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]Joint Institute for Regional Earth Systems Science and Engineering, University of California, Los Angeles, USA

Corresponding author: David C. Clarke (dave@daveclarke.ca)

Key Points:

- Continuous local regression is a compelling alternative to traditional IPCC trend estimation methods.
- Global warming assessed with full coverage land-ocean observational series reached 1.1°C in 2018 relative to 1850-1900.
- Global surface air temperature reached almost 1.2°C, implying a remaining 1.5°C carbon budget of ~275 $GtCO_2$ from 2019 on.

Abstract

Global mean surface temperature (GMST) is the most widely cited climate change indicator, with trends at multiple time scales figuring prominently in IPCC reports. Here we present an alternative non-linear continuous local regression (LOESS) method using multidecadal windows and evaluate GMST changes (ΔGMST) for five operational blended land-ocean surface temperature datasets. The best estimate of ΔGMST from pre-industrial (1850—1900) to 2018 is 1.12°C [0.93 – 1.27], based on three spatially complete global series. The IPCC's linear trend methodology applied to the three series assessed in IPCC AR5 yields 0.99°C [0.80 – 1.18], with much of the difference attributable to the trend methodology. LOESS yields lower estimates than linear over 1951-2018, and virtually identical results over 1979-2018. LOESS outperforms linear fits when validated against a 20- or 30-year averages relative to pre-industrial. We show that it reliably reproduces the known forced changes in ΔGMST when applied to output of a large model ensemble, except for years affected by large volcanic eruptions. Furthermore, our estimate of statistical uncertainties from a fit are reliable, by comparing against the ensemble spread. We also present a simple and easily updated remaining carbon budget to stay below 1.5 or 2°C, based on a global surface air temperature (SAT) estimate derived from model-based adjustment of blended full global GMST. Finally we perform a preliminary evaluation of recent short-term fluctuation. Continuous non-linear trend estimation offers a compelling alternative to linear trends for the assessment of long-term observational GMST series at multiple time scales.

1 Introduction

Global mean surface temperature (GMST) is arguably the key indicator of climate change (IPCC, 2013). GMST estimates and derived trends or changes, ΔGMST, have featured prominently in all IPCC assessments. Estimates of ΔGMST are a key component in IPCC 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, long-term IPCC ΔGMST estimates were a key scientific input to the Paris agreement to keep global surface temperature well below 2°C (UNFCCC, 2015).

This paper applies local regression (LOESS, Cleveland et al., 1992; Cleveland, 1979) for estimating forced changes, $\Delta GMST_F$. Conceptually, we decompose $\Delta GMST$ as:

$$\Delta GMST = \Delta GMST_F + \Delta GMST_{var} = \Delta GMST_{F,long} + \Delta GMST_{F,short} + \Delta GMST_{var} \quad (1)$$

where $\Delta GMST_{var}$ represents internal variability and we split $\Delta GMST_F$ into two components. We are primarily interested in $\Delta GMST_{F,long}$, which represents the ΔGMST in response to changes in long-lived forcing agents such as atmospheric CO_2. This contrasts with $\Delta GMST_{F,short}$, such as that due to volcanic eruptions. If $\Delta GMST_{F,short}$ is dominated by volcanism and average volcanism is constant, then for all long-term climate change relevant analyses $\Delta GMST_{F,short}$ is close to zero on average, enabling a best estimate of $\Delta GMST_{F,long}$. Methods of estimating ΔGMST may have different sensitivities to each component of Equation 1 and so may conflate them. We discuss how this affects our analysis; for example we show in Section 2.2.4 that our decomposition is easily related to an IPCC carbon budget calculation.

We argue for LOESS as an estimator of $\Delta GMST_{F,long}$ as it is conceptually simple, transparent, has quantifiable uncertainties and produces a continuous estimate. We show substantial advantages over other approaches used in the IPCC reports by applying it to a model large ensemble where we can validate against a reliable value for true $\Delta GMST_F$. We then apply it to observation-based datasets and show that the residual noise structure better matches that used in the calculation of uncertainties. Finally we calculate a best estimate of $\Delta GMST_F$, discuss recent internal variability and calculate an updated carbon budget.

All observation-based GMST series discussed herein merge land near-surface air temperatures (LSAT) from meteorological stations with sea surface temperatures (SST) from ship- and buoy-based measurements. Typically, monthly LSAT and SST analyses are generated for a regular longitude-latitude grid, and these are then merged to produce a GMST series. Before 2013, IPCC assessments relied solely or primarily on successive versions of the HadCRUT dataset, a collaboration of the UKMO Hadley Centre and UEA Climate Research Unit. The IPCC Fourth Assessment Report (IPCC AR4; Trenberth et al., 2007) used HadCRUT3 (Brohan et al., 2006) for its main estimate of long-term ΔGMST relative to a pre-industrial baseline of 1850-1900. IPCC AR4 also included GMST series from NASA GISS (Hansen et al., 2001) and NOAA NCDC (Smith and Reynolds, 2005), but only during 1900—2005. The NOAA and GISS series interpolate to better account for sparsely sampled areas; in contrast, HadCRUT3 and its successor HadCRUT4 (Morice et al., 2010) are strictly non-interpolated. However, HadCRUT provides an ensemble to robustly estimate some uncertainties, such as those associated with changing instrumentation.

By the IPCC Fifth Assessment Report (IPCC AR5; Hartmann et al., 2013a) the NOAA and NASA datasets stretched back to 1880 so the linear trend over 1880-2012 was introduced as a new “headline” estimate of warming since the 19th century, in addition to the intra-period estimate from HadCRUT4. Linear trends were also given for 1951-2012 and 1979-2012; all central estimates used ordinary least squares (OLS) with uncertainties adjusted to account for serial correlation in residuals by applying the Santer et al (2008) method to annual series (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 operational GMST series (both incorporating sophisticated statistical interpolation): Cowtan-Way (Cowtan and Way, 2014a; Cowtan and Way, 2014b; Cowtan et al., 2015) and Berkeley Earth (Rohde et al., 2011). Cowtan-Way was included in all SR1.5 main estimates of GMST change along with the three “traditional” series; these estimates included both intra-period and linear trend estimates of ΔGMST, with the four series mean from 1850-1900 to 2006-2015 serving as the primary metric.

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 model; 3) highly changeable estimates depending on the period selected; 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, since addressed by piecewise linear trend (Rahmstorf et al., 2017; Risbey et al., 2018). AR5 Box 2.2 presented a compelling continuous alternative for longer term ΔGMST estimation: a smoothing spline fit. Since AR5, other studies have presented alternative estimators for continuous long-term ΔGMST (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 intra-period or continuous trend estimates. For example, IPCC AR5 Box 2.2 estimated HadCRUT4 trends over 1900-2012 of 0.075 ± 0.013 °C decade^{-1} and 0.081 ± 0.010 °C decade^{-1} for linear OLS and smoothing spline trends respectively. SR15 table 1.2 shows a linear trend for Cowtan-Way of 1880-2015 of 0.93°C as opposed to an intra-period rise to 2006-2015 (i.e. centered at the end of 2010) of 0.91°C, implying a difference of 0.08°C to 2015, extending the period estimate by applying the SR1.5 assumption of 0.2°C per decade rise. Visser et al. (2018) compared linear trends to two multi-decadal "flexible" trend methods (integrated Random Walk and smoothing spline) for five GMST datasets over 1880-2016. The non-linear trends showed higher ΔGMSTand the two newer interpolated series, Cowtan-Way and Berkeley Earth had differences reaching ~0.1°C. Millar et al (2017a, 2017b) calculated a remaining carbon budget, based on an estimate of anthropogenic warming of 0.93°C to 2015 relative to 1860-1879, derived from HadCRUT4 by Otto et al (2015). The corresponding 1870—2015 linear trend was 0.84°C. Generally, linear trend estimates of long term GMST rise appear to be 0.05 – 0.10°C below estimates which do not assume a linear GMST progression.

In all these cases the ΔGMST estimates for each dataset fell within each other's 5-95% statistical uncertainties and the spread in ΔGMST estimates between different datasets is at least as wide as differences engendered by trend methodology. Nevertheless, as the IPCC enters the AR6 assessment, it may be prudent to consider whether new approaches should supplement or supplant the traditional linear trend approach. This work proposes LOESS with a fixed smoothing window of ± 20 years for the main multi-decadal trend analysis, resulting in trend evolution similar to smoothing spline and other techniques discussed above.

We include two components of uncertainty in our estimate of ΔGMST: statistical uncertainty from the LOESS fit including a correction for auto-correlation, which attempts to account for internal variability, plus dataset uncertainty derived from the spread between global temperature records.

The rest of the paper is structured as follows. Section 2.1 describes source data from observations (2.1.1), CMIP6 models (2.1.2) and a large model ensemble from (2.1.3). Section 2.2 covers methods, including trend estimation (2.2.1, trend methods and performance evaluation (2.2.2), large model ensemble evaluation (2.2.3) carbon budget calculation (2.2.4) and short term trend analysis (2.2.5). We present our results in Section 3, covering long-term trend analysis (3.1), large model ensemble analysis (3.2), remaining carbon budgets (3.3) and recent trends (3.4). Finally we discuss our results and issue recommendations in Section 4.

2 Source Data and Methods

2.1 Source Data

2.1.1 Global surface temperature data sets

Table 1 summarizes the five operational blended LSAT-SST series in widespread use. The first two columns show considerable overlap in the underlying datasets. There are two SST data sets: HadSST3 (Kennedy et al., 2011) from the UKMO, also used in Cowtan-Way and Berkeley Earth, and NOAA's ERSSTv5 (Huang et al., 2017), also used also by NASA GISTEMP. Similarly, the NOAA land station data set GHCNv4 (Menne et al., 2019) is also used by NASA GISTEMP, while CRUTEM4 (Jones et al., 2010) is used in Cowtan-Way. Even this description understates the overlap. For example both SST data sets rely primarily on the raw ungridded maritime observations from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman et al., 2016), albeit processed, filtered and supplemented in different ways.

Table 1. Five operational observational datasets.

Series	Land (LSAT)	Ocean (SST)	Interpolation	Averaging	Start year	Group(s)
HadCRUT4 (Morice et al., 2010)	CRUTEM4	HadSST3	None	Hemisphere average of gridboxes	1850	AR5_3 SR1.5_4
NOAA GlobalTemp v5 (Zhang et al., 2019)	GHCNv4	ERSSTv5	EOTs	Area weighted average	1880	AR5_3 SR1.5_4
NASA GISTEMP v4 (Lenssen et al., 2019)	GHCNv4	ERSSTv5	Distance weighting (to 1200 km)	80 zones x 100 sub-boxes	1880	AR5_3 SR1.5_4 Global_3
Cowtan-Way v2 (Cowtan & Way, 2014a; Cowtan & Way, 2014b; Cowtan et al., 2015)	CRUTEM4 (kriged)	HadSST3 (kriged)	Kriging (Complete)	Area weighted average	1850	SR1.5_4 Global_3
Berkeley Earth (Rohde et al., 2011)	Berkeley Earth	HadSST3 (reprocessed & kriged)	Kriging (to 1200 km)	Area weighted average	1850	Global_3

For this study's purposes, however, the differences in interpolation and averaging methods are more important. HadCRUT4 averages data within each 5°×5° gridbox and then calculates area-weighted hemispheric means with no interpolation. In contrast, NASA GISTEMP, Cowtan-Way and Berkeley Earth use extensive interpolation, and crucially, extrapolate land station surface temperatures over sea ice. Comparisons with temperature reanalyses, independent surface data and satellite retrievals show that this significantly reduces bias during the strong surface warming since the mid-twentieth century, although the evidence is mixed for earlier periods (Dodd et al., 2015; Cowtan et al., 2018a; Susskind et al., 2019).

GISTEMP and Berkeley Earth areal coverage is two to three times that of HadCRUT4 in the late 19th century, rising to virtually complete coverage since 1951 (See Figure S1, Supplementary Information). NOAA GlobalTemp interpolates via Empirical Orthogonal Transformations (EOTs), resulting in coverage between that of HadCRUT4 and NASA GISTEMP, but virtually no coverage at very high latitudes.

For trend analysis the datasets are assigned to various (overlapping) groups as seen in the last column of Table 1; groups are labeled by an abbreviation and the number of included series included. Thus, the Global_3 group includes three "full global" series identified above; two other groups (AR5_3 and SR1.5_4) identify series included in the last two IPCC surface temperature analyses, and OpAll_5 includes all 5 operational observational datasets.

For all series except Berkeley Earth, the published monthly anomaly series were used. However, there is a marked discrepancy between the Berkeley Earth's gridded dataset and the published monthly average over 1850-1950, so we use an area-weighted average of the gridded series instead (Supplementary Information, Figure S2). The three series starting in 1850 are baselined by subtracting the overall 1850-1900 mean from the original series. NASA GISTEMP and NOAA GlobalTemp are baselined such that their 1880—1900 mean matches that of the three longer-running datasets.

New versions of NASA GISTEMP and NOAA GlobalTemp were operationalized in 2019, so these should be stable for the foreseeable future. However, HadSST4 (Kennedy et al., 2019) was recently released so we also produce versions of HadCRUT4, Cowtan-Way and Berkeley Earth including HadSST4. We also perform a rudimentary sensitivity analysis of the difference between full and distance-limited interpolation by analyzing the impact on Cowtan-Way and CMIP6 ensemble trends when matching the reduced coverage of Berkeley Earth. We refer to such datasets as "masked", since we mask (i.e. remove from the calculation) grid cells in one series so as to match the lesser geographic coverage of another.

2.1.2 MPI-ESM Grand Ensemble

We only have one realization of real-world internal variability, and we do not know the true $\Delta GMST_F$. To address this we use output from the 100 historical simulations of the Max Planck Institute for Meteorology Grand Ensemble (MPI-GE, Maher et al., 2019)), taking the global mean near surface air temperature (SAT) over the full simulations (1850—2005) and baselining each to 1850—1900. Our approach is conceptually similar to that of Dessler et al. (2018), who used the MPI-GE to estimate how model internal variability can affect derived estimates of climate sensitivity.

By taking the ensemble mean as our best estimate of $GMST_F$, we can compare the performance of different estimators for $\Delta GMST_F$ as described in Section 2.2.2 below, and the ensemble spread provides an estimate of the uncertainty introduced by internal variability, conditional on the MPI-GE model's representation of $\Delta GMST_{var}$.

We also use the ensemble mean top of atmosphere net energy imbalance (ΔN_{TOA}) to flag years that are affected by volcanism and which will therefore contain a strong $\Delta GMST_{F,short}$ component. This is done by identifying all years where the year-over-year change in ΔN_{TOA} is equivalent to more than 0.3 W m^{-2} of cooling. Given the typical lifetime of volcanic effects on temperature, we exclude the identified years plus the two subsequent years. These are included in all calculations but separately discussed in some analyses. Note that we use global SAT only since we expect little effect of blending or masking in the comparison of derived ΔGMST to $\Delta GMST_F$ differences.

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 SAT over land or sea ice and SST over ocean (Eyring et al, 2016). These include near-surface air temperature ("tas") and sea surface temperature ("tos"), plus sea ice concentration ("sciconc" or "sciconca"). The simulations are 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* for each month, the blended temperature $T_{\mathrm{blend},i,j}$ is obtained as follows:

$$T_{\mathrm{blend},i,j} = w_{\mathrm{SAT},i,j}\, T_{\mathrm{SAT},i,j} + (1 - w_{\mathrm{SAT},i,j})\, T_{\mathrm{SST},i,j} \qquad (10)$$

where $w_{\mathrm{SAT},i,j}$ is the fraction of the grid cell that is land or sea ice, and $T_{\mathrm{SAT},i,j}$ and $T_{\mathrm{SST},i,j}$ are the local anomalies relative to 1850-1900. The global SAT series is calculated with $w_{\mathrm{SAT},i,j} = 1$ everywhere. For the blended series, $w_{\mathrm{SAT},i,j}$ is fixed for each calendar month by assigning all ocean area in a grid cell to sea ice if any of that calendar months over 1961-2014 has siconc > 3%.

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 their 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, and Section 2.2.5 the short-term trend analysis.

2.2.1 Trend calculations and their statistical uncertainty

The main analysis compares OLS linear trends to a continuous multidecadal LOESS trend (Cleveland et al., 1992), hereafter denoted $LOESS_{md}$. Estimates of ΔGMST are then easily obtained, for example the change from 1880—2018 is the fit evaluated in 2018 minus the 1880 value.

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

$$x_i = a + bt_i + e_i, \quad i = 1, \ldots, n \qquad (2)$$

where a and b are intercept and slope parameters to be fitted by OLS 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 $\mathrm{LOESS_{md}}$ 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). We choose local linear trend over quadratic, as this yields more stable end points.

An advantage of $\mathrm{LOESS_{md}}$ is that it is evaluated once over the whole series, and ΔGMST can then be estimated for any interval, whereas OLS trends must be evaluated anew for each interval and may have mismatched or highly changeable sub-interval trends.

Both methods assume statistically independent noise, necessitating a correction to the trend uncertainty 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) based on the general formula

$$n_e = \frac{n_t}{(1+2\sum_{j=1}^{n-1}\rho_j)} \tag{3}$$

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

$$1+2\sum_{j=1}^{n-1}\rho_j \approx 1+\frac{2\phi}{(1-\phi)} = \frac{(1+\phi)}{(1-\phi)} \tag{4}$$

where ϕ is estimated from the lag-one autocorrelation coefficient (Mitchell et al, 1966).

However, Foster and Rahmstorf (2011) demonstrated that the AR(1) model underestimated the autocorrelation of surface and tropospheric temperature trend residuals over 1979-2010, and proposed an autoregressive moving average, ARMA(1, 1) model in the form

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

$$\rho_j = \rho_1\phi^{j-1} \qquad j \geq 2 \tag{5}$$

Substituting (5) into (4) yields

$$1+2\sum_{j=1}^{n-1}\rho_j \approx 1+\frac{2\rho_1}{(1-\phi)} \tag{6}$$

Foster and Rahmstorf estimated the ARMA(1, 1) model in (5) from 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 first obtain both $\hat{\phi}$ and $\hat{\theta}$ and then estimated $\hat{\rho}_1$ according to (5). The MLE approach yields a more robust and efficient estimator $\hat{\phi}$, suitable for even very short series, as demonstrated by Monte Carlo simulations (see Figure S3).

Hausfather et al also introduced a bias correction to account for underestimated autocorrelation in shorter series. The bias correction is derived from the AR(1) in Tjøstheim and Paulsen (1996), extended to account for the positive difference between $\hat{\phi}$ and $\hat{\rho}_1$.

$$\begin{aligned} \hat{\phi}_{BC} &= \hat{\phi} + \left(1+4\left(2\hat{\phi}-\rho_1\right)\right)/n_t \\ \rho_{1\,BC} &= \rho_1 + \left(1+4\left(2\hat{\phi}-\rho_1\right)\right)/n_t \end{aligned} \tag{7}$$

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 (60 years) in length. A modified bias correction based on Nychka et al (2000) was also evaluated but was found to slightly overcorrect. For further details, see Figure S4.

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

$$n_e = \frac{n_t}{1+2\sum_{j=1}^{n-1}\rho_j} \approx \frac{n_t}{1+2\rho_{1BC}/(1-\hat{\phi})} \qquad (8)$$

In this study, corrections are estimated 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 required correction is applied directly to s_b, the standard error of the slope term b in (1), with $p = 2$.

$$s_b^{'} = s_b \frac{v}{v_e} = s_b \frac{n_t - 2}{n_e - 2} \qquad (9)$$

For non-parametric trend estimation such as LOESS, Monte Carlo simulations can be used to establish trend uncertainties, as in Visser et al (2016) for smoothing spline trends. Here we propose an alternative plausible heuristic uncertainty 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 reaches its maximum at the start and end points of the LOESS trend fit. If errors at these two points are independent, the corrected standard error $s_{\Delta T_n}^{'}$ ΔGMST$_n$ becomes

$$s_{\Delta T_n}^{'} = \sqrt{2}\,\max(s_e^{'}) = \sqrt{2}\,\max(s_e)\frac{n_t - p}{n_e - p} \qquad (10)$$

Monte Carlo simulations of trend plus simulated ARMA(1, 1) noise produces a trend probability distribution function nearly identical to that engendered by (10) for Cowtan-Way over 1880-2018 (see Figure S5). 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.

2.2.2 Estimates of observational ΔGMST, error components and performance tests

Following IPCC AR5, we assess OLS and LOESS$_{md}$ ΔGMST from 1880, 1951 and 1979 for each GMST series and our GMST groups. We also provide an additional “hybrid” LOESS$_{md}$ ΔGMST relative to the 1850-1900 baseline, which is simply LOESS$_{md}$ evaluated at a recent end point. We extend our calculations to 2018, the latest full year of data. Following IPCC SR1.5 we also calculate intra-period ΔGMST estimates by subtracting mean GMST over 1850—1900 from selected recent decades, for example that during 2009—2018. We compare “hybrid” $LOESS_{md}$ to intra-period ΔGMST by taking the central value of the end period fit, e.g. for 2009—2018 we evaluate LOESS at the beginning of 2014. $LOESS_{md}$ hybrid long term trends are also compared to selected to selected GMST-derived estimates of “human induced” warming (Haustein et al., 2017) and to CMIP6 outputs (see Section 2.2.4).

For each ΔGMST period we report statistical and observational uncertainty (where available). Firstly the statistical errors derived in Section 2.2.1, which are based on the fit residuals, so capture uncertainty introduced by internal variability and due to differences between the true GMST evolution and that assumed in the statistical model. For example, the OLS fit is linear, so any nonlinear components of $\Delta GMST_F$ will lead to larger residuals and increased statistical error. Secondly, for the observational uncertainty we report the 5—95 % range of ΔGMST values for OLS and $LOESS_{md}$ applied to each of the 100 member HadCRUT4 and Cowtan-Way ensembles. The HadCRUT4 ensemble uses a Monte-Carlo method to assess the fully correlated errors engendered by parametric uncertainty related to bias adjustments (Kennedy et al., 2011); Cowtan-Way reprocesses the HadCRUT4 the ensemble by the application of kriging to each ensemble member.

As well as comparing the temperature evolution, we compare the autocorrelation of the OLS and $LOESS_{md}$ residuals. Given that the statistical uncertainty calculation assumes ARMA(1,1) noise, the residual autocorrelation should follow ARMA(1,1) in order for the fit statistics to be considered reliable. Finally we assess the performance of the fit-derived ΔGMSTs against period mean differences for the Global_3 group. IPCC SR1.5 explicitly considered their main intra-period 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 1999-2018 linear trend through 2028. We argue that a smaller bias and root mean square error (RMSE) relative to the 20- and 30-year means represents better performance.

2.2.3 Large Ensemble Analysis for Method Validation and Uncertainty Calculation

The performance of each ΔGMST estimator is assessed by applying it to each of the MPI-GE members. $LOESS_{md}$ fits are calculated for each simulation’s annual output, as are linear OLS fits ending in 2005 from every start year from 1850—1980. We also use the “hybrid” calculation above, evaluating the fit at the end of 2000 to approximate the 1850—1900 to 1996—2005 ΔGMST, and compare it against the difference of period means. An advantage of this large ensemble is that we can estimate $GMST_F$ from the ensemble mean in each year and thereby compare each estimator’s performance against this. The distribution of ensemble member $\Delta GMST$-$\Delta GMST_F$ values then provides an estimate of the bias and uncertainties for each estimator and each period. In particular, for $LOESS_{md}$ the spread should be comparable to the statistical uncertainty from Section 2.2.2 provided that the residual variance is primarily driven by $\Delta GMST_{var}$.

2.2.4 CMIP6 comparisons, SAT adjustment and remaining carbon budget

LOESS series are generated for each CMIP6 SAT and blended SAT-SST series, and the ensemble is used to evaluate a median trend and uncertainty envelope. The blended series are then compared to the corresponding GMST observations. As several CMIP6 models have effective climate sensitivity (ECS) outside the IPCC's 1.5-4.5°C likely range, a subset of "likely ECS" models was also assessed (Forster et al., 2019).

The percentage increase in $LOESS_{md}$ ΔSAT relative to blended SAT-SST ΔGMST, $A_{blend} = (\Delta T_{SAT} - \Delta T_{blended})/\ \Delta T_{blended}$ was evaluated for each ensemble member. This yields an adjustment factor that can be applied to the blended observation series to estimate historical ΔSAT, a key input to the calculation of the remaining carbon budget.

The carbon budget calculation is based on the framework established in IPCC SR15 (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:

$$B_{lim} = \left(\Delta T_{lim} - \Delta T_{hist} - \Delta T_{nonCO_2,fut}\right)/\,TCRE \;-E_{Esfb} \quad (11)$$

where B_{lim} is the remaining carbon budget associated with a temperature limit ΔT_{lim} (1.5 or 2°C), with ΔT_{hist} the historical human-induced warming to date and $\Delta T_{nonCO_2,fut}$ 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. Building on the finding that observed and "human-induced" warming to date can be regarded as equivalent (Allen et al., 2018; Haustein et al., 2017), SR15 assessed ΔT_{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 difference between the equivalent CMIP5 blended-masked estimate (0.86°C) and global SAT (0.99°C).

In contrast, here we select the Global_3 GMST group and so do not need to rely on a model correction for the bias introduced by incomplete and changing geographic coverage. This means we are relying more heavily on observation-based statistics and less on climate model outputs, since the SAT adjustment factor is much smaller than the blended-masked adjustment. Our estimate for ΔT_{hist} is:

$$\Delta T_{hist} = A_{blend}\Delta T_{Global\_3} \quad (12)$$

where A_{blend} is the median of the A_{blend} values calculated for CMIP6 ensemble members and ΔT_{Global_3} is the $LOESS_{md}$ ΔGMST of the Global_3 group. ΔT_{Global_3} uncertainty is assessed by combining the 5—95% observational uncertainty of Cowtan-Way with the spread of central estimates of the Global_3 series and A_{blend} uncertainty is determined from the 5—95 % range of the ensemble. ΔT_{hist} 5—95% uncertainty is the sum of the relative uncertainties of ΔT_{Global_3} and A_{blend}.

As in Rogelj et al (2019), T_{nonCO_2} is estimated as 0.1°C (0.2°C) for $T_{\lim}$ of 1.5°C (2°C). TCRE uncertainty percentiles are based on AR5 likely range of 0.2–0.7°C per 1,000 Gt CO_2 (Collins et al., 2013), as in Nauels et al (2019). E_{Esfb} of 100 Gt CO_2 from permafrost thawing by until 2100 is included in one of the two primary analyses. SR1.5 also included alternative carbon budgets based on a lower T_{hist} from the average of the blended GMST datasets with no SAT adjustment. Our alternative is the Global_3 dataset average without the SAT adjustment. To contextualize the remaining budget against cumulative emissions to date we include data from the 2019 Global Carbon Budget (Friedlingstein et al., 2019).

2.2.5 Short term trend analysis

Finally, having validated $LOESS_{md}$, presented evidence for its advantages in estimating long-term ΔGMST and shown an example of its application to carbon budgets, we consider its implications for short-term trend analysis in more detail. In particular, we evaluate recent 15-year overlapping trends compared to the corresponding 30-year and 60-year trends. Such 15-year OLS trends were discussed in AR5 and are planned for inclusion in AR6.

For continuous non-linear 15-year trends, we apply a pentadal $LOESS_{pent}$ as for $LOESS_{md}$ but with span $\alpha_{pent} \pm 5$ years. We calculate ΔGMST as before, but over shorter intervals and express results in °C decade^{-1}. These 15-year trends are compared to the corresponding 30 and 60 year $LOESS_{md}$ trends. The $LOESS_{pent}$ trends can be overly sensitive to variability near the end points, so an end adjustment that modulates the $LOESS_{pent}$ by partial "return" to the long-term $LOESS_{md}$ trend was instituted. Two techniques were evaluated, and following superior validation (see Figure S6) results we selected a "first difference adjustment" which gradually matches the first difference of the $LOESS_{pent}$ trend line to that of the $LOESS_{md}$ trend. The 15-year OLS linear trends are evaluated conventionally, and are similarly compared to 30 and 60 year linear trends.

While the statistical uncertainty methodology described in Section 2.2.1 has been applied to very short term OLS trends (Hausfather et al., 2017) and could be extended to $LOESS_{pent}$, we defer this aspect for now. We do note that methods based on annual series (as in IPCC AR5) are ill-suited for 15 or even 30 year trends, as robust estimate of autocorrelation necessarily requires the higher sample numbers of monthly series. As well, observational uncertainties at very short time scales are dominated by partially correlated errors that are not captured in GMST ensembles (Kennedy et al., 2019; Hausfather et al., 2017), implying large underestimation of observational uncertainties in the HadCRUT4 15-year trend presented in Fyfe et al (2011) and IPCC AR5 (Flato et al., 2013).

Therefore for this preliminary analysis of very short term $LOESS_{pent}$ and OLS trends, we follow Fyfe et al (2016) and calculate central estimates of GMST series observational trends, and compare to the spread of CMIP6 ensemble trends.

3 Results

3.1 Long term trend analysis

The 1880—2018 Cowtan-Way ΔGMST estimates in Figure 1 demonstrate that the OLS and $LOESS_{md}$ central estimates lie outside each other's 5-95% uncertainty range according to statistical fit uncertainty (panel 1a) or observational uncertainty (panel 1b). The autocorrelation function of the residuals more closely matches ARMA(1,1) for LOESS (panel 1c) than OLS (panel 1d), supporting $LOESS_{md}$ over linear OLS and justifying our use of an error correction derived from ARMA(1,1) assumptions.

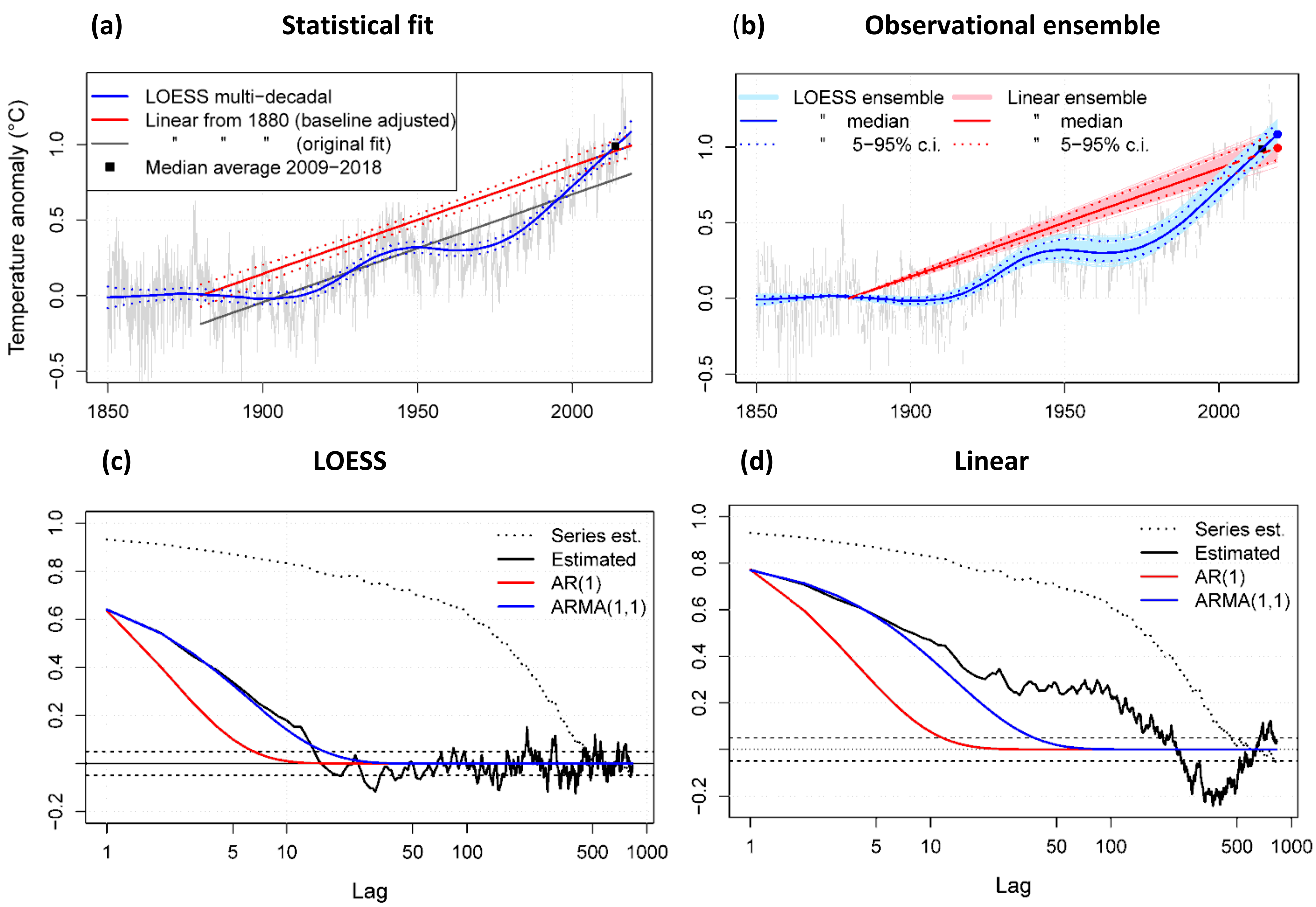


Figure 1: LOESS and OLS linear trend estimation 1880-2018. Top (a - b) Cowtan -Way monthly series (light gray) is shown with LOESS multi-decadal trend (blue), OLS linear trend (red) and 2009-2018 average (black square). The OLS linear trends have been shifted to zero start per IPCC methodologyz. **(a)** Trends are given with ARMA(1,1) corrected 5%-95% confidence interval (dotted lines). **(b)** LOESS (thin light blue lines) and OLS (thin pink lines) trends are derived from Cowtan and Way 100-member ensemble. **Bottom (c)** Autocorrelation function (ACF) of 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.

Estimates of ΔGMST for the observational series and groups, along with CMIP6 are given in Table 2. The datasets have similar rankings for both OLS and $LOESS_{md}$ over 1880-2018, with the highest being Berkeley Earth (1.17°C and 1.07°C) and the lowest HadCRUT4 (0.98°C, 0.94°C). All long-term $LOESS_{md}$ ΔGMST values are greater than the corresponding OLS estimates. The Global_3 series exhibit a greater relative difference than the non-global series; the difference between Berkeley Earth and HadCRUT4 in LOESS trend is ~0.2°C, but only 0.13°C for OLS. Thus OLS not only produces lower ΔGMST, but also de-emphasizes the differences between the datasets. It's also notable that the LOESS-OLS difference is higher for the three HadSST4 based series than for the two ERSSTv5 based series, with NOAA showing the smallest difference.

Table 2: 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 observational uncertainty where available (HadCRUT4, Cowtan & Way) and curly brackets denote CMIP6 ensemble spread. Best estimates from 3 full global series are denoted by *.

Period: / *Series:*	**1850-1900 to 2018**	**1880 - 2018**		**1951 - 2018**		**1979 - 2018**	
	LOESS	**LOESS**	**Linear**	**LOESS**	**Linear**	**LOESS**	**Linear**
HadCRUT4	1.00 [0.88 - 1.11] *(0.95 – 1.04)*	0.98 [0.86- 1.09] *(0.93 – 1.02)*	0.94 [0.80 - 1.08] *(0.88 – 1.01)*	0.73 [0.61 - 0.85] *(0.67 – 0.76)*	0.82 [0.67 - 0.97] *(0.76 – 0.88)*	0.67 [0.56 - 0.78] *(0.64 – 0.71)*	0.68 [0.56 - 0.80] *(0.65 – 0.72)*
NOAA GlobalTemp	1.06 [0.89 - 1.13]	1.03 [0.90 - 1.15]	1.02 [0.87 - 1.17]	0.84 [0.72 - 0.96]	0.94 [0.83 - 1.06]	0.70 [0.59 - 0.82]	0.68 [0.55 - 0.80]
NASA GISTEMP	1.09 [0.97 - 1.20]	1.06 [0.94 - 1.18]	1.02 [0.86 – 1.18]	0.90 [0.80 – 1.01]	1.00 [0.88 - 1.12]	0.75 [0.65 - 0.86]	0.73 [0.62 - 0.85]
IPCC AR5 (3 series)	**1.05 [0.88 – 1.20]**	**1.02 [0.85 – 1.18]**	**0.99 [0.80 - 1.18]**	**0.82 [0.61 - 0.97]**	**0.92 [0.67 - 1.12]**	**0.71 [0.56 - 0.86]**	**0.70 [0.54 - 0.85]**
Cowtan & Way	1.09 [0.98 – 1.21] *(1.00 – 1.19)*	1.11 [1.00 - 1.22] *(0.99 – 1.18)*	0.99 [0.86 – 1.12] *(0.88 – 1.09)*	0.78 [0.67 - 0.89] *(0.72 – 0.81)*	0.86 [0.70 - 1.01] *(0.80 – 0.92)*	0.73 [0.62 - 0.84] *(0.69 – 0.78)*	0.74 [0.62 - 0.86] *(0.71 – 0.78)*
IPCC SR15 (4 series)	**1.06 [0.88 – 1.20]**	**1.04 [0.85 – 1.22]**	**0.99 [0.80 - 1.18]**	**0.81 [0.59 - 0.97]**	**0.90 [0.67 - 1.12]**	**0.71 [0.56 - 0.86]**	**0.71 [0.54 - 0.86]**
Berkeley Earth	1.16 [1.05 - 1.27]	1.17 [1.06 - 1.28]	1.07 [0.94 - 1.20]	0.82 [0.72 - 0.92]	0.89 [0.75 - 1.03]	0.74 [0.64 - 0.83]	0.74 [0.63 - 0.85]
All Operational	**1.08 [0.88 - 1.27]**	**1.07 [0.85 - 1.28]**	**1.01 [0.80 - 1.20]**	**0.81 [0.59 - 0.97]**	**0.90 [0.67 - 1.12]**	**0.72 [0.56 - 0.86]**	**0.71 [0.54 - 0.86]**
Full Global (3 series) *	**1.12 * [0.97 - 1.27] *(1.00 – 1.25)***	**1.11 [0.94 - 1.28]**	**1.03 [0.86 - 1.20]**	**0.83 * [0.67 - 0.97]**	**0.91 [0.70- 1.12]**	**0.74 * [0.62 - 0.86]**	**0.74 [0.62 - 0.86]**
CMIP6 global SAT/SST	***1.04 {0.88 – 1.40}***	***1.08 {0.91 – 1.42}***	***0.82 {0.52 – 1.32}***	***0.99 {0.71 – 1.30}***	***0.96 {0.77 – 1.31}***	***0.87 {0.63 – 1.35}***	***0.97 {0.60 – 1.48}***
CMIP6 global SAT	*1.09 {0.91 – 1.45}*	*1.12 {0.94 – 1.48}*	*0.85 {0.56 – 1.37}*	*1.03 {0.72 – 1.33}*	*1.00 {0.78 – 1.36}*	*0.91 {0.65 – 1.37}*	*1.00 {0.63 – 1.53}*

The same patterns holds for LOESS changes from 1850—1900 to 2018, with ΔGMST ranging from 1.05°C [0.88-1.20] for the AR5 group up to 1.12°C [0.93-1.27] for the Global_3 series, which we report as our best estimate. This best estimate is 0.13 °C larger than the 0.99 °C from an OLS fit to the AR5 datasets through 1880—2018. Of the difference, 0.01 °C comes from the switch to an 1850—1900 baseline, 0.08 °C from the application of LOESS rather than OLS, and

0.04 °C from the use of the global datasets. Therefore we attribute most of the difference to the differing trend methodology, although some of this is due to the combined effect of how the OLS-LOESS difference increases for the Global_3 dataset relative to the non-global series. For

1951-2018, LOESS ΔGMSTs are lower than OLS, while 1979-2018 ΔGMSTs are almost identical, reflecting the near-linear rise since the mid-1970s.

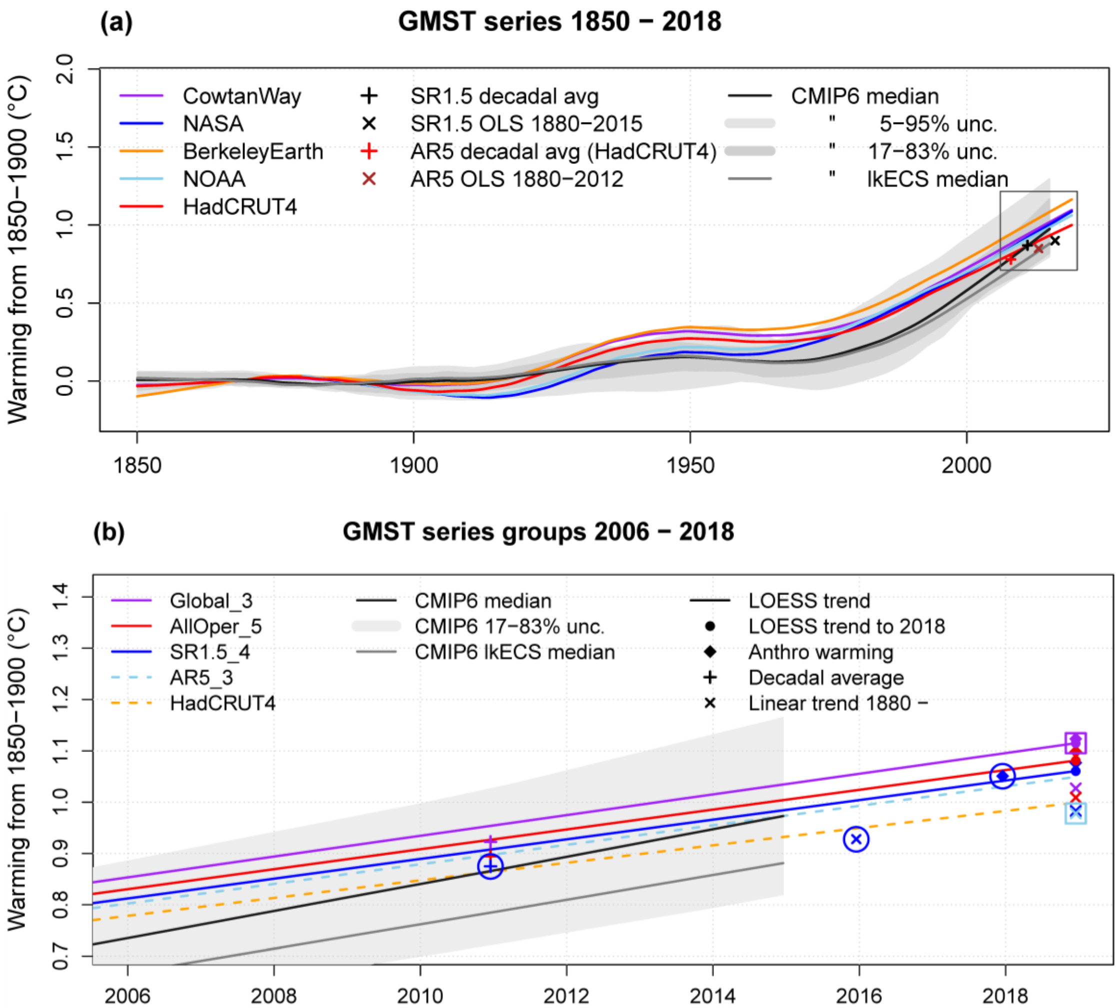


Figure 2: GMST series and group surface warming estimates. (a) Monthly series and multi-decadal LOESS trends (span ± 20 years) are shown for HadCRUT4 (red), NOAA GlobalTemp (light blue), NASA GISTEMP (blue), Cowtan and Way (purple) and Berkeley Earth (orange), together with OLS and period estimates from 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 trend (solid line with filled circle) is shown for each GMST grouping: Global_3 (purple), AllOper_5 (dark red), SR1.5_4 (dark blue), AR5_3 (light blue), along with HadCRUT4 (orange). AR5_3 and HadCRUT4 are shown as dashed lines to indicate these are now deprecated. 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). Updated IPCC SR15 estimates have been circled in dark blue. The AR_5 OLS trend and Global_3 LOESS trends to 2018 are highlighted by light blue and purple .squares respectively.

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

The difference in ΔGMST between blended 100% spatially complete and distance-limited interpolated series is negligible, when assessed by masking Cowtan-Way or the CMIP6 ensemble to Berkeley Earth coverage. This implies that the CMIP6 blended ensemble is directly comparable to the three global series. The Global_3 rise of 1.12°C is firmly in the upper half of the extended CMIP6 estimate extended to 2018, 1.04°C [0.88 – 1.44]. However, the Global_3 incremental trend of 0.20°C/decade is lower than the CMIP6 trend of 0.26°C/decade [0.18 – 0.38] or the CMIP6 likely ECS sub-ensemble 0.25°C/decade [0.18 – 0.29]. CMIP6 also shows more acceleration than observations since 1979 as evidenced by CMIP OLS-LOESS differential in this period (0.97°C versus 0.87°C).

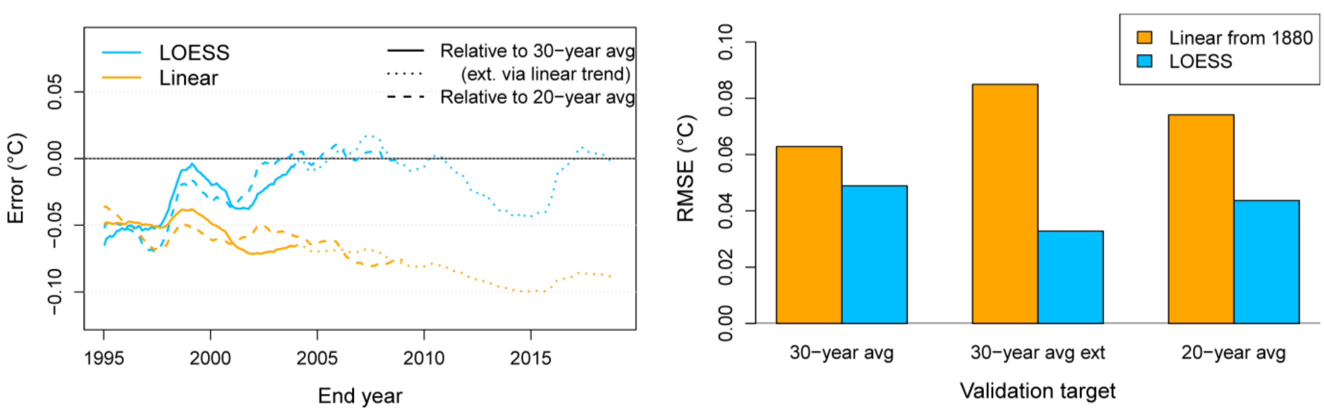
Figure 2(b) affords a closer view of ΔGMST estimates for different periods from models, observations and "human induced warming" from Haustein et al (2017). Our values are slightly higher than in SR1.5, as the most recent versions of the datasets are slightly warmer. As expected, the SR1.5 OLS estimates are below those from LOESS. Even the SR1.5 2006-2015 mean ΔGMST of 0.88°C (centered at the end of 2010) is slightly under the LOESS value at the same time of 0.91°C. This discrepancy may be related to internal variability, which suppressed early 2000s warming; the most recent available decade at the time of SR15's publication, 2008-2017, is virtually identical to the corresponding LOESS ΔGMST. The Haustein "human induced warming" estimate of 1.05°C to 2017 is slightly higher.

The Global_3 LOESS ΔGMST is more consistent with the corresponding 2008—2017 mean and Haustein estimates. As can be seen in Figure 3a, the LOESS-OLS difference arose in the early 2000s and has been entrenched since 2005. Comparison of these estimates to the corresponding centered 20-year or 30-year average in Figure 3b demonstrates that the LOESS trend has tracked the longer periods closely recently, and the comparison with 30-year "extended" average (which assumes continuation of the 30-year trend over the next 15 years) indicates that LOESS's smaller errors could continue for some time.

Use of HadSST4 instead of HadSST3 raises ΔGMST estimates of the three HadSST-based series, while LOESS-OLS differences remain similar. We conservatively estimate that Global_3 ΔGMST rises by 0.04°C to 1.16°C [1.97 – 1.31] (see Figure S7).

The LOESS and decadal intra-period estimates are more consistent with each other than with that of OLS. Figure 3(d) shows that the decadal mean and LOESS perform similarly, with slightly lower RMSE for LOESS since the late 1990s.

(a) (b)



(c) (d)

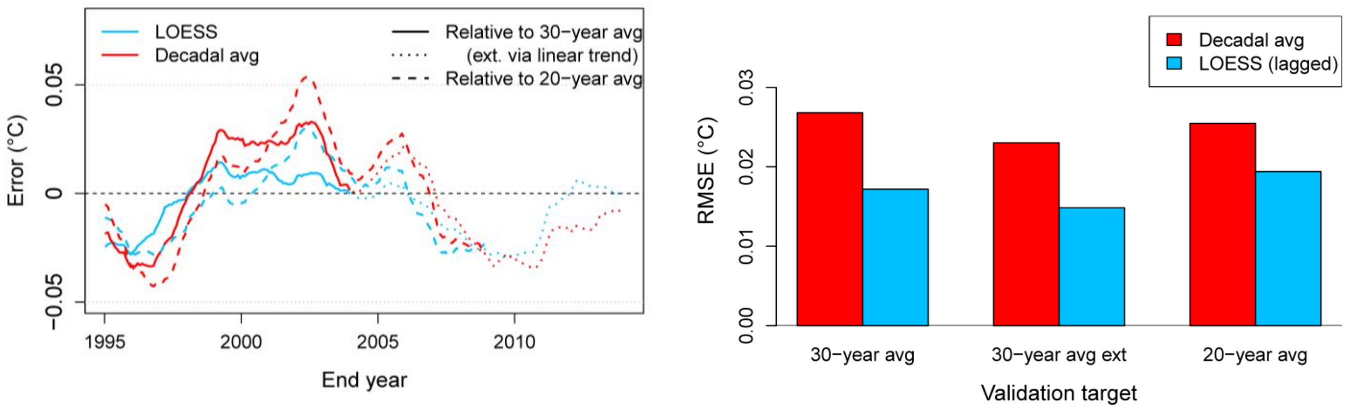


Figure 3: Trend estimation method validation based on average of 3 global series. (a-b) LOESS (light blue) versus linear trend (orange). **(c-d)** LOESS versus decadal period (red). Validation targets are 30-year average, 30-year average extended with linear trend and 20-year average.

3.2 Grand Ensemble Validation

Figure 4(a) shows the MPI-GE annual SAT range, LOESS fits and $GMST_F$ estimate while Figure 4(b) contains example LOESS and OLS fits to a single simulation. The continuity advantage of LOESS compared with OLS is obvious from this panel, as LOESS must only be calculated once for a given series. The forced, LOESS and OLS ΔGMST estimates through 2005 are shown for each start year from 1850—1980 in Figure 4(c).

The ΔGMST for LOESS and the forced series agree for all periods outside of those obviously affected with known volcanic eruptions. This suggests that the LOESS reliably estimates the forced change that is not associated with volcanoes, i.e. is close to the $\Delta GMST_{F,long}$ from Equation 1 that is primarily related to human-caused forcing changes. By contrast, the longer term OLS estimates are biased, with the true forced change commonly outside the 5—95 % range. More recently the OLS range is ~66 % larger, and while OLS better captures the volcano-driven excursion, this is not desirable for many assessments of long-term forced climate change. Furthermore, we argue that the LOESS fit reliably excludes the $GMST_{var}$ component: the correlation coefficient between the residuals from the LOESS fits and $GMST_F$ for all simulations is 0.88 when excluding values that fall within 3 years of a major volcanic eruption. This reinforces the findings of Takahashi et al (2019) who found that LOESS residuals and control simulation variability behaved similarly.

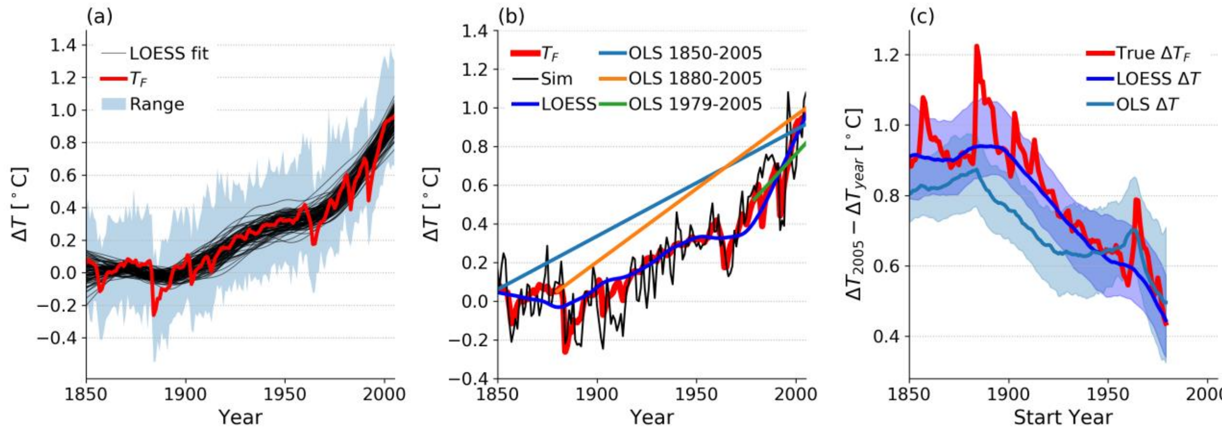


Figure 4. (a) MPI-GE SAT outputs, full ensemble range is shaded, each simulation's LOESS fit is in grey and the ensemble mean (our estimate of $GMST_F$) is in red. (b) example of fits applied to a single simulation (black) including LOESS (dark blue) and OLS over three different periods (straight lines) with $GMST_F$ in red. (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.

Table 3 contains estimates of ΔGMST from the Grand Ensemble. For differences between periods (e.g. 1850—1900 to 1996—2005), LOESS follows the hybrid method from Section 2.2.2 and OLS is fit between the middle of each period . This table reinforces the fact that OLS tends to underestimate the true forced warming since the late 19th century. Furthermore, LOESS is similar to the standard approach of differencing period means, with similar 5—95 % spread magnitude, albeit with all values shifted down by approximately 0.02 °C. This validates the LOESS calculation approach, and the latter columns show its advantage over period means since its calculation can be extended to the latest available year without greatly inflated uncertainty.

The 1880—2005 LOESS ensemble spread of 0.19 °C is in good agreement with the statistically derived 5—95 % uncertainties for observational datasets in Table 3. This provides support for our statistical estimates of uncertainty introduced in ΔGMST.

We propose that the absolute discrepancies between the LOESS and the $GMST_F$ period means may be largely explained by volcanism. Firstly, Figure 2 shows that LOESS fits are less sensitive to volcanic perturbations than $GMST_F$, and while the 1996—2005 is largely unaffected by volcanism, the ±20 year LOESS window captures some of the Pinatubo-induced cooling after 1991. By contrast, $\Delta GMST_F$ for 1880—2005 is 0.02 °C lower than that from LOESS. While the 1880 $GMST_F$ has no substantial volcanic cooling, the LOESS window now captures Krakatoa's large post-1883 cooling, which cools the 1880 LOESS estimate, and increases its 1880—2005 ΔGMST.

This Grand Ensemble analysis has:

(i) supported our LOESS-based statistical uncertainty estimates,

(ii) shown that LOESS has lower long-term bias and short-term uncertainty than OLS,

(iii) verified that LOESS reliably reproduces $\Delta GMST_F$, with bias magnitudes <0.05 °C depending on volcanism during the periods considered,

(iv) provided evidence that LOESS better estimates $\Delta GMST_{F,long}$, and is generally less sensitive to volcanism within a window.

Points (iii) and (iv) would not be possible with observational datasets since we cannot determine true $\Delta GMST_F$ so precisely. Point (iii) is strong evidence in support of LOESS, and point (iv) suggests that LOESS may better estimate the human-caused warming component for applications such as carbon budget calculation.

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.

ΔGMST Method	1850-1900 to 1996-2005 [°C]	1850-1900 to 2005 [°C]	1880 to 2005 [°C]
T_F	0.88 [0.87-0.89]	0.96 [0.94-0.98]	0.91 [0.88-0.94]
LOESS	0.86 [0.76-0.95]	0.93 [0.83-1.03]	0.93 [0.83-1.06]
OLS	0.75 [0.64-0.87]	0.78 [0.67-0.90]	0.86 [0.75-0.97]
Individual run means	0.88 [0.79-0.97]	0.88 [0.61-1.16]	0.90 [0.63-1.21]

3.3 Global SAT estimate and Remaining Carbon Budget

The percentage increase of the CMIP6 $LOESS_{md}$ SAT historical ensemble relative to the blended ensemble reaches 4.9% [3.2, 6.4] in 2014. The ratio stands at 6.4% in 1930, peaks at 8.6% in 1970 and descends thereafter. Our 4.9% is lower than Richardson et al (2018)'s 6.1%. However, Richardson et al. used CMIP5 with different periods, and when we select our likely ECS CMIP6 sub-ensemble with time periods matched to Richardson et al (2018) we find better agreement with a 5.7 % difference.

This ratio implies a central Global_3 ΔSAT estimate of 1.17°C [1.03 – 1.32] from 1850—1900 to 2018. 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.

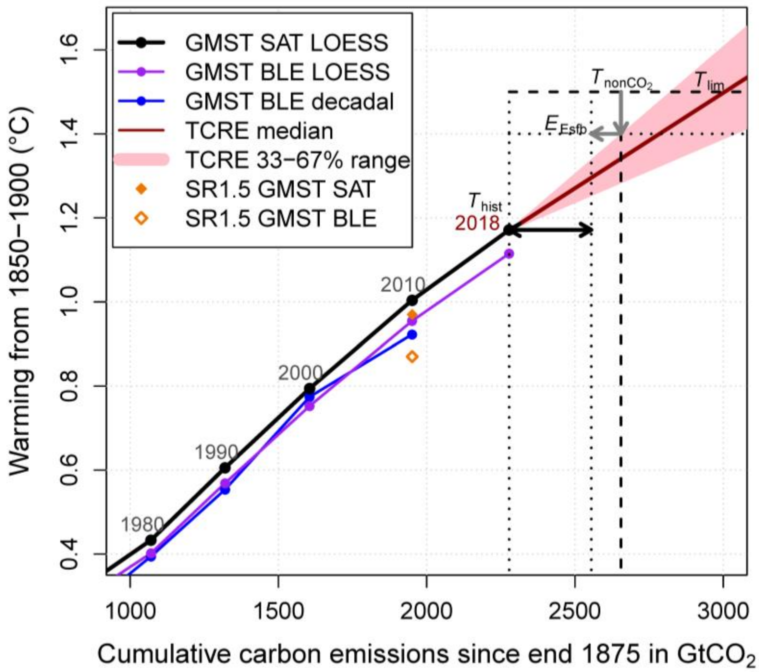


Figure 5: Global temperature change from 1850–1900 versus cumulative CO_2 emissions. The smoothed temperature response from the Global3 blended GMST group as decadal average (blue) and $LOESS_{md}$ trend (purple) are shown relative to cumulative CO_2 emissions from Friedlingsten et al (2019). The thick black line shows the Global3 GMST $LOESS_{md}$ trend, adjusted by the median difference between SAT 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.

The remaining carbon budgets from the start of 2019 for a 66% (50%) chance to stay below 1.5°C and 2.0°C are 275 (405) $GtCO_2$ and 935 (1225) $GtCO_2$ respectively (all numbers rounded to the nearest 5 $GtCO_2$). Given current annual emissions of just over 40 $GtCO_2$, the 66% 1.5°C remaining carbon budget is virtually identical to 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 ~30 $GtCO_2$ below those two studies. This follows from the slightly higher ΔT_{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 ΔT_{hist} reduces TCRE uncertainty, as there is less ΔT "to go".

SR1.5 also gave secondary carbon budgets for T_{hist} based on an unadjusted GMST four-series average over 2006-2015. We provide a corresponding budget based on unadjusted global GMST. Our 66% 1.5°C unadjusted GMST carbon budget is 360 $GtCO_2$ from 2019; the corresponding SR1.5 budget was 470 $GtCO_2$ from 2018. This large differential is to be expected as our unadjusted full global GMST estimate still accounts for coverage bias, whereas the SR1.5 group GMST does so only partially.

Following Rogelj et al (2019), all of the above estimates include E_{Esfb} approximate downward adjustment to account for Earth system feedbacks (release of CO_2 and CH_4 from warming wetland and permafrost thaw). Carbon budgets excluding this term would therefore be 100 $GtCO_2$ higher.

3.4 Recent trends

The recent trend evolution of Cowtan-Way can be seen in Figure 5a. The $LOESS_{pent}$ fluctuations around the smoothly rising $LOESS_{md}$ trend since ~1975 are characterized by surges and slowdowns. The first two brief slowdowns correspond to major volcanic eruptions, El Chichon in 1982 and Pinatubo in 1991. The early 2000s slowdown has since given away to a surge in GMST from 2012 to present. In our analysis, that surge has been slightly reduced (0.01°C) by our endpoint adjustment as described in Section 2.2.5. Figure 5b shows relatively good agreement between LOESS and OLS for overlapping 30- and 60-year trends, while illustrating the stark contrast between the more variable OLS and smoother LOESS over 15 years. The extreme OLS change from 1992-2006 (~0.3°C/decade) to 1998-2012 (~0.1°C/decade) is perhaps the clearest example of “broken” linear trends in the instrumental record.

The IPCC specifically pointed to a 1998-2012 trend between “a third to a half” of the 1951-2012 trend. However as seen in Figure 5c, only HadCRUT4 currently fulfils that criterion. The 15-year OLS trends of the updated versions of the other two AR5 series, NOAA GlobalTemp and NASA GISTEMP, now lie much closer to the 60-year trend, primarily due to an improved SST analysis. Meanwhile the two newer global series, Berkeley Earth and Cowtan-Way, have virtually identical 15-year and 60-year trends to 2012.

A different and clearer picture emerges with the continuous LOESS$_{pent}$ trends shown in Figure 5d. The 2012 trough in the 15-year LOESS$_{pent}$ trend of all five series is well above the 60-year LOESS$_{md}$ trend, and by 2018 most have returned to or above the 30-year trend.

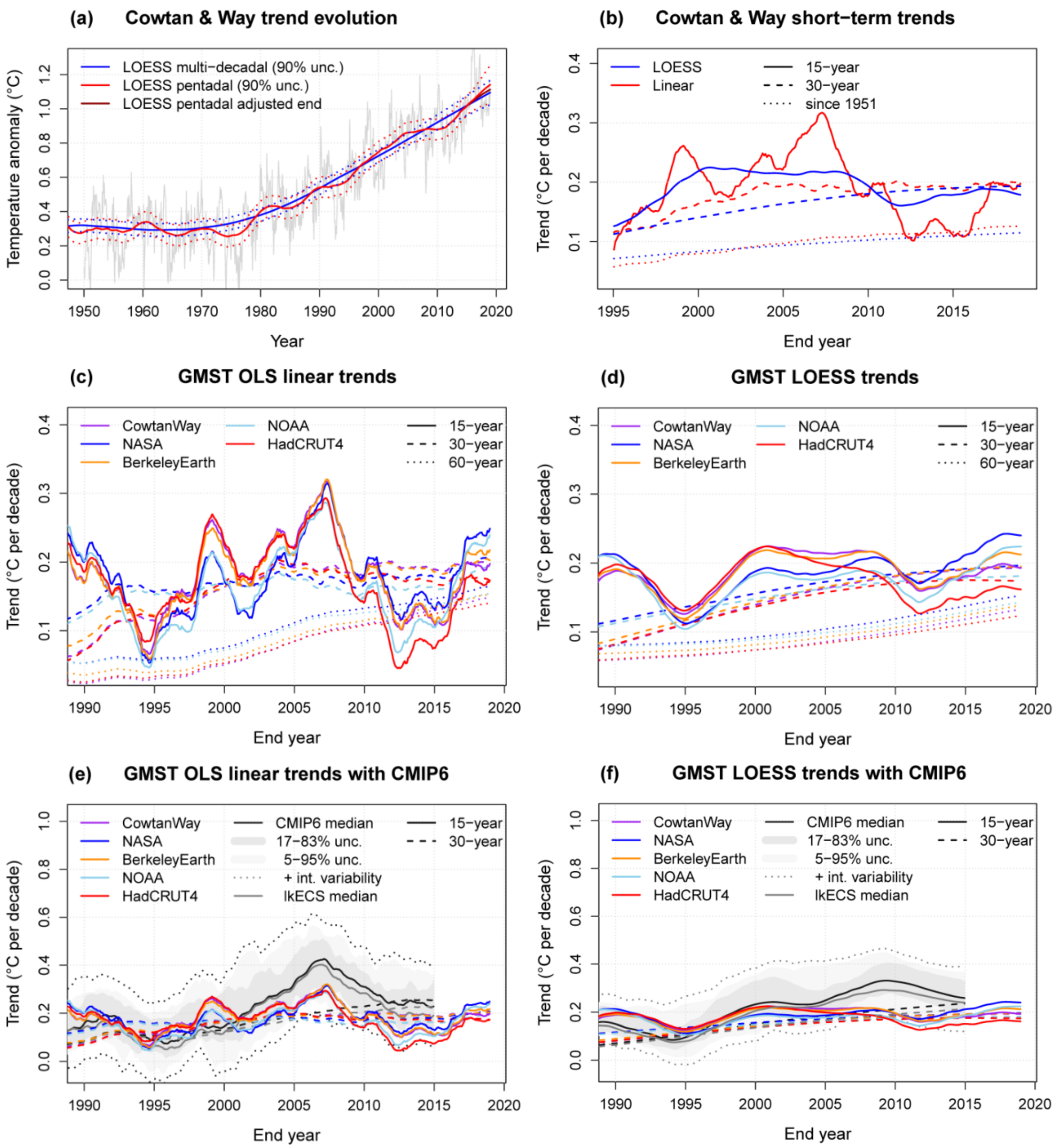


Figure 5: Short-term fluctuations versus medium-term trends. **(a)** Comparison of LOESS multi-decadal and pentadal trends over 1950-2018 with 5-95% uncertainty bands for Cowtan and Way monthly series. **(b)** Derived overlapping Cowtan and Way LOESS 15- and 30-year sub-trends compared to corresponding linear trends. **(c)** Overlapping OLS linear 15- and 30-year sub-trends derived from five operational GMST series. **(d)** Same as (c), except with overlapping LOESS 15- and 30-year sub-trends. **(e-f)** Same as (c-d), but with CMIP6 ensemble trends added (median with 5-95% spread).

This panel shows important differences between the datasets: of the Global_3 series, NASA GISTEMP's 15-year trends show the smallest slowdown and strongest recent rise, dipping from 0.21°C/decade in early 2008 to 0.17°C/decade and now back up to 0.24°C/decade, above the 30-year $LOESS_{md}$'s 0.19°C/decade. 15-year trends of Berkeley Earth and Cowtan-Way dip lower and have risen back to 0.21°C and 0.19°C per decade respectively. These differences can tentatively be attributed to an acknowledged cool bias in HadSST3 in recent years due to uncorrected changes in ship measurements, as confirmed by independent satellite and Argo float data (Karl et al., 2015; Hausfather et al., 2017). Meanwhile the recent HadCRUT4 trend evolution reflects both coverage bias (Cowtan and Way, 2013) and HadSST3 cool bias. The release of HadSST4 appears to have mitigated the latter issue and 15-year trends of the three HadSST-based series should move upward.

4 Discussion and Conclusions

Our analysis has explored the range of estimates of long-term GMST rise since the late 19th century in five observational series using two trend estimation methodologies. These estimates range from 0.94°C [0.80-1.06] (HadCRUT4 with OLS) up to 1.16°C [1.05 – 1.28] (Berkeley Earth with LOESS). Faced with a similar range of estimates, Vissar et al (2018) proposed that GMST be estimated as a grand average of all estimates. We recommend a very different approach.

Vissar et al argue that the spread due to trend method is minor compared to that of choice of GMST dataset. However, we show that LOESS is a reliable ΔGMST estimator and that the gap between OLS and LOESS estimates reaches 0.12°C for HadSST-based datasets with good spatial coverage such as Cowtan-Way over 1880—2018. This is comparable to the largest inter-dataset OLS difference of 0.13 °C between HadCRUT4 and Berkeley Earth. We have also demonstrated a clear lack of consistency of OLS in observational series, not only with LOESS, but also with intra-period estimates and regression-derived values of anthropogenic warming from Haustein et al (2017).

Furthermore, we validated LOESS against output from the MPI-GE, a large ensemble of a climate model. A large ensemble allows reliable determination of the forced component of ΔGMST, and this analysis demonstrated that LOESS consistently has very small bias as an estimator of forced ΔGMST relative to OLS, except for periods affected strongly by volcanic eruptions. For applications such as estimating anthropogenic warming or calculation of carbon budgets, this lower sensitivity to temporary volcanic perturbations in GMST is an advantage. The MPI-GE results also showed a similar spread of LOESS estimates to the statistical estimates obtained from the observation-based series, lending further support to LOESS. By contrast, OLS was consistently biased low when estimating forced ΔGMST relative to preindustrial, and had substantially larger uncertainty when calculating shorter term ΔGMST.

Whereas Visser et al. combined GMST series, we propose that only those that provide spatially complete GMST series should be used in the best estimates of ΔGMST. Firstly, global coverage is self-evidently more representative of global climate change, and secondly the past geographic extent of data coverage is arbitrary and may change as data rescue efforts digitize more historical data (Hawkins et al., 2019). While the infilled datasets will also change with the addition of this data, they should be less sensitive to such changes.

The selection of our Global_3 datasets leads to a substantial discrepancy of 0.12 °C relative to HadCRUT4 for ΔGMST from 1850-1900 to 2018. The differences since 1951 or 1979 are 0.10 °C and 0.07 °C respectively, i.e. they are smaller in absolute magnitude but larger in °C/decade. These divergences may grow, as the Global_3 $LOESS_{md}$ trend is now 0.03°C/decade higher than HadCRUT4 (0.20 versus 0.17 °C/decade). The recent divergence is likely attributable to bias coverage in HadCRUT4, as it implicitly assumes that areas without data coverage have the same mean temperature anomaly as areas with data coverage. By excluding fast-warming areas such as much of the Arctic, it under-reports recent warming according to independent satellite data and reanalyses (Dodd et al., 2015; Cowtan et al., 2018a; Susskind et al., 2019).

We also noted that recent SST dataset updates may be important for shorter term ΔGMST analyses, as the 1990s—2000s saw a transition from ship-based to buoy measurements of SST, and a change in the average properties of ships that recorded SST. The NOAA and NASA datasets use ERSST, whose version 4 was independently validated against satellite and Argo data, plus a buoy only dataset in Hausfather et al. (2017). As ERSST5 is similar to ERSST4 in recent decades, we judge that it reliably represents short-term SST changes. However, the Hausfather analysis discovered a slight cooling bias in HadSST3 that should be addressed in HadSST4 and thereby result in increased short-term ΔGMST estimates.

However, we cannot confirm whether SST updates or changes in data coverage during the pre-World War II (WWII) period will greatly affect the derived OLS-LOESS ΔGMST difference from the late 19th century or not. The substantial changes in ship-based measurements during WWII introduce numerous discontinuities that may result in errors in ΔGMST derived between pre- and post-WWII periods (Cowtan et al., 2018b). Similarly, it is not clear that pre-WWII sampling biases led to the same cooling bias that occurs in HadCRUT4 under recent warming. Future data updates may change the linearity across the full period and therefore the LOESS-OLS differences. Despite this, our proposed LOESS method is simple and transparent and may be quickly updated following any changes to the observation-based GMST series.

To summarize the ΔGMST analysis, we argue strongly in favor of using series that report near-global coverage and use the most up-to-date SST data available. This results in a current best estimate of warming from 1850—1900 to 2018 of 1.12 °C [1.00 – 1.25]. The fact that we can present an estimate of warming to 2018 with well-defined uncertainties is a substantial advantage over the IPCC's difference of period means approach: for example, the 1850—1900 to 2006—

2015 difference included in the SR15 was effectively years out of date, since it represented conditions in the middle of the 2006—2015 period.

As a final part of our ΔGMST analysis, we present an update to global near-surface air temperature, as opposed to the blended estimate of air and water temperatures that is provided by observational data. Our scaling was derived from CMIP6 climate models and found to be 4.9 [3.0—6.1] %, a much smaller correction than that required if we must also account for biases due to incomplete geographical coverage, as would be required if datasets like HadCRUT4 were included in the analysis. Despite our arguments against the use of HadCRUT4, its provision of an ensemble for the estimation of error introduced due to changing measurement technologies means that it still has a useful role in such analyses.

We find that our ΔGMST estimate implies a 2019-onwards carbon budget of 275 (405) GtCO for a 66% (50%) chance to stay below 1.5°C, implying less than a decade to exhaust the budget at the current rate of emissions.

Our proposed LOESS approach has shown promise in analysis of a large ensemble for disentangling long-term forced climate changes from internal variability, with caveats for the years following major volcanic eruptions. It provides a reliable estimate of forced ΔGMST and may be useful in future for the study of how long term trends and internal variability interact in observations and in upcoming analysis of the new CMIP6 model outputs. For example, our preliminary analysis shows that the high ECS CMIP6 models show non-ARMA(1,1) residual noise structures, in contrast to the observations. This may be a useful tool for investigating long-term modes of internal variability or changes in temperature driven by multi-decadal forcing variability, such as that due to anthropogenic aerosol.

Based on the evidence presented here we argue for the adoption of LOESS or a similarly flexible non-linear trend method such as smoothing spline as the primary trend estimation method for long-term GMST rise and trend evolution.

Acknowledgments and Data

The authors thank Andy Dessler for provision of MPI-GE series and Kevin Cowtan for suggestion of “first difference” enhancement for end-point adjustment. 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 data are available from

https://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html . HadSST4 data are available from https://www.metoffice.gov.uk/hadobs/hadsst4/data/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

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

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

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

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., D. Jacob, M. Taylor, M. Bindi, S. Brown, 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

IPCC, 2013: 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.

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

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

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

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.)].

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

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

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

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