Accounting for Changing Temperature Patterns Increases Historical Estimates of Climate Sensitivity

Timothy Andrews[1] [ID], Jonathan M. Gregory[1,2] [ID], David Paynter[3] [ID], Levi G. Silvers[4] [ID], Chen Zhou[5] [ID], Thorsten Mauritsen[6] [ID], Mark J. Webb[1] [ID], Kyle C. Armour[7] [ID], Piers M. Forster[8] [ID], and Holly Titchner[1]

[1]Met Office Hadley Centre, Exeter, UK, [2]NCAS-Climate, University of Reading, Reading, UK, [3]GFDL-NOAA, Princeton, NJ, USA, [4]Princeton University/GFDL, Princeton, NJ, USA, [5]Nanjing University, Nanjing, China, [6]Max Planck Institute for Meteorology, Hamburg, Germany, [7]University of Washington, Seattle, WA, USA, [8]School of Earth and Environment, University of Leeds, Leeds, UK

Abstract Eight atmospheric general circulation models (AGCMs) are forced with observed historical (1871–2010) monthly sea surface temperature and sea ice variations using the Atmospheric Model Intercomparison Project II data set. The AGCMs therefore have a similar temperature pattern and trend to that of observed historical climate change. The AGCMs simulate a spread in climate feedback similar to that seen in coupled simulations of the response to CO_2 quadrupling. However, the feedbacks are robustly more stabilizing and the effective climate sensitivity (EffCS) smaller. This is due to a *pattern effect*, whereby the pattern of observed historical sea surface temperature change gives rise to more negative cloud and longwave clear-sky feedbacks. Assuming the patterns of long-term temperature change simulated by models, and the radiative response to them, are credible; this implies that existing constraints on EffCS from historical energy budget variations give values that are too low and overly constrained, particularly at the upper end. For example, the pattern effect increases the long-term Otto et al. (2013, https://doi.org/10.1038/ngeo1836) EffCS median and 5–95% confidence interval from 1.9 K (0.9–5.0 K) to 3.2 K (1.5–8.1 K).

Plain Language Summary Recent decades have seen cooling over the eastern tropical Pacific and Southern Oceans while temperatures rise globally. Climate models indicate that these regional features, and others, are not expected to continue into the future under sustained forcing from atmospheric carbon dioxide increases. This matters because climate sensitivity depends on the pattern of warming, so if the past has warmed differently from what we expect in the future, then climate sensitivity estimated from the historical record may not apply to the future. We investigate this with a suite of climate models and show that climate sensitivity simulated for observed historical climate change is smaller than for long-term carbon dioxide increases. The results imply that historical energy budget changes only weakly constrain climate sensitivity.

1. Introduction

The relationship between global surface temperature change and the Earth's radiative response—a measure of the radiative feedbacks in the system and a key determinant of the Earth's climate sensitivity—can vary on timescales of decades to millennia. Thus, feedbacks governing warming over the observed historical record may be different from those acting on the Earth's long-term climate sensitivity to rising greenhouse gas concentrations (e.g., Armour, 2017; Gregory & Andrews, 2016; Marvel et al., 2018; Proistosescu & Huybers, 2017; Silvers et al., 2018; Zhou et al., 2016). This is in contrast to decades of studies that explicitly or implicitly assume that the relationship between historical temperature change and energy budget variations provides a direct constraint on long-term climate sensitivity (e.g., Gregory et al., 2002; Otto et al., 2013).

The primary reason why radiative feedback and sensitivity is not constant is because climate feedback depends on the spatial structure of surface temperature change (Andrews et al., 2015; Andrews & Webb, 2018; Armour et al., 2013; Ceppi & Gregory, 2017; Haugstad et al., 2017; Rose et al., 2014; Silvers et al., 2018; Zhou et al., 2016, 2017). This evolves on annual to decadal timescales with modes of unforced coupled atmosphere-ocean variability (e.g., Xie et al., 2016) and spatiotemporal variations in anthropogenic or natural forcings (e.g., Smith et al., 2016; Takahashi & Watanabe, 2016). It also evolves on decadal to centennial timescales in response to sustained anthropogenic forcing due to the intrinsic timescales of the climate response

(such as delayed warming in the eastern tropical Pacific and Southern Oceans; e.g., Andrews et al., 2015; Armour et al., 2016; Senior & Mitchell, 2000). Thus, the pattern of historical temperature change, and thus radiative feedback, is expected to be different from that in response to long-term CO_2 increases (see section 5). We refer to the dependency of radiative feedbacks on the evolving pattern of surface temperature change as a *pattern effect* (Stevens et al., 2016).

Most previous estimates of climate sensitivity based upon historical observations of Earth's energy budget have not allowed for a pattern effect between historical climate change and the long-term response to CO_2 (e.g., Otto et al., 2013). Armour (2017) found that the equilibrium climate sensitivity (ECS; the equilibrium near surface air temperature change in response to a CO_2 doubling) of atmosphere-ocean general circulation models (AOGCMs; estimated from simulations of abrupt CO_2 quadrupling [abrupt-4xCO$_2$]) was about 26% larger than climate sensitivity inferred from transient warming (1% CO_2 simulations, taken to be an analogue for historical climate change) due to pattern effects. Armour (2017) therefore concluded that energy budget estimates of Earth's ECS from the historical record should be increased by this amount. Lewis and Curry (2018) argue for a smaller pattern effect, highlighting ambiguities in the methodology when using idealized CO_2 experiments as an analogue for historical climate change. However, as noted in Armour (2017), the use of 1% CO_2 simulations as an analogue for historical climate change has important limitations in that it neglects the impact from non-CO_2 forcings and unforced climate variability that could have had a significant impact on the pattern of historical temperature change. In particular, under 1% CO_2, AOGCMs do not show cooling of the tropical eastern Pacific Ocean and Southern Ocean—features that have been observed over recent decades but are not expected in the long-term response to increased CO_2 (Zhou et al., 2016). These are regions where atmospheric feedbacks (in particular clouds) are sensitive to the patterns of surface temperature change due to their impact on local and remote atmospheric stability (e.g., Andrews & Webb, 2018; Zhou et al., 2017). This suggests that the magnitude of the pattern effect reported in Armour (2017) may be too low relative to historical climate change. This is an outstanding issue that we aim to address and quantify here.

Here we will show that a suite of atmospheric general circulation models (AGCMs) forced with historical (post 1870) sea surface temperatures (SSTs) and sea ice changes are ideal simulations for quantifying the relationship between historical climate sensitivity and idealized long-term model-derived ECS. They allow us, for the first time, to quantify the pattern effect associated with observed temperature patterns and so provide improved updates to estimates of climate sensitivity derived from historical energy budget constraints. The work builds upon individual studies (Andrews, 2014; Gregory & Andrews, 2016; Silvers et al., 2018; Zhou et al., 2016). Our aim is to (i) bring together these individual model results for an intercomparison of AGCMs forced with historical SST and sea ice variations, (ii) explore the dependence of the experimental design to the underlying SST and sea ice data set, (iii) explore how historical feedbacks in the AGCMs relate to feedbacks diagnosed from their parent AOGCM forced by abrupt-4xCO$_2$, (iv) quantify the pattern effect causing the difference between climate sensitivity under historical climate change and long-term CO_2 changes, and (v) use this pattern effect to update observed energy budget constraints on Earth's climate sensitivity.

2. Simulations, Models, and Data

Eight AGCMs (Table 1) are forced with monthly time-varying observationally derived fields of SST and sea ice from 1871 to 2010 using the Atmospheric Model Intercomparison Project (AMIP) II boundary condition data set (Gates et al., 1999; Hurrell et al., 2008; Taylor et al., 2000). All simulations have natural and anthropogenic forcings (e.g., greenhouse gases, aerosols, solar radiation, etc.) held constant at assumed preindustrial conditions (except CAM4, which used assumed constant present-day conditions; we assume the level of background forcing has no impact on the diagnosed feedback of the model). With *constant* forcings the variation in radiative fluxes comes about solely from the changing SST and sea ice boundary conditions, allowing radiative feedbacks to be accurately diagnosed directly from top-of-atmosphere (TOA) radiation fields (e.g., Haugstad et al., 2017). For details of individual simulations see Gregory and Andrews (2016) for HadGEM2 and HadAM3; Silvers et al. (2018) for GFDL-AM2.1, GFDL-AM3, and GFDL-AM4.0; and Zhou et al. (2016) for CAM4 and CAM5.3. We additionally include simulations from ECHAM6.3, which is closely related to the atmospheric component of the MPI-ESM 1.2 model to be used in CMIP6. This experiment, referred

Table 1

Feedback Parameters in amip-piForcing (λ_{amip}) and abrupt-4xCO$_2$ (λ_{4xCO2}) Atmospheric General Circulation Model and Atmosphere-Ocean General Circulation Model experiments

Model	λ_{amip} (W·m^{-2}·K^{-1})	λ_{4xCO2} (W·m^{-2}·K^{-1})	$S = \lambda_{4xCO2}/\lambda_{amip}$	$\Delta\lambda = \lambda_{4xCO2} - \lambda_{amip}$ (W·m^{-2}·K^{-1})	$EffCS_{amip}$ (K)	$EffCS_{4xCO2}$ (K)
CAM4	−2.27	−1.23	0.54	1.04	1.57	2.90
CAM5.3	−1.71	n/a	n/a	n/a	n/a	n/a
ECHAM6.3	−1.90	−1.36	0.72	0.54	2.17	3.01
GFDL-AM2.1	−1.67	−1.38	0.83	0.29	2.01	2.43
GFDL-AM3	−1.40	−0.75	0.53	0.65	2.13	3.99
GFDL-AM4.0	−1.91	n/a	n/a	n/a	n/a	n/a
HadAM3	−1.65	−1.04	0.63	0.61	2.14	3.38
HadGEM2	−1.37	−0.64	0.47	0.73	2.14	4.58
Mean(1.645*σ)	−1.74(0.48)	−1.07(0.52)	0.62(0.22)	0.64(0.40)	2.03(0.38)	3.38(1.29)

Note. S and $\Delta\lambda$ are the ratio and differences between λ_{4xCO2} and λ_{amip}, respectively. These are used to update feedback parameters derived from historical energy budget changes to account for the pattern effect between historical climate change and abrupt-4xCO$_2$. $EffCS_{amip} = -F_{2x}/\lambda_{amip}$ and $EffCS_{4xCO2} = -F_{2x}/\lambda_{4xCO2}$ are the effective climate sensitivities from the amip-piForcing and abrupt-4xCO$_2$ experiments, where F_{2x} is the model's effective radiative forcing for a doubling of CO$_2$ (calculated from the abrupt-4xCO$_2$ experiments using a linear regression technique as per Andrews et al., 2012).

to here as *amip-piForcing* (Gregory & Andrews, 2016), is included in the Cloud Feedback Model Intercomparison Project contribution to CMIP6 (Webb et al., 2017). The sensitivity of the results to the AMIP II boundary condition data set is explored with analogous experiments using the HadISST2.1 SST and sea ice data set (Titchner & Rayner, 2014; supporting information).

All simulations ran for 140 years from January 1871 through to December 2010, except for GFDL-AM2.1 and GFDL-AM3, which finished in December 2004. All data are global annual mean, and anomalies are presented relative to an 1871–1900 baseline. CAM4 and CAM5.3 results are single realizations; HadGEM2 and HadAM3 simulations are ensembles of four realizations each; ECHAM6.3, GFDL-AM2.1, and GFDL-AM4.0 have five realizations each; while GFDL-AM3 has six realizations. The HadGEM2 results are not identical to those presented in Gregory and Andrews (2016) because it has been discovered that land cover change was included in their HadGEM2 simulations. We have confirmed that the updated simulations used here, which have constant land cover, do not affect the main conclusions of Gregory and Andrews (2016). In fact the multidecadal variability in feedback in HadGEM2 is now found to be more consistent with their HadAM3 results (section 3).

For comparison to long-term climate sensitivity and feedback parameters we make use of an abrupt-4xCO$_2$ simulation of each AGCM's parent AOGCM. For CAM4, GFDL-AM2.1, GFDL-AM3, and HadGEM2 we use the CCSM4, GFDL-ESM2M, GFDL-CM3, and HadGEM2-ES CMIP5 abrupt-4xCO$_2$ simulations, respectively (Taylor et al., 2012). Feedbacks and associated effective climate sensitivity (EffCS; the equilibrium near surface air temperature change in response to a CO$_2$ doubling assuming constant feedback strength) are derived from the regression of global annual mean change in radiative flux dN against surface air temperature change dT for the 150 years of the simulation, according to $EffCS = -F_{2x}/\lambda$, where F_{2x}, the forcing from a doubling of CO$_2$, is equal to the dN axis intercept divided by 2 (to convert 4xCO$_2$ to 2xCO$_2$) and λ, the feedback parameter, is equal to the slope of the regression line (Andrews et al., 2012). We have similar simulations for ECHAM6.3 and HadAM3 using the MPI-ESM 1.1 and HadCM3 models, respectively, though these are not in the CMIP5 archive. The HadCM3 simulation is only 100 years long but is a mean of seven realizations. CAM5.3 and GFDL-AM4.0 do not yet have equivalent coupled 4xCO$_2$ simulations. We choose to use EffCS rather than the *true* ECS since few AOGCMs are run to equilibrium, and thus, the true ECS is not generally known. Paynter et al. (2018) showed that the actual ECS from multimillennial GFDL-ESM2M and GFDL-CM3 simulations was nearly 1 K higher than the EffCS we use here from abrupt-4xCO$_2$. Hence, the values we report for EffCS might be viewed as a lower bound on ECS if other models behave in a similar way.

3. Radiative Feedbacks and Sensitivities

Figure 1a shows the global annual mean near surface air temperature change (dT) of the eight individual AGCM amip-piForcing simulations in comparison to HadCRUT4 (Morice et al., 2012). As expected the models capture the observed variability and trends in dT well (the correlation coefficient, r, between observed and
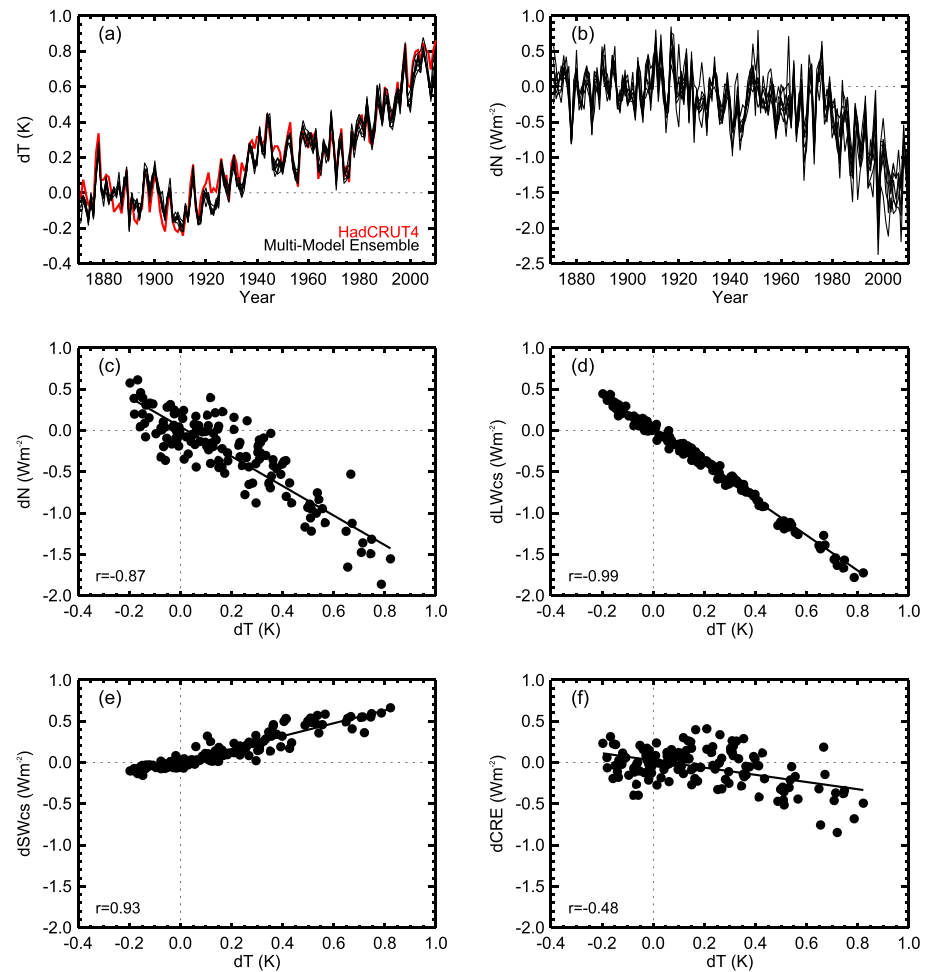
Figure 1. (a) Comparison of historical near-surface air temperature change (dT) simulated by the atmospheric general circulation models in amip-piForcing (individual black lines) against observed (HadCRUT4) variations (red). (b) Time series of the change in net top-of-atmosphere radiative flux (dN) in the individual atmospheric general circulation model experiments. (c–f) The relationship and correlation coefficient (r) between the multimodel ensemble mean (c) dN; (d) longwave clear-sky radiative flux change, dLWcs; (e) shortwave clear-sky radiative flux change, dSWcs; and (f) cloud radiative effect change, dCRE, against dT. All points are global annual means covering the historical period (1871–2010), and fluxes are positive downward. Changes are relative to an 1871–1900 baseline.

simulated dT is >0.95 for every model). However, the AGCMs omit the small part of the recent warming trend over land that arises as a direct adjustment to changes in CO_2 and other forcing agents (dT in HadCRUT4 averaged over 2000–2010 is 0.79 K, whereas it ranges from 0.66 to 0.76 K in the AGCMs; see also, Andrews, 2014; Gregory & Andrews, 2016). Figure 1b shows the net top-of-atmosphere radiative flux change, dN. It is generally negative because as dT increases positively the planet loses heat to space. This relationship is shown in Figure 1c for the multimodel ensemble mean. The slope of the regression line (ordinary least squares, over the annual mean 1871–2010 time series data) measures the feedback parameter λ_{amip} (in W·m^{-2}·K^{-1}), where subscript *amip* is used to indicate that the feedback parameter was derived from the amip-piForcing experiment. Individual model results are given in Table 1.

The equivalent feedback parameters derived from six available parent AOGCM abrupt-4xCO_2 simulations (λ_{4xCO2}) are compared to λ_{amip} in Figure 2 and Table 1. We find that λ_{amip} is more negative than λ_{4xCO2} in all models. In other words, AGCMs forced with historical SST and sea ice changes robustly simulate more stabilizing feedbacks (lower EffCS) than their parent AOGCM forced by long-term CO_2 changes. On average, the difference in λ between amip-piForcing and abrupt-4xCO_2 is $\Delta\lambda = \lambda_{4xCO2} - \lambda_{amip} = 0.64$ W·m^{-2}·K^{-1}, ranging from 0.29 to 1.04 W·m^{-2}·K^{-1} across the AGCMs (Table 1).
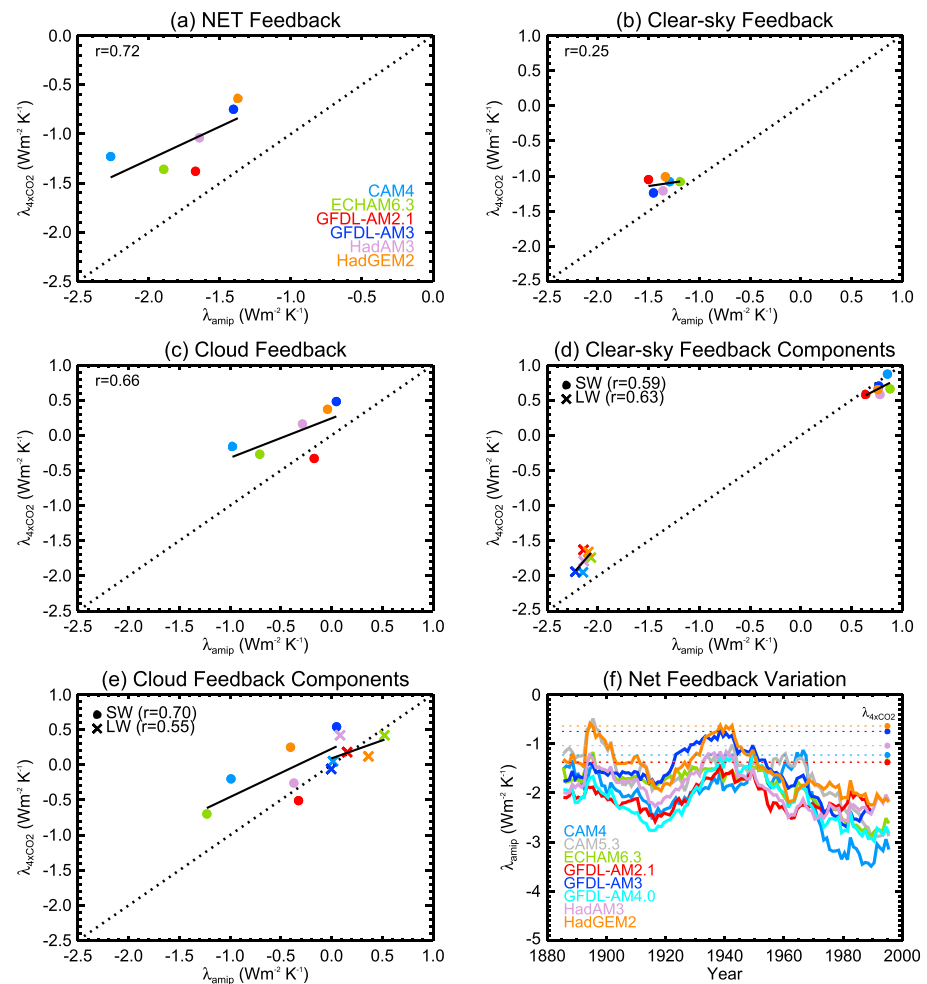
Figure 2. Relationship between the feedback parameter evaluated by regression of dN against dT over the historical period (1871–2010) in amip-piForcing (λ_{amip}) and 150 years of abrupt-4xCO$_2$ (λ_{4xCO2}) for (a) NET radiative feedback, (b) clear-sky component, (c) CRE component, (d) LW and SW clear-sky components, and (e) LW and SW CRE components. (f) Time series of λ_{amip} for individual AGCMs evaluated by linear regression of dN against dT in a sliding 30-year window in the amip-piForcing experiments, the year represents the center of the window. Colored circles in (f) with horizontal lines show the feedback parameter values from abrupt-4xCO$_2$. LW = longwave; SW = shortwave.

The source of $\Delta\lambda$ is shown in Figure 2. The clear-sky feedback (Figures 1d and 1e) is slightly (but robustly) more negative in amip-piForcing compared to abrupt-4xCO$_2$ (Figure 2b) due to differences in longwave (LW) clear-sky feedback processes that are partly offset by shortwave (SW) clear-sky feedback differences (Figure 2d). This difference in clear sky feedback between amip-piForcing and abrupt-4xCO$_2$ explains the relatively small change in net sensitivity between these experiments for the GFDL-AM2.1 model. For the other models, differences in cloud feedback (measured by changes in cloud radiative effect, CRE) (Figure 1f) are a larger source of the reduced sensitivity in amip-piForcing (Figure 2c). This mostly comes from SW cloud feedback processes, with historical LW cloud feedback processes generally being representative of that seen in abrupt-4xCO$_2$ (Figure 2e). These findings are consistent with process-orientated studies that suggest lapse-rate (which affect LW clear sky) and low-cloud (which affect SW, NET, and CRE) feedbacks vary the most with SST patterns, especially in the Pacific (see below and Andrews et al., 2015; Andrews & Webb, 2018; Ceppi & Gregory, 2017; Rose et al., 2014; Silvers et al., 2018; Zhou et al., 2016, 2017).

In amip-piForcing the model mean $EffCS_{amip} = -F_{2x}/\lambda_{amip}$ is ~2 K, ranging from 1.6 to 2.2 K across the AGCMs (Table 1). The narrowness of this $EffCS_{amip}$ range does not arise due to reduced uncertainty in λ_{amip} relative to λ_{4xCO2} (Table 1). On the contrary, the spread (measured by 1.645*σ) in λ_{amip} is almost the same size as the spread in λ_{4xCO2} (Table 1). The spread in $EffCS_{amip}$ is narrower primarily because λ_{amip} is on average more negative

than λ_{4xCO2}. Since EffCS depends on the reciprocal of λ, the same spread in λ, shifted to more negative numbers, will give rise to a narrower spread in EffCS (e.g., Roe, 2009). A similar spread in in λ_{amip} and λ_{4xCO2} suggests that different patterns of SST change across AOGCMs do not contribute significantly to the spread in atmospheric feedbacks in abrupt-4xCO2 experiments (see also Andrews & Webb, 2018; Ringer et al., 2014), which must therefore come about due to differences in atmospheric physics and parameterizations.

$EffCS_{4xCO2}$ (of the parent AOGCM) is in all cases larger than $EffCS_{amip}$, ranging from 2.4 to 4.6 K (Table 1). In the multimodel mean, $EffCS_{4xCO2}$ is ~67% larger than that implied from $EffCS_{amip}$. This model mean historical pattern effect is substantially larger than the 26% found by Armour (2017), supporting the hypothesis that the pattern effect is larger in the historical record than simulated in transient 1% CO_2 AOGCM simulations because the later miss key features of the observed warming pattern. This result is even more striking given that Armour (2017) used an EffCS definition from abrupt-4xCO2 that gives larger values than ours (they used years 21–150 of abrupt-4xCO2, whereas we use years 1–150).

It is also useful to study shorter time periods to help inform our understanding of the relationship between shorter-term variations in temperature and radiative fluxes, as have been used by many studies to estimate EffCS particularly since the satellite era (e.g., Forster, 2017). Figure 2f shows the feedback parameter for 30-year moving windows over the historical period in the AGCM simulations (calculated as per Gregory & Andrews, 2016), in comparison to λ_{4xCO2} (horizontal lines). There is a substantial multidecadal variability in the feedback parameter that is common to all models, with a peak in feedback parameter (higher EffCS) around the 1940s and a minimum (lower EffCS) in the most recent decades (post ~1980). Generally, λ_{amip} is always more negative than λ_{4xCO2}. There are only a few instances where the λ_{amip} is similar to λ_{4xCO2}, for example, ~1940 for HadGEM2 and GFDL-AM2.1, but no instances where λ_{amip} is substantially less negative than λ_{4xCO2}. The difference is greatest in the most recent decades, suggesting that energy budget constraints on ECS based on recent decades of satellite data will be most strongly biased low. This is consistent with process understanding of the pattern effect, since recent decades have shown substantial cooling in the eastern Pacific and Southern Oceans while warming in the west Pacific warm pool (e.g., Zhou et al., 2016). The cooling in the descent region of the tropical Pacific will favor increased cloudiness (a negative feedback), while warming in the west Pacific ascent region efficiently warms free tropospheric air (increasing the negative lapse-rate feedback widely across the tropics and midlatitudes) as well as further increasing the lower tropospheric stability and cloudiness in the marine low-cloud descent regions (Andrews & Webb, 2018; Ceppi & Gregory, 2017; Zhou et al., 2016).

Most of the multidecadal variation in feedback strength comes from changes in the strength of cloud feedback (the correlation between the NET and CRE feedback time series, calculated in a similar way, is >0.94 in each AGCM), while the clear-sky feedbacks show less variation (not shown). This, as well as atmospheric variability, helps explain why cloud feedback is not as linearly correlated to dT variations over the full historical period compared to clear-sky feedbacks (r = 0.48 for CRE compared to 0.99 and 0.93 for the clear-sky fluxes, Figures 1d–1f).

4. Constraints on Observed Estimates of Climate Sensitivity

The pattern effect causing the difference between simulated EffCS under historical climate change and long-term CO_2 increase implies that historical energy budget constraints on EffCS do not directly apply to long-term ECS. To account for this, we use the difference in λ between amip-piForcing and abrupt-4xCO2 as a measure of the pattern effect to update historical energy budget estimates of λ and EffCS. This is in contrast to Armour (2017) who had to use 1% CO_2 simulations as a surrogate for historical climate change. Here we are quantifying the pattern effect associated with patterns of temperature change that actually occurred in the real world, relative to those simulated by AOGCMs to long-term CO_2 increases. The pattern effect therefore assumes that long-term warming patterns in AOGCMs not yet seen in the historical record, and the radiative response to them, are credible (see section 5).

To illustrate the impact of the pattern effect we use the Otto et al. (2013) historical energy budget constraints as our starting point, though other data sets exist (see Forster, 2017) and clearly the EffCS estimates presented below will depend on this. First, we reproduce the historical EffCS estimates reported in Otto et al. (2013) using their best estimate and 5–95% confidence intervals for the historical (denoted by subscript *hist*) change in temperature (dT_{hist} = 0.48 ± 0.2 K), heat uptake (dN_{hist} = 0.35 ± 0.13 W/m^2) and radiative forcing
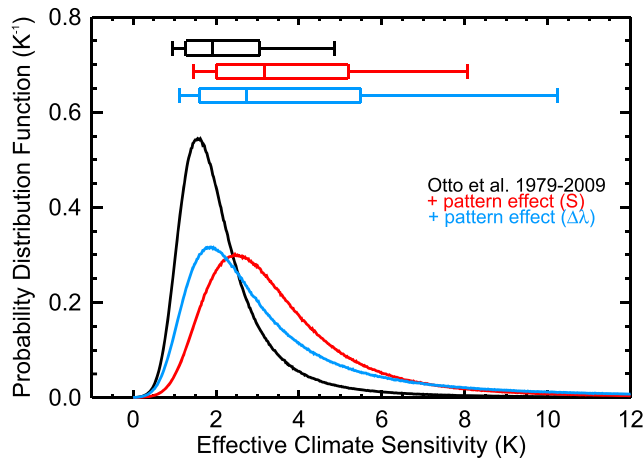
Figure 3. Comparison of the effective climate sensitivity probability distribution function from a historical energy budget constraint (Otto et al., 2013), before (black) and after (colors) accounting for the pattern effect between historical climate change and abrupt-4xCO_2. Red accounts for the pattern effect by scaling the historical feedback parameter λ_{hist} by the ratio ($S = \lambda_{4xCO2}/\lambda_{amip}$) of the feedbacks found in the amip-piForcing and abrupt-4xCO_2 simulations. Blue accounts for the pattern effect by adding the difference in feedbacks ($\Delta\lambda = \lambda_{4xCO2} - \lambda_{amip}$) to λ_{hist} (see section 4 and Table 1). Box plots show the 5–95% confidence interval (end bars), the 17–83% confidence interval (box ends), and the median (line in box).

(dF_{hist} = 1.21 ± 0.52 W/m^2) for the 40-year period 1970–2009 relative to preindustrial (which they define as 1860–1879; their Table S1, row 5). To be consistent with Otto et al. (2013) we also use their forcing and its uncertainty for a doubling of CO_2 (F_{2x} = 3.44 (±10%) W/m^2). We randomly sample (with replacement) 10 million times from the Gaussian distributions of dT_{hist}, dN_{hist}, dF_{hist}, and F_{2x} to calculate $\lambda_{hist} = (dN_{hist} - dF_{hist})/dT_{hist}$ and $EffCS_{hist} = -F_{2x}/\lambda_{hist}$. We assume the uncertainty in F_{2x} and the greenhouse gas component of dF_{hist} are correlated as in Otto et al. (2013). The resulting EffCS values are binned into intervals of 0.02 and normalized to produce a probability density function (PDF), excluding values less than 0 and greater than 20. The resulting PDF and percentiles (Figure 3, black lines) recovers the Otto et al. (2013) $EffCS_{hist}$ median (1.9 K) and 5–95% confidence interval (0.9–5.0 K) to within 0.1 K.

Following Armour (2017), we update the Otto et al. (2013) EffCS estimate for the pattern effect between historical climate change and abrupt-4xCO_2 using two methods. We first scale the historical feedback parameter λ_{hist} by the ratio of the feedbacks found in the amip-piForcing and abrupt-4xCO_2 simulations, so $\lambda = \lambda_{hist}*S$ where $S = \lambda_{4xCO2}/\lambda_{amip}$ (Table 1). EffCS is then given by $EffCS = -F_{2x}/\lambda = -F_{2x}/(\lambda_{hist}*S$; equivalent to Equation 4 in Armour, 2017). Alternatively, we update λ_{hist} by the difference in feedbacks, according to $\lambda = \lambda_{hist} + \Delta\lambda$, where $\Delta\lambda = \lambda_{4xCO2} - \lambda_{amip}$. EffCS is then given by $EffCS = -F_{2x}/\lambda = -F_{2x}/(\lambda_{hist} + \Delta\lambda$; equivalent to Equation 5 in Armour, 2017). We then calculate the EffCS PDF as above by randomly sampling from the F_{2x} and λ_{hist} distributions, along with S and $\Delta\lambda$ chosen randomly with equal likelihood from the individual model results (Table 1). Note that using the difference ($\Delta\lambda$) approach increases the likelihood of returning very large (or even negative) EffCS values, since $\lambda = \lambda_{hist} + \Delta\lambda$ can result in λ values close to 0 or even with a changed sign when sampling λ_{hist} values that are small. Hence, the results of this method are potentially sensitive to the assumption of excluding negative EffCS values or those greater than 20 K.

We compare the PDF of $EffCS_{hist}$ (which is an approximation of Otto et al., 2013) against its updated versions that accounts of the pattern effect in Figure 3. The Otto et al. (2013) median and 5–95% confidence interval increases from 1.9 K (0.9–5.0 K) to 3.2 K (1.5–8.1 K) using the ratio (S) approach (Figure 3, red lines) or 2.7 K (1.1–10.2 K) if we use the difference ($\Delta\lambda$) approach (Figure 3, blue lines). Alternatively, if we take the Otto et al. (2013) data relating to their most recent decade (2000–2009; their Table S1 row 4) then the Otto et al. (2013) estimate and 5–95% confidence interval increases from 2.0 K (1.2–3.9 K) to 3.3 K (1.8–6.8 K) using the ratio approach or 3.0 K (1.5–9.7 K) using the difference approach. Thus, either way and for different time periods, the pattern effect from amip-piForcing to abrupt-4xCO_2 results in a substantial median ECS increase, while the lowest values of ECS become less likely, and higher ECS values become much harder to rule out.

Another way of estimating the pattern effect is by comparing feedbacks in AOGCM historical simulations to abrupt-4xCO_2 (e.g., Marvel et al., 2018; Paynter & Frölicher, 2015). However, we believe amip-piForcing is superior, because (i) the diagnosed pattern effect in an AOGCM historical simulation will depend on its ability to correctly simulate the patterns of historical climate change, including the magnitude and timing of unforced variability, which they are not expected to simulate correctly (e.g., Mauritsen, 2016; Zhou et al., 2016) and (ii) determining feedbacks in AOGCM historical simulations requires knowledge of the time-varying effective radiative forcing of the model, something which is not routinely diagnosed and is difficult to assume because of model diversity in forcing, particularly from aerosols (Forster, 2017). The amip-piForcing approach alleviates both of the above issues.

Note that for simplicity in the above calculations we have assumed that λ_{amip} (calculated via linear regression over the amip-piForcing simulations, section 3) is appropriate to the time periods and methodology of Otto et al. (2013; who use finite differences, rather than linear regression, between decades to calculate changes). To check this we recompute λ_{amip} and the corresponding S and $\Delta\lambda$ values using the same method and time periods as Otto et al., that is, $\lambda_{amip} = dN/dT$, where dN and dT are averaged over the relevant decades (though

for 2000–2009 we use the 1995–2004 decade, since the GFDL runs finished in 2004). We cannot use an identical baseline as Otto et al. (2013) since our simulations begin in 1871 and their baseline begins in 1860. Regardless, for 1979–2009 or 2000–2009, the resulting updated EffCS PDF has a median and 5–95% confidence interval to within ±15% of the regression methods used above. Hence, in practice our conclusions are not sensitive on this assumption.

5. Summary and Discussion

An intercomparison of AGCMs forced with historical (post 1870) SSTs and sea ice from the AMIP II boundary condition data set reveal some common results:

1. When AGCMs are forced with historical SST and sea ice changes, the models agree on an EffCS of ~2 K, in line with best estimates from historical energy budget variations (e.g., Otto et al., 2013) but significantly lower than the EffCS of the corresponding parent AOGCMs when forced with abrupt-4xCO$_2$ (~2.4–4.6 K for the corresponding set of models).
2. The lower historical EffCS relative to abrupt-4xCO$_2$ is predominantly because LW clear-sky and cloud radiative feedbacks are less positive in response to historical SST and sea ice variations than in long-term climate sensitivity simulations. This is an example of what is called a pattern effect (Stevens et al., 2016) and is consistent with process understanding that suggests that lapse-rate and low-cloud feedbacks vary most with SST patterns, especially those in the tropical Pacific ascent/descent regions, which have large impacts on atmospheric stability (Andrews & Webb, 2018; Ceppi & Gregory, 2017; Zhou et al., 2016).
3. The models agree that the most recent decades (e.g., 1980–2010) generally give rise to the most negative feedbacks (lowest EffCS). Hence, the pattern effect will be largest for estimates of feedbacks and EffCS based on the satellite era. This is a period when the eastern tropical Pacific and Southern Oceans, regions important for the pattern effect, have been cooling but are not expected to continue to do so in the long-term response to increased CO$_2$ (e.g., Zhou et al., 2016).

The pattern effect causing the difference between EffCS under historical climate change and long-term CO$_2$ changes implies that current constraints on climate sensitivity that do not consider this give values that are too low and are overly constrained, particularly at the upper bound. We present an approach to adjust historical energy budget-derived EffCS estimates for the pattern effect. For example, the historical (1860–1879 to 1970–2009) observational EffCS estimate (median) and 5–95% confidence interval of Otto et al. (2013) increases from 1.9 K (0.9–5.0 K) to 3.2 K (1.5–8.1 K) using an approach that scales the historical feedback parameter by the ratio of the feedbacks found in amip-piForcing and abrupt-4xCO$_2$. Thus, the pattern effect increases historical EffCS median values, reduces the likelihood of the lowest EffCS values and makes higher values significantly harder to rule out. Determining whether values toward the extremes of these bounds are plausible would require further understanding of the pattern effect or assessing and combining other lines of evidence, such as from process understanding (see Stevens et al., 2016). This is important because a higher EffCS increases the risk of state-dependent feedbacks and large warmings (Bloch-Johnson et al., 2015).

The pattern effect between historical climate change and long-term CO$_2$ increase assumes that key aspects of long-term warming patterns simulated by AOGCMs not yet seen in the observational record, such as substantial warming of the Southern Ocean and eastern tropical Pacific, and the radiative response to them, are credible. Such patterns are consistent with paleo records (e.g., Fedorov et al., 2015; Masson-Delmotte et al., 2013) and basic physical understanding of the behavior and timescale of oceanic upwelling (e.g., Armour et al., 2016; Clement et al., 1996; Held et al., 2010), though they are difficult to observationally constrain (Mauritsen, 2016). To argue for a negligible pattern effect (e.g., Lewis & Curry, 2018) would require that atmospheric feedbacks are insensitive to patterns of temperature change or that the pattern of observed historical temperature change represents the equilibrated pattern response to increased CO$_2$. This is at odds with basic physical understanding and bodies of work on the role for unforced variability, transient effects, and non-CO$_2$ forcings such as aerosols on the pattern of historical climate change (e.g., Armour et al., 2016; Held et al., 2010; Jones et al., 2013; Takahashi & Watanabe, 2016; Xie et al., 2016). Further progress in constraining the pattern effect and EffCS will come from improved understanding of the causes and processes of surface temperature change patterns in observations and AOGCM projections, as well as the radiative response to them.

References

Andrews, T. (2014). Using an AGCM to diagnose historical effective radiative forcing and mechanisms of recent decadal climate change. *Journal of Climate*, *27*(3), 1193–1209. https://doi.org/10.1175/jcli-d-13-00336.1

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate*, *28*(4), 1630–1648. https://doi.org/10.1175/jcli-d-14-00545.1

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*, L09712. https://doi.org/10.1029/2012GL051607

Andrews, T., & Webb, M. J. (2018). The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical Pacific warming. *Journal of Climate*, *31*(2), 641–654. https://doi.org/10.1175/JCLI-D-17-0087

Armour, K. C. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, *7*(5), 331–335. https://doi.org/10.1038/nclimate3278

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, *26*(13), 4518–4534. https://doi.org/10.1175/JCLI-D-12-00544.1

Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A., & Newsom, E. R. (2016). Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geoscience*, *9*(7), 549–554. https://doi.org/10.1038/ngeo2731

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*, 4973–4980. https://doi.org/10.1002/2015GL064240

Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(50), 13,126–13,131. https://doi.org/10.1073/pnas.1714308114

Clement, A. C., Seager, R., Cane, M. A., & Zebiak, S. E. (1996). An ocean dynamical thermostat. *Journal of Climate*, *9*(9), 2190–2196. https://doi.org/10.1175/1520-0442(1996)009<2190:AODT>2.0.CO;2

Fedorov, A. V., Burls, N. J., Lawrence, K. T., & Peterson, L. C. (2015). Tightly linked zonal and meridional sea surface temperature gradients over the past five million years. *Nature Geoscience*, *9*(8), 568–569. https://doi.org/10.1038/ngeo2755

Forster, P. M. (2017). Inference of climate sensitivity from analysis of Earth's energy budget. *Annual Review of Earth and Planetary Sciences*, *44*(1), 85–106. https://doi.org/10.1146/annurev-earth-060614-105156

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., et al. (1999). An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, *80*(1), 29–55. https://doi.org/10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2

Gregory, J. M., & Andrews, T. (2016). Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, *43*, 3911–3920. https://doi.org/10.1002/2016GL068406

Gregory, J. M., Stouffer, R. J., Raper, S. C., Stott, P. A., & Rayner, N. A. (2002). An observationally based estimate of the climate sensitivity. *Journal of Climate*, *15*(22), 3117–3121. https://doi.org/10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2

Haugstad, A. D., Armour, K. C., Battisti, D. S., & Rose, B. E. J. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*, 7455–7463. https://doi.org/10.1002/2017GL074372

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), 2418–2427. https://doi.org/10.1175/2009JCLI3466.1

Hurrell, J., Hack, J., Shea, D., Caron, J., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *Journal of Climate*, *21*(19), 5145–5153. https://doi.org/10.1175/2008JCLI2292.1

Jones, G. S., Stott, P. A., & Christidis, N. (2013). Attribution of observed historical near–surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, *118*, 4001–4024. https://doi.org/10.1002/jgrd.50239

Lewis, N., & Curry, J. (2018). The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *Journal of Climate*, *31*(15), 6051–6071. https://doi.org/10.1175/JCLI-D-17-0667.1

Marvel, K., Pincus, R., Schmidt, G. A., & Miller, R. L. (2018). Internal variability and disequilibrium confound estimates of climate sensitivity from observations. *Geophysical Research Letters*, *45*, 1595–1601. https://doi.org/10.1002/2017GL076468

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, J., González, J. F., et al. (2013). Information from paleoclimate archives. In M. Collins, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 383–464). Cambridge, UK and New York: Cambridge University Press.

Mauritsen, T. (2016). Global warming: Clouds cooled the Earth. *Nature Geoscience*, *9*(12), 865–867. https://doi.org/10.1038/ngeo2838

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*, *117*, D08101. https://doi.org/10.1029/2011JD017187

Otto, A., Otto, F. E. L., Boucher, O., Church, J., Hegerl, G., Forster, P. M., et al. (2013). Energy budget constraints on climate response. *Nature Geoscience*, *6*(6), 415–416. https://doi.org/10.1038/ngeo1836

Paynter, D., & Frölicher, T. L. (2015). Sensitivity of radiative forcing, ocean heat uptake, and climate feedback to changes in anthropogenic greenhouse gases and aerosols. *Journal of Geophysical Research: Atmospheres*, *120*, 9837–9854. https://doi.org/10.1002/2015JD023364

Paynter, D., Frölicher, T. L., Horowitz, L. W., & Silvers, L. G. (2018). Equilibrium climate sensitivity obtained from multimillennial runs of two GFDL climate models. *Journal of Geophysical Research: Atmospheres*, *123*, 1921–1941. https://doi.org/10.1002/2017JD027885

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*(7), e1602821. https://doi.org/10.1126/sciadv.1602821

Ringer, M. A., Andrews, T., & Webb, M. J. (2014). Global-mean radiative feedbacks and forcing in atmosphere-only and fully-coupled climate change experiments. *Geophysical Research Letters*, *41*, 4035–4042. https://doi.org/10.1002/2014GL060347

Roe, G. (2009). Feedbacks, timescales, and seeing red. *Annual Review of Earth and Planetary Sciences*, *37*(1), 93–115. https://doi.org/10.1146/annurev.earth.061008.134734

Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., & Koll, D. D. B. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, *41*, 1071–1078. https://doi.org/10.1002/2013GL058955

Senior, C. A., & Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), 2685–2688. https://doi.org/10.1029/2000GL011373

Silvers, L. G., Paynter, D., & Zhao, M. (2018). The diversity of cloud responses to twentieth century sea surface temperatures. *Geophysical Research Letters*, *45*, 391–400. https://doi.org/10.1002/2017GL075583

Smith, D. M., Booth, B. B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., et al. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, 6(10), 936–940. https://doi.org/10.1038/nclimate3058

Stevens, B., Sherwood, S. C., Bony, S., & Webb, M. J. (2016). Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Future*, 4(11), 512–522. https://doi.org/10.1002/2016EF000376

Takahashi, C., & Watanabe, M. (2016). Pacific trade winds accelerated by aerosol forcing over the past two decades. *Nature Climate Change*, 6(8), 768–772. https://doi.org/10.1038/nclimate2996

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1

Taylor, K. E., Williamson, D., & Zwiers, F. (2000). The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations, PCMDI Report No. 60, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory.

Titchner, H. A., & Rayner, N. A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, 119, 2864–2889. https://doi.org/10.1002/2013JD020316

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, 10(1), 359–384. https://doi.org/10.5194/gmd-10-359-2017

Xie, S.-P., Kosaka, Y., & Okumura, Y. M. (2016). Distinct energy budgets for anthropogenic and natural changes during global warming hiatus. *Nature Geoscience*, 9(1), 29–33. https://doi.org/10.1038/ngeo2581

Zhou, C., Zelinka, M. D., & Klein, S. A. (2016). Impact of decadal cloud variations on the Earth's energy budget. *Nature Geoscience*, 9(12), 871–874. https://doi.org/10.1038/ngeo2828

Zhou, C., Zelinka, M. D., & Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *Journal of Advances in Modeling Earth Systems*, 9, 2174–2189. https://doi.org/10.1002/2017MS001096