The Benefits of Continuous Local Regression for Quantifying Global Warming

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

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

November 30, 2022

Abstract

Change in global mean surface temperature (ΔGMST) is a widely cited climate change indicator that figures prominently in IPCC reports, in which it was estimated via linear regression or differences between decade-plus period means. The Paris Agreement aims to limit warming since preindustrial (here approximated as 1850—1900) to "well below" 2 °C, and by knowing current ΔGMST it is possible to determine the remaining target-consistent warming and therefore a relevant remaining carbon budget. We propose non-linear continuous local regression (LOESS) using 40-year windows as a single method to derive ΔGMST with statistical uncertainty across all periods of interest. Using the three datasets with almost complete spatial coverage since the 1950s, we evaluate 1850—1900 to 2019 ΔGMST as 1.14 °C with likely (17—83 %) range of 1.05—1.25 °C, based on combined statistical and observational uncertainty, compared with 1880—2019 linear regression of 1.03 °C using all five operational datasets. In two model large ensembles LOESS, like period mean differences, is unbiased but provides a statistical error and gives warming through 2019, rather than a 2010—2019 average centred at the end of 2014. We compare observational and CMIP6 ΔGMST and estimate historical global surface air temperature change (ΔGSAT) using the CMIP6 ΔGSAT/ΔGMST ratio and its ensemble spread. Finally, we calculate remaining carbon budgets given our ΔGSAT of 1.21 °C with likely range of 1.11—1.32 °C. We argue that continuous non-linear trend estimation offers substantial advantages for assessment of long-term observational ΔGMST.

Hosted file

`clarke-richardson 2020 supplementary information 2020-06-26 s3.docx` available at https://authorea.com/users/560759/articles/608617-the-benefits-of-continuous-local-regression-for-quantifying-global-warming

Hosted file

`clarke-richardson 2020 supplementary information 2020-06-26 s2.docx` available at https://authorea.com/users/560759/articles/608617-the-benefits-of-continuous-local-regression-for-quantifying-global-warming

Hosted file

`clarke-richardson 2020 supplementary information 2020-06-26 s1.docx` available at https://authorea.com/users/560759/articles/608617-the-benefits-of-continuous-local-regression-for-quantifying-global-warming

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 (dave@daveclarke.ca)

Key Points:

- Continuous local regression is an alternative to traditional IPCC temperature change estimation methods.
- Global warming from near-global land-ocean observational series reached 1.14°C (likely range 1.05—1.25°C) in 2019 relative to 1850-1900.
- Global surface air temperature reached 1.21°C (likely range 1.11—1.32°C), for a remaining 1.5°C carbon budget of ~220 $GtCO_2$ from 2020 on.

Abstract

Change in global mean surface temperature (ΔGMST) is a widely cited climate change indicator that figures prominently in IPCC reports, in which it was estimated via linear regression or differences between decade-plus period means. The Paris Agreement aims to limit warming since preindustrial (here approximated as 1850—1900) to "well below" 2 °C, and by knowing current ΔGMST it is possible to determine the remaining target-consistent warming and therefore a relevant remaining carbon budget. We propose non-linear continuous local regression (LOESS) using 40-year windows as a single method to derive ΔGMST with statistical uncertainty across all periods of interest. Using the three datasets with almost complete spatial coverage since the 1950s, we evaluate 1850—1900 to 2019 ΔGMST as 1.14 °C with likely (17—83 %) range of 1.05—1.25 °C, based on combined statistical and observational uncertainty, compared with 1880—2019 linear regression of 1.03 °C using all five operational datasets. In two model large ensembles LOESS, like period mean differences, is unbiased but provides a statistical error and gives warming through 2019, rather than a 2010—2019 average centred at the end of 2014. We compare observational and CMIP6 ΔGMST and estimate historical global surface air temperature change (ΔGSAT) using the CMIP6 ΔGSAT/ΔGMST ratio and its ensemble spread. Finally, we calculate remaining carbon budgets given our ΔGSAT of 1.21 °C with likely range of 1.11—1.32 °C. 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) 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 Report (IPCC 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 began in 1880 and the 1880—2012 ordinary least squares (OLS) linear trend was a "headline" warming estimate along with the HadCRUT4 1850—1900 to 2003—2012 difference. 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 and Way, 2014a; Cowtan and Way, 2014b; Cowtan et al., 2015) and Berkeley Earth (Rohde et al., 2011). 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 measured air and ocean water temperature anomalies

(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 AR5 Box 2.2 discusses 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 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 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 of smoothing spline. Nevertheless, as the IPCC enters the AR6 assessment, 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 a statistical error and show that the fit residuals follow the assumed 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, and directly relate 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 (between 1850—1900 and latest decade), plus a global warming index which SR1.5 used as the main example of "more formal methods of quantifying externally driven warming" (Haustein et al., 2017). We also test the performance of our LOESS estimates using output from the two model large ensembles that begin in 1850. A 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 (2.1.1), CMIP6 models (2.1.2), two large model ensembles (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 we discuss our results and issue recommendations in Section 4.

2 Source Data and Methods

2.1 Source Data

IPCC discussions of temperature change and carbon budgets include multiple sources and approaches. We now remind the reader of our approach and justify each component. This Section lists data sources, including temperature datasets and the forcing datasets required to derive a global warming index referenced in SR1.5 as a potential alternative to ΔGMST for tracking anthropogenic warming. Two large ensembles are included to allow performance tests of each ΔGMST calculation method and CMIP6 data are added for updated model-observation ΔGMST comparisons and to derive an adjustment from ΔGMST to ΔGSAT.

2.1.1 Global surface temperature datasetsTypically, 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 BerkeleyEarth (Rohde et al., 2011). 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.

131

132 **Table 1.** Five operational observational datasets.

Series	Land (LSAT)	Ocean (SST)	Interpolation	Averaging	Start year
HadCRUT4 (Morice et al., 2012)	CRUTEM4	HadSST3	None	Hemisphere average of gridboxes	1850
NOAA GlobalTemp v5 (Zhang et al., 2019)	GHCNv4	ERSSTv5	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 et al., 2013)	Berkeley Earth	HadSST3 (reprocessed & kriged)	Kriging (to 1200 km)	Area weighted average	1850

133 Differences in spatial interpolation can affect calculated GMST. HadCRUT4 calculates area-
134 weighted hemispheric means with no interpolation between its 5°×5° grid boxes. In contrast,
135 NASA GISTEMP, Cowtan-Way and Berkeley Earth use extensive interpolation and, crucially,
136 extrapolate LSAT over sea ice. Comparisons with temperature reanalyses, independent surface
137 data and satellite retrievals show that this significantly reduces coverage bias arising from poor
138 sampling of the fastest warming areas, especially the Arctic, since the mid-twentieth century
139 (Dodd et al., 2015; Cowtan et al., 2018a; Susskind et al., 2019). Evidence is mixed for earlier
140 periods where reduced coverage leads to larger interpolation uncertainty (Cowtan et al., 2018)
141 and differences between underlying SST datasets are the largest source of discrepancies.
142 GISTEMP and Berkeley Earth's interpolated areal coverage is two to three times that of
143 HadCRUT4 in the late 19th century, and is virtually complete since 1951 (See Figure S1,
144 Supplementary Information). NOAA GlobalTemp's interpolation results in coverage between
145 that of HadCRUT4 and NASA GISTEMP, but largely misses very high latitudes and has no
146 coverage over Arctic sea ice.

147 We use the published monthly anomaly series, except for Berkeley Earth where we use the area-
148 weighted average of the gridded series, which diverges from the published series over 1850—
149 1950 (Supplementary Information, Figure S2, S3). For series starting in 1850 anomalies are
150 relative to 1850-1900 while NASA GISTEMP and NOAA GlobalTemp are baselined such that
151 their 1880-1900 mean matches that of the three longer-running datasets. This allows NASA and
152 NOAA ΔGMST estimates from 1850-1900 in a consistent manner, replacing the 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 datasets (OpAll group) and the subset of three datasets with near-global post-1950 coverage (Global_3 group). 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 from Haustein et al (2017). 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

Conceptually, we first decompose Δ*GMST* as:

$$\Delta GMST = \Delta GMST_F + \Delta GMST_{var} \quad (1)$$

where $\Delta GMST_{var}$ represents internal variability and $\Delta GMST_F$ the forced response. 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 $\Delta GMST_F$ through 2019 would be an appropriate estimate of human-induced warming, $\Delta GMST_{F,anthro}$, 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 $\Delta GMST_F$. We test whether our derived $\Delta GMST_{LOESS}$ approximates $\Delta GMST_F$, and consider the decomposition in an individual run to be:

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

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 $\Delta GMST_F$), and the magnitude of $\Delta GMST_{var}$ is small at 40-year timescales, then resultant $\Delta GMST_{LOESS} \approx \Delta GMST_{F,anthro}$.

These tests use 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.

2.1.3 Climate 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. 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 series: 1) global SAT and 2) global blended SAT-SST. At each grid cell *i, j*, the blended monthly temperature $T_{\mathrm{blend,i,j}}$ is:

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

where $w_{\mathrm{SAT,i,j}}$ is the land plus sea ice grid cell fraction, and $T_{\mathrm{SAT,i,j}}$ and $T_{\mathrm{SST,i,j}}$ are the local anomalies relative to 1850-1900. For global SAT $w_{\mathrm{SAT,i,j}} = 1$ everywhere, and for the blended series $w_{\mathrm{SAT,i,j}} = 1$ 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, Section 2.2.4 the CMIP6 comparison and carbon budget calculation.

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:

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

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 multidecadal LOESS point-to-point ($\mathrm{LOESS_{md}}$) ΔGMST is based on the LOESS fit from 1880—2019; for any starting point, ΔGMST to 2019 is the $\mathrm{LOESS_{md}}$ fit evaluated in 2019 minus the start value. We also introduce "baseline" LOESS ($\mathrm{LOESS_{bsln}}$) as our main ΔGMST estimate. $\mathrm{LOESS_{bsln}}$ 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.

Our $\mathrm{LOESS_{md}}$ 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 S6, S7) show 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 $\Delta GMST_F$ 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

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

where $\boldsymbol{\rho_j}$ 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

$$\rho_1 = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}$$

$$\rho_j = \rho_1\phi^{j-1} \qquad j \geq 2 \tag{7}$$

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 $\hat{\rho}_1$.

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

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.

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

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 $p = 2$.

$$s_b^{'} = s_b \frac{v}{v_e} = s_b \frac{n_t - 2}{n_e - 2} \quad (11)$$

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 errors at these two points are independent, the corrected standard error $s^{'}_{\Delta T_n}$ for $\Delta\mathrm{GMST_n}$ becomes

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

For both OLS and $\mathrm{LOESS_{md}}$ 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 $\mathrm{LOESS_{bsln}}$ we assume that the mean error during 1850—1900 is very small relative to the end point error and so its error is taken to be:

$$s^{'}_{\Delta T_n} = \max(s_e^{'}) = \max(s_e)\frac{n_t - p}{n_e - p} \quad (13)$$

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 (11) over 1880-2019 and by (12) from 1850—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 S13 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 ΔGMST 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. Finally, $LOESS_{md}$ is calculated over 1880—2019 for each simulation. The distribution of ensemble member $\Delta GMST\text{-}\Delta GMST_F$ 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 S14). The LOESS decomposition filters in time: $\Delta GMST_F$ excursions shorter than our window will inflate statistical error, while multidecadal $\Delta GMST_{var}$ 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 ΔGMST 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 ΔGMST 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 % ΔGMST total uncertainty and ΔGMST was then converted to ΔGSAT using a CMIP5-derived scaling. This Section describes the comparison with CMIP6 ΔGMST and conversion of observed ΔGMST to ΔGSAT, and then details the carbon budget calculation, which largely follows the IPCC SR1.5 recipe .

LOESS series are generated for each CMIP6 air-only (GSAT) and blended (GMST) series, with the blended series being comparable to 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 $A_{blend} = \Delta GSAT / \Delta GMST$, and the ensemble A_{blend} is used to scale observed ΔGMST to obtain historical ΔGSAT for calculating the remaining carbon budget. The carbon budget calculation 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:

$$B_{lim} = \left(\Delta GSAT_{lim} - \Delta GSAT_{hist} - \Delta GSAT_{nonCO_2,fut}\right) / TCRE \; - E_{Esfb} \qquad (13)$$

where B_{lim} is the remaining carbon budget associated with a temperature limit $\Delta GSAT_{lim}$ (1.5 or 2°C), with $\Delta GSAT_{hist}$ 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. From the finding that observed and "human-induced" warming to date are approximately equivalent (Allen et al., 2018; Haustein et al., 2017), SR15 assessed $\Delta GSAT_{hist}$ 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 GSAT (0.99°C).

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

$$\Delta GSAT_{hist} = A_{blend} \Delta GMST_{Global\_3} \quad (14)$$

where A_{blend} is the median value from CMIP6 ensemble members 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 (Cowtan et al., 2015; Richardson et al., 2018).

SR1.5's likely total uncertainty in ΔGMST and $\Delta GSAT_{hist}$ was ±0.12°C. Here we derive likely ΔGSAT 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 ΔGMST estimates are given in Table 2.

OLS ΔGMST is always lower than LOESS, with some even lying outside 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 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.

We identify two factors that appear to contribute to the increased long-term $LOESS_{md}$ ΔGMST relative to OLS: improved recent coverage (Global_3 being higher than OpAll), and those using HadSST relative to ERSSTv5. Improved coverage tends to increase recent trends, while the SST datasets differ most strongly before and during WWII.

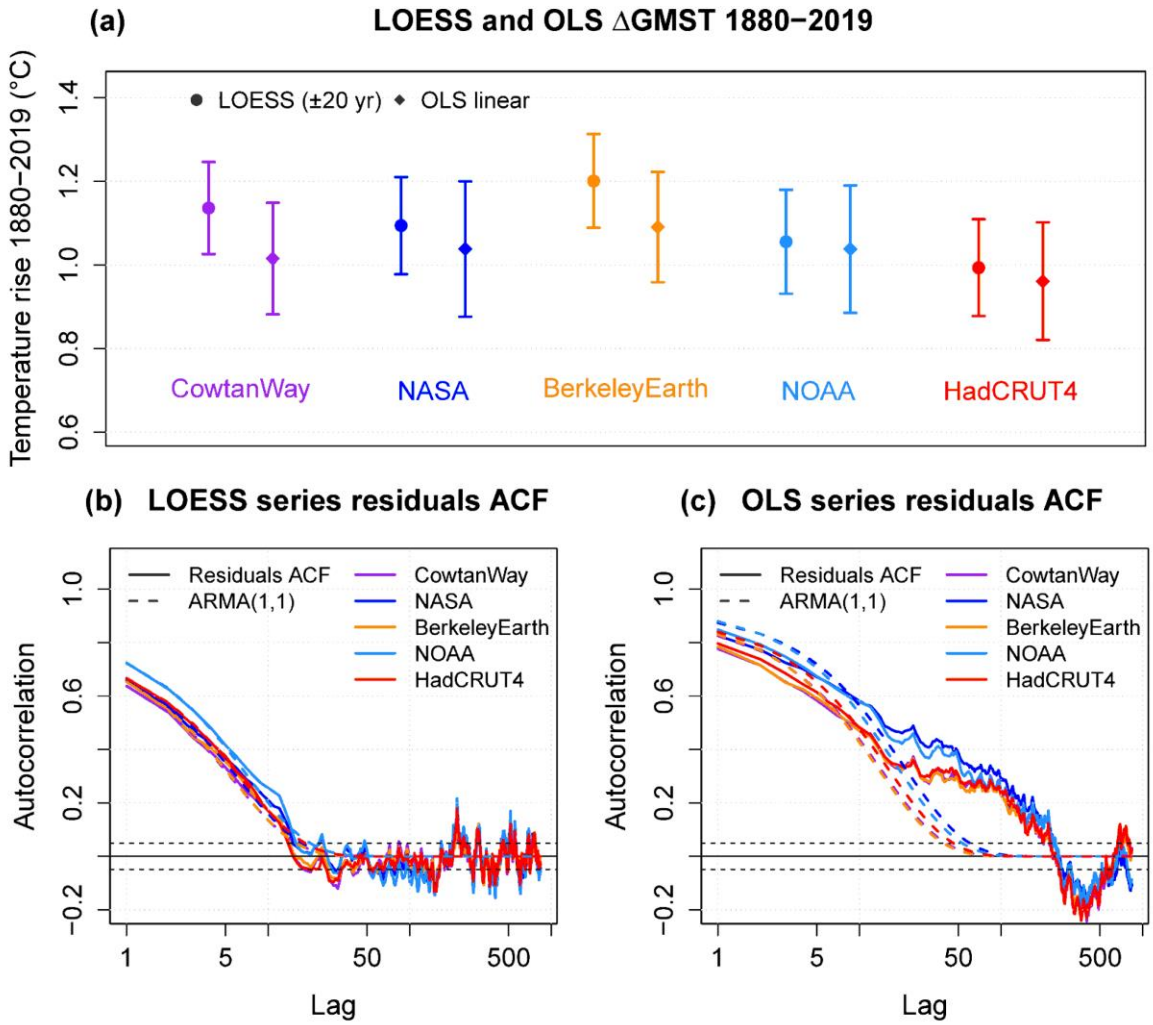

Figure 1: GMST series and 1880-2019 warming estimates. (a) LOESS (span ± 20 years) and OLS trends with 5-95% statistical fit uncertainty are shown for Cowtan and Way (purple), NASA GISTEMP (blue), Berkeley Earth (orange), NOAA GlobalTemp (light blue) and HadCRUT4 (red) over 1880-2019. **(b)** The autocorrelation 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 [0.94 - 1.10] *(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 [1.00 - 1.18]	0.99	1.06 [0.93 - 1.18]	1.04 [0.89 - 1.19]
NASA GISTEMP	1.12 [1.03 - 1.20]	1.01	1.09 [0.98 - 1.21]	1.04 [0.88 - 1.20]
Cowtan & Way	1.12 [1.04 - 1.20] (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 [1.11 - 1.26]	1.08	1.20 [1.09 - 1.31]	1.09 [0.96 - 1.22]
All Operational	**1.11** **[0.94 - 1.26]**	**1.00**	**1.10** **[0.88 - 1.31]**	**1.03** **[0.82 - 1.22]**
Full Global (3 series) *	**1.14 *** **[1.04 – 1.26]** ***(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).

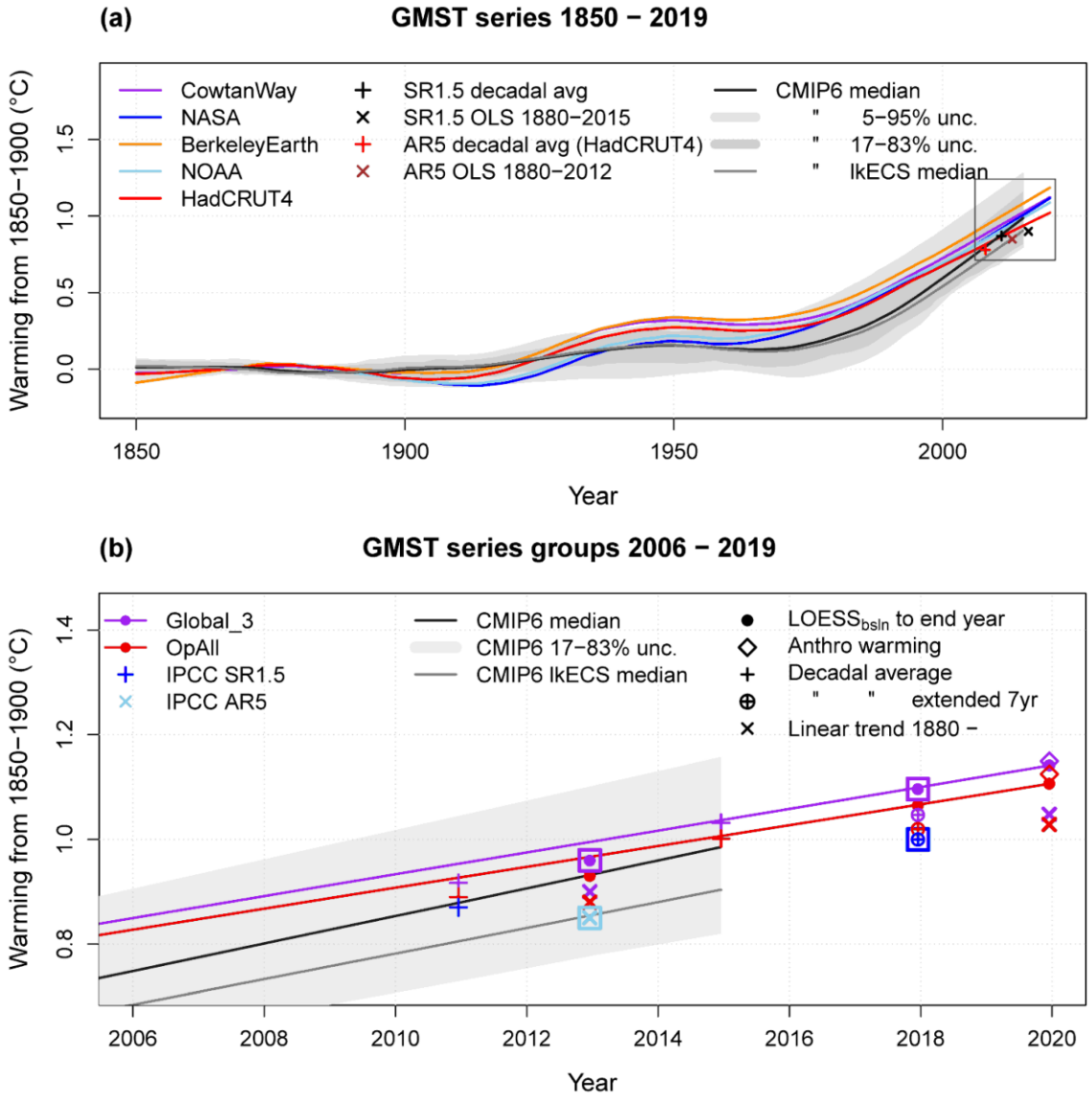

Figure 2: GMST series and group surface warming estimates. (a) Monthly series and multi-decadal $LOESS_{bsln}$ ΔGMST (span ± 20 years) are shown for HadCRUT4 (red), NOAA GlobalTemp (light blue), NASA GISTEMP (blue), Cowtan and Way (purple) and Berkeley Earth (orange), together with OLS and period estimates 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 21 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).

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

459 Separate tests showed that derived ΔGMST was similar when restricting CMIP6 spatial coverage
460 to that of Berkeley Earth, so we take the CMIP6 blended ensemble as directly comparable to the
461 Global_3 series (Figure S14). The Global_3 rise of 1.14°C is above the median CMIP6 estimate
462 extended linearly to 2019, 1.04°C [0.88 – 1.44]. However, the Global_3 incremental trend of
463 0.20°C/decade is lower than CMIP6's 0.26°C/decade [0.18 – 0.38] or the likely ECS sub-
464 ensemble's 0.25°C/decade [0.18 – 0.29].

465 In general, the observations are at or above recent IPCC long-term ΔGMST estimates. Figure
466 2(b) affords a closer view of recent ΔGMST estimates, including group $LOESS_{bsln}$ calculated to
467 2012 and 2017 for direct comparison to IPCC AR5 and SR1.5. As previously stated, AR5's main
468 estimate of 0.85°C was from linear OLS on the datasets available then. Since the mean 1880—
469 2012 OLS trend for OpAll is 0.89°C and $LOESS_{bsln}$ is 0.93°C, ΔGMST methodology accounts
470 for half of the discrepancy between AR5's 1880—2012 estimate and ours. The 2012 gap is even
471 wider for the Global_3 group; OLS to 2012 is 0.90°C and $LOESS_{bsln}$ is 0.96°C; that gap
472 continues to grow, reaching 0.09°C in 2019.

473 The SR1.5 2006-2015 mean ΔGMST of 0.87°C was extended to 2017 to provide an up-to-date
474 estimate of 1.0°C (Section 1.2.1.3 in Allen et al., 2018). The same adjustment applied to the
475 updated series shows a 0.03°C gap with $LOESS_{bsln}$. This discrepancy may be related to internal
476 variability suppressing early 2000s warming; taking 2008-2017 or 2010-2019 removes the
477 LOESS-period discrepancy. Both $LOESS_{bsln}$ and period estimates are in good agreement with the
478 slightly higher Haustein human-induced warming.

479 Figure 3 compares Global_3 $LOESS_{bsln}$ and period ΔGMST in more detail. Since IPCC SR1.5
480 explicitly considered the 2006-2015 mean as a proxy for the 1996-2025 average (relative to
481 1850-1900), we consider the centered 20-year average and a 30-year "extended" average
482 assuming the current linear 30-year trend continues over the next 15 years. Figure 3a shows that
483 $LOESS_{bsln}$ hews closer to the eventual average than the decade mean and confirms that 2006-
484 2015 was affected by an early 2000s slowdown. $LOESS_{bsln}$ has more stability relative to
485 anthropogenic warming estimates (Figure 3b) and has lower RMSE relative to the longer period
486 averages since the late 1990s (Figure 3c, 3d).

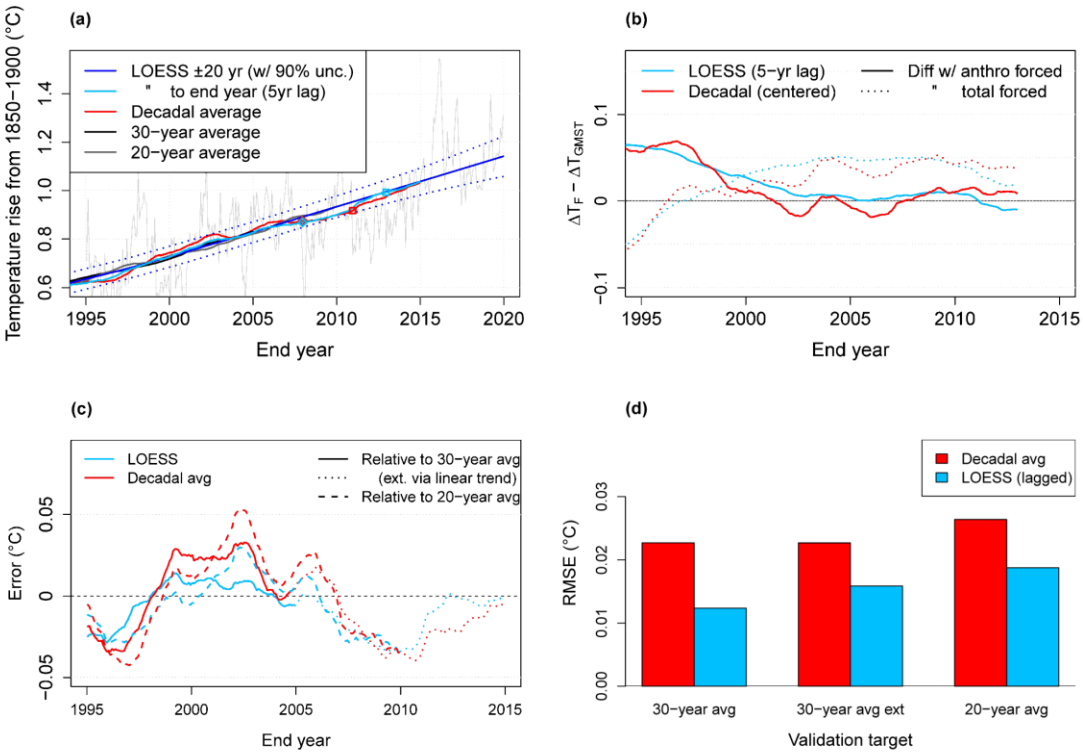

Figure 3: ΔGMST **estimation method validation based on average of 3 global series. (a) LOESS$_{bsln}$ to 2019 (blue) is shown with 5-year lagged LOESS (light blue), decadal average (red), 20-year average (light gray) and 30-year average (dark gray).** LOESS (light blue) versus decadal (red) differences 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).

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; recalculating the combined uncertainty following Section 2.2.4 yields 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 GMST$_F$ 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 ΔGMST estimates through 2019 for each start year from 1850—1980.

The ΔGMST$_F$ and LOESS ΔGMST agree well outside of periodic ΔGMST$_F$ spikes from volcanic eruptions, i.e. when the forced change is smooth over our ±20 year window, such that ΔGMST$_{LOESS}$≈ΔGMST$_{F.}$. OLS is biased relative to ΔGMST$_F$ 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.

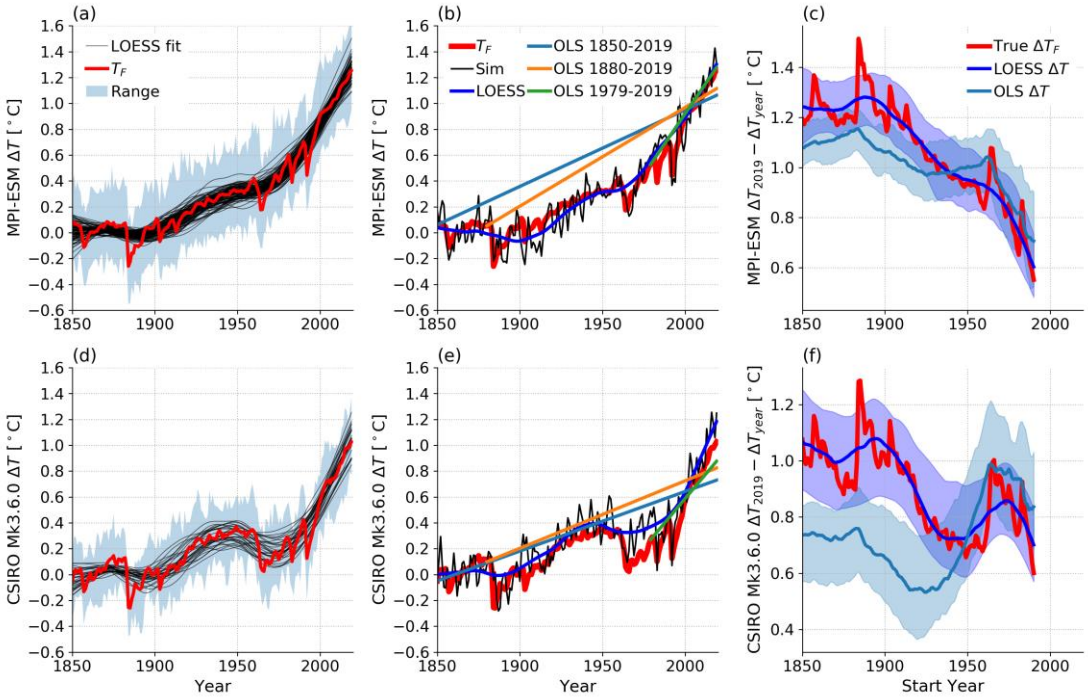

Figure 4. (a) MPI-GE SAT outputs, full ensemble range is shaded, each simulation's LOESS fit is in grey and the ensemble mean (our estimate of $GMST_F$) is in red. (b) example of fits applied to a single simulation (black) including LOESS (dark blue) and OLS over three different periods (straight lines) with $GMST_F$ in red. OLS lines are shifted up so that their end points correspond to the relevant ΔGMST for ease of comparison. (c) calculated ΔGMST for $GMST_F$ (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 ΔGMST 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 median bias magnitude <0.02 °C and an almost matching 5—95 % range. LOESS slightly outperforms centered period differences evaluated from 1850-1900 to end periods ranging from1986-1995 through 2010-2019 when validated against 30-year average (see Figure S15),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 for 1880—2019 LOESS-$GMST_F$ 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 analysis of these periods.

As our carbon budget calculations include an internal variability error component, we consider ensemble spread and statistical errors as candidates and compare the $LOESS_{bsln}$ ensemble 83rd minus 17th percentile and the statistical 17—83 % ranges for each run. The CSIRO Mk3.6.0 ensemble spread is 0.22 °C, equal to the largest individual run uncertainty (ensemble median

0.17 °C), while for MPI-ESM the ensemble spread (0.11 °C) and median statistical error (0.12 °C) almost match. The statistical errors are a reasonable representation of internal variability error in MPI, but underestimate that in CSIRO Mk 3.6.0. For the internal variability component of ΔGSAT uncertainty in our carbon budgets we present results both using statistical error (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 $\Delta GMST_F$ outside of years immediately following large volcanic eruptions, particularly supporting our $LOESS_{bsln}$ results.

Table 3. Long-term ΔGMST 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 ΔGMST[°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 ΔGMST of 1.14°C [1.05 – 1.25] (17—83% uncertainty) to an equivalent ΔGSAT. Our CMIP6 ensemble $LOESS_{bsln}$ A_{blend} ratio ΔGSAT/ΔGMST reflects an increase of ΔGSAT over ΔGMST of 5.8% [4.4, 7.2] in 2014.

Combining this ratio and its uncertainty in quadrature with our Global_3 ΔGMST, we obtain ΔGSAT 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 ΔGSAT 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 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 headline remaining carbon budget with a 66% chance to stay below 1.5°C, along with the historical cumulative CO_2 emissions and temperature change.

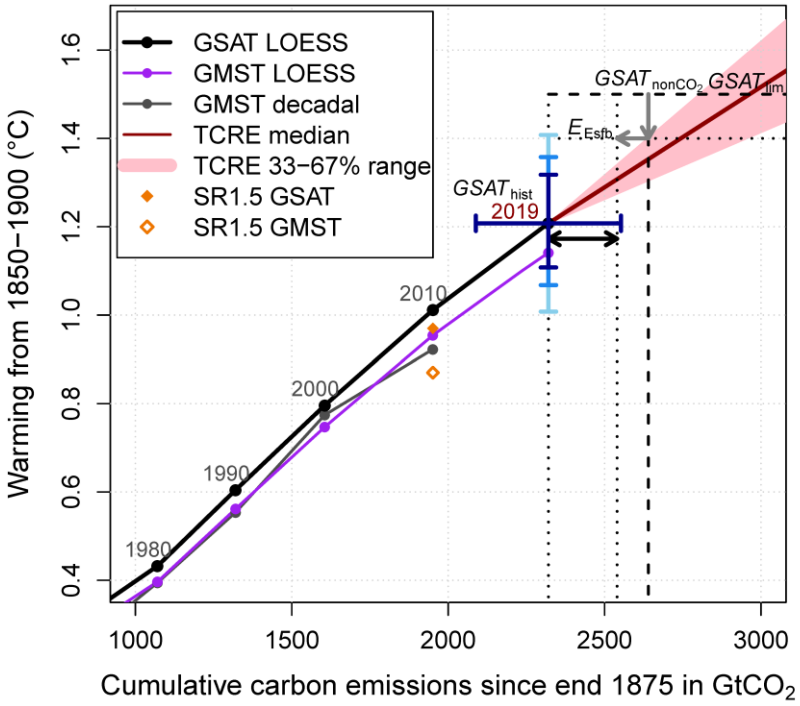

Figure 5: Global temperature change from 1850–1900 versus cumulative CO_2 emissions. The smoothed temperature response from the Global3 blended GMST group as decadal average (blue) and $LOESS_{md}$ trend (purple) are shown relative to cumulative CO_2 emissions from Friedlingsten et al (2019). The thick black line shows the Global3 GMST $LOESS_{md}$ trend, adjusted by the median difference between GSAT and blended historical runs from an ensemble of 21 CMIP5 models, again relative to cumulative CO_2 emissions. The pink shaded plume and dark red line are estimated temperature response to cumulative CO_2 emissions (TCRE) from 2019 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).

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 in SR1.5 (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 $\Delta GSAT_{hist}$ found in this study, combined with an identical TCRE spread starting in 2019 rather than a reference period centered at the start of 2011. In effect, the up-to-date estimate of $\Delta GSAT_{hist}$ 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 with a budget of 470 $GtCO_2$ from 2018. Our alternative budget using Global_3 ΔGMST instead of ΔGSAT 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 SR1.5 included HadCRUT4 and its more substantial coverage bias.

All estimates above account for Earth system feedbacks (CO_2 and CH_4 release from warming wetland and permafrost thaw) as in Rogelj et al. (2019): carbon budgets excluding this term would be 100 $GtCO_2$ higher.

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. 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 ΔGSAT 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 ΔGMST of 0.87°C to 2006-2015 using four datasets (1.0°C when extended to 2017) and estimated ΔGSAT 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 subsumed cumulative emissions up to 2017, necessitating an implicit extension of ΔGSAT to that date. The simplicity and coherence of our "up-to-date" ΔGMST and ΔGSAT 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 error derived using a noise model consistent with the data. Moreover, validation tests with observations and the large ensembles confirm $LOESS_{bsln}$ exhibits superior performance 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. These datasets are self-evidently more representative of global climate change and require smaller and less uncertain adjustments (~6%) to obtain ΔGSAT from ΔGMST, in contrast to the 17% adjustment applied to HadCRUT4 in IPCC SR1.5. The Global_3 datasets give 0.12 °C more warming than HadCRUT4 from 1850-1900 and the divergence related to incomplete 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.

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 fed further analyses such as future projections (Collins et al., 2013) and attribution (Bindoff et al., 2013).

If IPCC AR6 follows AR5, that would imply the three post-1850 datasets would form the basis for 2010-2019 period ΔGMST relative to 1850-1900. As noted above, $LOESS_{bsln}$ to 2019 offers a superior alternative. The case for excluding HadCRUT4 is compelling, although if the forthcoming HadCRUT5 represents quasi-global GMST then it should be included. 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 excluded according to our global coverage criterion. However, that case is less clear cut than HadCRUT4due to NOAA's complicated spatial coverage.

Since all observational datasets could be included, $LOESS_{bsln}$ ΔGMST removes a primary motivation for 1880-2019 ΔGMST in IPCC AR6. However, AR5 also compared ΔGMST trends from 1880 to short-term trends 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 ΔGMST uncertainty due to internal variability could be derived from the LOESS residuals. We combined this with observational uncertainty and carried it forward directly to ΔGSAT for carbon budget calculations, but it could also be used for other follow-on analyses. The median statistical errors from the large ensemble runs are within 25% of the ensemble spreads, and the residual autocorrelation structure implies potential for this approach.

However, models may not capture long-term internal variability. For example, recent Pacific changes may indicate stronger real-world multidecadal variability (e.g. England et al., 2014), although consensus is lacking (Seager et al., 2019). 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 artificially 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 ΔGMST and ΔGSAT estimates are close to observation-based anthropogenic warming, confirming a basic finding of IPCC SR1.5, we treat our ΔGSAT as an estimate of $\Delta GSAT_{F,anthro}$, albeit with appropriately wider uncertainties. In general, our approach yields straightforward and up-to-date

680 estimates of ΔGMST and ΔGSAT to inform remaining carbon budget calculations that
681 incorporate appropriate ΔGSAT uncertainties .

682 To summarize, we argue strongly in favor of $LOESS_{bsln}$ ΔGMST using series with near-global
683 coverage. Combining our statistical estimate of internal variability with dataset spread and
684 dataset parametric uncertainty results in a best estimate of warming from 1850—1900 to 2019 of
685 1.14 °C [1.05 – 1.25] (17-83% uncertainty). Not only is this updated through 2019, rather than
686 the prior-decade value of the IPCC's period mean difference, but it includes statistical error that
687 is not derivable for period mean differences.

688 Our CMIP6-derived GSAT adjustment yields corresponding ΔGSAT of 1.21°C [1.11–1.32] (17-
689 83% uncertainty), implying a remaining carbon budget of ~220 $GtCO_2$ for a 67% chance that
690 ΔGSAT since 1850-1900 remains below 1.5°C. This carbon budget is ~5.5 years of current
691 emissions. Our ΔGSAT estimate uncertainty can be adapted to a desired interpretation of
692 ΔGSAT, for example, as total or anthropogenic warming. All indices can be updated annually
693 and are only dependent on the temperature datasets, yielding a set of transparent and easily
694 communicated metrics to measure progress towards climate goals.

695

696

697 **Acknowledgments and Data**

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

700 MR's contribution was carried out at the Jet Propulsion Laboratory, California Institute of
701 Technology under a contract with the National Aeronautics and Space Administration
702 (80NM0018D004).

703 Berkeley Earth data are available from http://berkeleyearth.org/data/. Cowtan-Way data,
704 including merged HadSST4 series, are available from http://www-
705 users.york.ac.uk/~kdc3/papers/coverage2013/series.html. HadCRUT4 data are available from
706 https://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html . HadSST4 data are
707 available from https://www.metoffice.gov.uk/hadobs/hadsst4/data/download.html. NASA
708 GISTEMP data are available from https://data.giss.nasa.gov/gistemp/. NOAA GlobalTemp data
709 are available from https://www.ncei.noaa.gov/data/noaa-global-surface-
710 temperature/v5/access/timeseries/. CMIP6 data are available from https://esgf-
711 node.llnl.gov/search/cmip6/.

712

713 **References**

714 Allen, M.R., Dube, O.P., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S.,
715 Kainuma, M., et al. (2018). Framing and Context. In V. Masson-Delmotte, et al. (Eds.), Global
716 Warming of 1.5°C: An IPCC Special Report on the impacts of global warming of 1.5°C above
717 pre-industrial levels and related global greenhouse gas emission pathways, in the context of

718 strengthening the global response to the threat of climate change, sustainable development, and
719 efforts to eradicate poverty. Intergovernmental Panel on Climate Change.

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

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

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

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

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

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

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

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

747 Cowtan, K., Jacobs, P., Thorne, P., & Wilkinson, R. (2018a). Statistical analysis of coverage
748 error in simple global temperature estimators. Dynamics and Statistics of the Climate System,
749 3(1). https://doi.org/10.1093/climsys/dzy003

750 Cowtan, K., Rohde, R., & Hausfather, Z. (2018b). Evaluating biases in sea surface temperature
751 records using coastal weather stations. Quarterly Journal of the Royal Meteorological Society,
752 144(712), 670–681. https://doi.org/10.1002/qj.3235

753 Cowtan, K., & Way, R. G. (2014a). Coverage bias in the HadCRUT4 temperature series and its
754 impact on recent temperature trends. Quarterly Journal of the Royal Meteorological Society,
755 140(683), 1935–1944. https://doi.org/10.1002/qj.2297

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

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

762 Dessler, A. E., Mauritsen, T., & Stevens, B. (2018). The influence of internal variability on
763 Earth's energy balance framework and implications for estimating climate sensitivity.
764 Atmospheric Chemistry and Physics, 18(7), 5147–5155. https://doi.org/10.5194/acp-18-5147-
765 2018

766 England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., et al.
767 (2014). Recent intensification of wind-driven circulation in the Pacific and the ongoing warming
768 hiatus. Nature Climate Change, 4(3), 222–227. https://doi.org/10.1038/nclimate2106

769 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E.
770 (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental
771 design and organization. Geoscientific Model Development, 9(5), 1937–1958.
772 https://doi.org/10.5194/gmd-9-1937-2016

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

778 Forster, P. M., Maycock, A. C., McKenna, C. M., & Smith, C. J. (2019). Latest climate models
779 confirm need for urgent mitigation. Nature Climate Change. https://doi.org/10.1038/s41558-019-
780 0660-0

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

783 Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., et al.
784 (2016). ICOADS Release 3.0: a major update to the historical marine climate record.
785 International Journal of Climatology, 37(5), 2211–2232. https://doi.org/10.1002/joc.4775

786 Friedlingstein, P., Jones, M. W., O., Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., et al.
787 (2019). Global Carbon Budget 2019. Earth System Science Data, 11(4), 1783–1838.
788 https://doi.org/10.5194/essd-11-1783-2019

789 Goodwin, P., Katavouta, A., Roussenov, V. M., Foster, G. L., Rohling, E. J., & Williams, R. G.
790 (2018). Pathways to 1.5 °C and 2 °C warming based on observational and geological constraints.
791 Nature Geoscience, 11(2), 102–107. https://doi.org/10.1038/s41561-017-0054-8

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

795 Hartmann, D.L., A.M.G. Klein Tank, M. Rusticucci, L.V. Alexander, S. Brönnimann, et al.,
796 2013a: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science
797 Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental
798 Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., at al.
799 (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

800 Hartmann, D.L., A.M.G. Klein Tank, M. Rusticucci, L. Alexander, S. Brönnimann, et al.
801 (2013b): Observations: Atmosphere and Surface Supplementary Material. In: Climate Change
802 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment
803 Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-
804 K., Tignor, M., Allen, S.K., at al. (Eds.)]. Cambridge University Press, Cambridge, United
805 Kingdom and New York, NY, USA.

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

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

812 Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., & Thomas, S.
813 (2019). Hourly weather observations from the Scottish Highlands (1883–1904) rescued by
814 volunteer citizen scientists. Geoscience Data Journal, 6(2), 160–173.
815 https://doi.org/10.1002/gdj3.79

816 Hoegh-Guldberg, O., D. Jacob, M. Taylor, M. Bindi, S. Brown, et al., 2018: Impacts of
817 1.5ºC Global Warming on Natural and Human Systems. Global Warming of 1.5°C. An IPCC
818 Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related
819 global greenhouse gas emission pathways, in the context of strengthening the global response to
820 the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-
821 Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, et al. (eds.)].

822 Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al.
823 (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades,
824 Validations, and Intercomparisons. Journal of Climate, 30(20), 8179–8205.
825 https://doi.org/10.1175/jcli-d-16-0836.1

826 IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group
827 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.

Jeffrey, S. J., Rotstayn, L. D., Collier, M. A., Dravitzki, S. M., Hamalainen, C., Moeseneder, C., Wong, K. K., and J. I. Syktus (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

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

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

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., 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

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

905 Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019).
906 Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising
907 greenhouse gases. Nature Climate Change, 9(7), 517–522. https://doi.org/10.1038/s41558-019-
908 0505-x

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

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

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

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

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

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

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