal, Quebec, Canada
[2]Jet Propulsion Laboratory, California Institute of Technology, USA
[3]Colorado State University, Fort Collins, USA

Corresponding author: David C. Clarke (davidcclarke.research@gmail.com)

Key Points:

- Continuous local regression is an alternative to traditional IPCC temperature change estimation methods.
- Global warming, estimated from combined land and sea-surface temperature observational series with enhanced surface coverage, reached 1.14°C in 2019 relative to 1850—1900 (likely range 1.05—1.25°C).
- Global surface air temperature anomalies reached 1.21°C in 2019 relative to 1850—1900 (1.11—1.32°C), implying a remaining carbon budget of ~220 $GtCO_2$ to limit warming to 1.5°C.

Abstract

Change in global mean surface temperature (ΔGMST), based on a blend of land air and ocean water temperatures, is a widely cited climate change indicator that informs the Paris Agreement goal to limit global warming since preindustrial to "well below" 2°C. Assessment of current ΔGMST enables determination of remaining target-consistent warming and therefore a relevant remaining carbon budget. In recent IPCC reports, ΔGMST was estimated via linear regression or differences between decade-plus period means. We propose non-linear continuous local regression (LOESS) using ±20 year windows to derive ΔGMST across all periods of interest. Using the three observational GMST datasets with almost complete interpolated spatial coverage since the 1950s, we evaluate 1850—1900 to 2019 ΔGMST as 1.14°C with a likely (17—83 %) range of 1.05—1.25°C, based on combined statistical and observational uncertainty, compared with linear regression of 1.05°C over 1880—2019. Performance tests in observational datasets and two model large ensembles demonstrate that LOESS, like period mean differences, is unbiased. However, LOESS also provides a statistical uncertainty estimate and gives warming through 2019, rather than the 1850—1900 to 2010—2019 period mean difference centered at the end of 2014. We derive historical global near-surface air temperature change (ΔGSAT), using a subset of CMIP6 climate models to estimate the adjustment required to account for the difference between ocean water and ocean air temperatures. We find ΔGSAT of 1.21°C (1.11—1.32°C) and calculate remaining carbon budgets. We argue that continuous non-linear trend estimation offers substantial advantages for assessment of long-term observational ΔGMST.

1 Introduction

Estimates of global mean surface temperature anomalies (GMST), derived from a combination of near-surface air temperatures from land stations and sea surface temperatures over oceans, have long been a staple of climate study. GMST and derived trends or changes, ΔGMST, have featured prominently in IPCC reports, and are a key component in assessments of climate change attribution (Bindoff et al., 2013), climate model validation (Flato et al., 2013), global carbon budgets (Rogelj et al., 2018) and climate impacts (Hoegh-Guldberg et al., 2018). Perhaps most importantly, the IPCC's long-term ΔGMST estimate of 0.85°C, based on the 1880—2012 linear trend, was a key scientific input to the Paris agreement to keep global surface temperature change well below 2°C (IPCC, 2014; UNFCCC, 2015).

The IPCC Fifth Assessment Working Group I Report (IPCC WG1 AR5; Hartmann et al., 2013a) used three GMST datasets: HadCRUT4 (Morice et al., 2012), NASA GISTEMP (Hansen et al., 2010) and NOAA MLOST (Vose et al., 2010). While HadCRUT4 begins in 1850, the NOAA and NASA datasets only begin in 1880, so the 1880—2012 ordinary least squares (OLS) linear trend was presented as a "headline" warming estimate along with the HadCRUT4 1850—1900 to 2003—2012 difference in the Summary for Policymakers (IPCC, 2013). OLS trends for all datasets were also given for 1951—2012 and 1979—2012 with uncertainties adjusted to account for autocorrelated residuals (Santer et al., 2008; Hartmann et al., 2013b).

The IPCC Special Report on Global Warming of 1.5°C (IPCC SR1.5; Allen et al., 2018) included two new GMST datasets that incorporated sophisticated spatial interpolation: Cowtan-Way (Cowtan & Way, 2014a; Cowtan & Way, 2014b; Cowtan et al., 2015) and Berkeley Earth (Rohde et al., 2013). Reported ΔGMST was 0.87 ±0.12°C based on the average of HadCRUT4, NOAA, NASA and Cowtan-Way. An observation based estimate of Global Surface Air Temperature change (ΔGSAT) was introduced by adjusting HadCRUT4 ΔGMST to account for incomplete coverage and discrepancy in ocean air and sea-surface temperature anomalies, thus producing an estimate of air near-surface temperature at 2 m over the entire globe (Rogelj et al., 2018; Cowtan et al., 2015). The ΔGSAT estimate of 0.97°C in 2006—2015 implied lower remaining carbon budgets compared to preceding studies based on ΔGMST consistent with AR5's 0.85°C through 2012 (Millar et al., 2017a, 2017b; Goodwin et al., 2018; Richardson et al., 2018).

IPCC WG1 AR5 Box 2.2 discusses the following issues with linear trends for estimating ΔGMST: 1) poor approximation of trend evolution over time; 2) poor fit of residuals unamenable to correction via autoregressive or moving average models; 3) high sensitivity to selected period; and 4) divergent or even contradictory sub-period estimates relative to that of a larger encompassing interval. The latter two issues were particularly relevant in AR5 Section 2.4.3's discussion of the "observed reduction in warming trend" over 1998—2012 compared to 1951—2012 (Rahmstorf et al., 2017; Risbey et al., 2018). A smoothing spline non-linear trend fit was demonstrated to address these factors, and later studies presented alternative estimators for continuous long-term ΔGMST trends (Cahill et al., 2015; Peng-Fei et al., 2014; Mudelsee, 2019; Visser et al., 2018).

An issue of particular concern is that linear trends underestimate long-term (> 100 years) ΔGMST compared to other estimates. For example, IPCC AR5 Box 2.2 estimated HadCRUT4 1900—2012 trends of 0.075 ± 0.013 °C decade^{-1} and 0.081 ± 0.010 °C decade^{-1} for linear OLS and smoothing spline trends respectively. Generally, long-term linear fit ΔGMST from 1880 to present is 0.05—0.10°C below nonlinear estimates (SR15 table 1.2; Visser et al., 2018) although the spread in ΔGMST estimates between different datasets is commonly as wide as differences engendered by ΔGMST methodology. Ultimately, IPCC AR5 Box 2.2 recommended linear trends over non-linear estimates, noting that HadCRUT4 OLS-based long-term ΔGMST lay within the 5-95% uncertainty range from the smoothing spline. Nevertheless, as the IPCC enters the Sixth Assessment Report (AR6), a new method that supplements or supplants the traditional approaches could reduce known biases and address these shortcomings.

This work proposes a local regression technique (LOESS, Cleveland et al., 1992; Cleveland, 1979) with a ±20 year smoothing window for multi-decadal analysis. We also provide statistical uncertainty and show that the fit residuals follow the assumed ARMA(1, 1) autocorrelation structure. The framework can be extended to give self-consistent ΔGMST estimates with uncertainty over as little as 15 years, providing a potential alternative to linear fits over all intervals of interest.

However, here we focus on long-term ΔGMST and associated carbon budgets, directly relating our estimates to approaches discussed in AR5 and SR1.5. We compare against the IPCC approaches of OLS (1880—latest year) and period mean differences (from "preindustrial" reference period 1850—1900 to the latest decade), as well as a global warming index which SR1.5 used as the main estimate of "human-induced warming" (Haustein et al., 2017). We also test the performance of our LOESS estimates using output from the two model large ensembles with simulations that begin in 1850. Our final comparison is with the new CMIP6 model ensemble, and using a subset of this ensemble we derive a modest conversion factor to update our observation-based ΔGMST to ΔGSAT for carbon budget calculations.

The paper is structured as follows. Section 2.1 describes source data from observations and associated estimated radiative forcings (2.1.1), two large model ensembles (2.1.2) and CMIP6 models (2.1.3). Section 2.2 describes trend estimation (2.2.1), evaluation of ΔGMST methods and performance (2.2.2), large model ensemble evaluation (2.2.3) and ΔGSAT and carbon budget calculation (2.2.4). We present our results in Section 3, covering long-term ΔGMST analysis (3.1), large model ensemble analysis (3.2) and ΔGSAT and associated remaining carbon budgets (3.3). Finally in Section 4 we discuss our results and issue recommendations for the use of ΔGMST and ΔGSAT in future IPCC assessments.

2 Source Data and Methods

2.1.1 Global surface temperature datasets

Typically, gridded monthly land surface air temperature (LSAT) and sea surface temperature (SST) anomalies are generated then blended to produce GMST. Table 1 summarizes five blended LSAT-SST series in widespread use. There is considerable overlap in the underlying datasets. There are two SST data sets: HadSST3 (Kennedy et al., 2011) and NOAA's ERSSTv5 (Huang et al., 2017), and three LSAT datasets: GHCNv4 (Menne et al., 2019), CRUTEM4 (Jones et al., 2010), and Berkeley Earth (Rohde et al., 2013). Even this understates the overlap; for example, both SST datasets rely primarily on the comprehensive store of maritime observations from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman et al., 2016), albeit processed, filtered and supplemented in different ways. It is important to note, however, that there are important differences between each group's quality assurance and data homogenization procedures, and associated uncertainties, in both the land and SST datasets. In particular, bias adjustments of SST data to account for differences between buoy, engine intake and bucket measurements, can have a notable effect on long-term trends (Kennedy et al., 2019).

Table 1. Five operational observational datasets.

Series	Land (LSAT)	Ocean (SST)	Interpolation	Averaging	Start year
HadCRUT4 (Morice et al., 2012)	CRUTEM4	HadSST3	None	Simple average of hemispheric area-weighted averages	1850
NOAA GlobalTemp v5 (Zhang et al., 2019; Huang et al., 2020)	GHCNv4	ERSSTv5	Empircal orthogonal teleconnections (EOTs)	Area weighted average	1880
NASA GISTEMP v4 (Lenssen et al., 2019)	GHCNv4	ERSSTv5	Distance weighting (to 1200 km)	80 zones x 100 sub-boxes	1880
Cowtan-Way v2 (Cowtan & Way, 2014a; Cowtan & Way, 2014b; Cowtan et al., 2015)	CRUTEM4 (kriged)	HadSST3 (kriged)	Kriging (Complete)	Area weighted average	1850
Berkeley Earth (Rohde & Hausfather, 2020)	Berkeley Earth	HadSST3 (reprocessed & kriged)	Kriging (to ~2500 km)	Area weighted average	1850

Differences in spatial interpolation can affect calculated GMST. HadCRUT4 calculates area-weighted hemispheric means with no interpolation between its 5°×5° grid boxes, combined in a "simple" (equally-weighted) average. In contrast, NASA GISTEMP, Cowtan-Way and Berkeley Earth use extensive interpolation and, crucially, extrapolate LSAT over sea ice. Cowtan-Way interpolates HadCRUT4 to produce 100% apparent coverage, while GISTEMP and Berkeley Earth both interpolate up to 1200 km from observations, resulting in virtual areal coverage two to three times that of HadCRUT4 in the late 19th century. Nominal coverage in all three datasets is virtually complete since 1951 (see Figure S1, Supplementary Information). Reducing Cowtan-Way coverage to that of Berkeley Earth results in imperceptible differences in GMST even in the 19th century, indicating that distance-limited and unlimited kriging interpolation can be considered equivalent (See Figure S14, Supplementary Information). Spatial smoothing via empirical orthogonal teleconnections (EOTs; van den Dool et al., 2000) in NOAA GlobalTemp (and ERSSTv5) results in nominal coverage between that of HadCRUT4 and NASA GISTEMP, but largely misses very high latitudes and has no interpolated coverage over Arctic sea ice.

Comparisons with temperature reanalyses, independent surface data and satellite retrievals show that interpolation significantly mitigates coverage bias (and associated underestimation of warming) arising from poor sampling of the fastest warming areas, especially the Arctic, since the mid-twentieth century (Dodd et al., 2015; Cowtan et al., 2018a; Susskind et al., 2019; Lennsen et al., 2019). Evidence is mixed for earlier periods where reduced coverage leads to larger interpolation uncertainty (Cowtan et al., 2018a) and differences between underlying SST datasets are the largest source of discrepancies. Cowtan et al. (2018a) showed that both generalized least squares averaging and kriging interpolation mitigated errors engendered by "naïve" global or hemispheric averaging methods, such as those used in HadCRUT4, which implicitly set "missing" areas to the global average of sampled areas (Hansen et al., 2006). Thus, the three interpolated datasets are demonstrably more representative of global climate change.

We use the published monthly anomaly series, except for Berkeley Earth where we use the area-weighted average of the gridded series, which diverges from the published series over 1850—1950 (Supplementary Information, Figures S2, S3). For series starting in 1850 anomalies are relative to 1850—1900 while NASA GISTEMP and NOAA GlobalTemp are baselined such that their 1880—1900 mean matches that of the three longer-running datasets. These rebaselined NASA and NOAA series are used for all ΔGMST estimates calculated relative to 1850—1900 as outlined in Section 2.2.1. This streamlined and consistent scheme replaces multiple IPCC SR1.5 approaches based on scaling their 1880—2015 trends or matching to HadCRUT4 over 1880—1990. We also report the mean ΔGMST for all five operational datasets (OpAll group) and the subset of three datasets with near-global interpolated coverage post-1950 (Global_3 group), with the latter used as the basis for our main estimates. Group ΔGMST estimates are the mean of the individual estimates as in IPCC AR5.

We augment temperature data with summarized anthropogenic and natural radiative forcing data required to derive the "global warming index" referenced in SR1.5 as a potential alternative to ΔGMST for tracking anthropogenic warming (Haustein et al., 2017; Allen et al., 2018). These are used to estimate anthropogenic and natural forced changes, $\Delta GMST_{F,anthro}$ and $\Delta GMST_{F,nat}$, using a two-box impulse-response model with parameters derived from a least-squares-fit between observed temperatures and the modelled response (Otto et al., 2015; Haustein et al., 2017). These estimates are used to assess the characteristics of a particular LOESS window choice (Section 2.2.1) and as an additional comparator to long-term ΔGMST.

2.1.2 Model Large Ensembles

We perform tests using output from the large ensembles whose simulations begin in 1850: the Max Planck Institute for Meteorology Grand Ensemble (MPI-GE, N=100, Maher et al., 2019) and Commonwealth Scientific and Industrial Research Organisation Mk3.6.0 (CSIRO Mk3.6.0, N=30, Rotstayn et al., 2012; Jeffrey et al., 2013), taking their GSAT over historical-RCP8.5 simulations for 1850—2019 and baselining each to 1850—1900. We exclude five other large ensembles that start after 1850 (Deser et al, 2020), and our approach is conceptually similar to that in Dessler et al. (2018)'s estimation of how internal variability affects derived climate sensitivity in MPI-GE. The use of GSAT simplifies the calculations and since the year-to-year variability in GSAT-GMST difference is of order 0.01 °C in CMIP5 models (e.g. Figure 2 of Cowtan et al. 2015), we expect little effect of blending or masking on this particular analysis.

Conceptually, we first decompose $\Delta GSAT$ as:

$$\Delta GSAT_{total} = \Delta GSAT_F + \Delta GSAT_{var} \tag{1}$$

where $\Delta\text{GSAT}_{\text{var}}$ represents internal variability and $\Delta\text{GSAT}_{\text{F}}$ the forced response. The same decomposition would apply for ΔGMST. We adopt the IPCC SR1.5 argument that "[s]ince 2000, the estimated level of human-induced warming has been equal to the level of observed warming with a *likely* range of ±20%". From this it follows that a reliable estimate of $\Delta\text{GMST}_{\text{F}}$ through 2019 would be an appropriate estimate of human-induced warming, $\Delta\text{GMST}_{\text{F,anthro}}$, with relevance for temperature targets and carbon budgets. With just one realization of real-world internal variability we cannot perform this decomposition, but a large ensemble mean should approach that model's $\Delta\text{GMST}_{\text{F}}$. We test whether our derived $\Delta\text{GMST}_{\text{LOESS}}$ approximates $\Delta\text{GMST}_{\text{F}}$, and consider the decomposition in an individual run to be:

$$\Delta GMST_{total} = \Delta GMST_{LOESS} + \Delta GMST_{resid} \quad (2)$$

With a ±20-year window this effectively decomposes between short- and long-term ΔGMST. If periods are selected to minimize volcanism (which induces short-term $\Delta\text{GMST}_{\text{F}}$), and the magnitude of $\Delta\text{GMST}_{\text{var}}$ is small at 40-year timescales, then resultant $\Delta GMST_{LOESS} \approx \Delta GMST_{F,anthro}$ over the long-term intervals of interest.

2.1.3 Coupled Model Intercomparison Project, phase 6 (CMIP6) output

We include historical simulations over 1850—2014 from CMIP6 models which have the required fields for blending surface air temperatures (SAT) over land or sea ice and SST over ocean (Eyring et al, 2016), permitting "apples-to-apples" comparisons with land-ocean observational datasets and derivation of a $\Delta\text{GMST}_{\text{LOESS}}$ to $\Delta\text{GSAT}_{\text{LOESS}}$ adjustment. These include near-surface air temperature ("tas"), sea surface temperature ("tos") and sea ice concentration ("sciconc" or "sciconca", N=24 simulations listed in Table S1).

Following Cowtan et al (2015) and Richardson et al (2018), each simulation is processed to produce two area-weighted average series: 1) global SAT (i.e. GSAT) and 2) global blended SAT-SST (i.e. GMST). At each grid cell i, j, the blended monthly temperature $T_{\text{blend,i,j}}$ is:

$$T_{\text{blend,i,j}} = w_{\text{SAT,i,j}}\, T_{\text{SAT,i,j}} + (1 - w_{\text{SAT,i,j}})\, T_{\text{SST,i,j}} \quad (3)$$

where $w_{\text{SAT,i,j}}$ is the land plus sea ice grid cell fraction, and $T_{\text{SAT,i,j}}$ and $T_{\text{SST,i,j}}$ are the local anomalies relative to 1850—1900. For GSAT $w_{\text{SAT,i,j}} = 1$ everywhere, and for the blended GMST series $w_{\text{SAT,i,j}} = 1$ in ocean cells for a calendar month if any those months during 1961-2014 has siconc > 1%. This is similar to the Cowtan-Way blending algorithm and the "xaf" simulations in Cowtan et al. (2015).

2.2 Methods

Next we describe our approach to obtain ΔGMST, our uncertainty estimation, and the remaining carbon budget calculation. Section 2.2.1 explains the trend fits and errors; Section 2.2.2 explains the ΔGMST calculations, observational error and methods by which the fit quality are judged using observational data. Section 2.2.3 discusses the large ensemble methodology, and Section 2.2.4 the CMIP6 comparison and carbon budget calculation. We use ΔGMST and ΔGSAT to refer to a general change in global temperature, and use qualifiers or subscripts when referring to statistical estimation methods or its components. For example, $\text{LOESS}_{\text{bsln}}$ ΔGMST (or

ΔGMST$_{LOESS}$) refers to an estimate made with LOESS, while ΔGMST$_F$ refers to the forced component.

2.2.1 Trend calculations and their statistical uncertainty

For a series of n temperature observations x_i at time t_i, a linear trend is:

$$x_i = a + bt_i + e_i,\ i = 1, \ldots, n \qquad (4)$$

where a and b are intercept and slope parameters to be fitted and e_i are residual errors. The slope estimate $\hat{b}$ is used to obtain ΔGMST as $\hat{b}\,(t_n - t_i)$, with the uncertainty of $\hat{b}$ (and thus ΔGMST) determined as explained below.

Our multi-decadal LOESS point-to-point (LOESS$_{md}$) ΔGMST is based on the LOESS fit from 1880—2019; for any starting point, ΔGMST to 2019 is the LOESS$_{md}$ fit evaluated in 2019 minus the start value. We also introduce “baseline” LOESS (LOESS$_{bsln}$) as our main ΔGMST estimate. LOESS$_{bsln}$ is simply the same fit evaluated at the end year, yielding an estimate relative to 1850—1900 baseline, rather than to a given start year such as 1880. Although the central estimated fit is the same, the associated statistical fit uncertainties are quite different, as explained below.

Our LOESS$_{md}$ uses a fixed span α_{md} of ± 20 years, tricube weighting (the default) and a degree 1 smoothing parameter (i.e. locally weighted linear trend, which yields more stable end points). Tests with the Cowtan-Way series show that α of ±10 years captures internal decadal variability and has marked sensitivity to volcanic episodes early in the record and to a lesser extent over 1980—2019 (Figure S4). On the other hand, α of ±20 or ±30 years smooth out short-term variability and show similar warming from 1850—1900 to present: 1.12°C (±20 years) or 1.11°C (±30 years). Analysis of first differences for each LOESS window (Figures S5) shows large variance with α of ±5 years, which stabilises with α of ±20, ±25 or ±30 years. Large ensemble tests support this choice: α_{md} substantially smaller than ±20 years increases ΔGMST$_F$ discrepancy, while substantially longer than ±20 years introduces a low bias in 1850—2019 ΔGMST (Figures S6, S7). We therefore choose α_{md} = ±20 years to evaluate trends of length ≥30 years; LOESS$_{pent}$ (α = ±5 years) is reserved for future extension of our framework to cover very short-term trends of ≤15 years (see Figure S4, panel d).

Default methods assume statistically independent noise, necessitating an uncertainty correction if the fit residuals are autocorrelated. Santer et al (2000) presented a procedure for assessing an effective sample size (and associated reduction in degrees of freedom) from the general formula

$$n_e = \frac{n_t}{(1 + 2\sum_{j=1}^{n-1} \rho_j)} \qquad (5)$$

where ρ_j is the autocorrelation function of a noise model estimated from the fit residuals. If the noise follows an autoregressive(1) (AR(1)) process, then with $\rho_j = \phi^j$

$$1+2\sum_{j=1}^{n-1}\rho_j \approx 1+\frac{2\phi}{(1-\phi)}=\frac{(1+\phi)}{(1-\phi)} \tag{6}$$

where ϕ is estimated from the lag-one autocorrelation coefficient (Mitchell et al, 1966). However, Foster and Rahmstorf (2011) demonstrated that 1979-2010 GMST trend residuals were more consistent with an autoregressive moving average, ARMA(1, 1) model in the form

$$\begin{aligned} \rho_1 &= \frac{(\phi+\ \theta)(1+\phi\theta)}{1+2\phi\theta+\ \theta^2} \\ \rho_j &= \ \rho_1\phi^{j-1} \qquad j\geq 2 \end{aligned} \tag{7}$$

Substituting (6) into (5) yields

$$1+2\sum_{j=1}^{n-1}\rho_j \approx 1+\frac{2\rho_1}{(1-\phi)} \tag{8}$$

Foster and Rahmstorf used the Yule-Walker "method of moments" with $\hat{\phi}$ = $\hat{\rho}_1$ / $\hat{\rho}_2$. Hausfather et al. (2017) instead used Maximum Likelihood Estimation (MLE) to obtain $\hat{\phi}$ and $\hat{\theta}$ and then $\hat{\rho}_1$ via Eq. (6). Monte Carlo simulations show that MLE gives a more robust and efficient estimator $\hat{\phi}$, suitable for series as short as 8 years (see Figure S8). Hausfather et al. also introduced a bias correction to account for underestimated autocorrelation in shorter series, derived from AR(1) in Tjøstheim and Paulsen (1996) and extended to account for the positive difference between $\hat{\phi}$ and $\hat{\rho}_1$.

$$\begin{aligned} \hat{\phi}_{BC} &= \hat{\phi}+\left(1+4\left(2\hat{\phi}-\rho_1\right)\right)/\,n_t \\ \rho_{1\,BC} &= \rho_1+\left(1+4\left(2\hat{\phi}-\rho_1\right)\right)/\,n_t \end{aligned} \tag{9}$$

Although this bias correction is most pertinent for very short series, Monte Carlo simulations have demonstrated its relevance for highly autocorrelated series up to 720 months in length. We selected this bias correction after comparison with alternatives (e.g. Nychka et al., 2000; see Figure S9).

Substituting the bias corrected parameters and simplifying the correction term as in (5) yields the final effective length correction.

$$n_e=\frac{n_t}{1+2\sum_{j=1}^{n-1}\rho_j} \approx \frac{n_t}{1+2\rho_{1\,BC}\,/\,(1-\hat{\phi})} \tag{10}$$

We estimate corrections from the residuals of both LOESS and OLS. To apply this correction, we define nominal degrees of freedom v = $n_t - p$ and effective degrees of freedom v_e = $n_e - p$, where p is the number of actual or equivalent parameters of the trend fitting methodology.

In the linear case, the correction is applied directly to $s_{b,}$ the standard error of b in (1), with p = 2.

$$s_b^{'} = s_b \frac{v}{v_e} = s_b \sqrt{\frac{n_t - 2}{n_e - 2}} \tag{11}$$

For non-parametric trend estimation like LOESS, Monte Carlo simulations can establish uncertainties, as in Visser et al (2016) for smoothing spline trends. Here we propose a plausible heuristic method. First the above correction is applied to s_e, the standard errors of the residual fit, with p set to the equivalent number of parameters of the LOESS trend, derived from the trace of the LOESS projection matrix (Cleveland and Grosse, 1991); generally $p \approx 2/\alpha + 0.5$ for GMST datasets. For an equally spaced time series, s_e is maximum at the start and end of the LOESS fit. If statistical errors at these two points are independent, they may be combined in quadrature, by taking the square root of the sum of the squared standard errors, i.e. the square root of the sum of variances (see also Eq S4 in Karl et al., 2015). Then the corrected standard error $s^{'}_{\Delta T_n}$ for $\Delta GMST_n$ becomes

$$s^{'}_{\Delta T_n} = \sqrt{2}\,\max(s_e^{'}) = \sqrt{2}\,\max(s_e)\sqrt{\frac{n_t - p}{n_e - p}} \tag{12}$$

For both OLS and $LOESS_{md}$ we evaluate the sample autocorrelation function (ACF) of the fit residuals as well as the ACFs of the ARMA(1, 1) and AR(1) noise models fit to those residuals.

Finally, for $LOESS_{bsln}$ we assume that the mean error during the 1850—1900 baseline is small relative to the end point error. We are not aware of any formal method for calculating the required adjustment, so we generate an *ad hoc* correction tuned to perform well in Monte Carlo tests. To approximate the baseline uncertainty, we take the $LOESS_{md}$ start point uncertainty, $\max(s_e^{'})$, and reduce it according to the relative length of the baseline by applying an appropriate factor b_{adj}. This is similar in principle to the reduction of sample mean uncertainty with increasing sample size; in this case, b_{adj} is tuned to reproduce the results of Monte Carlo tests with Cowtan-Way data. For a baseline t_1 to t_b, with $b \leq n/2$, where n is the length of the full series we take (while also imposing a lower limit on b_{adj}):

$$b_{adj} = (t_{n/2} - t_b)/\,(t_{n/2} - t_1)\;\;;\;\; 0.5 \leq b_{adj} \leq 1 \tag{13}$$

Following quadrature the combined $LOESS_{bsln}$ error is then:

$$s^{'}_{\Delta T_n} = \sqrt{(b_{adj}{}^2 + 1)}\,\max(s_e^{'}) = \sqrt{(b_{adj}{}^2 + 1)}\,\max(s_e)\sqrt{\frac{n_t - p}{n_e - p}} \tag{14}$$

and (12) is a special case of (14) with a baseline of length 0 and b_{adj} = 1. Monte Carlo simulations of LOESS fits plus ARMA(1, 1) noise produce a probability distribution function nearly identical to that engendered in Cowtan-Way by (12) over 1880—2019 and by (14) from 1850—1900 and 1880—1900 to 2019 (Figures S10 and S11).

2.2.2 Estimates of observational ΔGMST, error components and performance tests

The main analysis focuses on long-term ΔGMST (results for other IPCC AR5 periods are in the Supplementary Information Table S2). In addition to OLS and $LOESS_{md}$ ΔGMST over 1880—2019, and $LOESS_{bsln}$ from 1850—1900 to 2019, we also calculate period difference ΔGMST estimates by subtracting mean GMST over 1850—1900 from the most recent decade, 2010—2019. The above are also compared to GMST-derived estimates of anthropogenic warming (Haustein et al., 2017; section 2.1.2) and to a CMIP6 ensemble (Section 2.2.4). Global_3 and OpAll group ΔGMST are the mean of individual dataset ΔGMST.

Following standard IPCC practice, we report the 5—95% statistical uncertainty range for LOESS and OLS ΔGMST estimates, as outlined in Section 2.2.1. Group uncertainties are reported conservatively and go from the smallest 5% to the largest 95% reported for any of their constituent datasets. We also report observational parametric uncertainty as the 5—95 % range of ΔGMST values derived from each of the 100-member HadCRUT4 and Cowtan-Way ensembles. These ensembles use a Monte-Carlo method to assess the fully correlated errors engendered by parametric uncertainty related to bias adjustments to individual temperature readings (Kennedy et al., 2011).

Figure S12 depicts these estimates and derived autocorrelation functions (ACF) for the Cowtan-Way monthly series with ARMA(1, 1) correction and for Cowtan-Way annual series with AR(1) correction (similar to IPCC AR5).

Finally we assess $LOESS_{bsln}$ ΔGMST against period mean differences for the Global_3 group by evaluating at the mid-point of the corresponding end decade; for example, $LOESS_{bsln}$ at the end of 2014 is comparable to the 1850—1900 to 2010—2019 period ΔGMST. IPCC SR1.5 explicitly considered their 1850—1900 to 2006—2015 ΔGMST estimate to be a proxy of the eventual 1996-2025 mean. We therefore compare the ΔGMST estimates for every year from 1995 against centered 20-year and 30-year means. We also compare to “extended” running 30-year periods, generated by assuming a continuation of the 1990—2019 linear trend through 2029. We argue that a smaller bias and root mean square error (RMSE) relative to the 20- and 30-year means represents better performance according to the IPCC’s own criterion.

2.2.3 Large Ensemble Analysis for Method Validation and Uncertainty Calculation

$LOESS_{bsln}$ is fit to the 1850—2019 annual output for each simulation, then the $\Delta GMST_{LOESS}$ through 2019 is evaluated from all start years 1850—1980. Separate linear OLS fits ending in 2019 are also obtained for those start years. We also evaluate $LOESS_{bsln}$ at the end of 2014 and compare with the 1850—1900 to 2010—2019 period ΔGMST (which we henceforth refer to as $\Delta GMST_{period}$). Finally, $LOESS_{md}$ is calculated over 1880—2019 for each simulation. The distribution of ensemble member $\Delta GMST$-$\Delta GMST_F$ provides an estimate of the bias and uncertainties for each estimator and each period, as argued in Section 3.2. If $\Delta GMST_{LOESS} \approx \Delta GMST_F$ then the LOESS residuals will be dominated by internal variability and our statistical uncertainty is related to error due to internal variability (we confirmed that the model residuals generally follow our assumed ARMA(1,1), Figure S13). The LOESS decomposition filters in time: $\Delta GMST_F$ excursions shorter than our window will inflate statistical uncertainty, while multi-decadal $\Delta GMST_{var}$ changes will be included in $\Delta GMST_{LOESS}$ and result in too small errors. We compare each run's statistical uncertainties with the ensemble 17—83 % and 5—95 % ranges to check for evidence that the observation-derived statistical uncertainties could represent internal variability in the 1850—1900 to 2019 $\Delta GMST_{LOESS}$ used for carbon budget calculations (see Section 2.2.4).

2.2.4 CMIP6 comparisons, GSAT adjustment and remaining carbon budget

IPCC SR15 reported remaining carbon budgets accounting for warming to date, but did not directly use the reported $\Delta GMST_{period}$ 5—95 % observational uncertainty from individual datasets. Instead AR5 5—95 % observational uncertainty through 1986-2005 was combined with additional uncertainties to produce a "likely" 17—83 % $\Delta GMST$ total uncertainty, and $\Delta GMST_{period}$ was then converted to $\Delta GSAT_{period}$ using a CMIP5-derived scaling. This Section describes the comparison with CMIP6 $\Delta GMST_{period}$ and conversion of observed $LOESS_{bsln}$ $\Delta GMST$ to $\Delta GSAT$, and then details the carbon budget calculation, which largely follows the IPCC SR1.5 methodology, as elaborated by Rogelj et al. (2019).

$LOESS_{bsln}$ series are generated for each of the 24 individual full-coverage CMIP6 air-only (GSAT) and blended (GMST) series described in Section 2.1.3, with the blended series being comparable to quasi-global GMST observations. We consider the full ensemble and also a sub ensemble of "likely ECS" models, excluding those with effective climate sensitivity (ECS) outside the CMIP5 1.9-4.5°C 90% ensemble range (Flato et al., 2013; Forster et al., 2019).

For each ensemble member's $LOESS_{bsln}$ changes we derive a "blending" factor A_{blend} = $\Delta GSAT_{LOESS}$ / $\Delta GMST_{LOESS}$, which represents the required adjustment to convert $\Delta GMST_{LOESS}$ to $\Delta GSAT_{LOESS}$, accounting for the difference between GSAT air temperatures and GMST "blending" of air and water temperatures. The median and ensemble distribution of A_{blend} scaling factors is applied to observed $\Delta GMST_{LOESS}$ to obtain historical observed $\Delta GSAT_{LOESS}$ with combined uncertainty for calculating the remaining carbon budget, as detailed below. The carbon budget calculation largely follows the framework established in IPCC SR1.5 (Rogelj et al., 2017), elaborated by Rogelj et al (2019) and implemented by Nauel et al (2019). We simplify the Rogelj et al (2019) remaining carbon budget equation to:

$$B_{lim} = \left(\Delta GSAT_{lim} - \Delta GSAT_{F,anthro} - \Delta GSAT_{nonCO_2,fut}\right) / TCRE \; - E_{Esfb} \qquad (15)$$

where B_{lim} is the remaining carbon budget associated with a temperature limit $\Delta GSAT_{lim}$ (1.5 or 2°C), with $\Delta GSAT_{F,anthro}$ (also referred to as $\Delta GSAT_{hist}$) the historical human-induced warming to date and $\Delta GSAT_{nonCO_2,fut}$ the expected future warming from non-CO_2 anthropogenic forcing. TCRE is the transient climate response to cumulative CO_2 emissions, while E_{Esfb} is an adjustment for Earth system feedbacks from permafrost thaw and warming wetlands. This is essentially the same framework as SR1.5, except that in SR1.5 non-CO_2 warming was not separate, but rather included in TCRE, and the earth-system feedback adjustment was incorporated in the results of SR1.5 Table 2.2, but not included in "headline" estimates in its Summary for Policymakers (IPCC, 2018).

In practice, observations based $\Delta GSAT_{obs}$ (whether $\Delta GSAT_{period}$, $\Delta GSAT_{LOESS}$ or using another statistical technique) is used as an approximation of $\Delta GSAT_{F,anthro}$, following from the finding that observed and "human-induced" warming to date are approximately equivalent (Allen et al., 2018; Haustein et al., 2017). Thus, SR15 assessed $\Delta GSAT_{F,anthro}$ as 0.97°C in 2006-2015 relative to 1850—1900, based on the HadCRUT4 average for that decade (0.84°C) adjusted by the ratio between the equivalent CMIP5 blended-masked estimate (0.86°C) and CMIP 5 ΔGSAT (0.99°C), as stated in Box 2 of Rogelj et al. (2019).

Here we select the Global_3 GMST group and so do not need to rely on a model correction for the additional bias introduced by HadCRUT4's incomplete and changing geographic coverage, which necessitates a correction substantially larger than A_{blend}. Our central estimate for $\Delta GSAT_{F,anthro}$ is:

$$\Delta GSAT_{F,anthro} = A_{blend\_med}\Delta GMST_{Global\_3} \tag{16}$$

where A_{blend_med} is the median value from CMIP6 A_{blend} ensemble and $\Delta GMST_{Global\_3}$ is the $LOESS_{bsln}$ ΔGMST of the Global_3 group (based on the mean of $LOESS_{bsln}$ applied to each of the three series). It should be noted this is a very conservative adjustment, as it may not fully account for coverage bias in the early part of the instrumental record, and ignores the "ice edge effect" cooling bias introduced by the variable sea ice mask in NASA GISTEMP and Berkeley Earth, which would add an additional ~3% (Cowtan et al., 2015; Richardson et al., 2018).

SR1.5's likely total uncertainty in $\Delta GMST_{obs}$ (and derived ΔGSAT) was ±0.12°C. Here we derive likely observation-based $\Delta GSAT_{LOESS}$ using Gaussian approximations to the observational, dataset spread and statistical fit uncertainties in the following steps (tests and details in Supplementary Table S3):

1. The Cowtan-Way ensemble spread is our best estimate of observational parametric ΔGMST uncertainty, so for each dataset its standard deviation is combined in quadrature separately with (i) the dataset-specific statistical 1σ uncertainty and (ii) the CSIRO Mk3.6.0 large ensemble standard deviation.

2. For ΔGSAT, the CMIP6 A_{blend} ensemble standard deviation is taken as the uncertainty value, and combined in quadrature with the results of 1.

3. We estimate a 17—83 % range by calculating those percentiles for each dataset following a Gaussian assumption, i.e. ±0.954σ from the mean, and then selecting the lowest 17 % and higher 83 % value from across the datasets.

There is no universally accepted method of accounting for dataset spread. We adopt step 3 as a conservative approach, however, by reporting the separate dataset uncertainties as described in Section 2.2.2 other groups can replicate or develop alternative uncertainty estimates.

We take Rogelj et al. (2019)'s, T_{nonCO_2} of 0.1°C (0.2°C) for T_{lim} of 1.5°C (2°C), and E_{Esfb} of 100 Gt CO_2 through 2100. TCRE percentiles are based on AR5's likely range of 0.2–0.7°C per 1,000 Gt CO_2 (Collins et al., 2013), as in Nauels et al (2019). SR1.5 included alternative carbon budgets using a lower T_{hist} from the average of the blended GMST datasets with no GSAT adjustment. Our alternative uses the Global_3 average without the GSAT adjustment. To contextualize the remaining budget against cumulative emissions to date we include data and uncertainties from the 2019 Global Carbon Budget (Friedlingstein et al., 2019).

3 Results

3.1 Long term ΔGMST analysis

Figure 1 compares $LOESS_{md}$ and OLS ΔGMST from 1880—2019 with associated 5—95% uncertainties (Fig. 1a). Figure 1b shows that the LOESS fit residuals follow our assumed ARMA(1, 1), which is necessary to justify our error correction and is not true for OLS (Figure 1c). Our full set of observational long-term ΔGMST estimates are given in Table 2.

$\Delta GMST_{OLS}$ is always lower than $\Delta GMST_{LOESS}$, with some central OLS ΔGMST estimates lying below the LOESS uncertainty range or nearly so (Cowtan-Way, Berkeley Earth). Datasets are similarly ranked for both OLS and $LOESS_{md}$ over 1880—2019, from HadCRUT4 (0.96, 0.99) to Berkeley Earth (1.05, 1.14). The Global_3 interpolated series exhibit a greater relative difference than the non-global series; the Berkeley Earth and HadCRUT4 $LOESS_{md}$ difference is 0.21°C, but only 0.13°C for OLS. Thus OLS not only renders lower ΔGMST, but also de-emphasizes the differences between the datasets.

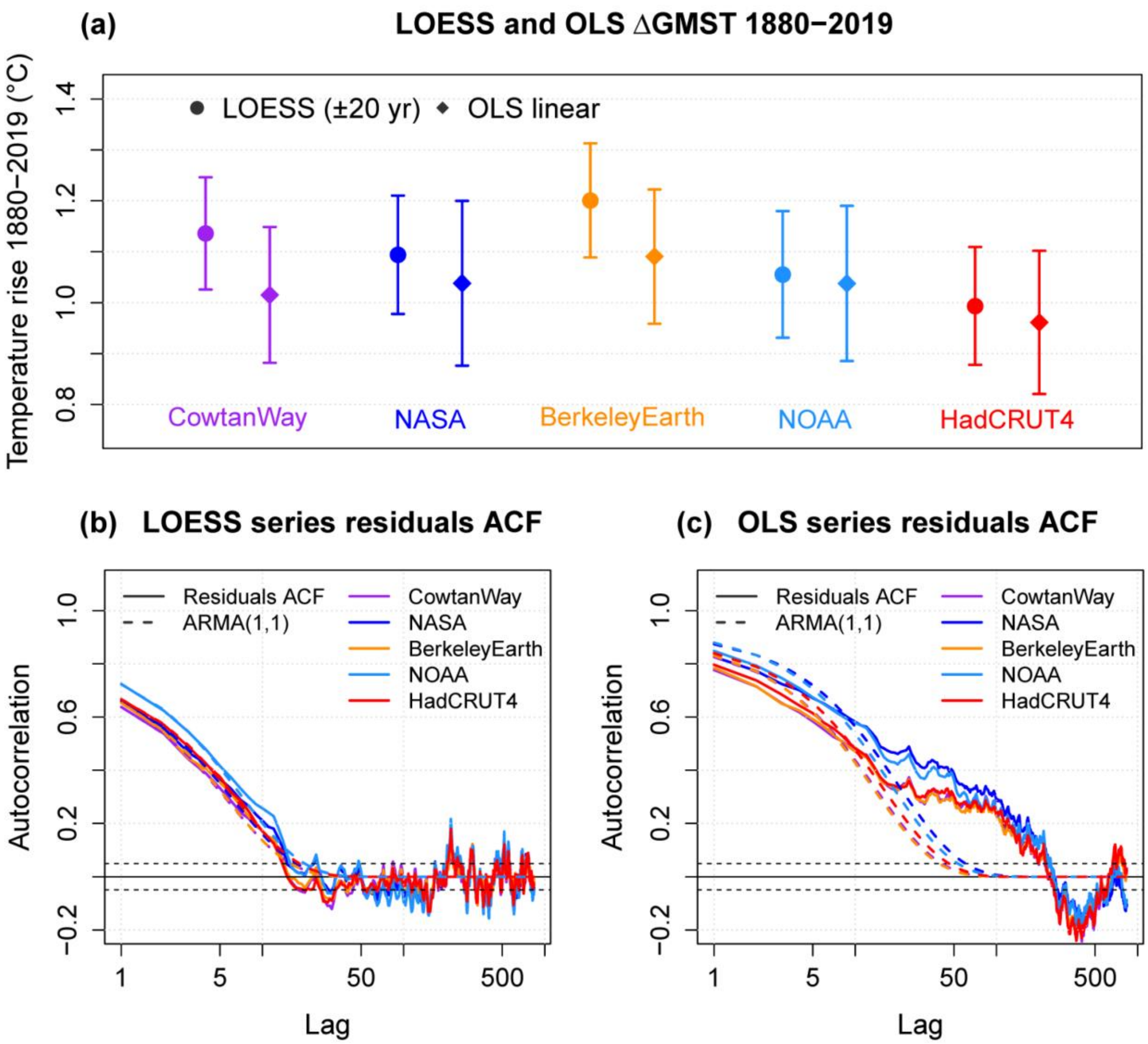

Figure 1: 1880-2019 warming estimates from five GMST series. (a) LOESS (span ± 20 years) and OLS trends with 5-95% statistical fit uncertainty are shown for Cowtan -Way (purple), NASA GISTEMP (blue), Berkeley Earth (orange), NOAA GlobalTemp (light blue) and HadCRUT4 (red) over 1880-2019. **(b)** The autocorrelation function (ACF) of the LOESS fit residuals are shown for each series (solid lines), along with the ACF of the estimated ARMA(1, 1) model used to correct for autocorrelation. **(c)** As in (b) except for OLS linear trend.

Table 2: Observed increase in GMST (°C) in datasets and dataset groupings. Numbers in square brackets correspond to 5–95% statistical fit uncertainty ranges, accounting for autocorrelation in fit residuals. Round brackets denote observational parametric uncertainty where available (HadCRUT4, Cowtan-Way). NOAA and NASA are each aligned to match 1880—1900 mean of the other three datasets. Best estimates from three full global series are denoted by *. Group mean estimates (in bold) are given with uncertainties encompassing the spread from lowest 5% to highest 95%. For the Global_3 group, the observational uncertainty is from Cowtan-Way, expanded by the spread of the three central estimates.

Period: / ***Series:***	**1850-1900 to 2019**	**1850-1900 to 2010-2019**	**1880 - 2019**	
	LOESS$_{bsln}$	**Latest decade**	**LOESS$_{md}$**	**Linear**
HadCRUT4	1.02 [0.93 - 1.11] *(0.97 – 1.07)*	0.93 (0.88 - 0.98)	0.99 [0.88 - 1.11] *(0.94 – 1.04)*	0.96 [0.82 - 1.10] *(0.92 – 1.03)*
NOAA GlobalTemp	1.09 [0.98 - 1.19]	0.99	1.06 [0.93 - 1.18]	1.04 [0.89 - 1.19]
NASA GISTEMP	1.12 [1.02 - 1.22]	1.01	1.09 [0.98 - 1.21]	1.04 [0.88 - 1.20]
Cowtan & Way	1.12 [1.04 - 1.21] *(1.05 – 1.19)*	1.01 (0.95 - 1.09)	1.14 [1.03 - 1.25] (1.08 – 1.21)	1.02 [0.88 - 1.15] *(0.94 – 1.09)*
Berkeley Earth	1.19 [1.10 - 1.27]	1.08	1.20 [1.09 - 1.31]	1.09 [0.96 - 1.22]
All Operational (OpAll)	**1.11** **[0.93 - 1.27]**	**1.00**	**1.10** **[0.88 - 1.31]**	**1.03** **[0.82 - 1.22]**
Full Global (Global_3) *	**1.14 *** **[1.02 – 1.27]** ***(1.05 – 1.26)***	**1.03**	**1.14** **[0.98 - 1.31]**	**1.05** **[0.88 - 1.22]**

For $LOESS_{bsln}$ to 2019, there are minor differences in assessed values but no changes in dataset rankings versus $LOESS_{md}$ 1880—2019. $LOESS_{bsln}$ is generally ~0.1 °C higher than 1850—1900 to 2010—2019 ΔGMST, reflecting the five-year offset and ~0.2 °C/decade recent warming (2010—2019 is centered at the end of 2014). At 1.14°C, Global_3 $LOESS_{bsln}$ ΔGMST to 2019 is 0.03°C higher than OpAll average, reflecting a 0.09°C difference with the mean of the two reduced coverage series from HadCRUT4 and NOAA GlobalTemp. The 1880—2019 $LOESS_{md}$ discrepancy is even wider: 0.09°C for NOAA and 0.15°C for HadCRUT4. $LOESS_{bsln}$ statistical fit uncertainties are smaller than $LOESS_{md}$ or OLS, reflecting the smaller uncertainty of departure from the 1850—1900 mean rather than a single point (as noted in Section 2.2.2).

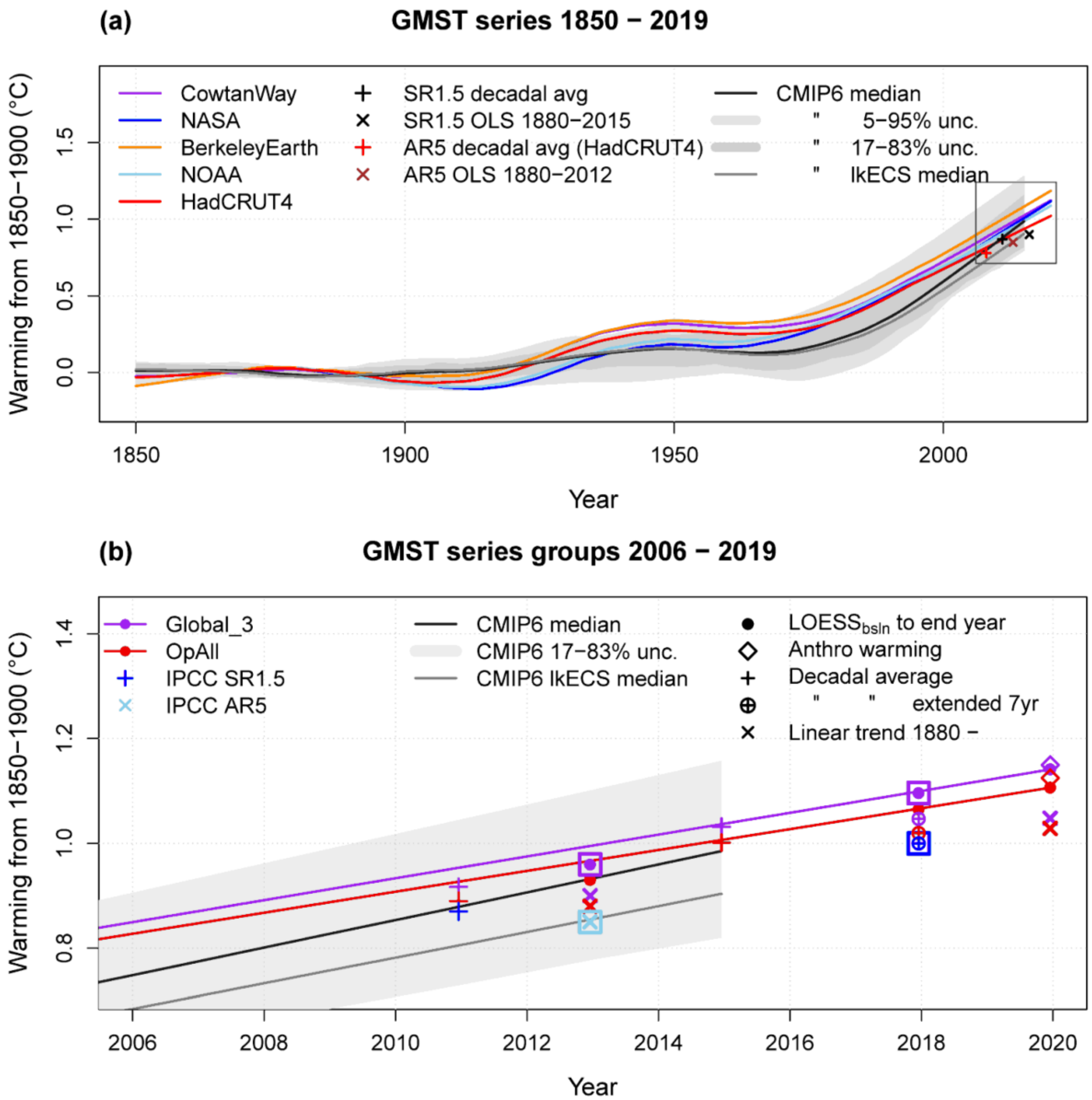

Figure 2: GMST series and group surface warming estimates. (a) Monthly series and multi-decadal $LOESS_{bsln}$ ΔGMST (span ± 20 years) are shown for HadCRUT4 (red), NOAA GlobalTemp (light blue), NASA GISTEMP (blue), Cowtan-Way (purple) and Berkeley Earth (orange), together with OLS and period estimates from IPCC AR5 and SR15. NOAA GlobalTemp and NASA GISTEMP have been matched to the longer datasets over the

overlapping 1880—1900 period. Also shown are 24 CMIP6 SAT-SST model runs, blended following Cowtan et al (2015) and Richardson et al (2018). **(b)** LOESS$_{bsln}$ (solid line with filled circle) is shown for two GMST groupings: Global_3 (purple) and OpAll (dark red).Also shown are selected additional warming estimates: anthropogenic following Haustein et al (2017) (diamonds), decadal average (crosses) and OLS linear trend from 1880 (x-crosses). Recent IPCC ΔGMST estimates are highlighted by large squares: AR5 OLS to 2012 (light blue) and SR1.5 2006-2015 mean extended to 2017 (blue), together with corresponding Global_3 LOESS$_{bsln}$ ΔGMST (purple).

The observation-based and CMIP6 blended ensemble LOESS$_{bsln}$ (Figure 2a) show broadly similar changes: a rise to 1950, a 1950—1975 flattening, and strong post-1975 warming. The observations show stronger 1920—1950 warming, especially in the three HadSST-based series, and weaker post-1975 warming.

Separate tests showed that derived ΔGMST$_{LOESS}$ was similar when restricting CMIP6 spatial coverage to that of Berkeley Earth, so we take the CMIP6 blended ensemble as directly comparable to the Global_3 series (Figure S14). The Global_3 rise of 1.14°C is just above the median CMIP6 estimate extended linearly to 2019, 1.12°C [0.91 – 1.41]. However, the Global_3 current trend of 0.20°C/decade (as estimated by the LOESS$_{bsln}$ slope at the 2019 end point) is lower than CMIP6's 0.26°C/decade [0.18 – 0.38] or the likely ECS sub-ensemble's 0.25°C/decade [0.18 – 0.29].

In general, Figure 2(a) shows LOESS$_{bsln}$ ΔGMST from the five updated observational datasets (coloured lines) are at or above recent IPCC long-term observational ΔGMST estimates (represented by crosses and x-crosses). Figure 2(b) affords a closer view of recent ΔGMST estimates, including group LOESS$_{bsln}$ calculated to 2012 and 2017 for direct comparison to IPCC AR5 and SR1.5. As previously stated, AR5's main estimate of 0.85°C was from linear OLS on the datasets available then. Since the mean 1880—2012 OLS trend for OpAll is 0.89°C and LOESS$_{bsln}$ is 0.93°C, ΔGMST methodology accounts for half of the discrepancy between AR5's 1880—2012 estimate and our OpAll based estimate. The 2012 gap is even wider for the Global_3 group. OLS to 2012 is 0.90°C and LOESS$_{bsln}$ is 0.96°C; that gap continues to grow, reaching 0.09°C in 2019.

The SR1.5 2006-2015 mean ΔGMST$_{period}$ of 0.87°C, centered at the end of 2010, was extended to the most recent year (2017) to provide a then current estimate of 1.0°C (Section 1.2.1.3 in Allen et al., 2018). The same extension to 2017 applied to the updated series shows a 0.03°C gap with LOESS$_{bsln}$ evaluated in 2017. This discrepancy may be related to internal variability suppressing early 2000s warming; the period difference estimate based on the most recent decade then available (2008-2017) shows no such discrepancy with LOESS$_{bsln}$. Both LOESS$_{bsln}$ and period estimates are in good agreement with the slightly higher Haustein human-induced warming ΔGMST$_{F,anthro}$ estimates.

Figure 3 compares Global_3 LOESS$_{bsln}$ and period ΔGMST in more detail. Since IPCC SR1.5 explicitly considered the 2006-2015 mean as a proxy for the 1996-2025 average (relative to 1850—1900), we consider the centered 20-year average and a 30-year "extended" average assuming the current linear 30-year trend continues over the next 15 years. We estimate that the 1979—2019 warming has been approximately linear (see Table S2 showing OLS-LOESS agreement over this period), and the large ensembles also imply minor errors from assuming linearity through 2025. Figure 3a shows that in general LOESS$_{bsln}$ departs less from the eventual 20 and 30 year average than the decade mean and confirms that 2006-2015 was affected by an early 2000s slowdown. $LOESS_{bsln}$ has more stability relative to anthropogenic warming estimates (Figure 3b) with near-identical concordance with $\Delta GMST_{F,anthro}$ since 2003, and has lower RMSE relative to the longer period averages since the late 1990s (Figure 3c, 3d).

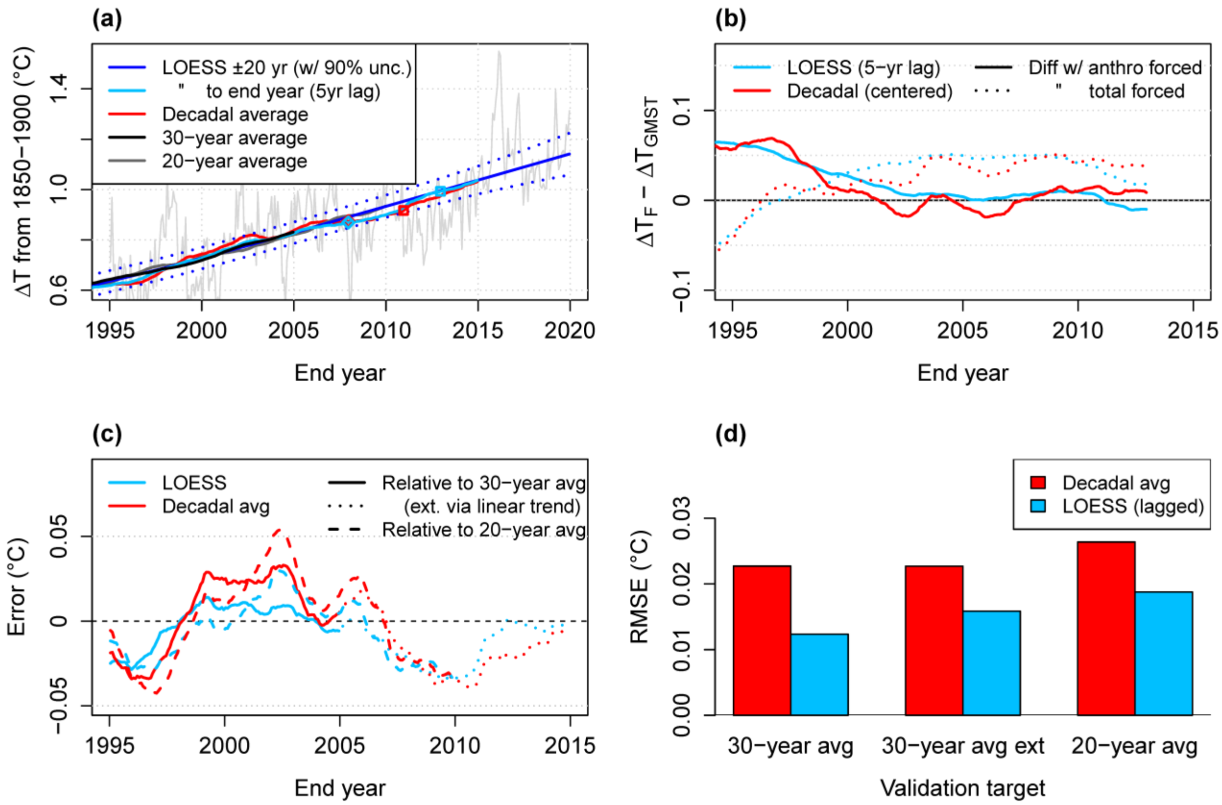

Figure 3: ΔGMST estimation method validation based on average of 3 global series. (a) $LOESS_{bsln}$ to 2019 (blue) is shown with 5-year lagged LOESS (light blue), decadal average (red), 20-year average (light gray) and 30-year average (black). LOESS (light blue) versus decadal (red) differences are shown with **(b)** forced warming estimates following Haustein et al. (2017) and **(c)** validation targets (30-year average, 30-year average extended with linear trend and 20-year average). **(d)** RMSE is calculated from errors shown in (c).

The equivalent performance evaluation of long-term Global_3 $LOESS_{bsln}$ versus OLS ΔGMST in Figure S15 shows a growing cool bias in OLS relative to the 20 and 30-year average from 1995 on (Figure S15a) and thus much higher RMSE than $LOESS_{bsln}$ relative to the longer period averages (Figure S15d).

Global_3 $LOESS_{bsln}$ ΔGMST to 2019 is our main input for subsequent analysis such as remaining carbon budget, for which combined 17—83 % uncertainty is required; the combined statistical and observational uncertainty calculated following the method outlined in Section 2.2.4 yields Global_3 ΔGMST of 1.14°C [1.05 – 1.25]**.**

3.2 Large Ensemble Validation

Figure 4(a,d) shows the MPI-GE and CSIRO Mk3.6.0 annual SAT range, individual $LOESS_{md}$ fits and $GSAT_F$ estimate, Figure 4(b,e) contains example LOESS and OLS fits to a single simulation and Figure 4(c,f) shows the forced, LOESS and OLS ΔGSAT estimates through 2019 for each start year from 1850—1980.

The $\Delta GSAT_F$ and LOESS ΔGSAT agree well outside of periodic $\Delta GSAT_F$ spikes from volcanic eruptions, i.e. when the forced change is smooth over our ±20 year window, such that $\Delta GSAT_{LOESS} \approx \Delta GSAT_F$. For changes from the 19th century to recently, the IPCC AR5 estimates of solar forcing change are negligible compared with anthropogenic forcing so longer-term $\Delta GSAT_F$ should approximate the $\Delta GSAT_{F,anthro}$ used in our later carbon budget calculation. Meanwhile, OLS is biased relative to $\Delta GSAT_F$ in the long term, and is more sensitive to internal variability in the short term, e.g. for 1990—2019 OLS ensemble spread is 62 % (MPI-ESM) or 26 % (CSIRO Mk3.6.0.) larger than LOESS ensemble spread.

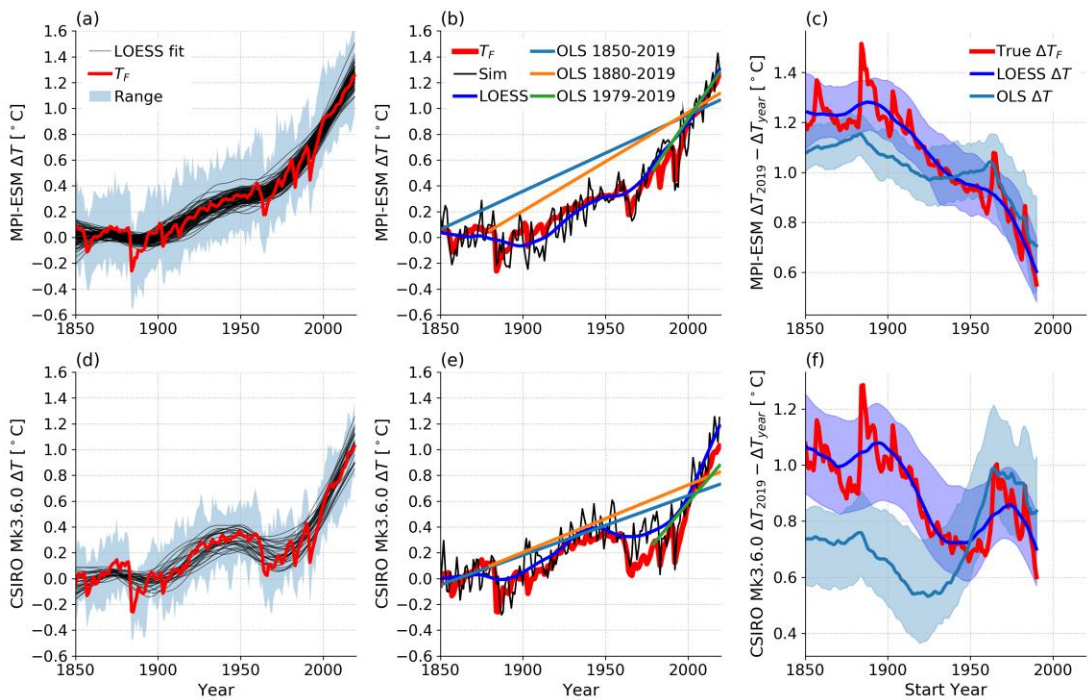

Figure 4. (a) MPI-GE SAT outputs, full ensemble range is shaded, each simulation's LOESS fit is in grey and the ensemble mean (our estimate of $GSAT_F$) is in red. (b) example of fits applied to a single simulation (black) including LOESS (dark blue) and OLS over three different periods (straight lines) with $GSAT_F$ in red. OLS lines are shifted up so that their end points correspond to the relevant ΔGMST for ease of comparison. (c) calculated ΔGSAT for $GSAT_F$ (red), based on the LOESS fit (dark blue) and based on OLS (cyan). For the fits, the lines are the ensemble median and the shaded regions the 5—95 % range.(d—f) as (a—c) but for the CSIRO Mk 3.6.0 ensemble.

Table 3 contains the large ensemble ΔGSAT estimates. For periods like 1850—1900 to 2010—2019, we use Section 2.2.2's $LOESS_{bsln}$ approach while OLS is fit between the middle of each period. In both ensembles LOESS performs similarly to the period difference with the 5^{th}, 50^{th} and 95^{th} percentiles of the ensemble LOESS and period difference calculations all agreeing to within ≤0.02 °C. LOESS slightly outperforms centered period differences evaluated from 1850—1900 to end periods ranging from 1986—1995 through 2010—2019 when validated against 30-year average (see Figure S16). This validates LOESS performance, and Table 3 shows an advantage over period means since its calculation can be extended to the latest available year without greatly inflated uncertainty. The 0.06—0.10 °C discrepancies in the third column of Table 3 for 1880—2019 LOESS-GSAT$_F$ are likely because the LOESS window centred at 1880 captures Krakatoa's large post-1883 cooling, thereby reducing the 1880 LOESS estimate and increasing its 1880—2019 ΔGMST. These results show that such biases are period-dependent, are indeed negligible for 1850—1900 to 2019 in these models, and support our choice of time periods in the analysis using observational datasets.

As our carbon budget calculations include an internal variability error component, we consider ensemble spread and statistical fit uncertainties as candidates and compare the LOESS$_{bsln}$ ensemble 83rd minus 17th percentile and the statistical 17—83 % ranges for each run over 1850—1900 to 2019. The CSIRO Mk3.6.0 17—83 % ensemble spread in GSAT LOESS$_{bsln}$ is 0.22 °C. This is larger than the median ensemble member's statistical range (0.18 °C) and similar to the largest individual ensemble member range (0.22 °C). For MPI-ESM the ensemble spread (0.11 °C) is smaller than the median statistical uncertainty (0.16 °C) and is marginally lower than the smallest member value (0.12 °C). For the internal variability component of ΔGSAT uncertainty in our carbon budgets we present results both using statistical uncertainty (derived only from observational data) and a more conservative estimate using the ±0.11 °C CSIRO Mk3.6.0 ensemble spread.

This large ensemble analysis has:

(i) provided limited support for our LOESS-based statistical uncertainty estimates being similar to model variability,

(ii) shown that LOESS matches or exceeds period difference performance while having lower long-term bias and short-term uncertainty than OLS,

(iii) verified that LOESS reliably reproduces ΔGSAT$_F$ outside of years immediately following large volcanic eruptions, particularly supporting our LOESS$_{bsln}$ results as an estimate of ΔGSAT$_{F,anthro}$.

Table 3. Long-term ΔGSAT estimated for various periods for the ensemble mean T$_F$, plus the ensemble medians and 5—95 % ranges for estimates based on LOESS, OLS or taking the mean of the raw SAT outputs. Uncertainties in T$_F$ differences are derived by treating T$_F$ as a sample mean and assuming the ensemble members follow a Gaussian distribution in any given year. The period errors are then combined in quadrature.

	MPI-ESM ΔGSAT[°C] median [5—95 %] *[17—83 %]*		
Method	1850-1900 to 2010-2019	1850-1900 to 2019	1880 to 2019
T$_F$	1.15 [1.15-1.16] *[1.15-1.16]*	1.25 [1.23-1.28] *[1.24-1.27]*	1.20 [1.17-1.23] *[1.18-1.22]*
LOESS	1.16 [1.07-1.24] *[1.11-1.21]*	1.25 [1.15-1.36] *[1.21-1.32]*	1.26 [1.15-1.36] *[1.20-1.31]*
OLS	1.02 [0.93-1.12] *[0.97-1.07]*	1.13 [1.04-1.23] *[1.08-1.18]*	1.15 [1.06-1.23] *[1.10-1.20]*
Individual runs	1.15 [1.07-1.24] *[1.11-1.20]*	1.24 [1.04-1.48] *[1.12-1.40]*	1.20 [0.92-1.50] *[1.04-1.39]*
	CSIRO Mk3.6.0 ΔGMST[°C]		
T$_F$	0.92 [0.90-0.93] *[0.91-0.92]*	1.03 [0.99-1.07] *[1.00-1.05]*	0.93 [0.88-0.98] *[0.90-0.96]*
LOESS	0.93 [0.79-1.04] *[0.82-1.01]*	1.05 [0.89-1.18] *[0.90-1.12]*	1.03 [0.84-1.16] *[0.91-1.10]*

OLS	0.63 [0.46-0.72] *[0.52-0.70]*	0.73 [0.56-0.85] *[0.61-0.82]*	0.75 [0.58-0.87] *[0.64-0.83]*
Individual runs	0.91 [0.78-1.04] *[0.83-1.00]*	1.03 [0.81-1.22] *[0.86-1.12]*	0.94 [0.66-1.15] *[0.76-1.05]*

3.3 Global SAT Estimate and Remaining Carbon Budget

We now convert our best estimate $\Delta GMST_{LOESS}$ of 1.14°C [1.05 – 1.25] (17—83% uncertainty) to an equivalent $\Delta GSAT_{LOESS}$ as outlined in Section 2.2.4. Our CMIP6 ensemble $LOESS_{bsln}$ A_{blend} ratio $\Delta GSAT_{LOESS}/\Delta GMST_{LOESS}$ reflects an increase of $\Delta GSAT_{LOESS}$ over full-coverage $\Delta GMST_{LOESS}$ of 5.8% [4.4, 7.2] in 2014, i.e. long-term near-surface air temperature warming is 5.8% greater than our blended estimate. This A_{blend} estimate is very similar to equivalent CMIP5-based estimates, but much lower than the 24% derived in CMIP5 for 1861—1880 to 2006—2015 using a HadCRUT4-like masking and blending algorithm (Richardson et al., 2016). This is due to the different handling of sea ice and the incorporation of complete (unadjusted) spatial coverage in the A_{blend} calculation**.**

Combining this ratio and its uncertainty with our Global_3 $\Delta GMST_{LOESS}$, as outlined in Section 2.2.4, we obtain $\Delta GSAT_{LOESS}$ of 1.21°C [1.11—1.32] from 1850—1900 to 2019, a lower uncertainty than the equivalent SR1.5 estimate of ±0.12°C (Section 1.2.1.2 in Allen et al., 2018). The conservative CSIRO-based internal variability yields a wider $\Delta GSAT_{LOESS}$ range of 1.07—1.37 °C. These estimates all represent uncertainty in total forced warming; however, uncertainty in anthropogenic warming was estimated to be still higher at ±0.2°C (Section 1.2.1.3 in Allen et al., 2018). The equivalent $LOESS_{bsln}$ HadCRUT4 estimate using the SR1.5 correction of ~15% yields slightly lower $\Delta GSAT_{obs}$ of 1.17°C, and the updated SR1.5 2006—2015 estimate extended to end of 2019 is 1.15°C. Finally, A_{blend} corrected $LOESS_{bsln}$ HadCRUT4 yields 1.08°C; the difference of 0.13°C with our main $\Delta GSAT_{LOESS}$ primarily reflects HadCRUT4 coverage bias, as well as a small sea ice edge effect. The other carbon budget calculation components also have large uncertainties. Cumulative emissions to end of 2019 are 2320 ±230 $GtCO_2$ (Friedlengstein et al., 2019), while non-CO_2 uncertainties are even higher (see Table 2.2 in Rogelj et al., 2018). Although no formal methods exist to combine these uncertainties, Rogelj et al (2018) estimated overall uncertainty of ±50% in SR1.5 remaining carbon budgets.

Figure 5 shows the calculation for the remaining carbon budget with a 66% chance to stay below 1.5°C, along with the historical cumulative CO_2 emissions and temperature change.

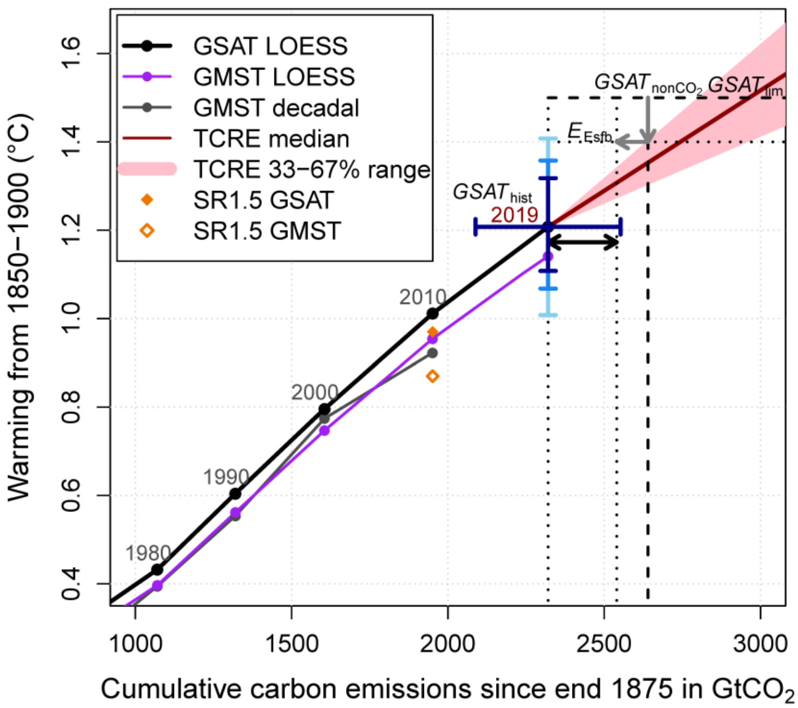

Figure 5: Global temperature change from 1850–1900 versus cumulative CO_2 emissions. The smoothed temperature response from the Global3 blended GMST group as decadal average (blue) and $LOESS_{bsln}$ trend (purple) are shown relative to cumulative CO_2 emissions from Friedlingsten et al (2019). The thick black line shows the Global3 GSAT $LOESS_{bsln}$ trend, obtained by adjusting GMST by the ratio of GSAT and blended GMST historical runs from an ensemble of 24 CMIP5 models. The pink shaded plume and dark red line are estimated temperature response to cumulative CO_2 emissions (TCRE) from the beginning of 2020 on. Also shown are other remaining carbon budget factors, T_{nonCO_2} and E_{Esfb} (gray arrows). The thick black double arrow represents the remaining carbon budget for 66% chance of remaining below 1.5°C. Vertical error bars show ΔGSAT combined observational and statistical uncertainty (dark blue), combined observational and internal variability derived from CSIRO ensemble (medium blue) and estimated uncertainty in anthropogenic warming (light blue).

Our remaining carbon budgets incorporate the SR1.5 Table 2.2 100 $GtCO_2$ adjustment for earth-system feedbacks (CO_2 and CH_4 release from warming wetland and permafrost thaw), following recent practice established in Rogelj et al. (2019) and Nauels et al. (2019). Carbon budgets excluding this term are therefore 100 $GtCO_2$ higher, as in the SR1.5 "headline" remaining carbon budget of 420 $GtCO_2$ (IPCC, 2018) to remain under 1.5°C (with 66% chance).

The remaining carbon budgets from the start of 2020 for a 66% (50%) chance to stay below 1.5°C and 2.0°C are 220 (350) $GtCO_2$ and 880 (1270) $GtCO_2$ respectively (rounded to nearest 5 $GtCO_2$). Given current annual emissions of just over 40 $GtCO_2$, the 66% 1.5°C remaining carbon budget is only ~15 $GtCO_2$ lower than the equivalent carbon budgets including earth-system feedbacks in SR1.5 Table 2.2 (320 $GtCO_2$ from 2018) and Nauels et al (235 $GtCO_2$ from 2020).

However, our 50% 1.5°C carbon budget is ~45 $GtCO_2$ below those two studies. This follows from the slightly higher $\Delta GSAT_{obs}$ found in this study, combined with an identical TCRE spread starting in 2020 rather than the SR1.5 reference period centered at the start of 2011. In effect, the up-to-date $LOESS_{bsln}$ estimate of $\Delta GSAT_{obs}$ reduces the contribution of TCRE uncertainty, as there is less ΔT "to go".

SR1.5's secondary carbon budgets used the average ΔGMST through 2006—2015 to obtain a 66 % chance of staying below 1.5 °C resulting in an equivalent budget of 470 $GtCO_2$ from 2018 (i.e. 385 $GtCO_2$ from 2020). Our alternative budget using Global_3 $\Delta GMST_{LOESS}$ instead of $\Delta GSAT_{LOESS}$ is 305 $GtCO_2$ from 2020. This large difference relative to SR1.5 is unsurprising as the Global_3 series show more historical warming whereas the SR1.5 $\Delta GMST_{period}$ average included HadCRUT4 and its more substantial coverage bias. We also note that an OLS 1880—2019 ΔGMST basis would imply even higher 1.5 °C 66% remaining carbon budgets of 455 $GtCO_2$ (Global_3) or 485 ($GtCO_2$ (OpAll).

4 Discussion and Conclusions

We have explored the range of warming estimates since the late 19th century across different observational series using multiple estimation methodologies, focusing on the Global_3 subset of extensively interpolated datasets (NASA GISTEMP, Cowtan-Way and Berkeley Earth). Our main $LOESS_{bsln}$ Global_3 ΔGMST since 1850—1900 is, to our knowledge, the first such estimator that (i) integrates robust statistical uncertainties, with fit residuals following the assumed noise process, (ii) has been extended to provide a corresponding $\Delta GSAT_{LOESS}$ since 1850—1900, including combined observational and internal variability uncertainties, and (iii) has been validated against output from model large ensembles.

IPCC SR1.5 reported $\Delta GMST_{period}$ of 0.87°C to 2006—2015 using four datasets (1.0°C when extended to 2017) and estimated $\Delta GSAT_{period}$ of 0.97°C by adjusting one dataset (HadCRUT4) for biases related to incomplete coverage and sea-air temperature differences, effectively discarding the other three. The ensuing carbon budget calculation included cumulative emissions up to 2017, necessitating an implicit extension of $\Delta GSAT_{period}$ to that date. The simplicity and coherence of our "up-to-date" $\Delta GMST_{LOESS}$ and $\Delta GSAT_{LOESS}$ estimates represent a clear advance over the IPCC ΔGMST period difference and ΔGSAT derivation methods. Not only is $LOESS_{bsln}$ generally an unbiased $\Delta GMST_F$ estimator outside periods of volcanism, but the method includes a more consistent and intuitive baseline alignment of datasets beginning in 1880 and maintains the previously stated advantage of including statistical uncertainty derived using a noise model consistent with the data. Moreover, validation tests with observations and the large ensembles confirm $LOESS_{bsln}$ results in lower biases relative to $\Delta GSAT_F$ and lower susceptibility to natural variation. None of this is surprising considering that the IPCC period difference method is essentially a 10-year moving average.

Another key difference with IPCC SR1.5 is our consistent use of the Global_3 datasets with extensive spatial interpolation. As previously noted in section 2.1.1, these datasets are demonstrably more representative of global climate change and require smaller and less uncertain adjustments (~6%) to obtain $\Delta GSAT_{LOESS}$ from $\Delta GMST_{LOESS}$, in contrast to the 15% adjustment applied to HadCRUT4 $\Delta GMST_{period}$ in IPCC SR1.5. The Global_3 datasets give 0.12

°C more warming than HadCRUT4 from 1850—1900 and the divergence related to unmitigated bias coverage may well grow, as the Global_3 $LOESS_{md}$ trend is now 0.03°C/decade higher than HadCRUT4's 0.17 °C/decade. Focusing on the three Global_3 datasets and our robust $LOESS_{bsln}$ estimator dramatically reduces the spread between ΔGMST estimates: the inter-dataset spread in Global_3 $LOESS_{bsln}$ 1850—1900 to 2019 ΔGMST is only 0.07°C. Including the non-global datasets increases the $LOESS_{bsln}$ spread to 0.17 °C, and including OLS and $LOESS_{md}$ trend methodologies increases the spread to 0.27 °C: from 0.93°C (OLS for HadCRUT4) to 1.20°C (Berkeley Earth $LOESS_{md}$).

SR1.5 also reported 1880—2012 and 1880—2015 linear trend ΔGMST, but mainly to provide "traceability" to the IPCC AR5. In contrast, AR5's main estimate of 0.85°C was based on the mean linear trend of available datasets, while HadCRUT4 2003—2012 period difference from 1850—1900 ΔGMST estimate was a primary input for further analyses such as future projections (Collins et al., 2013) and attribution (Bindoff et al., 2013).

If IPCC AR6 follows AR5 and provides both period difference and point-to-point trends for datasets beginning in 1850, that would imply the three post-1850 datasets would form the basis for 2010—2019 period $\Delta GMST_{obs}$ relative to 1850—1900. As noted above though, $LOESS_{bsln}$ to 2019 offers a superior alternative. Since HadCRUT4 clearly does not meet our "quasi-global" criterion, we omit it as a direct component of $\Delta GMST_{LOESS}$. Nevertheless, HadCRUT4 and its underlying land and ocean datasets (CRUTEM4 and HadSST3) form the essential basis of Cowtan-Way, and HadSST3 is also a key component of Berkeley Earth. Following the precedent set in IPCC SR1.5, the ERSSTv5 based datasets starting 1880 should also be considered, using baseline matching over 1880—1900. Our Global_3 group member, NASA GISTEMP is an obvious choice for inclusion, while NOAA GlobalTemp could be omitted according to our global coverage criterion. However, that case is less clear cut than HadCRUT4 due to NOAA's complicated spatial coverage. Once again, though, NOAA's GHCNv4 and ERSSTv5 datasets would still be present as they form the essential basis of NASA GISTEMP.

The recent release of HadCRUT5 (Morice et al., 2020) will certainly inform future regular updates of our main $\Delta GMST_{LOESS}$ and $\Delta GSAT_{LOESS}$ estimates. HadCRUT5 features sophisticated kriging interpolation, resulting in virtual coverage similar to Berkeley Earth, and incorporates updated datasets for land (CRUTEM5; Osborn et al., 2020) and ocean (HadSST4; Kennedy et al., 2019). We give a preliminary evaluation of the eventual effect of HadCRUT5 (and HadSST4) in Table S4. The incorporation of HadSST4 (instead of HadSST3) into Cowtan-Way and Berkeley Earth results in a noticeable increase in $\Delta GMST_{LOESS}$, while results for HadCRUT5 are nearly identical to Cowtan-Way/HadSST4.

Since observational datasets beginning in 1880, such as NASA GISTEMP, potentially could be included alongside the three datasets starting in 1850, $LOESS_{bsln}$ ΔGMST arguably renders 1880—2019 $\Delta GMST_{OLS}$ redundant in IPCC AR6. However, AR5 also compared long-term $\Delta GMST_{OLS}$ trends starting from 1880 to short-term trends starting from mid-century or later. Our results reinforce that 1880—2019 linear trend is inconsistent with $LOESS_{md}$ 1880—2019 ΔGMST. The bias of long-term OLS ΔGMST was confirmed in analysis of two large ensembles, which also showed that it has 26—62 % larger uncertainty than $LOESS_{md}$ for recent 30-year trends. As seen in Table S2, observed OLS trends from 1951 have wider uncertainty than the corresponding $LOESS_{md}$ estimates and show evidence of warm bias as well (for example the NASA GISTEMP 1951—2019 OLS is almost identical to 1880—2019). We therefore recommend $LOESS_{md}$ over linear trend for both long-term (> 120 years) and short-term (30-70 years) intervals.

$LOESS_{bsln}$ statistical uncertainties represent another opportunity for AR6. If $\Delta GMST_{LOESS}$ is close enough to $\Delta GMST_F$ then with an appropriate noise model the $\Delta GMST_{LOESS}$ uncertainty due to internal variability could be derived from the LOESS residuals. We combined this with observational uncertainty and carried it forward directly to $\Delta GSAT_{LOESS}$ for carbon budget calculations, but it could also be used for other follow-on analyses. The median statistical uncertainties from the large ensemble runs are within 25% of the ensemble spreads, and the residual autocorrelation structure implies potential for this approach.

However, global climate models may not capture long-term internal variability (Brown et al., 2015). For example, recent Pacific changes may indicate stronger real-world multi-decadal variability (e.g. England et al., 2014), although consensus is lacking (Seager et al., 2019). We take no position on the ability of models to generate this variability, only note that it has been studied in CMIP5 (e.g. Brown et al., 2015) and CMIP6 (e.g. Parsons et al., 2020) and report on how errors would affect our conclusions. Substantial internal variability on ±20 year timescales or longer would result in underestimated LOESS uncertainties. By contrast, large forced changes on shorter timescales, such as due to volcanism, would increase the uncertainties. Nevertheless, our method derives uncertainties directly from observations and so may have advantages over approaches that rely on model outputs or estimated forcings (Otto et al 2015; Haustein et al., 2017).

Given the above caveats we provided a more conservative ΔGSAT uncertainty incorporating the CSIRO model large ensemble spread and its pronounced internal variability. Since our $\Delta GMST_{LOESS}$ and $\Delta GSAT_{LOESS}$ estimates are close to observation-based anthropogenic warming, confirming a basic finding of IPCC SR1.5, we treat our $\Delta GSAT_{LOESS}$ as an estimate of $\Delta GSAT_{F,anthro}$, albeit with appropriately wider uncertainties. In general, our approach yields straightforward and up-to-date estimates of ΔGMST and ΔGSAT to inform remaining carbon budget calculations that incorporate appropriate ΔGSAT uncertainties .

To summarize, we argue strongly in favor of $LOESS_{bsln}$ ΔGMST using series with near-global coverage. Combining our statistical estimate of internal variability with dataset spread and dataset parametric uncertainty results in a best estimate of warming from 1850—1900 to 2019 of 1.14 °C [1.05 – 1.25] (17-83% uncertainty). Not only is this updated through 2019, rather than the prior-decade value of the IPCC's period mean difference, but it includes a potentially useful statistical fit uncertainty that is not readily or typically derived for period mean differences.

Our CMIP6-derived GSAT adjustment yields corresponding $\Delta GSAT_{LOESS}$ of 1.21°C [1.11–1.32] (17—83% uncertainty), implying a remaining carbon budget of ~220 $GtCO_2$ for a 66% chance that ΔGSAT since 1850—1900 remains below 1.5°C. This carbon budget is ~5.5 years of current emissions and is less than half the 455–485 $GtCO_2$ carbon budget implied by an OLS ΔGMST basis. Our ΔGSAT estimate uncertainty can be adapted to a desired interpretation of ΔGSAT, for example, as total or anthropogenic warming. All $\Delta GSAT_{LOESS}$ and $\Delta GMST_{LOESS}$ indices can be updated annually and are only dependent on the temperature datasets, yielding a set of transparent and easily communicated metrics to measure progress towards climate goals.

Acknowledgments and Data

The authors thank Andrew Dessler for provision of MPI-GE series. DCC thanks Shaun Lovejoy and Lenin Del Rio Amador for clarifying discussions.

MR's contribution was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration (80NM0018D004).

Berkeley Earth data are available from http://berkeleyearth.org/data/. Cowtan-Way data, including merged HadSST4 series, are available from http://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html. HadCRUT4.6 data are available from https://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html. HadCRUT5 data are available from https://www.metoffice.gov.uk/hadobs/hadcrut5/data/current/download.html. NASA GISTEMP data are available from https://data.giss.nasa.gov/gistemp/. NOAA GlobalTemp data are available from https://www.ncei.noaa.gov/data/noaa-global-surface-temperature/v5/access/timeseries/. CMIP6 data are available from https://esgf-node.llnl.gov/search/cmip6/.

References

Allen, M.R., Dube, O.P., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S., Kainuma, M., et al. (2018). Framing and Context. In V. Masson-Delmotte, et al. (Eds.), Global Warming of 1.5°C: An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Intergovernmental Panel on Climate Change.

Bindoff, N.L., Stott, P.A., AchutaRao, K.M., Allen, M.R., Gillett, N., et al., 2013: Detection and Attribution of Climate Change: from Global to Regional. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., at al. (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. Journal of Geophysical Research, 111(D12). https://doi.org/10.1029/2005jd006548

Brown, P. T., Li, W., & Xie, S.-P. (2015). Regions of significant influence on unforced global mean surface air temperature variability in climate models. Journal of Geophysical Research: Atmospheres, 120(2), 480–494. https://doi.org/10.1002/2014jd022576

Cahill, N., Rahmstorf, S., & Parnell, A. C. (2015). Change points of global temperature. Environmental Research Letters, 10(8), 84002. https://doi.org/10.1088/1748-9326/10/8/084002

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association, 74(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cleveland, W. S., & Grosse, E. (1991). Computational methods for local regression. Statistics and Computing, 1(1), 47–62. https://doi.org/10.1007/bf01890836

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1991), Local regression models. In J.M. Chambers, & T. Hastie (Eds.), *Statistical Models in S*, Wadsworth, Pacific Grove, Calif.

Collins M, et al. (2013) Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, eds Stocker TF, et al. (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA).

Cowtan, K., Hausfather, Z., Hawkins, E., Jacobs, P., Mann, M. E., Miller, S. K., et al. (2015). Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. Geophysical Research Letters, 42(15), 6526–6534. https://doi.org/10.1002/2015GL064888

Cowtan, K., Jacobs, P., Thorne, P., & Wilkinson, R. (2018a). Statistical analysis of coverage error in simple global temperature estimators. Dynamics and Statistics of the Climate System, 3(1). https://doi.org/10.1093/climsys/dzy003

Cowtan, K., Rohde, R., & Hausfather, Z. (2018b). Evaluating biases in sea surface temperature records using coastal weather stations. Quarterly Journal of the Royal Meteorological Society, 144(712), 670–681. https://doi.org/10.1002/qj.3235

Cowtan, K., & Way, R. G. (2014a). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. Quarterly Journal of the Royal Meteorological Society, 140(683), 1935–1944. https://doi.org/10.1002/qj.2297

Cowtan, K., & Way, R. G. (2014b). Update to "Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends". Temperature reconstruction by domain: version 2.0 temperature series. https://doi.org/10.13140/RG.2.1.4728.0727

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., et al. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. Nature Climate Change, 10(4), 277–286. https://doi.org/10.1038/s41558-020-0731-2

Dessler, A. E., Mauritsen, T., & Stevens, B. (2018). The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity. Atmospheric Chemistry and Physics, 18(7), 5147–5155. https://doi.org/10.5194/acp-18-5147-2018

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., et al. (2014). Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. Nature Climate Change, 4(3), 222–227. https://doi.org/10.1038/nclimate2106

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., et al. (2013). Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., at al. (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Forster, P. M., Maycock, A. C., McKenna, C. M., & Smith, C. J. (2019). Latest climate models confirm need for urgent mitigation. Nature Climate Change. https://doi.org/10.1038/s41558-019-0660-0

Foster, G., & Rahmstorf, S. (2011). Global temperature evolution 1979–2010, Environmental Research Letters, 6(4), 44022. https://doi.org/10.1088/1748-9326/6/4/044022

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., et al. (2016). ICOADS Release 3.0: a major update to the historical marine climate record. International Journal of Climatology, 37(5), 2211–2232. https://doi.org/10.1002/joc.4775

Friedlingstein, P., Jones, M. W., O., Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., et al. (2019). Global Carbon Budget 2019. Earth System Science Data, 11(4), 1783–1838. https://doi.org/10.5194/essd-11-1783-2019

Goodwin, P., Katavouta, A., Roussenov, V. M., Foster, G. L., Rohling, E. J., & Williams, R. G. (2018). Pathways to 1.5 °C and 2 °C warming based on observational and geological constraints. Nature Geoscience, 11(2), 102–107. https://doi.org/10.1038/s41561-017-0054-8

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., & Medina-Elizade, M. (2006). Global temperature change. Proceedings of the National Academy of Sciences, 103(39), 14288–14293. https://doi.org/10.1073/pnas.0606291103

Hansen, J., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., et al. (2001). A closer look at United States and global surface temperature change. Journal of Geophysical Research: Atmospheres, 106(D20), 23947–23963. https://doi.org/10.1029/2001jd000354

Hartmann, D.L., A.M.G. Klein Tank, M. Rusticucci, L.V. Alexander, S. Brönnimann, et al., 2013a: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., at al. (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hartmann, D.L., A.M.G. Klein Tank, M. Rusticucci, L. Alexander, S. Brönnimann, et al. (2013b): Observations: Atmosphere and Surface Supplementary Material. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., at al. (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hausfather, Z., Cowtan, K., Clarke, D. C., Jacobs, P., Richardson, M., & Rohde, R. (2017). Assessing recent warming using instrumentally homogeneous sea surface temperature records. Science Advances, 3(1), e1601207. https://doi.org/10.1126/sciadv.1601207.

Haustein, K., Allen, M. R., Forster, P. M., Otto, F. E. L., Mitchell, D. M., Matthews, H. D., & Frame, D. J. (2017). A real-time Global Warming Index. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-14828-5

Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., & Thomas, S. (2019). Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. Geoscience Data Journal, 6(2), 160–173. https://doi.org/10.1002/gdj3.79

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., et al., 2018: Impacts of 1.5ºC Global Warming on Natural and Human Systems. Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, et al. (eds.)].

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. Journal of Climate, 30(20), 8179–8205. https://doi.org/10.1175/jcli-d-16-0836.1

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., et al. (2020). Uncertainty Estimates for Sea Surface Temperature and Land Surface Air Temperature in NOAAGlobalTemp Version 5. Journal of Climate, 33(4), 1351–1379. https://doi.org/10.1175/jcli-d-19-0395.1

IPCC, 2013: Summary for Policymakers. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., at al. (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland.

IPCC, 2013: Summary for Policymakers. In Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., et al. (eds.)]. World Meteorological Organization, Geneva, Switzerland, 32 pp.

Jeffrey, S. J., Rotstayn, L. D., Collier, M. A., Dravitzki, S. M., Hamalainen, C., Moeseneder, C., Wong, K. K., & Syktus, J. I. (2013). Australia's CMIP5 submission using the CSIRO Mk3.6 model, Australian Meteorological and Oceanographic Journal, 63, 1–13, http://www.bom.gov.au/amoj/docs/2013/jeffrey_hres.pdf

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. Journal of Geophysical Research: Atmospheres, 117(D5), n/a-n/a. https://doi.org/10.1029/2011jd017139

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., et al. (2015). Possible artifacts of data biases in the recent global surface warming hiatus. Science, 348(6242), 1469–1472. https://doi.org/10.1126/science.aaa5632

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., & Killick, R. E. (2019). An Ensemble Data Set of Sea Surface Temperature Change From 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set. Journal of Geophysical Research: Atmospheres, 124(14), 7719–7763. https://doi.org/10.1029/2018jd029867

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., & Saunby, M. (2011). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. Journal of Geophysical Research, 116(D14). https://doi.org/10.1029/2010jd015218

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., & Saunby, M. (2011). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. Journal of Geophysical Research, 116(D14). https://doi.org/10.1029/2010jd015220

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. Journal of Geophysical Research: Atmospheres, 124(12), 6307–6326. https://doi.org/10.1029/2018JD02952.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. Journal of Advances in Modeling Earth Systems, 11(7), 2050–2069. https://doi.org/10.1029/2019ms001639

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. Journal of Climate, 31(24), 9835–9854. https://doi.org/10.1175/jcli-d-18-0094.1

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. Journal of Geophysical Research: Atmospheres, 117(D8). https://doi.org/10.1029/2011jd017187

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., et al. (2020). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset, submitted to Journal of Geophysical Research: Atmospheres, in review

Mudelsee, M. (2019). Trend analysis of climate time series: A review of methods. Earth-Science Reviews, 190, 310–322. https://doi.org/10.1016/j.earscirev.2018.12.005

Nauels, A., Rosen, D., Mauritsen, T., Maycock, A., McKenna, C., Roegli, J., et al. (2019). ZERO IN ON the remaining carbon budget and decadal warming rates. The CONSTRAIN Project Annual Report 2019. University of Leeds. https://doi.org/10.5518/100/20

Nychka, D., Buchberger, R., Wigley, T. M. L., Santer, B. D., Taylor, K. E., & Jones, R. (2000). Confidence intervals for trend estimates with autocorrelated observations. NCAR manuscript. https://opensky.ucar.edu/islandora/object/manuscripts:881

Osborn, T.J., Jones, P.D., Lister, D.H,, Morice, C.P,, Simpson, I.R, Winn, J, , et al . (2020). Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 dataset. Journal of Geophysical Research, accepted.

Parsons, L. A., Brennan, M. K., Wills, R. C. J., & Proistosescu, C. (2020). Magnitudes and Spatial Patterns of Interdecadal Temperature Variability in CMIP6. Geophysical Research Letters, 47(7). https://doi.org/10.1029/2019gl086588

Peng-Fei, L., F. Xiao-Li, & L. Juan-Juan (2015), Historical Trends in Surface Air Temperature Estimated by Ensemble Empirical Mode Decomposition and Least Squares Linear Fitting, Atmospheric and Oceanic Science Letters: 8:1, 10-16. https://doi.org/ 10.3878/AOSL20140064

Rahmstorf, S., Foster, G., & Cahill, N. (2017). Global temperature evolution: recent trends and some pitfalls. Environmental Research Letters, 12(5), 54001. https://doi.org/10.1088/1748-9326/aa6825

Richardson, M., Cowtan, K., & Millar, R. J. (2018). Global temperature definition affects achievement of long-term climate goals. Environmental Research Letters, 13(5), 54004. https://doi.org/10.1088/1748-9326/aab305

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., et al., 2018: Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, et al. (eds.)].

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., & Séférian, R. (2019). Estimating and tracking the remaining carbon budget for stringent climate targets. Nature, 571(7765), 335–342. https://doi.org/10.1038/s41586-019-1368-z

Risbey, J. S., Lewandowsky, S., Cowtan, K., Oreskes, N., Rahmstorf, S., Jokimäki, A., & Foster, G. (2018). A fluctuation in surface temperature in historical context: reassessment and retrospective on the evidence. Environmental Research Letters, 13(12), 123008. https://doi.org/10.1088/1748-9326/aaf342

Rohde, R. A., & Hausfather, Z. (2020). The Berkeley Earth Land/Ocean Temperature Record. Earth System Science Data, 12(4), 3469–3479. https://doi.org/10.5194/essd-12-3469-2020

Rohde, R., Muller, R.A., Jacobsen, R., Muller, E., Perlmutter, C., et al. (2013). A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. Geoinformatics & Geostatistics: An Overview, 1(1). https://doi.org/10.4172/2327-4581.1000101

Rotstayn, L. D., Jeffrey, S. J., Collier, M. A., Dravitzki, S. M., Hirst, A. C., Syktus, J. I., & Wong, K. K. (2012). Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: a study using single-forcing climate simulations. Atmospheric Chemistry and Physics, 12(14), 6377–6404. https://doi.org/10.5194/acp-12-6377-2012

Santer, B. D., Thorne, P. W., Haimberger, L., Taylor, K. E., Wigley, T. M. L., Lanzante, J. R., et al. (2008). Consistency of modelled and observed temperature trends in the tropical troposphere. International Journal of Climatology, 28(13), 1703–1722. https://doi.org/ 10.1002/joc.1756

Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019). Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. Nature Climate Change, 9(7), 517–522. https://doi.org/10.1038/s41558-019-0505-x

Smith, T.M., and R.W. Reynolds, 2005: A global merged land and sea surface temperature reconstruction based on historical observations (1880–1997). J. Clim., 18, 2021–2036

Susskind, J., Schmidt, G. A., Lee, J. N., & Iredell, L. (2019). Recent global warming as confirmed by AIRS. Environmental Research Letters, 14(4), 44030. https://doi.org/10.1088/1748-9326/aafd4e

Trenberth, K.E., Jones, P.D., Ambenje, P., Bojariu, R., Easterling, D., et al., 2007: Observations: Surface and Atmospheric Climate Change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

UNFCCC, 2015, Report on the Structured Expert Dialogue on the 2013–2015 Review FCCC/SB/2015/INF.1

Takahashi, H., Lebsock, M. D., Richardson, M., Marchand, R., & Kay, J. E. (2019). When Will Spaceborne Cloud Radar Detect Upward Shifts in Cloud Heights? Journal of Geophysical Research: Atmospheres. https://doi.org/10.1029/2018jd030242

van den Dool, H. M., Saha, S., & Johansson, Å. (2000). Empirical Orthogonal Teleconnections. Journal of Climate, 13(8), 1421–1435. https://doi.org/10.1175/1520-0442(2000)013<1421:eot>2.0.co;2

Visser, H., Dangendorf, S., van Vuuren, D. P., Bregman, B., & Petersen, A. C. (2018). Signal detection in global mean temperatures after "Paris": an uncertainty and sensitivity analysis. Climate of the Past, 14(2), 139–155. https://doi.org/10.5194/cp-14-139-2018

Zhang, H.-M., Lawrimore, J., Huang, B., Menne, M., Yin, X., Sanchez-Lugo, A., Gleason, B.E., et al. (2019). Updated Temperature Data Give a Sharper View of Climate Trends. Eos, 100. https://doi.org/10.1029/2019eo128229

Earth and Space Science

Supporting Information for

The benefits of local regression for quantifying global warming

David C. Clarke[1], Mark Richardson[2,3]

[1]Independent Researcher, Montreal, Quebec, Canada

[2]Jet Propulsion Laboratory, California Institute of Technology, USA [3]Joint Institute for Regional Earth Systems Science and Engineering, University of California, Los Angeles, USA

Contents of this file

Tables S1 to S4

Introduction

- Table S1 provides information about the CMIP6 models used in this study.
- Table S2 compares $LOESS_{bsln}$ and OLS ΔGMST s over 1880-2010, 151-2019 and 1979-2019.
- Table S3 presents combined ΔGMST and ΔGSAT uncertainties.
- Table S4 shows impact of incorporating HadCRUT5 and HadSST4 on $LOESS_{bsln}$ ΔGMST and ΔGSAT

Table S1: CMIP6 model ensemble. CMIP6 model runs employed in this study are listed, along with a preliminary evaluation of ECS assessed by 4xCO_2 abrupt experiments, with the resulting ECS_{lk} flag (set to "Y" if ECS is within the CMIP5 5-95% range of 1.9-4.5°C). The SSP ext column lists Shared Scenario Pathway continuations of interest that were available in February, 2020.

CAMS	CAMS-CSM1-0	r1i1p1f1	2.3	Y	245, 370
CAS	FGOALS-F3-L	r1i1p1f1	4.7	Y	245, 370
CCCma	CanESM5	r1i1p1f1	5.6	N	245, 370
CNRM-CERFACS	CNRM-CM6-1	r1i1p1f2	4.8	N	245, 370
CNRM-CERFACS	CNRM-ESM2-1	r1i1p1f2	4.7	N	245, 370
CISRO	ACCESS-ESM1-5	r1i1p1f1	3.5	Y	245, 370
IPSL	IPSL-CM6A-LR	r1i1p1f1	4.8	N	245, 370
MIROC	MIROC6	r1i1p1f1	2.6	Y	245, 370
MIROC	MIROC-ES2L	r1i1p1f2	2.7	Y	245, 370
MOHC	HadGEM3-GC31-LL	r1i1p1f3	5.5	N	245, 370
MOHC	UKESM1-0-LL	r1i1p1f2	5.4	N	245, 370
MPI-M	MPI-ESM1-2-LR	r1i1p1f1	2.8	Y	245, 370
MRI	MRI-ESM2-0	r1i1p1f1	3.1	Y	245, 370
NASA-GISS	GISS-E2-1-G	r1i1p1f1	2.7	Y	
NASA-GISS	GISS-E2-1-G-CC	r1i1p1f1	*3.2*	Y	
NASA-GISS	GISS-E2-1-H	r1i1p1f1	*3*	Y	
NCAR	CESM2	r1i1p1f1	5.2	N	245, 370
NCAR	CESM2-WACCM	r1i1p1f1	4.7	N	245, 370
NCC	NORCPM1	r1i1p1f1	*3*	Y	
NCC	NorESM2-LM	r1i1p1f1	*2.9*	Y	
NOAA-GFDL	GFDL-CM4	r1i1p1f1	3.9	Y	245
NOAA-GFDL	GFDL-ESM4	r1i1p1f1	3.2	Y	245
NUIST	NESM3	r1i1p1f1	4.7	N	245
SNU	SAM0-UNICON	r1i1p1f1	3.8	Y	

Figure S12: Cowtan-Way ΔGMST to 2019. Top (a - b) Cowtan-Way monthly series (light gray) is shown with $LOESS_{bsln}$ (blue) and 2010-2019 average (black square) relative to 1850-1900, along with OLS linear trend over 1880-2019 (red). The OLS linear trend central estimates and uncertainty have been shifted upward to provide direct comparison to the other two estimates. **(a)** Trends are given with ARMA(1,1) corrected 5%-95% confidence interval (dotted lines). **(b)** $LOESS_{md}$ (thin light blue lines) and OLS (thin pink lines) trends are derived from Cowtan and Way 100-member ensemble. **Middle (c)** Autocorrelation function (ACF) of $LOESS_{md}$ statistical fit residuals (black), compared to that estimated with ARMA(1, 1) model (blue) and AR(1) model (red) for LOESS trend. **(d)** As in (c), except for OLS linear trend. **Bottom (e)** ACF for $LOESS_{md}$ fit residuals for Cowtan-Way annual series, compared to AR(1) model (red) for LOESS trend. **(f)** As in e), except for OLS linear trend.

Table S2: Observed increase in GMST (°C) in datasets and dataset groupings. Numbers in square brackets correspond to 5–95% statistical trend fit uncertainty ranges, accounting for autocorrelation in fit residuals. Round brackets denote 5–95% observational parametric uncertainty where available (HadCRUT4, Cowtan-Way). Best estimates from 3 full global (NASA GISTEMP, Cowtan-Way and Berkeley Earth series are denoted by *.

Period: / ***Series:***	**1880 - 2019**		**1951 - 2019**		**1979 - 2019**	
	LOESS_{md}	**Linear**	**LOESS_{md}**	**Linear**	**LOESS_{md}**	**Linear**
HadCRUT4	0.99 [0.88 - 1.11] *(0.94 – 1.04)*	0.96 [0.82 - 1.10] *(0.92 – 1.03)*	0.75 [0.63 - 0.87] *(0.67 – 0.76)*	0.84 [0.69 - 1.00] *(0.76 – 0.88)*	0.70 [0.58 - 0.81] *(0.64 – 0.71)*	0.70 [0.59 - 0.82] *(0.65 – 0.72)*
NOAA GlobalTemp	1.06 [0.93 - 1.18]	1.04 [0.89 - 1.19]	0.87 [0.75 - 0.99]	0.97 [0.83 - 1.10]	0.74 [0.62 - 0.85]	0.71 [0.58 - 0.84]
NASA GISTEMP	1.09 [0.98 - 1.21]	1.04 [0.88 - 1.20]	0.94 [0.83 - 1.04]	1.03 [0.90 - 1.15]	0.79 [0.69 - 0.89]	0.77 [0.65 - 0.88]
Cowtan & Way	1.14 [1.03 - 1.25] (1.08 – 1.21)	1.02 [0.88 - 1.15] (0.94 – 1.09)	0.81 [0.70 - 0.91] (0.75 – 0.87)	0.88 [0.73 - 1.04] (0.83 – 0.94)	0.75 [0.65 - 0.86] (0.70 – 0.79)	0.77 [0.66 - 0.88] (0.74– 0.81)
Berkeley Earth	1.20 [1.09 - 1.31]	1.09 [0.96 - 1.22]	0.85 [0.74 - 0.95]	0.92 [0.78 - 1.06]	0.77 [0.67 - 0.86]	0.78 [0.67 - 0.88]
All Operational	**1.10 [0.88 - 1.31]**	**1.03 [0.82 - 1.22]**	**0.84 [0.63 – 1.04]**	**0.93 [0.69 - 1.15]**	**0.75 [0.58 - 0.89]**	**0.75 [0.58 - 0.88]**
Near Global (3 series) *	**1.14 * [0.98 - 1.31]**	**1.05 [0.88 - 1.22]**	**0.83 * [0.70 – 1.04]**	**0.91 [0.70- 1.12]**	**0.74 * [0.65 - 0.89]**	**0.74 [0.65 - 0.88]**

Table S3: Combined GMST and GSAT changes and uncertainty ranges for each dataset, group and combination of uncertainties. As described in main manuscript: individual dataset ΔGMST combine in quadrature Cowtan & Way ensemble uncertainty and either statistical error ("stat") or CSIRO ensemble standard deviation ("CSIRO"). ΔGSAT combines fractional ΔGMST and A_{blend} uncertainties in quadrature. We justify quadrature combinations as the Shapiro-Wilks test does not reject normality in any case: for Cowtan & Way ensemble (p 0.27), the CSIRO ensemble (p=0.48) or CMIP6 ensemble A_{blend} (p=0.17). Group_3 uncertainty ranges are lowest minimum percentile to highest maximum percentile from across the datasets. This means that the 5—95 % and 17—83 % are not consistent according to any standard formal PDF distribution.

			Statistical σ		**CSIRO ensemble σ**	
		Mean	**17—83 %**	**5—95 %**	**17—83 %**	**5—95 %**
ΔGMST [°C]	CowtanWay	1.12	1.06—1.18	1.02—1.23	1.02—1.22	0.95—1.30
	GISTEMP	1.12	1.05—1.18	1.01—1.22	1.01—1.22	0.94—1.29
	Berkeley	1.19	1.12—1.25	1.08—1.29	1.08—1.29	1.01—1.36
	Group_3	1.14	1.05—1.25	1.01—1.29	1.01—1.29	0.94—1.36
ΔGSAT [°C]	CowtanWay	1.19	1.12—1.25	1.07—1.30	1.08—1.30	1.00—1.38
	GISTEMP	1.18	1.11—1.25	1.07—1.30	1.07—1.29	0.99—1.37
	Berkeley	1.25	1.19—1.32	1.14—1.37	1.14—1.36	1.06—1.44
		1.21	1.11—1.32	1.07—1.37	1.07—1.36	0.99—1.44

Table S4: Impact of HadSST4 and HadCRUT5 on observational ΔGMST and ΔGSAT in °C. The Cowtan-Way/HadSST4 and HadCRUT5 datasets have been extended to the end of 2019, by assuming the same monthly temperature innovations as observed over 2019 as in the published Cowtan-Way (with HadSST3) dataset. Berkeley Earth/HadSST4 $LOESS_{bsln}$ $\Delta GMST_{LOESS}$ is estimated by applying the difference between Cowtan-Way/HadSST4 and Cowtan-Way/HadSST3 $\Delta GMST_{LOESS}$ to BerkeleyEarth/HadSST3 $\Delta GMST_{LOESS}$. Numbers in square brackets correspond to 5–95% statistical trend fit uncertainty ranges, accounting for autocorrelation in fit residuals.

Period/metric: / ***Series:***	**$LOESS_{bsln}$ w/HadSST3 (*) 1850-1900 to 2019**		**$LOESS_{bsln}$ w/HadSST4 (**) 1850-1900 to 2019**	
	ΔGMST	**ΔGSAT**	**ΔGMST**	**ΔGSAT**
NASA GISTEMP	1.12 [1.02 - 1.22]	1.18	1.12 [1.02 - 1.22]	1.18
Cowtan-Way	1.12 * [1.04 - 1.21]	1.19 *	1.19 ** [1.08 - 1.30]	1.26 **
Berkeley Earth	1.19 * [1.10 - 1.27]	1.26 *	1.26 ** [1.17 - 1.34]	1.33 **
Full Global (3 series)	**1.14 * [1.02 – 1.27]**	**1.21 ***	**1.19 ** [1.02 – 1.34]**	**1.26 ****
HadCRUT5	N/A	N/A	1.20 ** [1.09 - 1.32]	1.27 **
Full Global (4 series)	**N/A**	**N/A**	**1.19 ** [1.02 – 1.34]**	**1.26 ****

AGU PUBLICATIONS

Earth and Space Science

Supporting Information for

The benefits of local regression for quantifying global warming

David C. Clarke[1], Mark Richardson[2,3]

[1]Independent Researcher, Montreal, Quebec, Canada

[2]Jet Propulsion Laboratory, California Institute of Technology, USA [3]Joint Institute for Regional Earth Systems Science and Engineering, University of California, Los Angeles, USA

Contents of this file

Figures S12 to S16

Introduction

- Figures S12 and S13 are supplementary to Section 2.2 and provide additional information and provide further information on LOESS and OLS statistical fit of observational series and model large ensembles.
- Figures S14 explores the sensitivity of observations and model output to distance-limited masking.
- Figure S15 shows performance of $LOESS_{bsln}$ and OLS from 1880 ΔGMST against 20 and 30-year average in Global_3 .
- Figure S16 shows performance of $LOESS_{bsln}$ and 10-year "period" ΔGMST against 30-year average in MPI and CSIRO model ensembles.

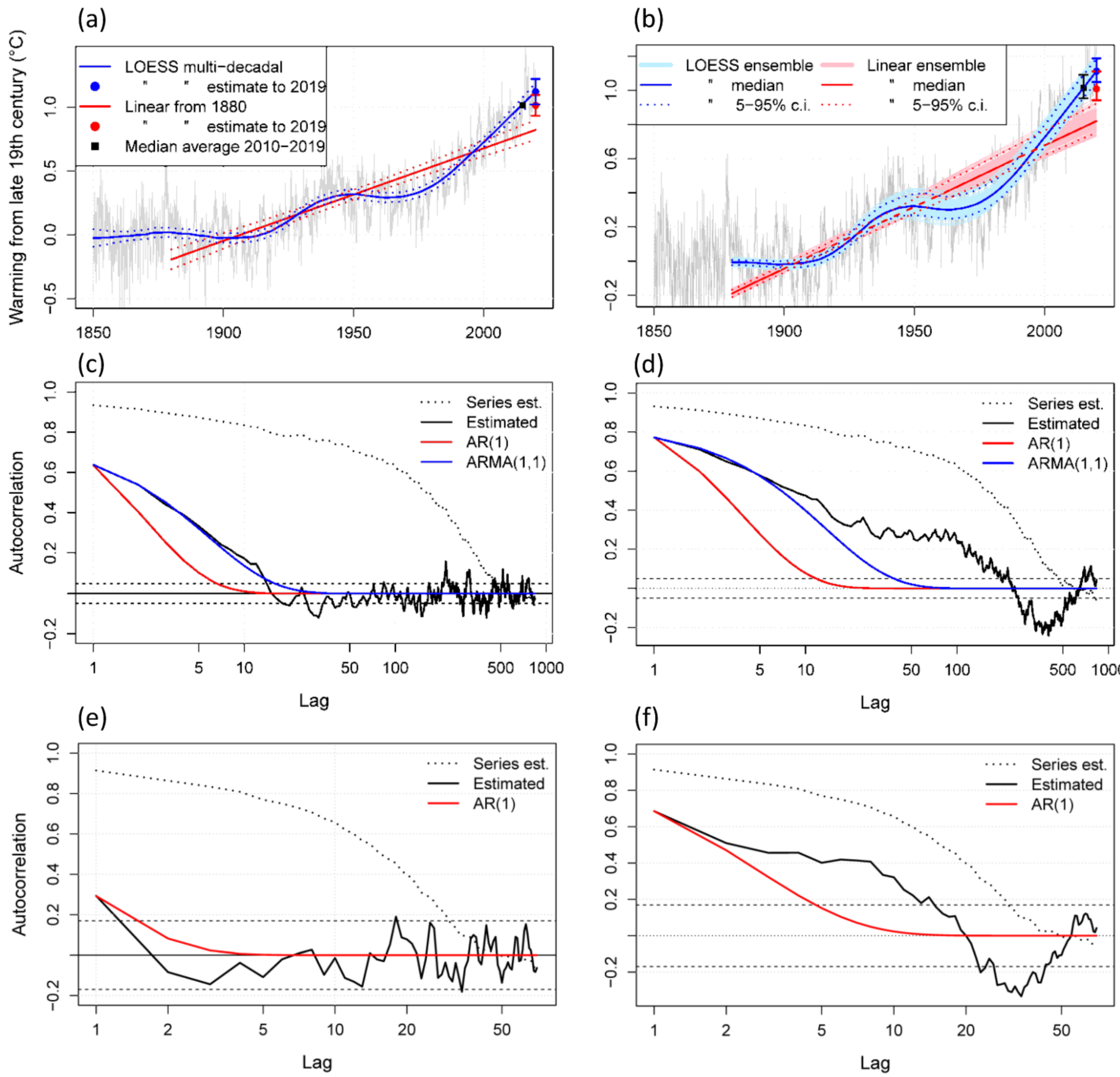

Figure S12: Cowtan-Way ΔGMST to 2019. Top (a - b) Cowtan-Way monthly series (light gray) is shown with $LOESS_{bsln}$ (blue) and 2010-2019 average (black square) relative to 1850-1900, along with OLS linear trend over 1880-2019 (red). The OLS linear trend central estimates and uncertainty have been shifted upward to provide direct comparison to the other two estimates. **(a)** Trends are given with ARMA(1,1) corrected 5%-95% confidence interval (dotted lines). **(b)** $LOESS_{md}$ (thin light blue lines) and OLS (thin pink lines) trends are derived from Cowtan and Way 100-member ensemble. **Middle (c)** Autocorrelation function (ACF) of $LOESS_{md}$ statistical fit residuals (black), compared to that estimated with ARMA(1, 1) model (blue) and AR(1) model (red) for LOESS trend. **(d)** As in (c), except for OLS linear trend. **Bottom (e)** ACF for $LOESS_{md}$ fit residuals for Cowtan-Way annual series, compared to AR(1) model (red) for LOESS trend. **(f)** As in e), except for OLS linear trend.

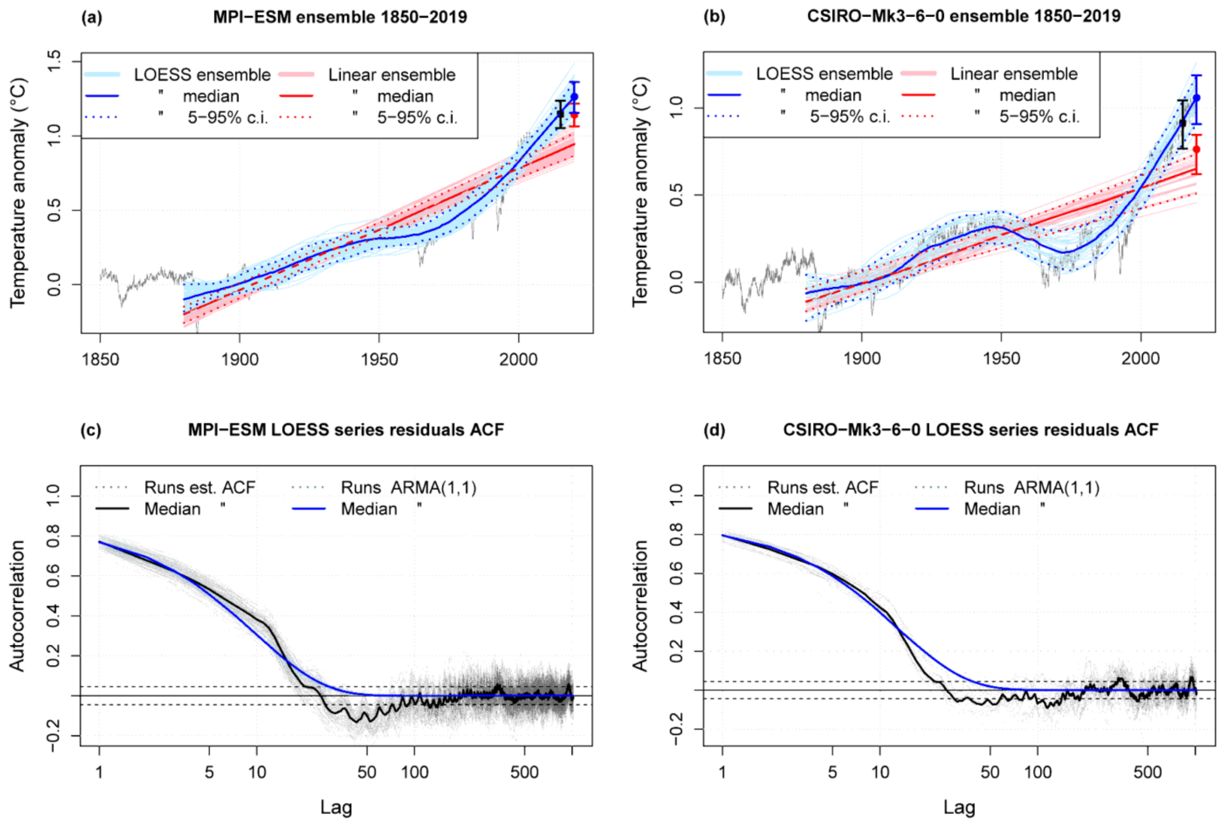

Figure S13: MPI and CSIRO Large Ensemble ΔGMST to 2019. (a) MPI ensemble median monthly series (light gray) is shown with $\mathrm{LOESS}_{\mathrm{bsln}}$ (blue) and 2010-2019 average (black square) relative to 1850-1900, along with OLS linear trend over 1880-2019 (red). The OLS linear trend central estimates and uncertainty have been shifted upward to provide direct comparison to the other two estimates. $\mathrm{LOESS}_{\mathrm{bsln}}$ (thin light blue lines) and OLS trends (thin pink lines) are derived from each ensemble member. **(b)** As in (a), except for CSIRO ensemble. **(c)** Autocorrelation function (ACF) of LOESS statistical fit residuals (black), compared to that estimated with ARMA(1, 1) model (blue) for LOESS trend. **(d)** As in (c), except for CSIRO ensemble..

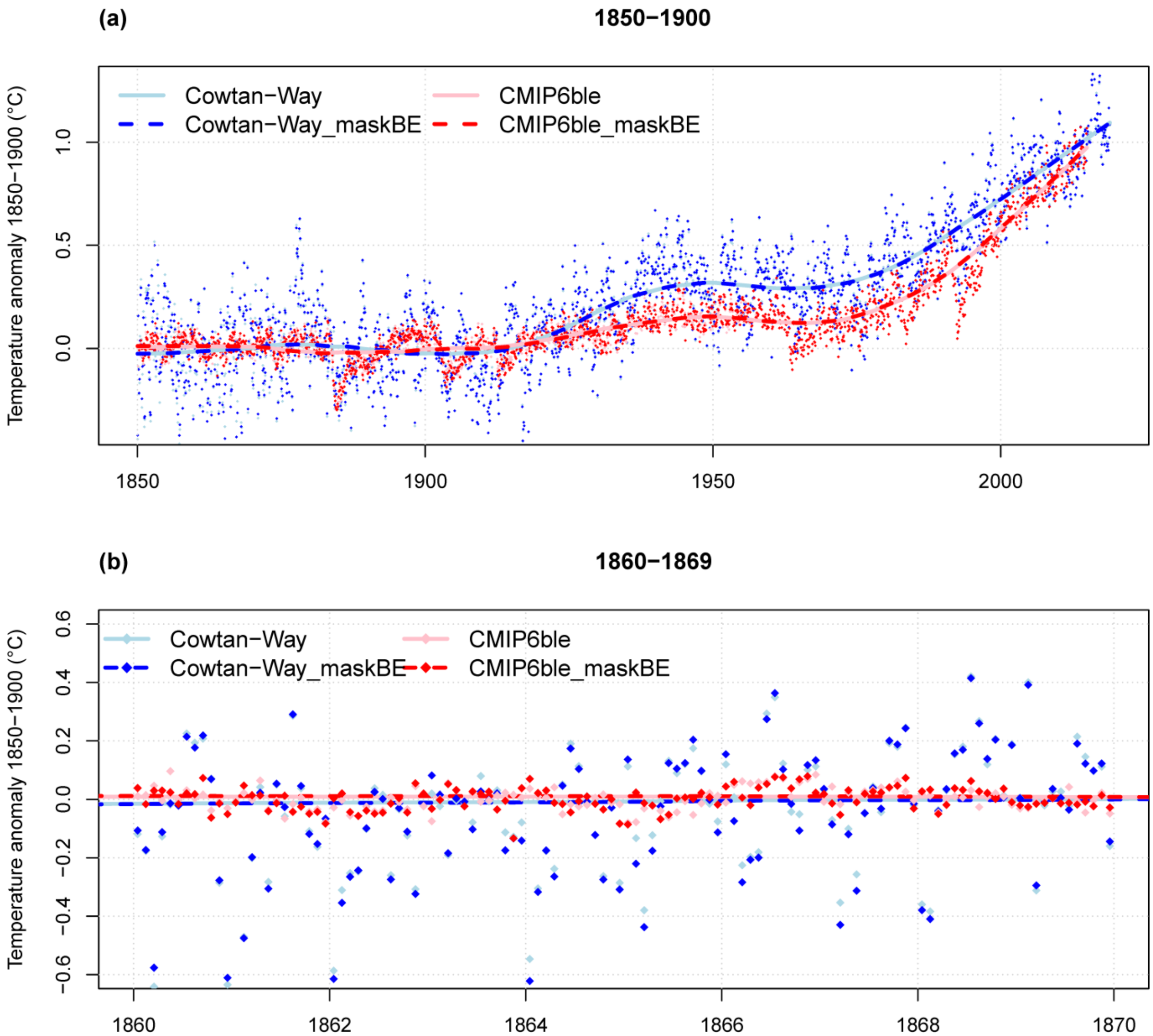

Figure S14: Evaluation of distance-limited interpolation. (a) Shown over 1850-2019 are monthly anomalies (small diamonds) and multi-decadal LOESS trends (full as solid lines, masked as dotted lines) for Cowtan-Way (light blue), Cowtan-Way masked to Berkeley Earth coverage (blue), CMIP6 blended ensemble median (pink) and CMIP6 masked to Berkeley Earth coverage (red). **(b)** As (a), but over 1860-1869.

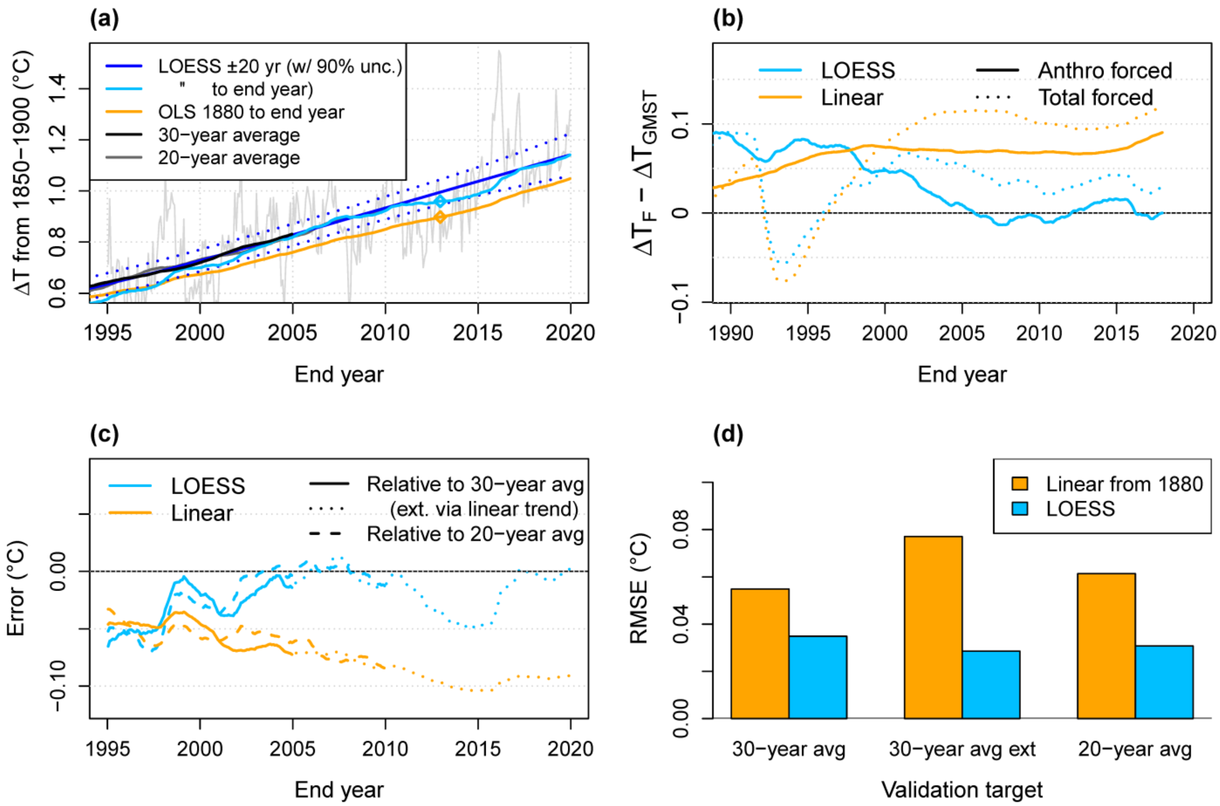

Figure S15: ΔGMST estimation method validation based on average of 3 global series. (a) $LOESS_{bsln}$ to 2019 (blue) is shown with LOESS (light blue), OLS from 1880 (orange), 20-year average (light gray) and 30-year average (black). LOESS (light blue) versus OLS (orange) differences are shown with **(b)** forced warming estimates following Haustein et al. (2017) and **(c)** validation targets (30-year average, 30-year average extended with linear trend and 20-year average). **(d)** RMSE is calculated from errors shown in (c).

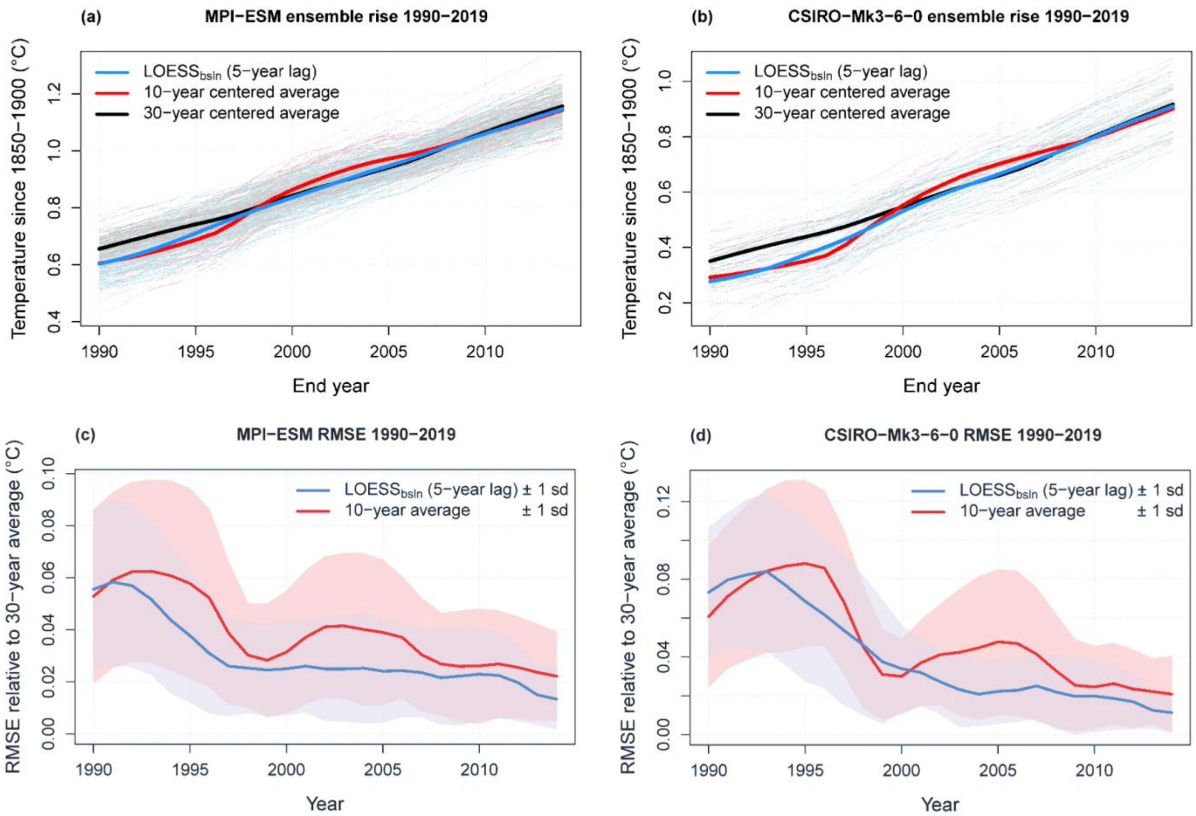

Figure S16: ΔGMST method validation based on large MPI and CSIRO ensembles. (a) 5-year lagged LOESS$_{bsln}$ (blue), centered 10-year average (red), and centered 30-year averages (dark gray) are shown from 1850-1900 baseline to indicated end year for each MPI ensemble member (thin lines) and averaged over ensemble (thick lines). The 30-year average is extended from 2004 to 2014, by extending each ensemble member over 2020-2030 with the continuation of its 1990-2019 linear trend. **(b)** As in (a), but for CSIRO ensemble. **(c)** MPI ensemble RMSE is calculated against 30-year average validation target for LOESS$_{bsln}$ (blue) and 10-year average (red). **(d)** As in (c), but for CSIRO ensemble.

Earth and Space Science

Supporting Information for

The benefits of local regression for quantifying global warming

David C. Clarke[1], Mark Richardson[2,3]

[1]Independent Researcher, Montreal, Quebec, Canada

[2]Jet Propulsion Laboratory, California Institute of Technology, USA [3]Joint Institute for Regional Earth Systems Science and Engineering, University of California, Los Angeles, USA

Contents of this file

Figures S4 to S12

Introduction

- Figures S4 to S7 are supplementary to Section 2.2.1 and provide additional information concerning LOESS window size performance.
- Figures S8 to S11 are supplementary to Section 2.2.1 and provide additional information concerning the methods used to assess statistical fit uncertainty.

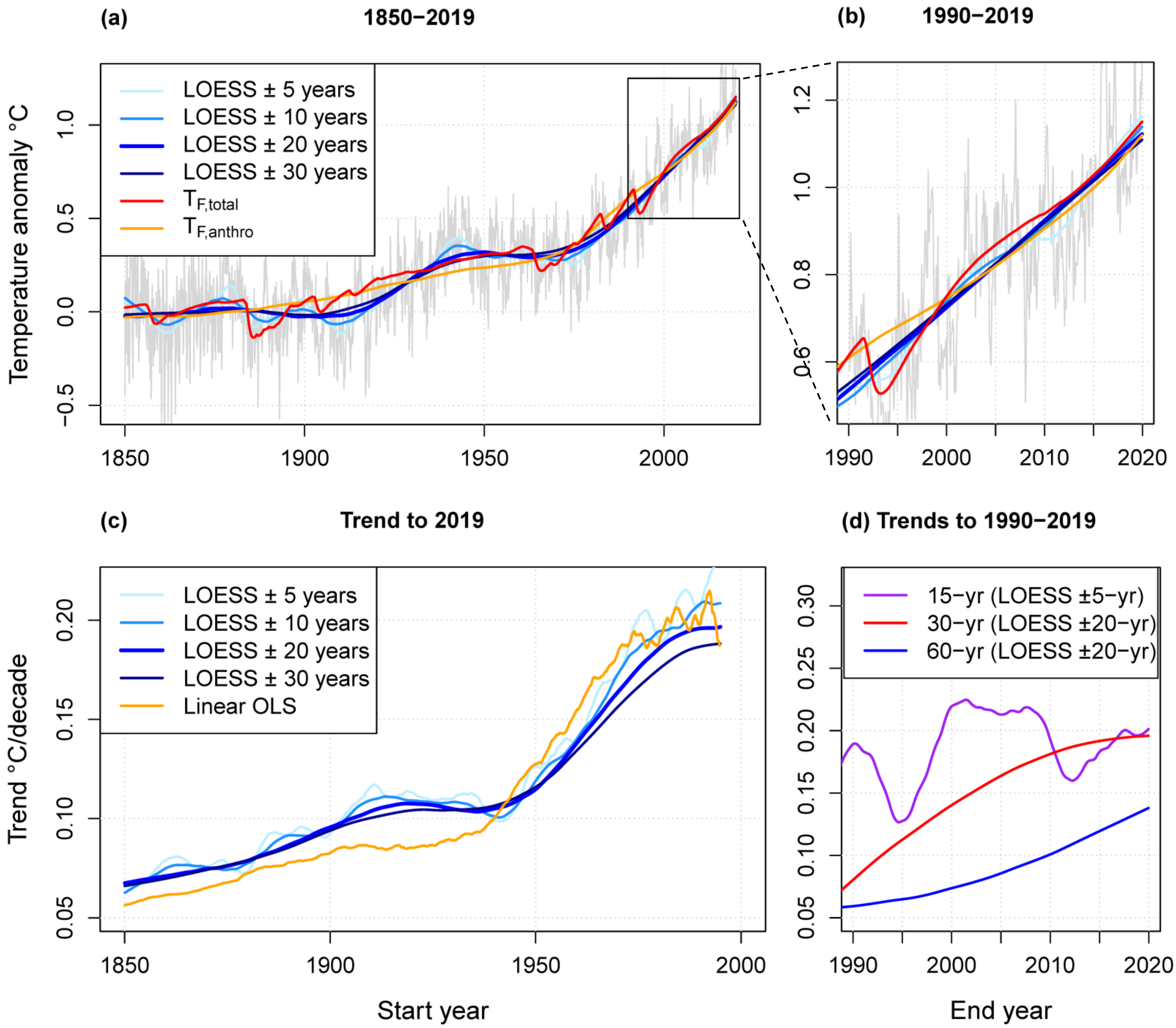

Figure S4: LOESS window selection and trend methodology. (a) Cowtan-Way monthly global average temperature series over 1850-2019 (light gray line) is shown with LOESS smooth (blue lines), with windows ranging from ±30 years down to ±5 years. Also shown are total and anthropogenic forced temperature (red and dark orange lines respectively), estimated from two-box model forced response regressed against Cowtan-Way series following Otto et al. (2015) and Haustein et al. (2017). **(b)** Same as (a), except over 1990-2019. **(c)** Overlapping LOESS trends (blue lines) and OLS trends (orange lines) to 2019 are shown, with trend start points of 1850 to 1995. **(d)** Overlapping fixed length LOESS trends ending in years 1990-2019 are shown.

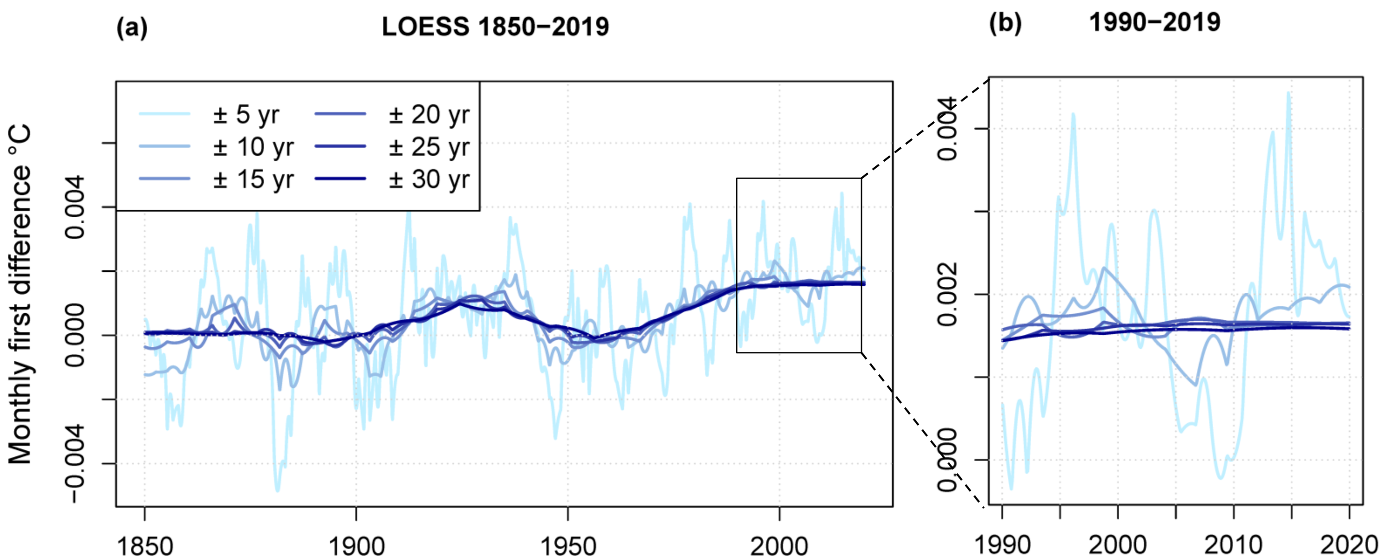

Figure S5: LOESS window smoothing characteristics. (a) Monthly first differences over 1850-2019 (light gray line) are shown for LOESS smooths applied to Cowtan-Way temperature series, with windows ranging from ±5 years (light blue) to ±30 years (dark blue). **(b)** Same as (a), except over 1990-2019.

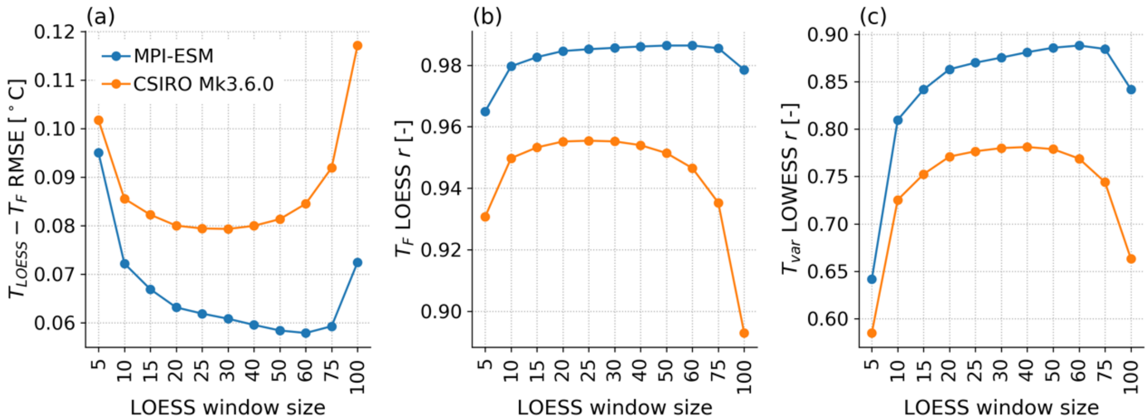

Figure S6. Large ensemble statistics for non-volcanic year LOESS with different window lengths relative to forced temperature change T_F (as assessed by ensemble mean at each time step) or their respective residuals (a) RMSE for T_F versus LOESS, (b) Pearson's *r* for LOESS versus T_F, (c) Pearson's *r* for LOESS residuals versus T_F residuals.

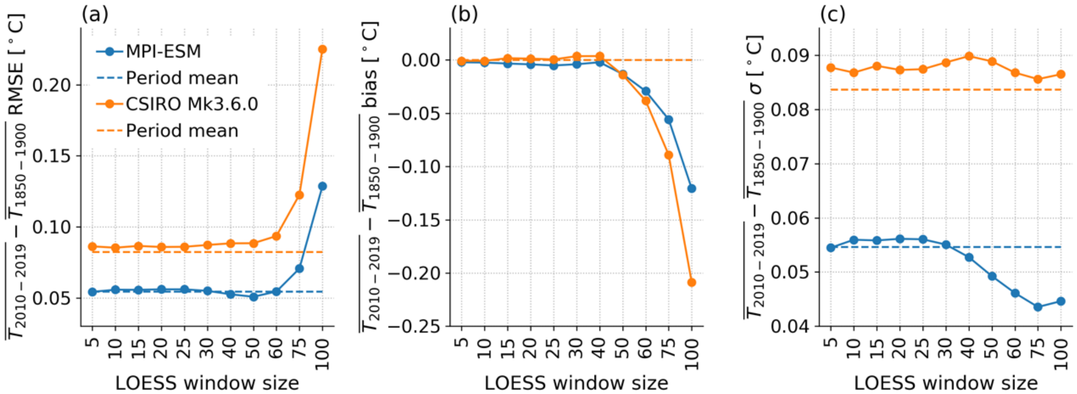

Figure S7. Ensemble performance statistics for the derived temperature change from 1850—1900 to 2010—2019. The RMSE and bias are calculated relative to the same value calculated from the ensemble mean T_F estimates. Solid lines with points are from LOESS fits with different window sizes (where a size of 10 is ±5 years) and dashed lines are those derived from taking the individual run period mean differences.

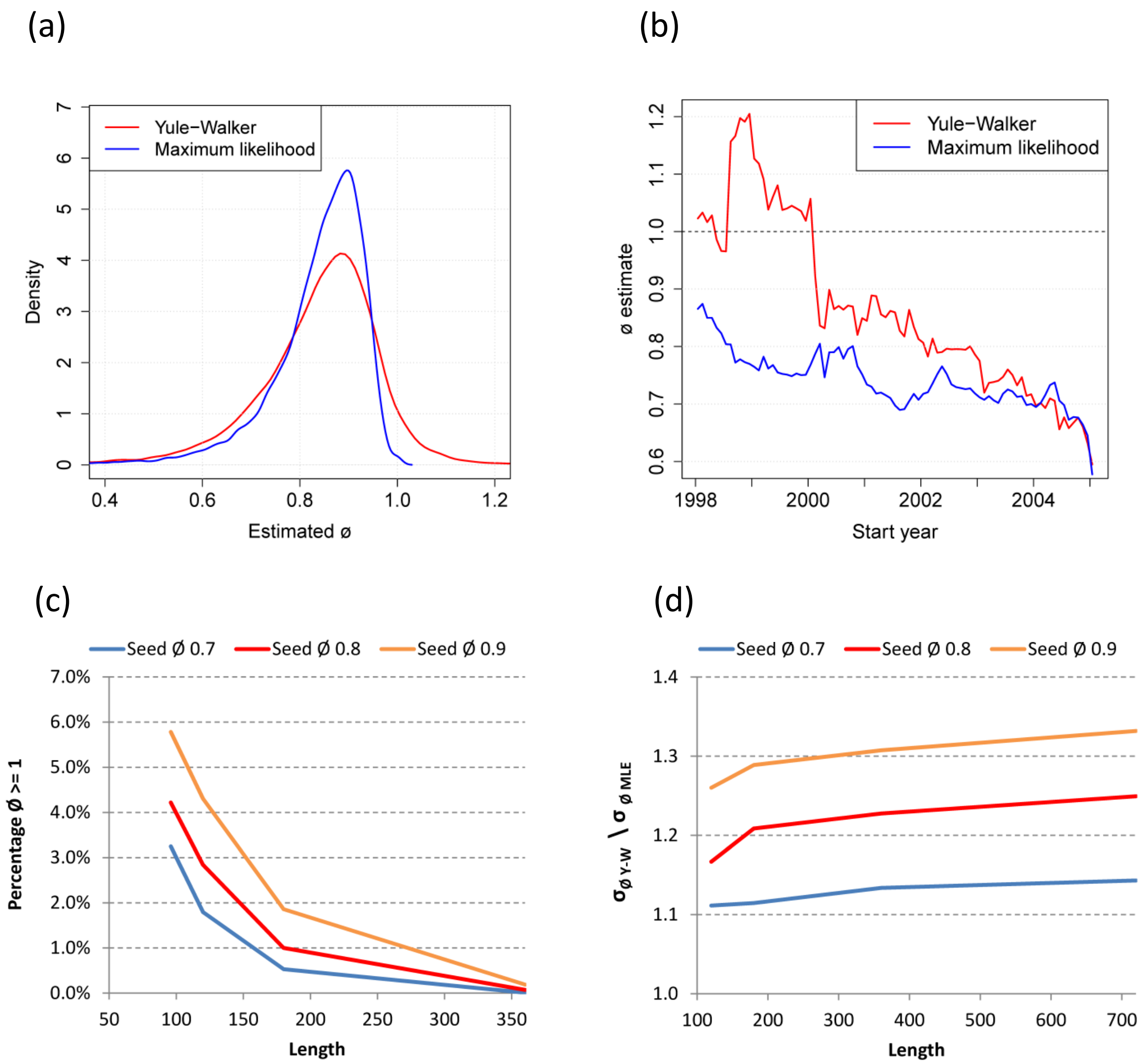

Figure S8: Yule-Walker (Y-W) vs Maximum Likelihood Estimation (MLE). (a) PDF of $\hat{\phi}$ in simulated ARMA(1, 1) 8-year (96-month) series with seed ϕ = 0.9. **(b)** $\hat{\phi}$ estimates derived from residuals of 8-year linear trends in Cowtan & Way over 1998-2012. **(c)** Percentage of simulated series with Y-W $\hat{\phi}$ > 1 by seed and length. **(d)** Efficiency of MLE relative to Y-W by seed and length .

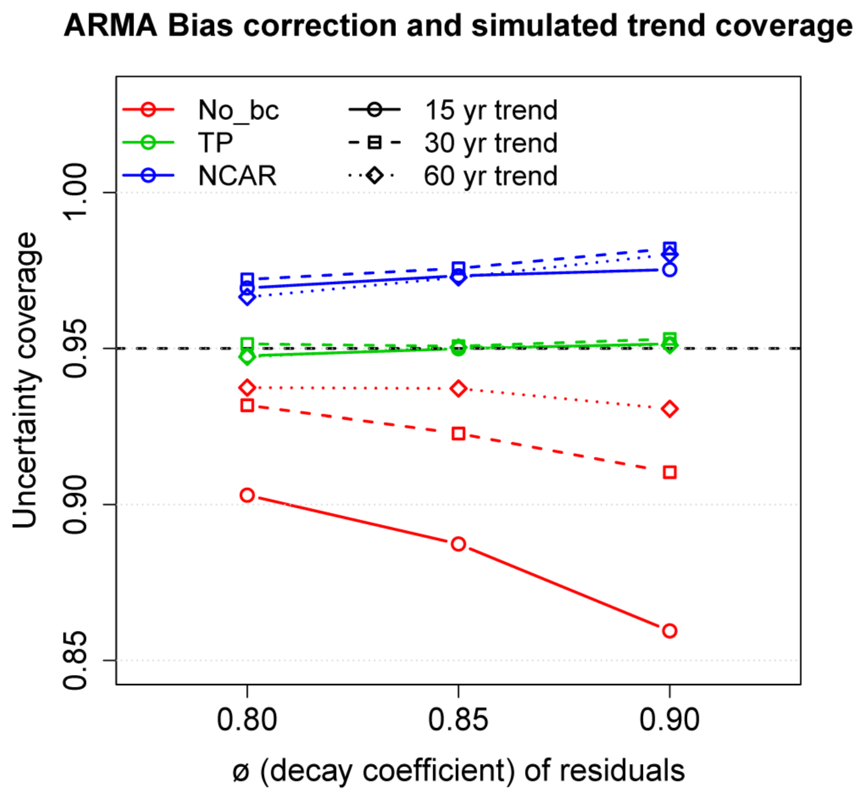

Figure S9: ARMA (1, 1) bias correction. Simulated 15-year (solid lines and circles) and 30-year (dashed lines and open circles) trends were generated assuming positively-correlated ARMA(1, 1) noise for three different levels of φ (phi) and three different bias correction schemes: No bias correction (red), bias correction derived from Tjøstheim and Paulsen (1996) as used in this study (TP, green), and an alternative bias correction derived from Nychka et al (2001) (NCAR, blue). See section 2.2.2 for details of the bias correction methodology.

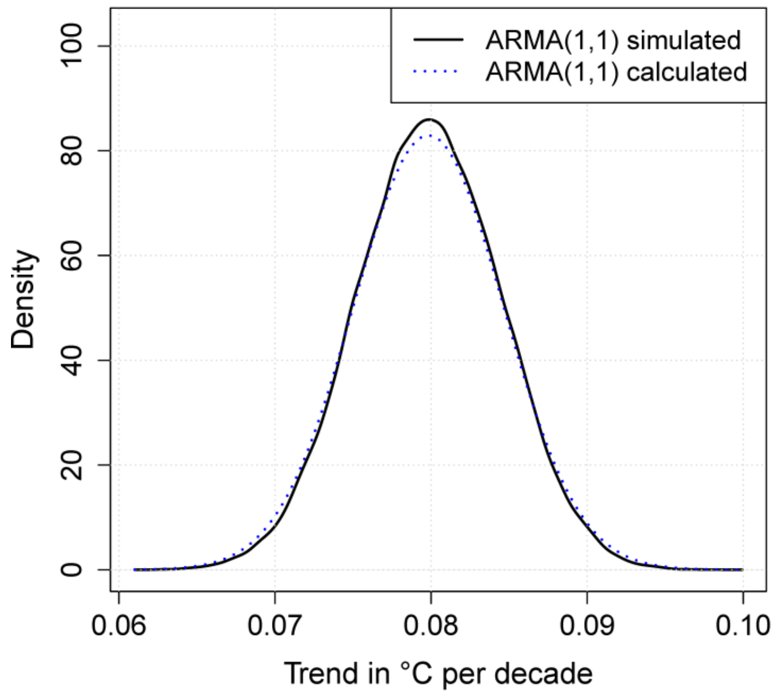

Figure S10: Uncertainty of LOESS$_{md}$ trends. ΔGMST trends over 1880-2019, expressed as change in °C per decade, were simulated by generating a Monte Carlo ensemble of 200K simulations from the Cowtan-Way observational series. Each realization is composed of a central estimate of the trend from Cowtan-Way with added ARMA(1, 1) noise according to the noise model assessed from the fit residuals, as detailed in section 2.2.2. The PDF of the simulated ensemble trend (solid line) is compared to the calculated trend uncertainty (dotted line).

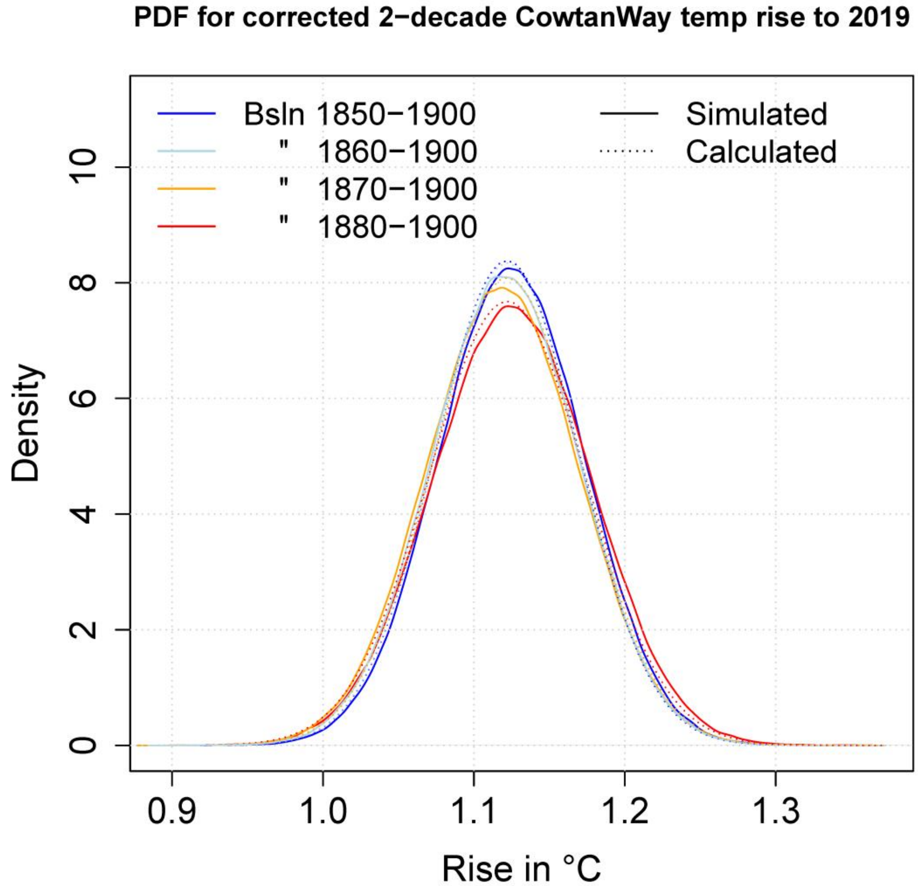

Figure S11: Uncertainty of $\mathrm{LOESS}_{\mathrm{bssln}}$ ΔGMST. $\mathrm{LOESS}_{\mathrm{bssln}}$ ΔGMST from various baselines to 2019, expressed as change in °C, were simulated by generating a Monte Carlo ensemble of 350K simulations from the Cowtan-Way observational series. Each realization is composed of a central estimate of the temperature rise from Cowtan-Way with added ARMA(1, 1) noise according to the noise model assessed from the fit residuals, as detailed in section 2.2.2. The PDF of the simulated ensemble trend (solid line) is compared to the calculated uncertainty (dotted line).

AGU PUBLICATIONS

Earth and Space Science

Supporting Information for

The benefits of local regression for quantifying global warming

David C. Clarke[1], Mark Richardson[2,3]

[1]Independent Researcher, Montreal, Quebec, Canada

[2]Jet Propulsion Laboratory, California Institute of Technology, USA [3]Joint Institute for Regional Earth Systems Science and Engineering, University of California, Los Angeles, USA

Contents of this file

Figures S1 to S3

Introduction

- Figures S1 to S3 are supplementary to Section 2.1.1 and provide additional information about the described observational GMST data series.

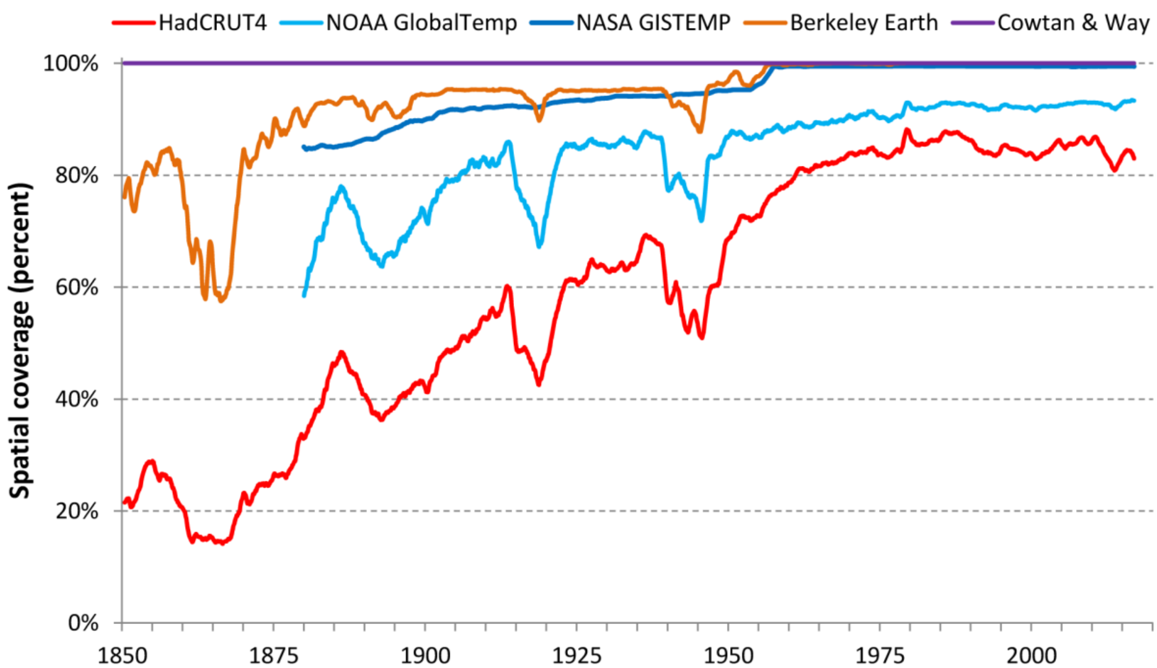

Figure S1: GMST spatial coverage 1850 – 2019. Monthly spatial coverage is shown for the five data series assessed in this study: HadCRUT4 (red), NOAA GlobalTemp (light blue), NASA GISTEMP (dark blue), Cowtan-Way (purple) and Berkeley Earth (orange).

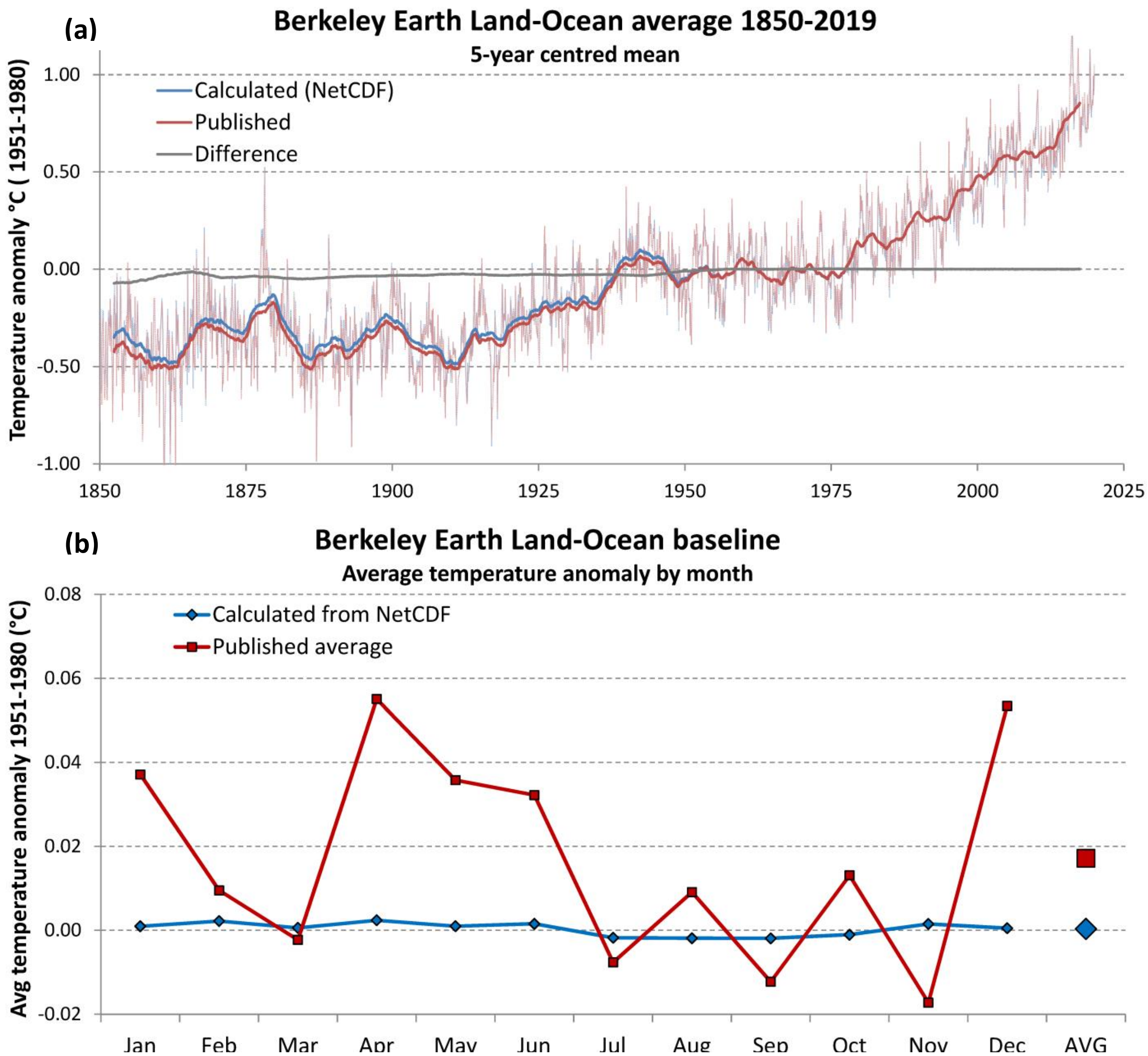

Figure S2: Berkeley Earth GMST 5-yr average and baseline annual cycle. (a) Shown are the published 5-year centred running average (red) and that calculated from area-weighted gridded average (blue). The difference series (gray) is 0 under full coverage after ~1955, but shows noticeable differences before then, especially over 1850-1900 (mean difference of ~0.04 °C). **(b)** Shown are the annual cycle in published baseline monthly averages (red squares) and that calculated from gridded data (blue diamonds).

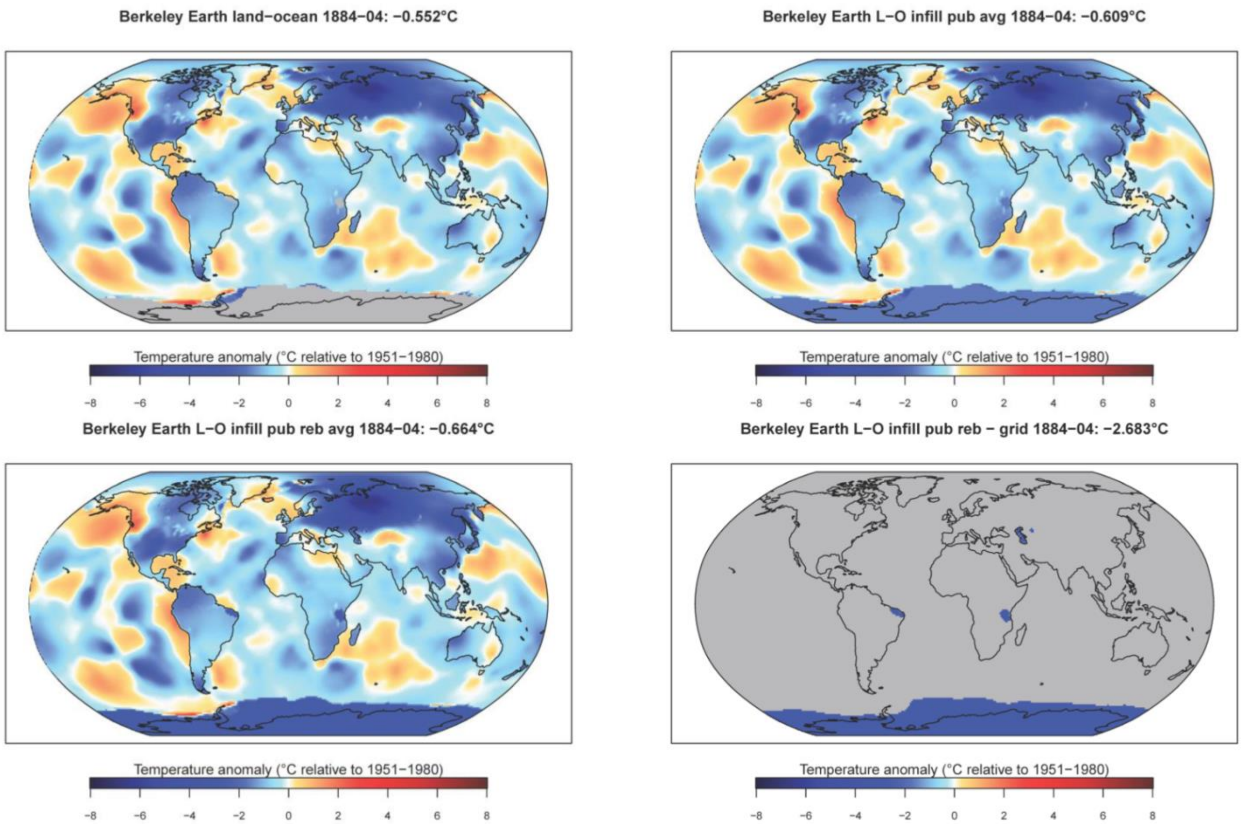

Figure S3: Berkeley Earth land-ocean 1884-04. (a) 1x1 gridded land-ocean anomaly data with area-weighted average**. (b)** Same as (a), except infilled so that resulting average matches Berkeley Earth published average. **(c)** Same as (a), except infilled so that resulting average matches Berkeley Earth rebaselined published average (i.e. April 1951-1980 average = 0). **(d)** The difference between (a) and (c), demonstrating that missing areas must average ~-2.7°C in order for the overall weighted average to match the rebaselined published average.