

## Reply to “Comment on ‘The Impact of Recent Forcing and Ocean Heat Uptake Data on Estimates of Climate Sensitivity’”<sup>1</sup>

NICHOLAS LEWIS

*Bath, United Kingdom*

JUDITH CURRY

*Climate Forecast Applications Network, Reno, Nevada*

(Manuscript received 11 October 2018, in final form 4 November 2019)

### ABSTRACT

Cowtan and Jacobs assert that the method used by Lewis and Curry in 2018 (LC18) to estimate the climate system’s transient climate response (TCR) from changes between two time windows is less robust—in particular against sea surface temperature bias correction uncertainty—than a method that uses the entire historical record. We demonstrate that TCR estimated using all data from the temperature record is closely in line with that estimated using the LC18 windows, as is the median TCR estimate using all pairs of individual years. We also show that the median TCR estimate from all pairs of decade-plus-length windows is closely in line with that estimated using the LC18 windows and that incorporating window selection uncertainty would make little difference to total uncertainty in TCR estimation. We find that, when differences in the evolution of forcing are accounted for, the relationship over time between warming in CMIP5 models and observations is consistent with the relationship between CMIP5 TCR and LC18’s TCR estimate but fluctuates as a result of multidecadal internal variability and volcanism. We also show that various other matters raised by Cowtan and Jacobs have negligible implications for TCR estimation in LC18.

### 1. Introduction

Cowtan and Jacobs (2020, hereinafter CJ20) argue that transient climate response (TCR) estimation using relatively short time windows, as in Lewis and Curry (2018, hereinafter LC18), can be affected by uncertainty in bias corrections to sea surface temperature data. They argue that use of the whole historical record can mitigate the impacts of short time windows on estimation of TCR, particularly with respect to the early part of the record.

Here we investigate the effects of window selection and find that including uncertainty arising from it would at most slightly increase the total uncertainty in LC18’s TCR estimate. Although the LC18 TCR estimate is

based on selected, relatively short, time windows, we find no evidence that it is biased relative to estimates using information from the whole historical record.

Moreover, two fundamental issues confound CJ20’s analysis of the comparative evolution of warming in observations and climate models. First, their claims are based on comparing temperature changes in historical simulations by CMIP5 climate models and observations and not on comparing the ratio of temperature and forcing changes (on which ratio LC18’s TCR estimation is based) in CMIP5 models with that in observations. The two types of comparisons are equivalent only if forcing in CMIP5 models on average evolves identically to its estimated actual evolution. This is not the case, and therefore these two approaches are not equivalent. Moreover, even ignoring forcing evolution differences, observed temperature would evolve differently from that in CMIP5 models unless they accurately simulate the response of the real climate system to forcing—an assumption that is contrary to LC18’s results. Second, even in the absence of

---

<sup>1</sup> Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-18-0669.s1>.

---

Corresponding author: Nicholas Lewis, [nhlewis@btinternet.com](mailto:nhlewis@btinternet.com)

time-varying biases in temperature and forcing estimates, it is expected that different window choices will lead to somewhat different estimates of TCR, because of differences in the influences of multidecadal internal climate variability and episodic volcanic forcing. CJ20's comparison of observed and CMIP5-simulated warming analysis does not account for the effect of multidecadal internal variability, in particular that due to the Atlantic multidecadal oscillation (AMO), which noticeably affects the observed global temperature record but not the CMIP5 mean. Booth et al. (2012) concluded that aerosols are a prime driver of twentieth-century North Atlantic Ocean climate variability, but Zhang et al. (2013) found major discrepancies between Booth et al.'s simulations and observations, casting considerable doubt on their claim. Although the debate on internal variability versus external forcing continues, a recent comprehensive review (Zhang et al. 2019) found strong observational and modeling evidence that a crucial driver of the observed Atlantic multidecadal variability is multidecadal Atlantic meridional overturning circulation internal variability, rather than external forcing. Further, Lin et al. (2019) and Yan et al. (2019) showed that coupled models did not reproduce observed Atlantic multidecadal variability.

As discussed in LC18, some sensitivity of TCR estimates to choice of window is inevitable: the window method will not give unbiased estimates when the early window (base period) and late window (final period) are affected very differently by multidecadal internal variability. What CJ20 regard as lack of robustness against choice of window period is in fact a key advantage of the window method: selection of the base and final periods enables minimization of the influence of internal variability, as well as uncertainty in volcanic forcing and its effects, while simultaneously obtaining the large change in total forcing needed for well-constrained TCR estimation. LC18's preferred 1869–82 early window and 2007–16 late window were selected with regard to these factors. LC18 found low sensitivity to alternative choices of both early and late windows, including windows several decades long, that were consistent with the matching criteria.

The important question is not whether window selection affects TCR estimation but whether the chosen windows provide unbiased estimation of TCR in context of the full information available from the historical period, and whether adequate allowance is made for temperature-related uncertainty. We first address these questions and then examine the reasons for the evolution of historical period warming differing between observations and CMIP5 models.

## 2. Are TCR estimates that are based on the LC18 window periods representative of the historical period?

We investigate, using the infilled globally complete “Had4\_krig\_v2” temperature record (Cowtan and Way 2014a,b,c,d), how TCR estimation using the LC18 data and method is affected when employing approaches that do not involve window selection. We first estimate TCR from changes between pairs of years, initially using all pairs with broadly comparable influence from multidecadal internal variability. We accordingly select all year-pairs during 1850–2016 that are separated by either 55–75 or 120–140 years, periods that are bands around an integer multiple of the approximately 65-yr AMO cycle length during the historical period (see section 4 in LC18). We ameliorate the effects of mismatched volcanic forcing by scaling AR5-based volcanic forcing by 0.55 to account for its low efficacy (Lewis and Curry 2015, 2018; Gregory et al. 2016). We compute total forcing by taking 500 000 samples from the LC18 2011 uncertainty distributions for each forcing component (efficacy-adjusted where relevant) and using them to scale the LC18 best-estimate forcing time series, the uncertainty distributions and best estimates being based on data from the IPCC Fifth Assessment Report (Myhre et al. 2013; Prather et al. 2013). For each sample set, we sum the scaled forcing component time series and divide them by the corresponding sampled forcing from a doubling of pre-industrial CO<sub>2</sub> concentration ( $F_{2\times\text{CO}_2}$ ), thus deriving 500 000 annual time series of total forcing relative to  $F_{2\times\text{CO}_2}$  ( $F_{\text{rel.55vol}}^{\text{LC18/AR5}}$ ). We use annual values of the ensemble of Had4\_krig\_v2 realizations (which sample systematic bias and parameter uncertainty) to measure temperature ( $T^{\text{obs}}$ ), repeating the 100 samples 5000 times. We add to each of the resulting 500 000 temperature time series a 167-yr-long sequence of random draws from a set of 167 normal distributions having standard deviations  $\sigma$  equal to the sums (adding in quadrature) of 1- $\sigma$  sampling and measurement uncertainties (Morice et al. 2012a,b) and coverage uncertainty (Cowtan and Way 2014b,c,d) for each year over 1850–2016. We compute the changes  $\Delta T^{\text{obs}}$  in sampled temperature, and  $\Delta F_{\text{rel.55vol}}^{\text{LC18/AR5}}$  in sampled total forcing relative to  $F_{2\times\text{CO}_2}$ , for every pair of years.

TCR estimates are computed using two slightly different methods. In the first (“aggregating”) method we compute TCR as the median over all samples of  $\sum \Delta T^{\text{obs}} / \sum \Delta F_{\text{rel.55vol}}^{\text{LC18/AR5}}$ , the sums being over all year-pairs. Doing so weights the influence of each year-pair according to the magnitude of its  $\Delta T^{\text{obs}}$  and  $\Delta F_{\text{rel.55vol}}^{\text{LC18/AR5}}$ , preventing year-pairs involving very small  $\Delta F_{\text{rel.55vol}}^{\text{LC18/AR5}}$  from unduly influencing the TCR estimation. In the second (“nonaggregating”) method we compute TCR by

TABLE 1. Quantiles for TCR estimates from all pairs of equal-length early and late windows at least a decade long, at varying minimum levels of forcing increase. Each TCR estimate is the quotient of the differences in window period means for the median  $T$  (Had4\_krig\_v2) and  $F_{\text{rel}}^{\text{LC18/AR5}}$  LC18 AR5-based time series. Figures in italics are based on  $F_{\text{rel},55\text{vol}}^{\text{LC18/AR5}}$ , calculated with volcanic forcing scaled by 0.55. To apply the minimum interwindow forcing change requirement, the  $F_{\text{rel}}^{\text{LC18/AR5}}$  change in ERF relative to that for a doubling of  $\text{CO}_2$  concentration is converted into a forcing change by multiplying by  $F_{2\times\text{CO}_2}$ .

Min required period-avg interwindow forcing increase ( $\text{W m}^{-2}$ )	No. of window- pair samples	TCR 5th percentile (K)	TCR 17th percentile (K)	TCR 50th percentile (K)	TCR 83rd percentile (K)	TCR 95th percentile (K)
1.0	98 338	0.97	1.16	1.34	1.54	1.71
		<i>1.10</i>	<i>1.21</i>	<i>1.34</i>	<i>1.50</i>	<i>1.61</i>
1.5	34 833	1.06	1.19	1.30	1.48	1.62
		<i>1.16</i>	<i>1.23</i>	<i>1.32</i>	<i>1.48</i>	<i>1.60</i>
2.0	11 242	1.08	1.17	1.29	1.42	1.54
		<i>1.20</i>	<i>1.26</i>	<i>1.33</i>	<i>1.44</i>	<i>1.59</i>

calculating the median TCR over all samples for each year pair (setting TCR to  $+\infty$  where  $\Delta F_{\text{rel},55\text{vol}}^{\text{LC18/AR5}} < 0$  but  $\Delta T^{\text{obs}} > 0$ ) and then taking their median over all year-pairs. The median TCR estimates using the two methods are respectively 1.35 and 1.34 K, both of which are almost identical to the LC18-preferred Had4\_krig\_v2-based estimate of 1.33 K.

We likewise compute a TCR estimate from all pairs of years in 1850–2016 without adjusting volcanic forcing (producing  $\Delta F_{\text{rel}}^{\text{LC18/AR5}}$  and not  $\Delta F_{\text{rel},55\text{vol}}^{\text{LC18/AR5}}$ ). Since changes in volcanic forcing and in the LC18 AMO index then both average to almost zero, doing so largely sidesteps the influence of volcanic forcing and of multidecadal internal variability. The resulting median TCR estimates using the aforementioned two methods are 1.36 and 1.33 K.

For more direct comparison with LC18, we also compute TCR estimates using all pairs of equal-length<sup>1</sup> windows a decade or more long during the period 1850–2016. Doing so samples uncertainty realizations arising from time-varying errors in SST and land temperature measurements and from their combination into median global temperature estimates, and from misestimation of the time profile of evolving forcing, as well as from internal variability and from the influence of episodic volcanism, but does not sample uncertainty in present-day forcing. Table 1 shows quantiles for the resulting estimates<sup>2</sup> with differing minimum required levels of

interwindow median forcing increase. Window combinations for which the median interwindow forcing increase is small contain little relevant information and cannot provide meaningful TCR estimates; for the preferred LC18 estimate the increase was  $2.52 \text{ W m}^{-2}$ . The median TCR estimates are insensitive to the minimum required forcing increase and are all very close to the LC18 preferred estimate. For estimates with the highest ( $2.0 \text{ W m}^{-2}$ ) minimum forcing increase, which are most relevant to LC18's TCR estimate, the 5%–95% TCR uncertainty range arising from random window selection is 1.08–1.54 K, or 1.20–1.59 K using 0.55-scaled volcanic forcing. The width of these ranges—0.103 and 0.073, respectively, in fractional standard deviation terms<sup>3</sup>—reflects the fact that many of the window combinations involve mismatched influences from internal variability and/or volcanism. These window selection uncertainty ranges do not imply that LC18 underestimated uncertainty in global temperature change: the 1- $\sigma$  fractional uncertainty in LC18's preferred TCR estimate attributable to temperature change uncertainty (including that from internal variability) alone was 0.103.<sup>4</sup> Moreover, even if no allowance is made for double counting of temperature change uncertainty, estimated overall TCR uncertainty would increase little if window selection uncertainty were added. Adding (in quadrature) the 0.103 or 0.073 1- $\sigma$  fractional uncertainty in TCR from window selection to the 1- $\sigma$  fractional uncertainty of the preferred LC18 TCR estimate would only increase it to 1.13 times its original

<sup>1</sup> Because of computational limitations.

<sup>2</sup> When computing TCR estimates using the windows method, we use median  $F_{\text{rel}}^{\text{LC18/AR5}}$  (or  $F_{\text{rel},55\text{vol}}^{\text{LC18/AR5}}$ ) and  $T^{\text{obs}}$  time series to derive the TCR estimate rather than taking the median of the sample-derived TCR estimates. We found that this more computationally tractable approach produced windows-based TCR best estimates that are essentially identical to those computed from sampled time series. We also employ this approach when estimating TCR by regression.

<sup>3</sup> So as to be able readily to combine uncertainties, we work with 1-standard deviation fractional uncertainties, here derived by scaling from 17%–83% ranges and medians in Table 1.

<sup>4</sup> Scaling from the 5%–95% range and median for Had4\_krig\_v2  $\Delta T$  in Table 2 of LC18. If temperature uncertainty alone is incorporated, the fractional uncertainty in TCR is equal to that in  $\Delta T$ .

level, or to 1.07 times that level if using 0.55-scaled volcanic forcing.<sup>5</sup>

CJ20 state that a more robust approach than selecting particular windows would be to use the entire temperature record. LC18 did so as part of its sensitivity testing. When AR5 volcanic forcing is scaled by 0.55, regression of median annual-mean temperature on forcing over 1850–2016 gives a Had4\_krig\_v2-based TCR estimate of 1.27 K, which is marginally lower than LC18's two-window-based preferred estimate of 1.33 K. Regressing pentadal means (over 1852–2016) significantly improves the fit (to an  $R^2$  of 0.92, where  $R$  is correlation coefficient) and gives a TCR estimate of 1.33 K. Using such pentadal-mean regression on each of the 500 000 pairs of samples of temperature and forcing time series gives a 5%–95% TCR range of 0.91–1.84 K, marginally lower and narrower than the LC18 preferred estimate range. Regression using unscaled volcanic forcing produces substantially lower median TCR estimates.

Regression over the full historical period makes the most complete use of the available information. However, sensitivity to the treatment of volcanic forcing means that it is more difficult to be confident that volcanic forcing is not biasing TCR estimation when using regression, even of pentadal mean data, than when using the windows method and matching mean volcanic forcing. Barnes and Barnes (2015) found that if the windows method were adopted, then it was generally best to use windows at the start and end of the record each of approximately one-third of its length. That points to using an 1850–1904 early window and a 1962–2016 late window. Fortuitously, these have well-matched mean volcanic forcing. The TCR estimate using those windows is 1.32 K.

CJ20 also raise issues regarding temperature measurement. They state that coverage of the “water hemisphere” was almost nonexistent in the 1860s. However, the 1869–82 primary early window used in LC18 avoids the 1860s (except for 1869, when global coverage was highest). Moreover, during 1869–82 observational coverage, although limited, was slightly higher in the (land sparse) Southern Hemisphere than the Northern Hemisphere. CJ20 additionally say that nineteenth-century temperatures are dependent on large “bucket corrections” to sea surface temperature (SST) observations, but these were relatively small during 1850–82 (Folland and

Parker 1995; Kent et al. 2017). Indeed, CJ20 suggest that the change from wooden buckets to poorly insulated canvas buckets requiring a large bias correction occurred primarily during 1890–1910.

CJ20 question the 1930–50 early window period used for one LC18 TCR estimate because it spans World War II and is the subject of sizeable discrepancies between SST products. However, those discrepancies only became sizeable in 1941. Restricting the 1930–50 base period to 1930–40 would barely change that particular LC18 Had4\_krig\_v2-based TCR estimate.

CJ20 claim that a residual (negative) bias in recent SST observations affects windows starting after 2005. CJ20 accordingly base their analysis on LC18's 1995–2016, rather than its 2007–16, late window. However, LC18's Had4\_krig\_v2-based TCR estimate using the 1995–2016 window is 0.01 K lower than when using 2007–16.

Significant uncertainties in SST data certainly exist, with data coverage and quality limitations in the nineteenth century of particular relevance for LC18. Total global temperature uncertainty was quite large during 1869–82, with coverage uncertainty being the largest component. However, fractional uncertainty in forcing change dominates the uncertainty in temperature change (see LC18's Table 2) when estimating TCR. LC18's temperature uncertainty ranges incorporated the dataset providers' uncertainty estimates for the University of East Anglia Climatic Research Unit–Hadley Centre global land-plus-ocean temperature dataset, version 4 (HadCRUT4), (Morice et al. 2012a,b) and Had4\_krig\_v2 global mean temperature products, which allowed for coverage, bias and parameter, and measurement and intragrid cell sampling uncertainties. CJ20 cite no evidence indicating that the dataset providers' uncertainty estimates were inadequate during LC18's 1869–82 early window. Further, there is a close match, particularly over 1850–1940, between SST evolution in the global “HadOST” product (Haustein et al. 2019) and in scaled land and ocean coastal observations. Since global temperature evolved very similarly in HadOST and HadCRUT4, this close match bolsters confidence that there was no major bias in the 1869–82 temperature data used in LC18.

CJ20 claim that comparison of modeled and observed temperatures for late windows starting after 2005 is affected by overestimation of forcings in models. Since LC18 did not make any comparisons of modeled and observed temperatures over the historical period, the only issue of relevance to LC18 is whether it misestimated recent forcing. None of the three supporting studies that CJ20 cite indicates that LC18 misestimated recent forcing. Tatebe et al. (2019) do not directly discuss the recent evolution of forcings. Volodin and Gritsun (2018) suggest that the slower 2000–14 warming in the INM-CM5 model

<sup>5</sup> Scaling from the 17%–83% range in Table 3 of LC18, giving a fractional standard deviation of 0.193 for the preferred LC18 TCR estimate. Uncertainties are taken to be normally distributed and independent for the purposes of deriving their standard deviations and combining them. Adding in quadrature a fractional standard deviation of 0.103 or 0.073 to the original level of 0.193 respectively increases it to 0.219 or 0.207.

than in INM-CM4 is primarily due to (downward) revisions between CMIP5 and CMIP6 in post-2000 solar irradiance estimates. However, the solar forcing changes used in LC18 are closely in line with those in CMIP6. Huber and Knutti (2014) likewise point to CMIP5 twenty-first-century solar forcing changes being misestimated, and also to post-2000 stratospheric aerosol (volcanic) forcing being incorrect, in the representative concentration pathway (RCP) used for CMIP5 model projections. This point is likewise inapplicable to LC18, which used the same updated stratospheric aerosol optical dataset as used by Huber and Knutti (2014). Moreover, the more comprehensive Outten et al. (2015) study found, in a CMIP5 model, that since the mid-2000s underestimation of changes in other forcing agents more than counteracted overestimation of changes in solar and volcanic forcing. None of these studies addressed bias in CMIP5 model forcing that already existed by their start dates, of 1980 or later.

CJ20 claim that previous studies have identified differences in inferred forcings and in the temperature impact of historical versus transient forcing changes as potential explanatory factors for recent observational energy-budget TCR estimates being lower than average climate model TCR values. None of the three supporting studies that they cite supports either contention. Storelvmo et al. (2016) is an observation-based TCR study that ignores all forcings other than CO<sub>2</sub> and surface downwelling solar radiation (DSRS). Moreover, it uses changes in DSRS as a proxy for aerosol forcing changes, despite correlation between DSRS and global sulfur dioxide emissions being insignificant over their analysis period. Armour (2017) did not address observational TCR estimation, and moreover considered the temperature impact of historical forcing evolution to be very similar to that of transient ramp CO<sub>2</sub> forcing. Richardson et al. (2016) addresses comparing temperature in observations and models.

### 3. Differences between observed and CMIP5 model-simulated historical warming

We compare  $T^{\text{obs}}$  warming from Had4\_krig\_v2, which uses SST over the open ocean, with the standard global surface air temperature (“tas”) measure of warming in models. LC18 (their section 7e) concluded from observational and reanalysis evidence that in the real climate system, tas warmed at most a few percent more than a blend of tas and “tos” (model top ocean layer temperature), a substantially smaller difference than that claimed by CJ20. Indeed, the 1979-onward ERA-Interim reanalysis globally complete surface air temperature record, adjusted for inhomogeneities in

their SST source (Simmons et al. 2017), shows slightly lower warming over 1979–2016 than does Had4\_krig\_v2. Moreover, CJ20’s claim that LC18 “argue that this field [tos] is not the top layer of the bulk ocean surface temperature” is incorrect. Rather, LC18 argued that the tas/tos warming difference reflects the model-simulated warming difference between tas and ocean skin temperature, which will warm differently from SST.

We compare observed and model-simulation warming as follows. We form a 25-member ensemble comprising all CMIP5 models in LC18’s Table S2 except CESM1-CAM5.1-FV2, CNRM-CM5.2, FGOALS-s2, GISS-E2-H-p3, GISS-E2-R-p3, and MPI-ESM-P (the expansions of models can be found at <https://www.ametsoc.org/PubsAcronymList>). The excluded models either do not have all of the required simulation data available or are nonstandard physics variants.<sup>6</sup> We create anomalies from model simulation data by subtracting matching sections of linear fits over their preindustrial control simulations. We form model-ensemble-mean global surface temperature (tas)  $T^j$  and top-of-atmosphere radiative imbalance  $N^j$  time series ( $j = 1, 2, \dots, 25$ ) from merged 1861–2005 historical and 2006–16 RCP4.5 simulation data and then average  $T^j$  to give a CMIP5-mean time series,  $T^{\text{CMIP5}}$ . We also divide each model’s  $T^j$  by its estimated TCR (derived from averaging over years 60–80 of its “1pctCO2” simulation tas anomalies) to give  $\Delta T^j_{\text{rel}}$ , being simulated historical warming relative to TCR, and average these to give the CMIP5-mean time series,  $T^{\text{CMIP5}}_{\text{rel}}$ .

We compute for each model an estimated 1861–2016 time series of historical effective radiative forcing (ERF) relative to ERF for a doubling of preindustrial CO<sub>2</sub> concentration ( $F_{2\times\text{CO}_2}$ ), as  $F^j_{\text{rel}} = (\lambda^j_{1-50}\Delta T^j + \Delta N^j)/F^j_{2\times\text{CO}_2}$ , with  $\lambda^j_{1-50}$  and  $F^j_{2\times\text{CO}_2}$  being estimated from the model’s abrupt4xCO2 simulation data, and derive their mean,  $F^{\text{CMIP5}}_{\text{rel}}$ . This method provides satisfactory estimates (see the online supplemental material; Forster et al. 2013).

A TCR-relevant comparison between observed and model-simulated warming requires the removal of the response to volcanic forcing, the efficacy of which likely differs between the real climate system and model simulations. We take the volcanic forcing component of the LC18 AR5-based ERF time series,  $F^{\text{LC18/AR5}}_{\text{Vol}}$ , and compute a 15-yr running mean of  $F^{\text{CMIP5}}_{\text{rel}}$  with data for all years in which  $F^{\text{LC18/AR5}}_{\text{Vol}} < -0.5 \text{ W m}^{-2}$  being ignored,

<sup>6</sup>We exclude the GISS-E2-H-p3 and GISS-E2-R-p3 nonstandard physics variants because otherwise four GISS-E2 model variants would be included, composing 15% of the ensemble (as reduced by excluding the models with insufficient data), which is considered to be excessive.

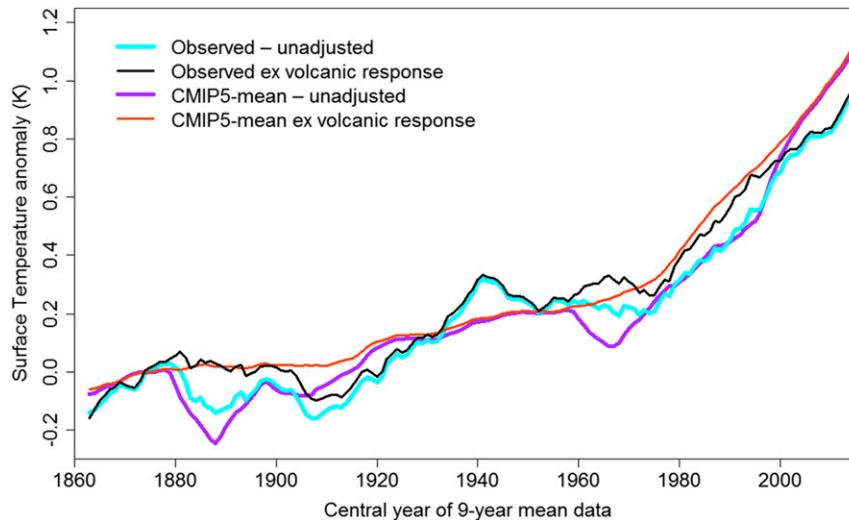


FIG. 1. CMIP5-mean and observed global mean warming before and after removing the response to volcanism. CMIP5 historical simulation data have been extended using RCP4.5 simulation data. The values plotted are centered 9-yr running means of  $T^{\text{obs}}$ ,  $T_{\text{exVol}}^{\text{obs}}$ ,  $T^{\text{CMIP5}}$ , and  $T_{\text{exVol}}^{\text{CMIP5}}$  anomalies relative to the 1869–82 mean. Values for years for which the running mean is formed from fewer than 5 years are not plotted.

reducing the averaging period near the beginning and end. We convolve both the resulting ex-volcanic forcing 15-yr running mean and  $F_{\text{Vol}}^{\text{LC18/AR5}}$  with an exponential response function, and multiply regress  $T^{\text{CMIP5}}$  and  $T_{\text{rel}}^{\text{CMIP5}}$  in turn on the two resulting time series. We use a fit-determined 2.5-yr  $e$ -folding time. The regressor time series derived from  $F_{\text{Vol}}^{\text{LC18/AR5}}$  is scaled by its coefficient in the first regression and subtracted from  $T^{\text{CMIP5}}$  to give  $T_{\text{exVol}}^{\text{CMIP5}}$ . The regressor time series derived from  $F_{\text{Vol}}^{\text{LC18/AR5}}$  is also scaled by its coefficient in the second regression and subtracted from  $T_{\text{rel}}^{\text{CMIP5}}$  to give  $T_{\text{rel,exVol}}^{\text{CMIP5}}$ . The volcanic signal is removed from  $T^{\text{obs}}$  using the same approach and time constant, but  $F_{\text{Vol}}^{\text{LC18/AR5}}$  and not  $F_{\text{rel}}^{\text{CMIP5}}$ , to form  $T_{\text{exVol}}^{\text{obs}}$ . We remove the volcanic signal from  $F_{\text{rel}}^{\text{CMIP5}}$  similarly, without using an exponential response function, to form  $F_{\text{rel,exVol}}^{\text{CMIP5}}$ . Note that  $F_{\text{rel,exVol}}^{\text{LC18/AR5}}$  is a weighted combination of  $F_{\text{rel}}^{\text{LC18/AR5}}$  and  $F_{\text{rel,55vol}}^{\text{LC18/AR5}}$  that eliminates volcanic forcing.

On decadal time scales, the mean evolution of warming of CMIP5 models over the historical period broadly matches that of observed warming until 2000, with some fluctuation (Fig. 1, thick purple and cyan lines). When the fitted response to volcanic forcing is removed (Fig. 1, black and orange-red lines), CMIP5-mean historical/RCP4.5 warming exceeds observed warming by the mid-1980s, with the gap widening from the mid-1990s.

The varying relationship between CMIP5-mean and observed warming, minus the volcanic response, reflects three factors:  $F_{\text{rel,exVol}}$  evolving differently in CMIP5 models from its estimated evolution in the real climate system, the ratio of CMIP5-mean TCR to TCR in the real

climate system, and internal variability affecting observed warming.

Figure 2 compares the evolution of  $F_{\text{rel}}^{\text{CMIP5}}$  with that of  $F_{\text{rel}}^{\text{LC18/AR5}}$  (red and black lines).<sup>7</sup> Figure 2 also shows that  $F_{\text{rel,exVol}}^{\text{CMIP5}}/F_{\text{rel,exVol}}^{\text{LC18/AR5}}$ , the ratio of CMIP5-mean to AR5-based nonvolcanic ERF/ $F_{2\times\text{CO}_2}$ , (the “relative forcing ratio”; blue line) is (using smoothed data) approximately 0.6 from 1925 to 1940, declines from then until circa 1960, and thereafter climbs to reach a plateau circa 1990. Forcing prior to 1925 is small. The changes in the relative forcing ratio are due principally to CMIP5-mean aerosol forcing being substantially stronger than the aerosol forcing estimates used in LC18 and, as a fraction of total anthropogenic forcing, rising from 1940, peaking in the early 1960s and thereafter declining. Comparison of the RCP4.5 dataset (Meinshausen et al. 2011) and LC18 anthropogenic forcings suggests that differing post-1990 trends in tropospheric ozone, greenhouse gas, and aerosol forcing (all of which were revised in LC18 from AR5 best estimates to reflect more recent evidence) account for  $F_{\text{rel,exVol}}^{\text{CMIP5}}/F_{\text{rel,exVol}}^{\text{LC18/AR5}}$  remaining stable thereafter at 0.84–0.86. This value is close to the 0.86 ratio in Otto et al. (2013) of estimated CMIP5-mean ERF in 2010 before and after adjusting for the models’ stronger than observationally estimated aerosol forcing.

<sup>7</sup> The volcanic components of the two relative ERF time series are not equivalent since the AR5-based volcanic forcing is not efficacy-adjusted, whereas by construction the CMIP5 ERF time series is efficacy-adjusted.

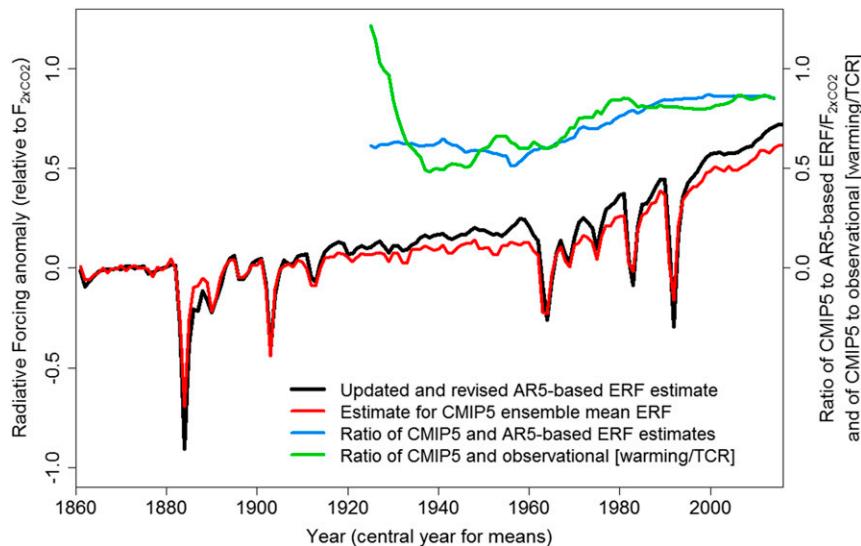


FIG. 2. CMIP5-mean and AR5-based/LC18 ERF relative to  $F_{2\times\text{CO}_2}$  over 1861–2016, the ratio of their ex-volcanic versions (the ratio of  $F_{\text{rel\_exVol}}^{\text{CMIP5}}$  to  $F_{\text{rel\_exVol}}^{\text{LC18/AR5}}$ ), and the corresponding ratio of CMIP5-mean and observational warming relative to respectively model and observational TCR estimates [the ratio of  $T_{\text{rel\_exVol}}^{\text{CMIP5}}$  to  $(T_{\text{exVol}}^{\text{obs}}/1.33)$ ]; 1.33 K is the LC18 preferred median TCR estimate when using Had4\_krig\_v2. The CMIP5-ensemble mean TCR is 1.82 K. Data are anomalies from the 1869–82 mean. Relative ERF and relative warming ratios are calculated model by model before computing CMIP5 means. The ratios are of centered 15-yr running means (shortened to 5 yr by the final year plotted, 2014).

Figure 2 also shows the ratio of smoothed  $T_{\text{rel\_exVol}}^{\text{CMIP5}}$  to smoothed  $T_{\text{rel\_exVol}}^{\text{obs}}/\text{TCR}^{\text{obs}} - \text{TCR}^{\text{obs}}$  being LC18's TCR estimate—since 1925 (the “relative warming ratio”; green line). When the green line is above the blue line, CMIP5-mean warming relative to that observed is greater than predicted by their respective TCR and  $F_{\text{rel\_exVol}}$  estimates, and vice versa. The relative warming ratio starts off much higher than the relative forcing ratio, reflecting the unusually cold first quarter of the twentieth century, before falling below the relative forcing ratio during the warm period centered around 1940, when the AMO was positive. From the late 1950s until circa 1990, the relative warming ratio largely tracks the rising relative forcing ratio, but generally exceeds it as the negative phase of the AMO, which reached its nadir in the 1970s, was associated with cooler global temperature. After 1990 the relative warming ratio remains close to the relative forcing ratio, as is to be expected if the LC18 TCR estimate is accurate. Incomplete removal of the volcanic signal might also contribute to the fluctuations in the two ratios between the mid-1950s and late 1990s.

#### 4. Conclusions

Our analyses show that the windows used in LC18 gave TCR estimates in line with those using information from all historical period data, including from all window

combinations, and that window selection contributes little to total uncertainty in TCR estimation. The differing evolution of temperature in observations versus models is consistent with the substantially different observationally based and CMIP5-mean TCR estimates once differences in the evolution of estimated forcing and in the effects of volcanism and multidecadal internal variability are accounted for.

Computer code used in this paper is included in the online supplemental material. It obtains data from publicly accessible datasets.

#### REFERENCES

- Armour, K. C., 2017: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nat. Climate Change*, **7**, 331–335, <https://doi.org/10.1038/nclimate3278>.
- Barnes, E. A., and R. J. Barnes, 2015: Estimating linear trends: Simple linear regression versus epoch differences. *J. Climate*, **28**, 9969–9976, <https://doi.org/10.1175/JCLI-D-15-0032.1>.
- Booth, B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228–232, <https://doi.org/10.1038/nature10946>.
- Cowan, K., and R. G. Way, 2014a: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, <https://doi.org/10.1002/qj.2297>.
- , and —, 2014b: Coverage bias in the HadCrut4 temperature record and its impact on recent temperature trends. Update—Temperature reconstruction by domain: Version 2.0 temperature

- series. Update Rep., 9 pp., <http://www-users.york.ac.uk/~7Ekdc3/papers/coverage2013/update.140106.pdf>.
- , and —, 2014c: Annual data for “Coverage bias in the HadCrut4 temperature record and its impact on recent temperature trends.” University of York Department of Chemistry, accessed 2 November 2019, [http://www-users.york.ac.uk/~7Ekdc3/papers/coverage2013/had4\\_krig\\_annual\\_v2\\_0\\_0.txt](http://www-users.york.ac.uk/~7Ekdc3/papers/coverage2013/had4_krig_annual_v2_0_0.txt).
- , and —, 2014d: Ensemble data for “Coverage bias in the HadCrut4 temperature record and its impact on recent temperature trends.” University of York Department of Chemistry, accessed 2 November 2019, [http://www-users.york.ac.uk/~7Ekdc3/papers/coverage2013/had4\\_krig\\_ensemble\\_v2\\_0\\_0.txt](http://www-users.york.ac.uk/~7Ekdc3/papers/coverage2013/had4_krig_ensemble_v2_0_0.txt).
- , and P. Jacobs, 2020: Comment on “The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity.” *J. Climate*, **33**, 391–396, <https://doi.org/10.1175/JCLI-D-18-0316>.
- Folland, C., and D. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319–367, <https://doi.org/10.1002/qj.49712152206>.
- Forster, P. M., and Coauthors, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.*, **118**, 1139–1150, <https://doi.org/10.1002/jgrd.50174>.
- Gregory, J. M., T. Andrews, P. Good, T. Mauritsen, and P. M. Forster, 2016: Small global-mean cooling due to volcanic radiative forcing. *Climate Dyn.*, **47**, 3979–3991, <https://doi.org/10.1007/s00382-016-3055-1>.
- Haustein, K., and Coauthors, 2019: A limited role for unforced internal variability in twentieth-century warming. *J. Climate*, **32**, 4893–4917, <https://doi.org/10.1175/JCLI-D-18-0555.1>.
- Huber, M., and R. Knutti, 2014: Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.*, **7**, 651–656, <https://doi.org/10.1038/ngeo2228>.
- Kent, E. C., and Coauthors, 2017: A call for new approaches to quantifying biases in observations of sea surface temperature. *Bull. Amer. Meteor. Soc.*, **98**, 1601–1616, <https://doi.org/10.1175/BAMS-D-15-00251.1>.
- Lewis, N., and J. A. Curry, 2015: The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dyn.*, **45**, 1009–1023, <https://doi.org/10.1007/s00382-014-2342-y>.
- , and —, 2018: The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *J. Climate*, **31**, 6051–6071, <https://doi.org/10.1175/JCLI-D-17-0667.1>.
- Lin, P., Z. Yu, J. Lü, M. Ding, A. Hu, and H. Liu, 2019: Two regimes of Atlantic multidecadal oscillation: Cross-basin dependent or Atlantic-intrinsic. *Sci. Bull.*, **64**, 198–204, <https://doi.org/10.1016/j.scib.2018.12.027>.
- Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, <https://doi.org/10.1007/s10584-011-0156-z>.
- Morice, C. P., J. J. Kennedy, N. A. Rayner and P. D. Jones, 2012a: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res. Atmos.*, **117**, D08101, <https://doi.org/10.1029/2011JD017187>.
- , —, —, and —, 2012b: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set [data]. UK Met Office, accessed 2 November 2019, [https://www.metoffice.gov.uk/hadobs/hadcrut4/data/4.5.0.0/time\\_series/HadCRUT.4.5.0.0.annual\\_ns\\_avg\\_realisations.zip](https://www.metoffice.gov.uk/hadobs/hadcrut4/data/4.5.0.0/time_series/HadCRUT.4.5.0.0.annual_ns_avg_realisations.zip).
- Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 659–740.
- Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nat. Geosci.*, **6**, 415–416, <https://doi.org/10.1038/ngeo1836>.
- Outten, S., P. Thorne, I. Bethke, and Ø. Seland, 2015: Investigating the recent apparent hiatus in surface temperature increases: 1. Construction of two 30-member Earth System Model ensembles. *J. Geophys. Res.*, **120**, 8575–8596, <https://doi.org/10.1002/2015JD023859>.
- Prather, M., and Coauthors, Eds., 2013: Annex II: Climate System Scenario Tables. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1395–1445, accessed 2 November 2019, [http://www.climatechange2013.org/images/report/WG1AR5\\_AIIISM\\_Datafiles.xlsx](http://www.climatechange2013.org/images/report/WG1AR5_AIIISM_Datafiles.xlsx).
- Richardson, M., K. Cowtan, E. Hawkins, and M. Stolpe, 2016: Reconciled climate response estimates from climate models and the energy budget of Earth. *Nat. Climate Change*, **6**, 931–935, <https://doi.org/10.1038/nclimate3066>.
- Simmons, A. J., P. Berrisford, D. P. Dee, H. Hersbach, S. Hirahara, and J.-N. Thépaut, 2017: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quart. J. Roy. Meteor. Soc.*, **143**, 101–119, <https://doi.org/10.1002/qj.2949>.
- Storelvmo, T., T. Leirvik, U. Lohmann, P. Phillips, and M. Wild, 2016: Disentangling greenhouse warming and aerosol cooling to reveal Earth’s climate sensitivity. *Nat. Geosci.*, **9**, 286–289, <https://doi.org/10.1038/ngeo2670>.
- Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727–2765, <https://doi.org/10.5194/gmd-12-2727-2019>.
- Volodin, E., and A. Gritsun, 2018: Simulation of observed climate changes in 1850–2014 with climate model INM-CM5. *Earth Syst. Dyn.*, **9**, 1235–1242, <https://doi.org/10.5194/esd-9-1235-2018>.
- Yan, X., R. Zhang, and T. R. Knutson, 2019: A multivariate AMV index and associated discrepancies between observed and CMIP5 externally forced AMV. *Geophys. Res. Lett.*, **46**, 4421–4431, <https://doi.org/10.1029/2019GL082787>.
- Zhang, R., and Coauthors, 2013: Have aerosols caused the observed Atlantic multidecadal variability? *J. Atmos. Sci.*, **70**, 1135–1144, <https://doi.org/10.1175/JAS-D-12-0331.1>.
- , and Coauthors, 2019: A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Rev. Geophys.*, **57**, 316–375, <https://doi.org/10.1029/2019RG000644>.