On the effect of historical SST patterns on radiative feedback

On the effect of historical SST patterns on radiative feedback

Timothy Andrews[1] and Alejandro Bodas-Salcedo

Met Office Hadley Centre, Exeter, UK.

Jonathan M. Gregory

Met Office Hadley Centre, Exeter, UK and National Centre for Atmospheric Science, University of Reading, Reading, UK.

Yue Dong

Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA.

Now at Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA.

Kyle C. Armour

Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA and School of Oceanography, University of Washington, Seattle, WA, USA.

David Paynter

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, NJ, USA.

Pu Lin

Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA.

Angshuman Modak and Thorsten Mauritsen

University of Stockholm, Department of Meteorology, Stockholm, Sweden.

Jason N.S. Cole

Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, BC, Canada.

Brian Medeiros

National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, Boulder, CO, USA.

James J. Benedict

Rosenstiel School of Marine and Atmospheric Science, University of Miami, Coral Gables, FL, USA.

Now at Los Alamos National Laboratory, Los Alamos, NM, USA.

Hervé Douville and Romain Roehrig

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France.

[1] Correspondence to Timothy Andrews (timothy.andrews@metoffice.gov.uk)

Tsuyoshi Koshiro and Hideaki Kawai

Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan.

Tomoo Ogura

National Institute for Environmental Studies, Tsukuba, Japan.

Jean-Louis Dufresne

Laboratoire de Météorologie Dynamique/IPSL, CNRS, Sorbonne Université, École Normale Supérieure, PSL Research University, École Polytechnique, Paris, France.

Richard P. Allan

National Centre for Earth Observation, University of Reading, Reading, UK.

Department of Meteorology, University of Reading, Reading, UK.

Chunlei Liu

South China Sea Institute of Marine Meteorology, Guangdong Ocean University, Zhanjiang, China.

Submitted to *Journal of Geophysical Research*

21st February 2022; Revised June 2022 & August 2022.

Abstract

We investigate the dependence of radiative feedback on the pattern of sea-surface temperature (SST) change in fourteen Atmospheric General Circulation Models (AGCMs) forced with observed variations in SST and sea-ice over the historical record from 1871 to near-present. We find that over 1871-1980, the Earth warmed with feedbacks largely consistent and strongly correlated with long-term climate sensitivity feedbacks (diagnosed from corresponding atmosphere-ocean GCM abrupt-4xCO2 simulations). Post 1980 however, the Earth warmed with unusual trends in tropical Pacific SSTs (enhanced warming in the west, cooling in the east) and cooling in the Southern Ocean that drove climate feedback to be uncorrelated with – and indicating much lower climate sensitivity than – that expected for long-term CO_2 increase. We show that these conclusions are not strongly dependent on the AMIP II SST dataset used to force the AGCMs, though the magnitude of feedback post 1980 is generally smaller in nine AGCMs forced with alternative HadISST1 SST boundary conditions. We quantify a 'pattern effect' (defined as the difference between historical and long-term CO_2 feedback) equal to 0.48 ± 0.47 [5-95%] W m^{-2} K^{-1} for the time-period 1871-2010 when the AGCMs are forced with HadISST1 SSTs, or 0.70 ± 0.47 [5-95%] W m^{-2} K^{-1} when forced with AMIP II SSTs. Assessed changes in the Earth's historical energy budget agree with the AGCM feedback estimates. Furthermore satellite observations of changes in top-of-atmosphere radiative fluxes since 1985 suggest that the pattern effect was particularly strong over recent decades but may be waning post 2014.

1. Introduction

1.1. Background

A common starting point for quantifying the sensitivity of the Earth's climate to external perturbations is consideration of the global-mean energy budget, $N = F + \lambda T$, where N is the net downward radiative flux at the top-of-atmosphere (TOA) (units W m^{-2}), F the effective radiative forcing (ERF, units W m^{-2}), λ the climate feedback parameter (units W m^{-2} K^{-1}, a negative number in this paper, but the opposite convention is also used) and T the surface-air-temperature change (units K) relative to an unperturbed steady state in which N=F=0. Applied to non-steady states, such as the Earth's historical record (since the 1800s), λ is determined via either (i) differences (denoted by Δ) between two climate states (often present-day and pre-industrial) according to $\lambda = (\Delta N - \Delta F)/\Delta T$ (e.g. Gregory et al., 2002; Otto et al., 2013; Sherwood et al., 2020), or (ii) regression in the differential form $\lambda = d(N - F)/dT$ if the timeseries of N, F and T are known (Gregory et al. 2004; Gregory et al. 2020).

Until recently it was often assumed that λ was - to a good approximation - a constant property of the climate system, such that feedbacks that applied over the historical record also applied to the Earth's long-term response, as quantified by the canonical equilibrium climate sensitivity (ECS, units K) to a forcing from a doubling of CO_2 (F_{2x}) over pre-industrial levels. Thus ECS was estimated directly from historical changes in N, T and F, according to ECS = $-F_{2x} / \lambda = - F_{2x}\, \Delta T / (\Delta N - \Delta F)$ (e.g. Gregory et al, 2002; Otto et al., 2013, amongst many others).

However, it is now recognised that λ varies in time since a forcing is applied and with the strength and/or type of that forcing (e.g. Senior and Mitchell, 2000; Hansen et al., 2005; Andrews et al. 2012; Armour et al., 2013; Geoffroy et al., 2013; Rose et al. 2014; Gregory et al. 2015; Andrews et al. 2015; Marvel et al. 2016; Rugenstein et al. 2016; Richardson et al., 2019; Dong et al. 2020; Bloch-Johnson et al., 2021; Rugenstein and Armour, 2021). Hence λ is an 'effective feedback parameter' that applies only to the climate change over which it was calculated. More specifically, over the historical record λ is thought to be more stabilizing (more negative, climate sensitivity smaller) than might operate in the long-term future, and so λ estimated from historical climate change would understate ECS (e.g. Gregory and Andrews, 2016; Zhou et al., 2016; Armour, 2017; Proistosescu & Huybers, 2017; Andrews et al., 2018; Marvel et al., 2018; Silvers et al., 2018; Lewis and Curry, 2018; Gregory et al. 2020; Sherwood et al. 2020; Dong et al. 2021).

The reason for the underestimate of long-term ECS is that climate feedbacks setting λ, such as cloud and lapse-rate changes, vary with the pattern of surface warming. Proxy reconstructions of past equilibrium climates and atmosphere-ocean general circulation model (AOGCM) simulations of long-term climate change show an 'ENSO-like' temperature pattern with strong temperature changes in the eastern Pacific as well as the Southern Ocean, whereas observed historical warming shows more pronounced warming in the western equatorial Pacific relative to the tropical mean and cooling in the eastern Pacific and Southern Ocean over recent decades (e.g. Collins et al., 2013; Li et al., 2013; Andrews et al., 2015; Gregory and Andrews, 2016; Zhou et al., 2016; Dong et al., 2019; Sherwood et al., 2020; Rugenstein et al. 2020; Olonscheck et al., 2020; Fueglistaler and Silvers, 2021; Watanabe et al. 2021; Power et al. 2021; Tierney et al. 2019; 2020).

Thus, more-stabilizing feedbacks have occurred over the historical record because enhanced warming in the western Pacific warm pool – a region of deep ascent and convection – results in a stronger negative lapse-rate feedback widely across the tropics due to efficient warming of the free troposphere, which in turn causes increased cloudiness (a negative cloud feedback) in the eastern tropical Pacific due to remotely controlled increased lower tropospheric stability. In contrast, less-stabilizing feedbacks are expected in the future as enhanced warming in the eastern Pacific – characterised by descending air and marine low cloud decks which are capped under a temperature inversion and form over the relatively cool sea-surface-temperatures (SSTs) – results in a positive cloud feedback, without an accompanying negative lapse-rate feedback since the warming is 'trapped' in the boundary layer (e.g. Zhou et al., 2016, Andrews and Webb, 2018, Ceppi and Gregory, 2017; Dong et al. 2019).

The dependence of radiative feedback on the pattern of surface temperature change has been termed a 'pattern effect' (Stevens et al., 2016), which distinguishes it from other feedback variations that might occur for example as a function of the magnitude of ΔT (e.g. Block & Mauritsen, 2013; Caballero and Huber, 2013; Bloch-Johnson et al., 2021). While the term 'pattern effect' could be applied to any change in SST pattern and associated change in radiative feedback, here we will use it to mean (unless explicitly stated) the pattern effect that arises due to the difference in warming pattern between historical climate change and long-term ECS.

Armour (2017) and Andrews et al. (2018) proposed a method to account for the pattern effect in estimates of ECS derived from historical climate changes via a modification of the energy budget approach. Their method requires an estimate of the difference in feedback, $\Delta\lambda$, due to the pattern effect that arises between historical climate change and long-term ECS, so that ECS=- $F_{2x}/(\lambda_{hist} + \Delta\lambda)$, where λ_{hist} is the historical value. Since $\Delta\lambda$ is found to be positive, it increases the best estimate of ECS and substantially lifts the upper uncertainty bound, but has only a small impact on the lower bound (Armour, 2017; Andrews et al., 2018; Sherwood et al. 2020).

One way of defining the pattern effect, $\Delta\lambda$, is to contrast λ_{hist} in an Atmospheric GCM (AGCM) simulation forced by observed historical SST and sea-ice variations (termed an *amip-piForcing* simulation, see Section 2) with λ_{4xCO2} from 150 years of a coupled AOGCM *abrupt-4xCO2* simulation with the same AGCM, so that $\Delta\lambda = \lambda_{4xCO2} - \lambda_{hist}$ (Andrews et al. 2018). Hence our quantification of $\Delta\lambda$ not only depends on λ_{hist} but also on the (somewhat arbitrary) time frame and method used to calculate λ_{4xCO2}. Ideally we would use the feedback parameter directly associated with ECS rather than λ_{4xCO2}, but this is difficult to calculate in AOGCMs due to the millennial timescales required to equilibrate the deep ocean. Hence feedbacks calculated from 150 years of *abrupt-4xCO2* are often used as a surrogate for long-term ECS feedbacks (Andrews et al., 2012). Technically this is still an 'effective feedback parameter' and associated 'effective climate sensitivity' (EffCS), rather than definitive ECS, but in practice it is found to provide a suitable analogue for long-term feedbacks in climate projections (Grose et al., 2018) and ECS (Sherwood et al. 2020), hence the distinction between EffCS and ECS is not considered further (see Rugenstein et al. (2020) and Rugenstein and Armour (2021) for further discussion).

We assume other impacts on λ, such as the nature of the forcing agent – so called 'efficacies' (Hansen et al., 2005; Marvel et al. 2016; Richardson et al., 2019) – primarily occur due to forcing-specific impacts on historical SST patterns that will be included in the historical record, rather than any dependence on the actual forcing agent concentration in the atmosphere (which will be excluded in our design, because forcing levels are fixed at pre-industrial levels in *amip-piForcing*) (Haugstad et al., 2017). On the other hand, *abrupt-4xCO2* experiments contain larger warming than the historical record, so any state dependence on T (e.g. Block & Mauritsen, 2013; Caballero and Huber, 2013; Bloch-Johnson et al., 2021) might erroneously be diagnosed as a pattern effect using our method. Bloch-Johnson et al. (2021) estimated that λ might vary with T by ~ +0.029 W m^{-2} K^{-2} (multi-model-mean) in step CO_2 experiments relative to pre-industrial level temperature feedbacks, but with substantial uncertainty in the both the magnitude and in some cases even the sign of this state dependence (model range −0.14 to 0.109 to W m^{-2} K^{-2}). While this may play some role in our diagnosed Δλ, we assume it to be small since both Gregory and Andrews (2016) and Andrews and Webb (2018) showed that the pattern effect is large in experiments with identical *T* but contrasting historical and *abrupt-4xCO2* SST patterns.

The principal advantage of using *amip-piForcing* simulations in the calculation of the pattern effect is that λ_{hist} will be consistent with the SST patterns that occurred over the historical record. In contrast, one could use AOGCM historical simulations for λ_{hist}, but when AOGCMs are free to simulate their own historical SST patterns they struggle to reproduce the observed recent decadal trends in tropical Pacific SST patterns (Gregory et al. 2020; Fueglistaler and Silvers, 2021; Watanabe et al. 2021; Dong et al., 2021) and the associated magnitude of λ_{hist}, thus underestimating the pattern effect (Gregory et al., 2020; Dong et al. 2021). This AOGCM bias in the pattern effect has important implications, which we return to in the Discussion, but our focus in this manuscript is on the historical pattern effect as simulated by AGCMs *given* the observed SSTs, thus avoiding the issue of AOGCM biases in historical SST patterns. Note that while our focus is on the atmospheric response to a given SST pattern, causality can work in both directions. For example cloud feedback has been shown to have an impact on the pattern of tropical Pacific SST changes in models (Chalmers et al., 2022).

amip-piForcing simulations also show multi-decadal variations in λ_{hist} (Gregory and Andrews 2016; Zhou et al., 2016; Andrews et al., 2018; Fueglistaler and Silvers, 2021; Dong et al. 2021). In particular λ_{hist} is generally most negative (pattern effect largest) over the most recent decades. This is because variations in atmospheric feedback are well explained by changes in SSTs in regions of tropical deep convection relative to the tropical-mean (Fueglistaler and Silvers, 2021) or global-mean (Dong et al. 2019). Since the late 1970s, regions of deep convection have warmed by about +50% more than the tropical-mean (Fueglistaler and Silvers, 2021), and the eastern Pacific has cooled despite temperatures increasing globally (e.g. Hartmann et al. 2013; Power et al. 2021; and see our Figures 4 and 9). Hence under this configuration of tropical Pacific SST change, we would expect negative feedback from the mechanisms described above (e.g. Zhou et al., 2016, Andrews and Webb, 2018, Ceppi and Gregory, 2017; Dong et al. 2019).

A limitation of the *amip-piForcing* experiment for quantifying λ_{hist} is that it may include a structural dependence on the underlying SST patterns and sea-ice in the Atmospheric Model Intercomparison Project (AMIP) II boundary condition data set (Gates et al., 1999; Hurrell et al., 2008; Taylor et al., 2000) used to force the *amip-piForcing* simulations (Andrews et al., 2018; Lewis and Mauritsen, 2021; Zhou et al., 2021; Fueglistaler and Silvers, 2021). Different SST reconstructions have slightly different patterns of SST change over the historical period, and λ_{hist} may be affected. Indeed Lewis and Mauritsen (2021) and Fueglistaler and Silvers (2021) showed that warming in the tropical western Pacific relative to the tropical-mean is less pronounced in other SST datasets, and so we might expect less negative feedbacks (Δλ less positive) if the AGCMs were forced with non-AMIP II datasets.

Consistent with this expectation, Andrews et al. (2018) noted that in one AGCM the magnitude of λ_{hist} was reduced by ~ 0.2 W m^{-2} K^{-1} when the AMIP II SSTs were replaced by HadISST2.1 SSTs (sea-ice remaining unchanged) in an *amip-piForcing* simulation. Partly because of this, Sherwood et al. (2020) and Forster et al. (2021) assessed the historical pattern effect to be smaller and more uncertain (Δλ = 0.5 ± 0.5 W m^{-2}) than simply taking the *amip-piForcing* based model distribution reported by Andrews et al. (2018) (Δλ = 0.64 ± 0.40 W m^{-2}). Subsequently, Lewis and Mauritsen (2021) and Zhou et al. (2021) also found λ_{hist} to be less negative (Δλ smaller) when using other SST datasets than AMIP II used in *amip-piForcing* simulations discussed here.

1.2. Aims and motivating questions

Andrews et al. (2018) provides much of the published quantitative analysis on λ_{hist} to observed SST patterns and $\Delta\lambda$, but only six AGCMs from only four different modelling centres were considered. Hence, a first motivation of this manuscript is to revisit their numbers with a broader set of models by utilizing the new *amip-piForcing* simulations from the Cloud Feedback Model Intercomparison Project phase 3 (CFMIP, Webb et al. 2017) contribution to the Coupled Model Intercomparison Project phase 6 (CMIP6, Eyring et al., 2016). The larger ensemble totalling 14 models when combined will provide a more robust quantification of the magnitude and spread of λ_{hist} and $\Delta\lambda$ to a broader set of model physics and climate sensitivities (Zelinka et al. 2020; Meehl et al. 2020; Flynn and Mauritsen, 2020).

Secondly, the limited set of models in Andrews et al. (2018) prevented them from robustly exploring and quantifying the relationship between λ_{hist} and λ_{4xCO2} across models. In other words, it is not known whether feedbacks acting over the historical record in AGCMs are correlated to feedbacks acting on long-term ECS. For example is there a relationship between the two that could form the basis of an emergent constraint? Do different parts of the historical record relate better to feedbacks acting on long-term ECS than other parts, and why? As we will show, feedbacks over different parts of the historical record have different relationships to λ_{4xCO2}, and this is important for understanding what can and cannot be directly constrained from the historical record.

Thirdly, λ_{hist} and $\Delta\lambda$ have been shown to vary substantially on decadal timescales with λ_{hist} being most negative (pattern effect largest) over recent decades since ~1980 (Gregory and Andrews 2016; Zhou et al., 2016; Andrews et al., 2018; Gregory et al. 2020; Dong et al. 2021). This is consistent with the findings of Fueglistaler and Silvers (2021), who identified ~1980 as the point in which the Earth begins to warm with a particular (even "*peculiar*") configuration of tropical Pacific SSTs where "*regions of deep convection warm about +50% more than the tropical average*" driving large negative cloud feedbacks. Hence we are motivated to separate λ_{hist} and $\Delta\lambda$ into a 'before' and 'after' 1980. This separation leads into our next motivating question.

Fourthly, are observations of recent decadal warming and TOA radiative fluxes since the 1980s in agreement with the strongly negative λ values simulated by the AGCMs? If so, what would such a strongly stabilizing feedback parameter (and large pattern effect) in the presence of a substantial rate of observed global warming (~0.19 K dec^{-1}, Tokarska et al., 2020) imply for the efficiency of ocean heat uptake and is there any relationship between them? Are any of these relationships affected by the most recent data in which Loeb et al. (2020; 2021) identified a marked change in the Earth's radiation budget associated with the 2015/2016 El Niño event and a change in sign in the Pacific Decadal Oscillation (PDO) index. Such a shift in tropical Pacific SST patterns (a shift to warming in the eastern Pacific) should favour more positive feedbacks (Loeb et al., 2020).

Finally and fifthly, a limitation of the *amip-piForcing* approach, as discussed in Section 1.1, is that λ_{hist} and $\Delta\lambda$ derived from these experiments includes a structural dependence on the SST patterns and sea-ice in the AMIP II boundary condition data set used to force the AGCMs (Andrews et al., 2018; Lewis and Mauritsen, 2021; Zhou et al., 2021; Fueglistaler and Silvers, 2021). To investigate this further, we supplement the *amip-piForcing* simulations with sensitivity tests with nine AGCMs forced with historical HadISST1 (Rayner et al., 2003) SSTs as per Lewis and Mauritsen (2021).

In summary, previous studies have shown that historical climate feedback (λ_{hist}) varies on decadal timescales in *amip-piForcing* simulations and is larger in magnitude (climate sensitivity smaller) than that seen in long-term *abrupt-4xCO2* simulations associated with ECS, giving rise to a 'pattern effect'. This is accentuated over recent decadal climate change. Here we make use of observations of the Earth's energy budget from 1985 and a new suite of *amip-piForcing* simulations from CFMIP3/CMIP6 (giving us a combined ensemble of 14 models), as well as targeted HadISST1 versus AMIP II SST dataset sensitivity tests with nine AGCMs, to address the above questions.

The manuscript is organised as follows: Section 2 describes the model and observational data. Section 3 presents the model results. Section 4 brings in the observational data. Section 5 presents a summary, discussion and outlook.

2. Methods and Data

2.1 amip-piForcing

To provide estimates of λ_{hist} consistent with the observed variations in SST patterns we turn to AGCMs forced with observed monthly variations in SSTs and sea-ice, while keeping all forcing agents such as greenhouse gases and aerosols etc. constant at pre-industrial levels. Since the radiative forcing is constant (ΔF=dF=0) by construction, λ_{hist} can be diagnosed via λ_{hist} = dN/dT (or $\Delta N/\Delta T$ if using finite differences between climate states) (Andrews, 2014; Gregory and Andrews, 2016, Zhou et al., 2016; Silvers et al., 2018; Andrews et al., 2018). Such an experimental design is now referred to as *amip-piForcing* (Gregory and Andrews, 2016). The experimental protocol builds on the Atmospheric Model Intercomparison Project (AMIP) design (Gates et al. 1999) that has long been used in climate modelling, but extends back to 1870 (rather than 1979 in AMIP) and forcing agents are kept at pre-industrial levels. As per AMIP, the underlying SST and sea-ice dataset used to force the AGCMs is the AMIP II boundary condition data set (Gates et al., 1999; Hurrell et al., 2008; Taylor et al., 2000). A description of the *amip-piForcing* protocol for CFMIP3/CMIP6 is given in Webb et al. (2017). When forced with observed monthly SSTs and sea-ice, AGCMs generally reproduce the observed relationships between surface temperature patterns, cloudiness and radiative fluxes well (Allan et al., 2014; Loeb et al. 2020), lending some credibility to the radiative effects of their simulated pattern effects to different SST patterns.

The *amip-piForcing* simulations used in this study are summarised in Table 1. They reflect a combination of new CFMIP3/CMIP6 simulations with the latest generation of models archived in the CMIP6 database and those used in Andrews et al. (2018) with some updates (see below). The exception is MPI-ESM1-2-LR (Mauritsen et al., 2019); this is a CMIP6 generation model but its *amip-piForcing* simulation is not currently included in the CMIP6 database. Note that this model contains the ECHAM6.3 atmospheric model, so the results ought to be very similar to the older ECHAM6.3 simulations used in Andrews et al. (2018) and Lewis and Mauritsen (2021), though the models are not identical owing to differences in atmospheric composition and land surface properties (see Mauritsen et al., 2019, regarding the transition from MPI-ESM1.1 to MPI-ESM1.2). Furthermore, the newer MPI-ESM1-2-LR simulations include a longer time-period than the ECHAM6.3 simulations (Table 1).

The CFMIP3/CMIP6 *amip-piForcing* simulations begin in year 1870, but we discard the first year to be consistent with the earlier Andrews et al. (2018) ensemble which started in January 1871. The CFMIP3/CMIP6 simulations end in Dec 2014, whereas the simulations in the original Andrews et al. (2018) ensemble (largely) ended in Dec 2010. In part to address this, some of the Andrews et al. (2018) simulations have been rerun, including CAM4, GFDL-AM3 and GFDL-AM4 simulations, which now end in Dec 2014 or later (see Table 1). Another difference to Andrews et al. (2018) is that we now have an *abrupt-4xCO2* AOGCM simulation with GFDL-AM4 which they did not consider, to permit a quantification of the pattern effect in that model. In contrast, we exclude the Andrews et al. (2018) CAM5.3 simulation from our analysis since there is no *abrupt-4xCO2* AOGCM simulation to compare against.

The models used, time-periods covered and number of ensembles are detailed in Table 1. Where ensembles exist, an ensemble-mean dT and dN is created before analysis. Note that it makes little difference to λ if, alternatively, individual members are first analysed and then the results ensemble-meaned (Gregory and Andrews, 2016; Lewis and Mauritsen, 2021). All models share a common 1871-2010 time-period and so the principal analysis is restricted to this time-period, but we consider the additional years to 2014 too. All data are global-annual-ensemble-means and expressed as anomalies relative to an 1871-1900 baseline and the timeseries data has been made available (see Data Availability Section).

Unless otherwise stated all uncertainties in multi model ensemble-mean results represent a 5-95% confidence interval, calculated as 1.645σ across models assuming a gaussian distribution. We do not attempt to adjust our uncertainty for the number of independent models, *n*, used in the ensemble (i.e. dividing by square root of *n*). Our approach is similar to a "statistical indistinguishable ensemble" approach (Annan and Hargraves, 2011; 2017) though likely overstates the uncertainty in the true value if the ensemble shares characteristics of a "truth centred paradigm" (Sanderson and Knutti, 2012).

2.2 HadSST-piForcing

To test the sensitivity of the *amip-piForcing* results to the underlying SST dataset, we repeat the *amip-piForcing* simulations with nine AGCMs (see Table 1) but replace the AMIP II boundary condition SST dataset with HadISST1 (Rayner et al. 2003). All other aspects of the simulations, including sea-ice, are identical to the *amip-piForcing* simulations. This is the same experimental design as Lewis and Mauritsen (2021), and we include their ECHAM6.3 simulations here (which again ought to be similar to the MPI-ESM1-2-LR simulations). The simulations cover a common time-period across models of 1871-2010, like in *amip-piForcing*, but some models are also extended further (see Table 1). We refer to these simulations as *hadSST-piForcing*, but note only the SSTs are from the HadISST1 dataset (hence 'hadSST' rather than 'hadISST'), the sea-ice remains as per *amip-piForcing*. Like *amip-piForcing*, all data are global-annual-ensemble-means and expressed as anomalies relative to an 1871-1900 baseline, and the timeseries data has been made available (see Data Availability Section).

Lewis and Mauritsen (2021) provide a summary of the source observational inputs used to construct the AMIP II and HadISST1 SST datasets and how they differ. In addition, we note that AMIP II uses HadISST1 SSTs (Rayner et al. 2003) prior to November 1981 and version 2 of the National Oceanic and Atmospheric Administration (NOAA) weekly optimum interpolation (OI.v2) SST analysis (Reynolds et al. 2002) thereafter. The merging procedure rebases the HadISST1 SSTs to avoid discontinuities in the merged dataset (Hurrell et al. 2008). Hence AMIP II and HadISST1 might be expected to be more similar before 1981, and diverge afterwards.

2.3 abrupt-4xCO2

A corresponding *abrupt-4xCO2* simulation using each AGCM's coupled AOGCM is used to determine the model's long-term sensitivity metrics (F_{4x}, λ_{4xCO2} and ECS = -0.5*F_{4x}/ λ_{4xCO2}) from regression of global-annual-mean dN against dT over 150 years of the simulations (see Andrews et al., 2012). We also use λ_{4xCO2} diagnosed from years 1-20 and years 21-150 of the *abrupt-4xCO2* simulation following Andrews et al. (2015), which approximately separates the two principal timescales of the climate response: the mixed-layer and deep-ocean (see Geoffroy et al. 2013 and Andrews et al. 2015). *abrupt-4xCO2 data* is available on the CMIP5 database (Taylor et al., 2012) for CCSM4, GFDL-CM3 and HadGEM2-ES. All other abrupt-4xCO2 data is available on the CMIP6 database (Eyring et al., 2016), except for HadCM3 and MPI-ESM1.1. For ECHAM6.3/MPI-ESM1.1, *abrupt-4xCO2* global-annual mean d*N* and d*T* timeseries data are provided by Andrews et al. (2018). HadAM3 data is taken from Andrews et al. (2018) and Andrews et al. (2015); while a mean of seven realizations, this simulation is only 100 years long so the calculations are over years 1-100 for λ_{4xCO2} and years 1-20 or 21-100 for the separation of timescales in this model.

Note when aligning each AGCM to its AOGCM, sometimes the AGCM and AOGCM model names differ in the literature. We indicate where this is applicable in Table 1. This does not apply to the newer CFMIP3/CMIP6 simulations which publish their AGCM and AOGCM simulations under consistent names.

2.4 Observations of recent decadal climate change

To understand Earth's recent decadal climate change since ~1985 we turn to its observed global-mean energy budget (i.e. d*T*, d*N* and d*F*). For d*T* we use the HadCRUT5 analysis dataset (Morice et al. 2021) (the current version is HadCRUT.5.0.1.0). This is an improvement on previous HadCRUT products and extends coverage in data sparse regions (see Morice et al. 2021). For d*F* we use the best estimate historical ERF timeseries produced by IPCC AR6 (Forster et al. 2021; Smith et al. 2021). For d*N* we use various versions of the DEEP-C satellite based reconstruction of the Earth's radiation balance from 1985 to near-present. These are described in detail in Allan et al. (2014) and Liu et al. (2015; 2017; 2020), but as we will use various versions of this product we give a brief overview here.

The DEEP-C dataset is derived by merging satellite observations of top-of-atmosphere radiative flux timeseries from ERBE WFOV (Earth Radiation Budget Experiment Satellite wide field of view) and ECMWF reanalysis (ERA-Interim/ERA5) since 1985 with CERES (Clouds and the Earth's Radiant Energy System) satellite observed fluxes since March 2000. Hence prior to March 2000 it is largely informed by ERBE WFOV and ERA reanalysis, then aligns with CERES from March 2000. AMIP and high resolution AGCM simulations and reanalyses are used in the merging process to bridge the gaps between products and avoid discontinuities in the timeseries, including a gap in the satellite record during 1993 and 1999 (Allan et al. 2014). It is important to note that substantial uncertainty in decadal changes in d*N* associated with the merging process affects the record and this is conservatively estimated to be as high as 0.5 Wm^{-2} for changes applying across the whole record (Liu et al. 2020). However, uncertainty in the CERES period since March 2000 is much smaller based on assessment of instrument drift (Loeb et al. 2021). Various versions of the DEEP-C dataset exist which parallel updates to the underlying products and update the merging process. We use the latest version (DEEP-C v5, Liu and Allan 2022) for our principal analysis, which is based on CERES EBAF v4.1 and ERBS WFOV v3, alongside ERA5 reanalysis and AMIP6 simulations (Liu and Allan, 2022). To illustrate structural uncertainties in our analysis we also use previous versions (v2, v3 and v4) of the DEEP-C datasets. The availability of datasets is provided in the Data Availability Section.

3. Historical feedback and pattern effect in amip-piForcing and hadSST-piForcing simulations

Figure 1a shows the multi-model ensemble mean d*T* timeseries in the *amip-piForcing* and *hadSST-piForcing* simulations, alongside an observed estimate from HadCRUT5 analysis dataset. The AGCM design reproduces the observed historical d*T* variability well (the correlation coefficient, *r*, between observed and both simulated d*T* timeseries is 0.97). However the AGCMs do not reproduce the observed trends precisely, notably omitting some observed warming particularly in the most recent decades (Figure 1a). This is because the AGCM design omits a small component of warming associated with land surface temperature change (which is not prescribed in the models) that arises as a direct consequence of increases in greenhouse gases and other forcing agents independent of SST change (this is often considered as part of the ERF rather than feedback) (see Andrews, 2014; Gregory and Andrews, 2016; Andrews et al., 2018). This will be included in the observed record but not in the simulated dT because greenhouse gases and other forcing agents are kept constant at pre-industrial levels in *amip-piForcing* and *hadSST-piForcing*.

As dT increases, dN reduces (Figure 1b), i.e. the climate loses more heat to space as a consequence of the climate response and feedbacks in the system. Figure 1c and 1d show the difference in the dT and dN timeseries between the *amip-piForcing* and *hadSST-piForcing* ensemble-mean response. For most of the time the differences vary approximately about zero. However, larger differences are evident from 1981 onwards, when the dN response in *amip-piForcing* is substantially larger than that in *hadSST-piForcing* (Figure 1b and 1d), up to ~0.5 W m^{-2} in some years (Figure 1d). This is consistent with 1981 being the year in which the AMIPII boundary condition source dataset switches from HadISST1 to OI.v2 SST (see Section 3.2).This motivates us to separate the historical record into two time-periods either side of 1980, i.e. 1871-1980 and 1981-2010 (Section 3.2).

However, we first consider feedback and the pattern effect that arises when calculated over the historical record as a whole, rather than any time-period within. This is useful for informing studies that use the entire observed historical record to estimate ECS via energy budget constraints (e.g. Andrews et al., 2018; Sherwood et al. 2020; Forster et al. 2021). It also allows a direct comparison of our results using a broad ensemble of models to the narrower range of model results reported by Andrews et al. (2018) and Lewis and Mauritsen (2021).

3.1 Considering the historical record as a whole

Figures 1e and 1f show the λ_{hist} = dN/dT relationship in the ensemble-mean *amip-piForcing* and *hadSST-piForcing* simulation for 1871-2010. λ_{hist} is determined from ordinary least square linear regression on global-annual-mean dN and dT timeseries data. λ_{hist} values for individual models are given in Table 2 alongside their *abrupt-4xCO2* sensitivity metrics. Across the fourteen model ensemble of *amip-piForcing* simulations λ_{hist} = -1.65 ± 0.46 W m^{-2} K^{-1}, slightly smaller in magnitude but with similar spread to the Andrews et al. (2018) ensemble (they reported λ_{hist} = -1.74 ± 0.48 W m^{-2} K^{-1}). Like in Andrews et al. (2018), the spread in λ_{hist} is extremely similar to the spread in λ_{4xCO2} from the coupled AOGCM *abrupt-4xCO$_2$* ensemble (Table 2) (this is also true for the individual feedback terms, see below). The pattern effect, $\Delta\lambda = \lambda_{4xCO2} - \lambda_{hist}$ between *amip-piForcing* and *abrupt-4xCO2* (with λ_{4xCO2} from years 1-150 of abrupt-4xCO2) is $\Delta\lambda$ = 0.70 ± 0.47 W m^{-2} K^{-1} across the ensemble (Table 3), which is slightly larger in magnitude but with more spread than that reported by Andrews et al. (2018) (0.64 ± 0.40 W m^{-2} K^{-1}).

Tables 2 and 3 also present the equivalent λ_{hist} and $\Delta\lambda$ values when the AGCMs are forced with HadISST1 SSTs instead (*hadSST-piForcing*) and Figure 2 shows the relationship to *amip-piForcing*. λ_{hist} = -1.43 ± 0.41 W m^{-2} K^{-1} in *hadSST-piForcing* (Table 2), which is smaller in magnitude but with similar spread to the *amip-piForcing* results above. Subsetting to the nine AGCMs with both simulations, λ_{hist} is 0.28 ± 0.17 W m^{-2} K^{-1} smaller in magnitude in *hadSST-piForcing* but well correlated (r=0.93) with *amip-piForcing* values (Figure 2a, red points). The regression slopes of the red line in Figures 2a (slope = 0.84 ± 0.21) and 2b (slope = 0.84 ± 0.26) are statistically consistent with unity, implying there is little AGCM dependence in the difference between λ_{hist} from *amip-piForcing* and *hadSST-piForcing*. Hence, given the strong correlation and close approximation of being parallel to the one-to-one line (Figure 2, red points), we suggest a simple offset given by the difference (0.28 ± 0.17 W m^{-2} K^{-1}, Table 3) well approximates the relationship between λ_{hist} over 1871-2010 in *amip-piForcing* and *hadSST-piForcing*.

Despite λ_{hist} being smaller in magnitude in *hadSST-piForcing*, $\Delta\lambda$ = 0.48 ± 0.36 W m^{-2} K^{-1} is still large and positive across the *hadSST-piForcing* ensemble (Table 3). The smaller uncertainty than the *amip-piForcing* pattern effect likely reflects the narrower diversity of model physics in the smaller *hadSST-piForcing* ensemble, for example we do not have *hadSST-piForcing* experiments for the model (MIROC6) with the smallest pattern effect in *amip-piForcing*. If we subset the *amip-piForcing* ensemble to just those nine models with corresponding *hadSST-piForcing* experiments (Fig 2b, red points), then the spread (as measured by 1.645σ) across models in $\Delta\lambda$ reduces from 0.47 to 0.38, which is similar to the spread found in *hadSST-piForcing*.

That a large pattern effect is present in the *hadSST-piForcing* simulation over the historical record is not in contradiction with the results of Lewis and Mauritsen 2021 (LM2021), who reported a '*negligible unforced historical pattern effect*' with ECHAM6.3 when forced with HadISST1 SSTs. This is because LM2021 calculated their pattern effect by comparing λ from *hadSST-piForcing* to λ derived from a coupled AOGCM historical simulation, or approximations of it from years 1-70 of *1%CO2* or years 1-50 of *abrupt-4xCO2* simulations. This necessarily gives a smaller pattern effect because it excludes many of the SST variations and patterns effects seen on longer timescales in CO_2 forced simulations (Senior and Mitchell, 2000; Gregory et al. 2004; Andrews et al. 2012; Armour et al., 2013; Geoffroy et al., 2013; Andrews et al. 2015; Rugenstein et al. 2016). While this might be useful for trying to quantify different mechanisms of the pattern effect (e.g. forced or unforced, see Dessler, 2020), it is a quantity we are less interested in, as we want to know the λ of relevance to long-term ECS and projections of the late 21st century. Therefore contrasting to λ_{4xCO2} from years 1-150 is the most relevant metric (Sherwood et al., 2020), as we have done here.

Following Andrews et al. (2018) we decompose λ into its component longwave (LW) clear-sky, shortwave (SW) clear-sky and cloud radiative effect (CRE, equal to all-sky minus clear-sky fluxes) terms in Figure 3. Deviations away from the one-to-one line indicate a difference in *amip-piForcing* and *abrupt-4xCO2* λ (i.e. the pattern effect). Tables of the individual model results are given in the Supplementary Tables 1 - 3. It confirms the basic premise that historical LW clear-sky and cloud feedbacks are more stabilizing than under *abrupt-4xCO$_2$*, consistent with the mechanistic and process understanding that the pattern effect arises predominantly from a lapse-rate (which affects LW clear-sky fluxes) and cloud feedback dependence on SST patterns (e.g. Zhou et al., 2016, Andrews and Webb, 2018, Ceppi and Gregory, 2017; Dong et al. 2019). Figure 3 and Supplementary Tables 1 - 3 show that the inter-model spread in feedback in both *amip-piForcing* and *abrupt-4xCO2* is dominated by cloud rather than clear-sky feedbacks. Figure 3 also suggests there is a small compensation to the total pattern effect from SW clear-sky feedbacks, likely from sea-ice. That is, AGCMs forced with AMIP II boundary condition sea-ice changes have a slightly more positive feedback than found in their coupled *abrupt-4xCO2* simulations, though the difference is small (Figure 3). Consequently, a simple attribution of the difference in total feedback between *amip-piForcing* and *abrupt-4xCO2* to an SST driven pattern effect (as we have done here) will slightly understate the actual effect, though the term is small and we neglect it from now on. We discuss sea-ice uncertainties further below.

MIROC6 is the only model in the *amip-piForcing* ensemble to have near zero pattern effect (Table 3 and note the single black dot on the one-to-one line in Figure 3). The reason for this different behaviour remains unclear. One could speculate that there is a relationship between a model's climate sensitivity and its pattern effect, given that MIROC6 has the lowest ECS of all models considered here (ECS=2.6K, Table 2). However, we note that there is little correlation between ECS and Δλ across models (*r*=0.4) and that several other models with low ECS have large Δλ.

Alternatively, it could be that MIROC6's atmospheric physics are largely insensitive to different SST patterns and/or that its AOGCM *abrupt-4xCO2* warming pattern is more similar to the historical record than other models. Both are potentially possible. For example, λ_{hist} for 1871-1980 and 1980-2010 separately (next Section and Table 2) shows that MIROC6 does simulate a pattern effect, but achieves a near zero pattern effect over the historical record as a whole by having a smaller (relative to other models) pattern effect over recent decades, offset by a negative pattern effect over the earlier period. In addition - and in contrast to other models - MIROC6 simulates a negative LW clear-sky pattern effect (red dot below the one-to-one line, Figure 3) which offsets its positive cloud feedback pattern effect.

The model with the largest pattern effect is CESM2 (Table 3). This occurs because of a particularly large cloud feedback sensitivity to SST patterns (grey dot furthest from the one-to-one line, Figure 3). Zhu et al. (2022) argue that an issue in CESM2's cloud microphysics related to cloud ice number leads to an unrealistically large cloud sensitivity to warming in this model. Whether this is responsible for the model's large pattern effect is unclear. Mixed-phase clouds have not typically been associated with the pattern effect, though might be of relevance to pattern effects over the Southern Ocean (Dong et al. 2020; Bjordal et al. 2020). It would be interesting in future work to identify the different cloud types associated with the pattern effect and conduct sensitivity experiments with CESM2 to investigate which aspects of the cloud feedback change with different cloud microphysics schemes.

Many of our *amip-piForcing* simulations (eleven models) continue to Dec 2014 (Table 1), and six have corresponding *hadSST-piForcing* simulation, so we consider how this extended period affects the overall assessment of the historical pattern effect. In the eleven *amip-piForcing* simulations, λ_{hist} = -1.65 ± 0.48 W m^{-2} K^{-1} over 1871-2010, but this increases in magnitude so that λ_{hist} = -1.71 ± 0.51 W m^{-2} K^{-1} if calculated over 1871-2014 (Supplementary Table 4). An increase occurs in every model and the magnitude of change across the ensemble is 0.07 ± 0.06 W m^{-2} K^{-1} (Supplementary Table 4). In the six corresponding *hadSST-piForcing* simulations, λ_{hist} = -1.48 ± 0.41 W m^{-2} K^{-1} over 1871-2010, but this increases in magnitude so that λ_{hist} = -1.53 ± 0.39 W m^{-2} K^{-1} if calculated over 1871-2014 (Supplementary Table 4). The magnitude of the increase (0.05 ± 0.05 W m^{-2} K^{-1}) is thus slightly smaller in this dataset (Supplementary Table 4).

While we have focused on the SST driven pattern effect, a remaining structural uncertainty in assessing total feedback differences between λ_{4xCO2} and λ_{hist} relates to the sea-ice dataset used to force the AGCMs. Andrews et al. (2018) provided a sensitivity test (see their Supplementary Material) by repeating the *amip-piForcing* simulation in two AGCMs but forced with HadISST2.1 (Titchner and Rayner, 2014) SSTs and sea-ice. They found that the historical feedback parameter increased by ~0.6 W m^{-2} K^{-1} when forced with HadISST2.1 compared to AMIP II, and attributed most of this change to differences in the sea-ice datasets rather than SST. They noted that HadISST2.1 has substantially more pre-industrial Antarctic sea-ice concentration (see Titchner and Rayner, 2014), and so generated more sea-ice loss (more positive feedback) over the historical period (Andrews et al. 2018), as well containing large discontinuities in the timeseries. The historical sea-ice trends and associated feedbacks over the Southern Ocean in the HadISST2.1 dataset are difficult to reconcile with those found in AOGCMs and our physical understanding of them (Schneider et al. 2018). We do not pursue this further, but simply highlight that dataset assumptions made about pre-industrial sea-ice concentrations in Antarctica can have substantial impacts on diagnosed feedbacks in AGCMs and remains an outstanding uncertainty in assessing total feedback differences. Fortunately, in *amip-*

piForcing the difference in SW clear-sky feedback (which will be strongly impacted on by sea-ice feedbacks) is similar to that seen in λ_{4xCO2} (Figure 3) so this can be ignored if the focus is solely on SST driven feedbacks in the atmosphere.

In summary, for warming since the 1800s (using either 1871-2010 or 1871-2014), both *amip-piForcing* and *hadSST-piForcing* suggest a substantial pattern effect between radiative feedbacks operating over historical climate change and long-term ECS.

3.2 *Considering the historical record before and after 1980*

We now return to the divergence in d*N* response between *amip-piForcing* and *hadSST-piForcing* simulations around 1980 (Figure 1d). As well as the change in behaviour discussed above, 1980 provides a convenient separation of historical feedbacks and the pattern effect for two other motivating reasons: (i) Fueglistaler and Silvers (2021) identify ~1980 as the point in which the Earth begins to warm with a particular configuration of tropical Pacific SSTs where regions of deep convection warm substantially more than the tropical mean, driving large negative cloud feedbacks and consistent with a large pattern effect over this period (Gregory and Andrews 2016; Zhou et al., 2016; Andrews et al., 2018; Gregory et al. 2020); and (ii) Fueglistaler and Silvers (2021) also identify ~1980 as a useful approximation of when the satellite era was integrated into the global observing system, and so developing an understanding of feedbacks and the pattern effect specifically from 1980 onwards will aid interpretation of our most comprehensive observations of climate change and how they might relate to the future change (next Section).

Figure 4 compares the surface temperature pattern over the two time-periods 1871-1980 and 1981-2010 in *amip-piForcing* and *hadSST-piForcing*. Differences between the two SST reconstructions are extremely subtle. For the earlier 1871-1980 time period, warming is more uniform, in part because of the longer time-period considered which will smooth out variability. Since 1981 there has been western Pacific warming with cooling in the Southern Ocean and off equatorial eastern Pacific (which are regions of marine low clouds), despite temperatures increasing in the global mean. Hence, we might expect a small pattern effect prior to 1980 and a large pattern effect post 1980 (e.g. Gregory and Andrews, 2016; Zhou et al., 2016, Andrews and Webb, 2018, Ceppi and Gregory, 2017; Dong et al. 2019, Fueglistaler and Silvers 2021).

Figures 1g and 1h show the $\lambda_{hist} = dN/dT$ relationship in the ensemble-mean *amip-piForcing* and *hadSST-piForcing* simulation for 1871-1980 (grey points) and 1981-2010 (blue points). Results for individual models are given in Table 2. Figures 1g and 1h confirms the basic premise that λ_{hist} strengthens in magnitude post 1980, consistent with the change in SST patterns (Figure 4).

For the earlier time-period, 1871-1980, $\lambda_{hist} = -1.14 \pm 0.33$ W m^{-2} K^{-1} in *amip-piForcing* is similar to $\lambda_{hist} = -1.21 \pm 0.38$ W m^{-2} K^{-1} in *hadSST-piForcing* (Table 2) – suggesting little sensitivity of the results to these two SST datasets over this time period. This is unsurprising given that the datasets are similar (though not identical) prior to this period (Section 2.2 and Figure 4). For the nine AGCMs that performed both simulations Figure 2a shows the relationship between λ_{hist} in *amip-piForing* and *hadSST-piForcing*. For all time-periods λ_{hist} in *amip-piForcing* and *hadSST-piForcing* is found to be well correlated ($r \geq 0.87$, Figure 2a). For the earlier 1871-1980 results, the λ_{hist} values fall close to the one-to-one line (blue dots, Figure 2) and within the range of λ_{4xCO2} (grey shaded areas in Figure 2). This suggests that for 1871-1980 λ_{hist} is broadly independent of the two SST datasets (consistent with their common basis) and that the pattern effect is small for this time period. Indeed, the 1871-1980 pattern effect is small but positive ($\Delta\lambda = 0.19 \pm 0.35$ W m^{-2} K^{-1} in *amip-piForcing* and 0.26 ± 0.26 W m^{-2} K^{-1} in *hadSST-piForcing*, Table 3 and Figure 2b).

In contrast, for 1981 onwards (i.e. 1981-2010), λ_{hist} is generally far from the λ_{4xCO2} range (i.e. a large pattern effect) and away from the one-to-one line (i.e. a dependence on the SST dataset) (Figure 2a; grey points). Indeed, λ_{hist} over 1981-2010 is substantially stronger in magnitude than over 1871-1980 (λ_{hist} = -2.33 ± 0.72 W m^{-2} K^{-1} in *amip-piForcing* over 1981-2010, Table 2; Figure 2a) and the pattern effect is large ($\Delta\lambda$ = 1.38 ± 0.75 W m^{-2} K^{-1}, Table 3; Figure 2b), although somewhat weaker in magnitude in *hadSST-piForcing* ($\Delta\lambda$ = 1.24 ± 0.88 W m^{-2} K^{-1}, Table 3; Figure 2b). For 1981-2010, λ_{hist} is generally weaker in *hadSST-piForcing* (Table 2; Figure 3a) by 0.24 ± 0.46 W m^{-2} K^{-1} across the nine AGCMs using both SST datasets.

These results are generally consistent with Fueglistaler and Silvers (2021) and Lewis and Mauritsen (2021) who both point to the AMIP II SST dataset as having larger (relative) western tropical Pacific warming than in other SST datasets, and hence from the process understanding we would expect a more negative feedback (and larger pattern effect) in *amip-piForcing*, as found above. The one exception is GFDL-AM4, which simulates a more negative λ_{hist} under HadISST1 SSTs than AMIP II from 1981-2010, and so a larger pattern-effect over this period under HadISST1 SSTs (Tables 2 and 3 and the single grey dots in Figures 2a and 2b which sit on the other side of the one-to-one line from the other models). The reasons for this remain unclear.

In summary we have shown that a division around 1980 usefully separates historical climate change into two time-periods: (i) pre 1981 the Earth warmed over most of the historical record with an averaged warming pattern that is relatively uniform, and feedbacks largely consistent with long-term ECS feedbacks (i.e. a relatively small pattern effect), and (ii) post 1980 where the Earth warmed with a particular configuration of strong SST gradients that drove feedbacks much more stabilizing than those seen in long-term ECS feedbacks (i.e. large pattern effect), albeit with a sensitivity of the magnitude of this result to the SST dataset considered.

3.3 Relationships between historical and ECS feedbacks

We now consider whether feedbacks over the historical period in *amip-piForcing* are related to λ_{4xCO2}. This is in contrast to the previous sections which only quantified their difference (i.e. the pattern effect).

Firstly, we note that the spread in feedback across models over the earlier (1871-1980) time-period in *amip-piForcing* is well correlated with the spread in feedback across models in *abrupt-4xCO2* (r=0.69, Figure 5a). In contrast, feedbacks over the most recent decades (1981-2010) are only weakly correlated with λ_{4xCO2} (r=0.27). Secondly, feedback over the full historical record (1871-2010) is only weakly correlated with feedback from the 1871-1980 time-period (r=0.45, Figure 5b). In contrast, 1871-2010 feedback is strongly correlated with feedback over the most recent 1980-2010 decades (r=0.91, Figure 5b). This strong correlation between 1981-2010 and the 1871-2010 feedback arises because the spread for 1871-2010 is dominated by the spread for 1981-2010.

Given that the feedbacks applying in 1871-1980 and in 1981-2010 are different, we infer that the SST patterns over these two periods are driven by different mechanisms. Because the feedbacks of 1871-1980 are correlated with *abrupt-4xCO2*, the difference between the two periods could be explained by CO_2 being the dominant influence in 1871-1980 SST patterns, while something else (e.g. perhaps variability, aerosol, volcanism) dominates during 1981-2010. This is only a hypothesis, because these experiments do not provide a way to attribute the observed SST changes to causes.

The result is that the spread in feedbacks over the full historical record are only weakly correlated with λ_{4xCO2} (r=0.51, Figure 3), because of the strong pattern effect post 1980. Hence, we can say little about future λ_{4xCO2} directly from climate change post 1980 or even the full historical record without adjusting for a pattern effect. In contrast, the feedbacks operating over the earlier 1871-1980 time-period are correlated with λ_{4xCO2} (r=0.69, Figure 5a).

That recent decadal feedbacks are the most unrepresentative of the long-term climate sensitivity is unfortunate, not just because it coincides with the advent of the satellite record and so is extremely well observed, but also because climate change since ~1980 ought to provide the best constraint on ECS (e.g. Jiménez-de-la-Cuesta and Mauritsen, 2019). This is because it offers a strong global warming signal, which AOGCMs attribute to greenhouse gas increases, while avoiding the large uncertainty associated with global-mean aerosol radiative forcing in energy budget estimates of ECS. However the role of aerosols should not be discounted entirely, since strong compensating regional changes may have impacted on SST patterns (e.g. Smith et al. 2015; Takahashi & Watanabe, 2016; Moseid et al., 2020). In contrast, although feedbacks operating over the earlier 1871-1980 part of the historical record are correlated with long-term CO_2 induced feedbacks, a reliable observational constraint is harder because the climate change signal is smaller and the observations poorer. We discuss this further in the Discussion section.

Up to now we have only considered a comparison of *amip-piForcing* feedbacks to a single definition of *abrupt-4xCO2* feedbacks (i.e. feedbacks diagnosed over years 1-150 in *abrupt-4xCO2*). Here we briefly consider separating λ_{4xCO2} into the two principal timescales of the *abrupt-4xCO2* response following Andrews et al. (2015) by calculating λ_{4xCO2} over years 1-20 (a fast timescale) and 21-150 (a slow timescale) (Table 2). The rationale is that 20 years is approximately the timescale required for the mixed-layer to equilibrate in response to step forcing, and any subsequent climate response scaling with the slower deep-ocean timescale, as approximated by two-layer models (Held et al., 2010; Geoffroy et al., 2013; Gregory et al., 2015).

Figure 5c shows λ_{hist} from 1871-1980 is largely scattered about the one-to-one line with λ_{4xCO2} from years 1-20, suggesting little to no pattern effect between these two. This is potentially consistent with the historical record largely being the result of the faster timescale responses (Held et al. 2010; Proistosescu & Huybers, 2017). In contrast, post-1980 λ_{hist} is far from the one-to-one line (i.e. large pattern effect to years 1-20 of *abrupt-4xCO2*, Figure 5c) but is marginally correlated (r=0.53), suggesting recent decades do contain some information relevant to the feedback seen in the fast timescale response to CO_2. However, the longer-term feedbacks associated with the slow timescale response to CO_2 (years 21-150 of *abrupt-4xCO2*, Figure 5d) have no correlation with λ_{hist} post-1980 (r=-0.06, Figure 5d).

3.4 Decadal variability in feedbacks and the pattern effect

In this final section of GCM results we briefly comment how λ_{hist} and the pattern effect varies on decadal timescales in the *amip-piForcing* and *hadSST-piForcing* simulations.

Following Gregory and Andrews (2016) we calculate $\lambda_{hist} = dN/dT$ over a moving 30 year window in the *amip-piForcing* and *hadSST-piForcing* simulations (Figure 6a and b). For example λ_{hist} calculated over the 30 year period 1925 to 1954 is presented at year 1939.5 in Figure 6. In Figures 6c-h the LW and SW clear-sky and cloud radiative effect of the feedback are also shown. The correlation coefficient between the *amip-piForcing* and *hadSST-piForcing* multi-model-mean λ_{hist} timeseries is 0.84, suggesting the broad features of the decadal λ_{hist} variations are robust to the SST datasets. In particular λ_{hist} peaks (least negative, smallest pattern effect) around 1940 while generally being large in magnitude (large pattern effect) over recent decades (see also Gregory and Andrews, 2016; Zhou et al. 2016; Andrews et al. 2018; Gregory et al. 2020). The clear sky feedbacks (Figures 6c-f) are largely stable, while the variation in λ_{hist} is almost entirely explained by variation in cloud feedback (Figures 6g-h), consistent with previous findings (e.g. Zhou et al. 2016; Andrews et al. 2018).

In Section 5, we discuss further the reasons for the decadal variations in SST patterns and λ_{hist}, i.e. whether they are the result of spatiotemporal changes in forcings such as aerosols or volcanic forcing or due to unforced variability.

4. Observed climate change

We next consider whether the radiative feedback and pattern effects simulated by the GCMs are consistent with observed variations in the Earth's energy budget. Gregory et al. (2020) asked a similar question for the post 1980 period and suggested they are (see their Figure 5c), but here we go a few steps further. Specifically, not only do we consider the post 1980 period, but also assess changes in the Earth's energy budget back to the 1800s. Furthermore we investigate the implications of a strongly negatively feedback parameter (large pattern effect) since 1985 on the observed rate of global warming.

The observations also provide an opportunity to bring our λ_{hist} and pattern effect estimate up to date with the most recently observed data (up to and including 2019), whereas our GCM analysis generally finished in 2014. The observations post 2014 period are of particular interest given they include the major El-Nino event of 2015/2016 that was associated with eastern-pacific warming and marked changes in the observed radiation budget (Loeb et al. 2020; 2021). We expect these post 2014 years to have an impact λ_{hist} and the pattern effect, given the process understanding discussed previously (e.g. Zhou et al., 2016, Andrews and Webb, 2018, Ceppi and Gregory, 2017; Dong et al. 2019).

4.1 Comparison of AGCM results to observed estimates

We first validate the AGCM λ_{hist} estimates over recent decades. To do this we use a merged satellite dataset (ERBE WFOV + CERES) (Allan et al. 2014) that provides an observational estimate of dN variations from 1985 to 2019. For dT we use the HadCRUT5 analysis dataset (Morice et al. 2021). For dF we use the IPCC AR6 (Forster et al. 2021; Smith et al., 2021) best estimate historical ERF changes. These datasets are described in further detail in Section 2.4. We first consider the 30-year period 1985 to 2014, consistent with many of the AGCMs.

Figure 7a and 7b show the dT, dN and dF timeseries over this period. The 1985-2014 'observed' $-\lambda_{hist} = \mathrm{d}(F - N)/\mathrm{d}T \sim 2.0 \pm 0.7$ W m^{-2} K^{-1} relationship is shown in Figure 7d. Note the stated 5-95% uncertainty is $\pm 1.645\sigma$ from the standard error of the linear fit, with no allowance for systematic uncertainties. As discussed in Section 2.4, observed multi-decadal changes in dN are subject to a substantial uncertainty (up to 0.5 Wm^{-2}) primarily related to the breaks in the record prior to 2000, though are considerably smaller afterwards (Liu et al. 2020). Note also that years 1991-2 are excluded from the calculation as these years are identified as being strongly impacted by the volcanic forcing from the Pinatubo eruption (Figure 7b). Whilst λ_{hist} is robust to this (we get just the same $\lambda_{hist} \sim -2.0 \pm 0.7$ W m^{-2} K^{-1} if we include these years), including these years has an impact on the ocean heat uptake efficiency estimate (see Section 4.3). The observed 1985-2014 λ_{hist} estimate is shown on Figure 6a and 6b (red line) as an illustration in comparison to the AGCM decadal variations in λ_{hist}. The observed λ_{hist} best estimate agrees exceptionally well with the AGCM multi-model mean, and nearly all models are within the 5-95% uncertainty estimate as they approach the 1985-2014 value (Figure 6a and 6b).

A more rigorous comparison of individual AGCM results to the observed estimate is shown in Figure 8. Here the AGCM λ_{hist} estimates from *amip-piForcing* and *hadSST-piForcing* have been calculated in the same way as the observations, i.e. over 1985-2014 excluding 1991-2. The overlap between the model and observed estimates points to broad consistency between the models and observations in the recent decadal value of λ_{hist} (Figure 8). The large uncertainties (which are likely underestimated since we have not accounted for structural errors) inhibit a more precise validation of individual models against the observed estimate.

For the full the historical record we estimate λ_{hist} from IPCC AR6 assessed changes in T, N and F. Forster et al. (2021) give these as $\Delta T = 1.03 \pm 0.20$ K, $\Delta N = 0.59 \pm 0.35$ W m^{-2} and $\Delta F = 2.20$ [1.53 to 2.91] W m^{-2} for the time-period 1850-1900 to 2006-2019. For simplicity we assume $\Delta F = 2.20 \pm 0.7$ W m^{-2}, where we have approximated the uncertainty in ΔF as a Gaussian. Randomly sampling (with replacement) from the Gaussian distributions in ΔN, ΔF and ΔT gives $\lambda_{hist} = (\Delta N - \Delta F)/\Delta T = -1.6 \pm 0.8$ W m^{-2} K^{-1}. This is again in agreement with the *amip-piForcing* ($\lambda_{hist} = -1.65 \pm 0.46$ W m^{-2} K^{-1}, Table 2) and *hadSST-piForcing* ($\lambda_{hist} = -1.43 \pm 0.41$ W m^{-2} K^{-1}, Table 2) 1871-2010 ensembles, though an exact match is not expected given the slightly different time-periods and methods (e.g. finite differences versus regression) used. Still, the agreement provides further confidence in the GCM's simulated radiative response to observed SST and sea-ice variations over the historical record, and strengthens the conclusion that λ_{hist} has become more negative over recent decades compared to the longer 1871-2010 time-period.

Finally, IPCC AR6 assessed the long-term ECS relevant feedback parameter (analogous to our λ_{4xCO2}) to be -1.16 ± 0.65 W m^{-2} K^{-1} (Forster et al., 2021) by combining lines of evidence from observations, theory, process models and GCMs on individual climate feedback processes. Combining this with our observed λ_{hist} estimates above gives an estimate of the pattern effect independently of our GCM ensemble. This gives an estimated pattern effect of ~0.8 ± 1.0 W m^{-2} K^{-1} for 1985-2015 and ~0.4 ± 1.1 W m^{-2} K^{-1} for the full historical record (the 1850-1900 to 2006-2019 changes). While the uncertainties are substantial, there is again agreement with our GCM results.

4.2 Recent observed trends and the efficiency of ocean heat uptake

We have seen that both models and observed variations in the Earth's energy budget agree on the Earth having had strongly stabilizing feedbacks over recent decades relative to AOGCM feedbacks under long-term CO_2 forced climate change. Quantifying this in a different way, a feedback parameter of ~ -2.0 Wm^{-2} K^{-1} suggests an EffCS = $-F_{2x}/\lambda_{hist}$ as low as ~ 4.0/2.0 ~ 2.0 K operating over 1985-2014, assuming $F_{2x} = 4.0$ W m^{-2} (Sherwood et al. 2020). From this it seems possible that the rate of global warming over this period (~0.19 K dec^{-1}, Tokarska et al., 2020) might have been larger had the Earth warmed over this period with a pattern of SST associated with more positive feedbacks, as found in earlier parts of the historical record (Section 3). However, we also investigate the possibility that changes in ocean heat uptake efficiency may have compensated the changes in feedbacks and low EffCS to maintain a higher warming rate over this period than would be expected without this compensation.

To do this we turn to the 'climate resistance' (ρ, units W m^{-2} K^{-1}) "zero-layer" model of Gregory and Forster (2008) to analyse the ocean heat uptake efficiency (κ, units W m^{-2} K^{-1}). This is expressed as dF=ρ dT, where $\rho = \kappa - \lambda$, and κ is defined as $\kappa = dN/dT$ and is found to be strongly related to the thermal coupling constant (γ, units W m^{-2} K^{-1}) between the upper and lower ocean in the two-layer model (Gregory et al. 2015; see their Figure 8). While initially proposed to describe scenarios with steadily increasing forcing, it is also been applied to ~30 year timescales to usefully describe or interpret the energy balance (Gregory and Forster, 2008; Watanabe et al., 2013). Despite being a gross simplification of the climate system (we discuss potential limitations below), dF=ρ dT is found to be an excellent approximation (r=0.86) over 1985 – 2014 (excluding the 1991-2 Pinatubo years, see below) in our data (Figure 7c). From this relationship we deduce ρ = dF/dT ~ 2.4 ± 0.5 W m^{-2} K^{-1} over 1985-2014 (Figure 7c) and similarly κ = dN/dT ~ 0.4 ± 0.8 W m^{-2} K^{-1}. In contrast, AOGCM simulations of steady increasing CO_2 generally have a larger ocean heat uptake efficiency (κ = 0.73 ± 0.18 W m^{-2} K^{-1} for years 61-80 of CMIP5 1%CO2 AOGCM simulations, Gregory et al., 2015).

Another effect on surface temperature to consider is the possibility that the pattern of surface warming and/or atmospheric circulation may change the efficiency of global heat uptake (and vice versa), thus not only is λ inconstant, but κ may also vary. Using passive ocean uptake experiments wherein ocean circulation cannot change, Newsom et al. (2020) find that ocean heat uptake efficiency can be expected to be smaller when warming is enhanced in the tropics (where deep ocean ventilation is small) and larger when warming is enhanced in the high latitudes (where deep ocean ventilation is large). With relatively small warming in the southern high latitudes, this suggests that the surface/ocean-mixed layer might have been less efficient at fluxing heat into the deep ocean over the same period as the large pattern effect, potentially enhancing global surface warming and muting some of the impact of feedback changes. However, stronger trade winds, as have been observed over 1981-2010, can also be expected to accelerate subtropical cells, enhancing ocean heat uptake efficiency and slowing global surface warming (England et al. 2014), an effect not accounted for in the passive ocean heat uptake experiments of Newsom et al. (2020). Thus, variations in both radiative feedbacks and ocean heat uptake appear to be physically linked through SST patterns and may even to some extent co-vary (Newsom et al. 2020).

As our dN timeseries does not predate 1985 we cannot investigate whether κ has varied in a way that would counter changes in λ_{hist} prior to 1985. Instead, we go forward in time exploiting the datasets up to and including 2019. This includes the major El-Nino event of 2015/2016 and marked changes in the observed radiation budget (Loeb et al. 2020; 2021). Figure 9 illustrates the impact of this event on the pattern of decadal surface warming. Over 1985-2014 there is marked cooling over the eastern Pacific (Figure 9a) which is much reduced when the pattern is calculated over 1987-2016 (Figure 9b) to include the peak 2015-16 El-Nino years. The difference (Figure 9c) shows the warming event of the 2015-16 El-Nino on the eastern Pacific, while cooling in the western Pacific, as well as a slight reduction in Southern Ocean cooling. This is precisely the pattern of SST change we'd expect to have an impact on λ.

Table 4 shows the impact on 30-year derived ρ, λ and κ values moving forward in time from 2014, up to and including 1990-2019. Figure 7 (red crosses) shows these additional 5 years in comparison to the 1985-2014 ρ and λ relationships. Post 2014, λ reduces in magnitude (Table 4) and all the red crosses fall below the 1985-2014 λ relationship in Figure 7d. λ is approximately 25% smaller in magnitude over 1990-2019 compared to 1985-2014 (Table 4). This is consistent with process based arguments that a shift to eastern Pacific warming post 2014 ought to drive more positive feedbacks and consequently a reduction of the pattern effect over these years. It is also consistent with Loeb et al. (2020) who performed a similar analysis but over 2001-2014 compared to 2001-2017. They also showed that AGCMs were able to capture this change in radiative response. It would be useful for future analysis if *amip-piForcing* type simulations were extended to at least 2019 to capture the largest change in λ (Table 4), and ideally right up to the most recent SST and sea-ice data available.

In contrast to λ, ρ is relatively stable to these additional years (Table 4) and the 1985-2014 ρ relationship is found to be an excellent predictor for 2015-2019 (red crosses fall on or close to the line, Figure 7c). A consequence of ρ being well approximated as constant but λ not, is that κ (equal to ρ + λ) must compensate for the change in λ. Thus beyond 2014, the pattern effect declines but its impact on surface temperature is buffered by a change in ocean heat uptake efficiency. This is consistent with the original hypothesis that variations in SST patterns affect both heat loss to space (radiative feedbacks) and the efficiency of heat uptake into the deep-ocean in a way that might co-vary (Newsom et al., 2020). However, the extent of any anti-correlation is unclear, it may simply apply to short term variability. It clearly does not apply to longer-term forced changes, given that Gregory et al. (2015) found substantial variations in ρ, which would not occur if κ and λ were strongly anti-correlated.

While the zero-layer model appears to work well on this short timescale (Figure 7c) we caution against assuming all changes in ocean heat content are driven by global T, as assumed by the dN = κdT relationship. This is because, especially on short timescales, other influences that do not correlate with global T, such as wind-driven ocean circulation changes perhaps, will also alter ocean heat content (England et al., 2014). In such a situation, it would be reasonable to write $N = \kappa T + U$ where U is an additional term to the heat balance, not related to global T. This implies $\kappa = N/T - U/T$, and including this term in the forced heat balance, $N = F + \lambda T + U$, gives λ = (N-F)/T – U/T. Thus U/T would perturb the estimate of κ (a positive number) and λ (a negative number) in opposite directions, as we see in our data. Hence our results are potentially evidence for variation in ocean heat content not driven by global T, but we cannot say exactly what it is – other than it does not scale with global T.

We caution that structural errors could impact on our diagnosis. Specifically, both κ and λ are related to dN and so any bias or error in the observed dN trend would bias κ and λ in opposite directions. Moreover ρ=dF/dT would be unaffected by any bias or error in dN, and so the anti-correlation would compensate to leave $\rho = \kappa - \lambda$ unaffected. We illustrate this in Table 4, which shows these quantities calculated over 1985-2014 using 5 available different versions of the DEEP-C dN datasets (see Section 2.4). Differences in the results emerge (λ reduces in magnitude from ~-2.2 Wm^{-2} K^{-1} to ~ -2.0 Wm^{-2} K^{-1}, with a compensating increase in κ) as the DEEP-C datasets transition from v3 to v4 (i.e. v2 and v3 give the same results, as do v4 and v5), highlighting the impact of potential structural errors in these results. We do not purse the cause of the difference in the results, but it is likely due to changes between v3 and v4 in how the DEEP-C method bridges the gap between satellite products in the 1990s (a longer adjustment period and a different modelling ensemble is used) (Liu et al., 2020). However it is also important to note that the observational record since 2000, applying the CERES dataset, is subject to much smaller structural uncertainty than the earlier record implying a greater confidence in our analysis of the anomalous N variations post 2014.

4.3 Effect of the Pinatubo volcanic eruption

Finally, we comment on the effect of the Pinatubo volcanic eruption on these results. There is a large negative spike in dF and dN around 1991 and 1992 (Figure 7b). While we found no impact of these years on our estimate of 1985 – 2014 λ_{hist}, they have a strong impact on ρ and κ. Including these years in the regression analysis, we find ρ = dF/dT ~ 2.9 ± 0.7 W m^{-2} K^{-1} and κ = dN/dT ~ 0.8 ± 0.9 W m^{-2} K^{-1}, much larger than when these years are excluded from the analysis as above. This is consistent with Gregory et al. (2015) who found the 'transient climate response parameter' (equal to 1/ρ, units K W^{-1} m^{2}) to explosive eruptions to be smaller (ρ larger) than that evaluated in AOGCMs under steadily increasing CO_2, principally because the surface/mixed-layer readily gives up heat (κ larger) in response to a short-lived forcing like an explosive volcanic eruption. Hence if the time-period under consideration contains large volcanic eruptions then the "zero-layer" model (dF=ρ dT) is found to be a poor approximation (i.e. ρ not constant) over the entire time-period because it neglects the importance of the upper-ocean heat capacity on short timescales (Gregory and Forster, 2008; Held et al. 2010; Gregory et al., 2016). This manifests itself as a sensitivity of ρ and κ to the inclusion or exclusion of volcanic years, as we have found here.

5. Summary, Discussion and Conclusions

5.1 Historical feedbacks and the pattern effect

The dependence of radiative feedback on the pattern of SST change was investigated in fourteen Atmospheric General Circulation Models (AGCMs) forced with observed variations in sea-surface-temperature (SST) and sea-ice over the historical record from 1871 to near-present (*amip-piForcing* experiment). We found that the pattern effect identified in a previous model intercomparison (Andrews et al, 2018) is largely robust to a wider set of new generation AGCMs with a broader range of atmospheric physics and climate sensitivities. Our qualitative conclusions were not strongly dependent on the AMIP II SST dataset used to force the AGCMs; indeed, the feedbacks in nine AGCMs using SSTs from HadISST1 (*hadSST-piForcing*) were found to be strongly correlated with feedbacks in *amip-piForcing*, though the magnitude of the pattern effect post 1980 was found to be smaller under HadISST1 SSTs (see also Andrews et al., 2018; Lewis and Mauritsen, 2021; Zhou et al., 2021; Fueglistaler and Silvers, 2021).

Separating the historical record at 1980, we found that over 1871-1980 the Earth warmed with a relatively uniform warming pattern and feedbacks largely consistent and strongly correlated with long-term *abrupt-4xCO2* feedbacks (i.e. with relatively small pattern effect - Figures 2 and 5). In contrast, post 1980 the Earth warmed with a strong tropical Pacific SST gradient (Figure 4) where regions of deep convection warm substantially more than the tropical mean (Fueglistaler and Silvers, 2021). This drove large negative feedbacks and pattern effects in both our *amip-piForcing* and *hadSST-piForcing* simulations, consistent with the physical understanding of how lapse-rate and cloud feedbacks depend on tropical Pacific SST patterns (Zhou et al., 2016; Andrews and Webb, 2018; Ceppi and Gregory, 2017; Dong et al., 2019).

As well as a large pattern effect, feedbacks post 1980 were found to be uncorrelated with long term CO_2 driven feedbacks (Figure 5). This is unfortunate, because the feedback inferred from this period therefore does not constrain the CO_2 feedback or ECS. It is also surprising, because the period since ~1980 contains a well observed large global temperature response, which AOGCMs attribute to increasing greenhouse gases, and it avoids the aerosol forcing uncertainty issue which is small in energy budget estimates of ECS over this period (at least in the global-mean; regional aerosol forcing could still impact on SST patterns and feedbacks) (Jiménez-de-la-Cuesta and Mauritsen, 2019). Despite this, it turns out to be the worst period for inferring the Earth's long-term CO_2 climate sensitivity from the observed global energy balance. Conversely, feedbacks acting earlier in the record (1871-1980) are representative of the long-term response (i.e. smaller pattern effect) and do correlate with λ_{4xCO2} across models, yet this period has a smaller climate change signal and is not as well observed, containing much larger uncertainties relative to the climate change signal (e.g. Otto et al., 2013), as well as a large forcing uncertainty. Hence the usefulness of this time-period is limited for setting a constraint on λ_{hist}.

Considering the historical record as a whole is useful for informing studies that use the entire observed record to estimate ECS via energy budget constraints (e.g. Sherwood et al. 2020). We found that the pattern effect over 1871-2010 to be $\Delta\lambda = 0.70 \pm 0.47$ W m^{-2} K^{-1} in our *amip-piForcing* ensemble and $\Delta\lambda = 0.48 \pm 0.36$ W m^{-2} K^{-1} in *hadSST-piForcing*, where the smaller uncertainty in *hadSST-piForcing* likely reflects the narrower set of model physics in this smaller ensemble (for example we do not have *hadSST-piForcing* experiments for the model (MIROC6) with the smallest pattern effect in *amip-piForcing*). The question therefore arises as to which of these estimates ought to be used for adjusting historical energy budget constraints on ECS for pattern effects.

Both Lewis and Mauritsen (2021) and Fueglistaler and Silvers (2021) showed that the AMIP II dataset had the largest warm pool trends relative to the tropical-mean of all SST reconstructions they considered. Hence one interpretation of our results is that the pattern effect in *amip-piForcing* might usefully be regarded as an upper bound on the structural uncertainty of the experimental design to observational uncertainty in SST reconstructions. A best estimate might place more weight on the *hadSST-piForcing* pattern effects, which have warm pool trends (relative to the tropical-mean) closer to the middle of the range of SST reconstructions (Fueglistaler and Silvers, 2021; Lewis and Mauritsen, 2021). In that case, a best estimate of the historical pattern effect could be 0.48 ± 0.47 W m^{-2} K^{-1} for the time-period 1871-2010, which represents the pattern effect from *hadSST-piForcing* but retaining the larger uncertainty from the (larger ensemble) *amip-piForcing* results. If calculated over 1871-2014 the pattern effect increases by 0.05 ± 0.05 W m^{-2} K^{-1} according to the *hadSST-piForcing* ensemble. This best estimate of the historical pattern effect is close to that used in Sherwood et al. (2020), who assumed a value of 0.5 ± 0.5 W m^{-2} K^{-1} (they were informed by Andrews et al. (2018) who used *amip-piForcing* but allowed for a potentially smaller pattern effect than that study based on expert judgement). On the other hand, just because the AMIP II SST trends are at one end of the range of SST reconstructions does not necessarily mean they are more erroneous. Indeed, Zhou et al. (2021) showed that TOA radiative fluxes simulated by CAM5.3 correlated better with CERES observations when forced with AMIP II SSTs rather than HadISST SSTs, suggesting the results from *amip-piForcing* may be more reliable. In this case, the 1871-2010 pattern effect is 0.70 ± 0.47 W m^{-2} K^{-1}. In the future, a model intercomparison of the pattern effect to a broader range of SST reconstructions would be useful to address any outstanding structural uncertainty to SST reconstructions.

To provide independent evidence for the historical pattern effect, we used IPCC AR6 assessed changes in T, N and F between 1850-1900 to 2006-2019 (Forster et al. 2021) to estimate a historical feedback parameter of $\lambda_{hist} = (\Delta N - \Delta F)/\Delta T = -1.6 \pm 0.8$ W m^{-2} K^{-1}. This was found to be in agreement with the *amip-piForcing* and *hadSST-piForcing* ensembles. IPCC AR6 also assessed the long-term ECS relevant feedback parameter (-1.16 ± 0.65 W m^{-2} K^{-1}, Forster et al., 2021) from combining lines of evidence from observations, theory, process models and GCMs on individual climate feedback processes. Contrasting this with the λ_{hist} estimate above gives an estimate of the pattern effect of 0.4 ± 1.1 W m^{-2} K^{-1} for historical changes between 1850-1900 to 2006-2019. While the uncertainties are substantial, this is in agreement with our GCM based estimate of the historical pattern effect.

5.2 Observed climate change since 1985 and ocean heat uptake efficiency

Satellite based reconstructions of the Earth's energy balance over 1985 to 2014 suggest a feedback parameter of ~-2.0 ± 0.7 W m^{-2} K^{-1}, in agreement with our *amip-piForcing* and *hadSST-piForcing* ensembles. Evidence is also emerging from satellite records in support of the physical processes and mechanisms of the pattern effect between surface temperature, atmospheric stability, cloudiness and radiative fluxes over recent decades (e.g. Zhou et al., 2016; Ceppi and Gregory, 2017; Loeb et al., 2020; Fueglistaler and Silvers, 2021; Ceppi and Fueglistaler, 2021).

Extending our analysis post 2014 included the major El-Nino event of 2015/2016 that was associated with eastern-pacific warming and marked changes in the observed radiation budget (Loeb et al. 2020; 2021). Including these post 2014 years (up to and including 2019) reduced the magnitude of the observed λ estimate by up to ~25%, consistent with eastern Pacific warming driving more positive feedbacks (as also suggested in Loeb et al., 2020). This suggests the pattern effect that has existed over recent decades may be waning if a shift from western to eastern Pacific warming is maintained in the longer term, as might be expected from a change in the PDO index identified by Loeb et al. (2021).

Given the substantial rate of global warming since 1985, what does the presence of a large pattern effect imply for ocean heat uptake efficiency (κ)? We estimated $\kappa = dN/dT \sim 0.4 \pm 0.8$ W m^{-2} K^{-1} over 1985-2014, which is smaller (but not necessarily inconsistent) with AOGCM simulations of steady increasing CO_2 ($\kappa = 0.73 \pm 0.18$ W m^{-2} K^{-1} for years 61-80 of CMIP5 1%CO2 AOGCM simulations, Gregory et al. 2015). It raises the possibility that the pattern of surface warming and/or atmospheric circulation may also change the efficiency of global heat uptake, thus both λ and κ might vary and to some extent be related (Newsom et al., 2020). If an anti-correlation existed, it could buffer the impact of a large pattern-effect on transient climate change.

We found that despite the change in radiative feedback post 2014 when the eastern Pacific warmed, the climate resistance $\rho = dF/dT = \kappa - \lambda$ remained approximately constant, suggesting that κ and λ co-varied. We showed that this result is potential evidence for a change in ocean heat content not driven by global T. While this result is suggestive, the extent of this compensation and timescales it applies to remains unclear. It may simply apply to short term variability and clearly does not apply to longer-term forced changes (e.g. Gregory et al., 2015). Future research investigating how ocean uptake and atmospheric radiative feedbacks are linked through patterns of SST change would be useful.

5.3 Outlook and Implications for AOGCMs

Our results raise important questions for studies that have used emergent relationships from AOGCMs to constrain ECS from recently observed decadal warming since ~1980 (e.g. Jiménez-de-la-Cuesta and Mauritsen, 2019; Tokarska et al., 2020; Nijsse et al., 2020).

Firstly, how is it possible that AOGCMs produce an emergent relationship between their recent decadal warming trends and their ECS, while our results suggest that recent decadal feedbacks ought to be unrelated to ECS? One solution to this conundrum is provided by Fueglistaler and Silvers (2021), who showed that AOGCMs typically do not simulate the recent configuration of tropical Pacific SST patterns that gave rise to the recent pattern effect (though some models do have broad agreements, e.g. Olonscheck et al. 2021, Watanabe et al. 2021). Instead, the pattern of warming in AOGCMs (and thus feedbacks) over recent decades is more similar to that seen in their *abrupt-4xCO2* simulations (Gregory et al., 2020; Dong et al. 2021). Hence AOGCMs are generally biased in their simulation of the recent decadal feedbacks and the pattern effect, compared to their equivalent AGCMs forced with observed SST variations, as shown in Gregory et al. (2020) and Dong et al. (2021).

If AOGCMs are biased in their simulation of recent decadal feedbacks and the pattern effect, it suggests they may be biased toward simulating recent decadal temperature trends that are too high; in turn, this would bias emergent constraints that use them toward values of ECS that are too low. Alternatively, those models that do match the observed warming trend may do so via a compensation of processes: too small a pattern effect balanced against too large a heat uptake into the deep-ocean. Some evidence for the potential of this compensating behaviour is provided by Hedemenn et al. (2017). Analysing the origins of decadal temperature variability in models, they demonstrated an anti-correlation between the TOA radiative flux and deep-ocean (defined as below 100m) flux contributions to the model's surface layer and decadal temperature trends (see their

Figure 3). In other words, when the TOA radiative flux is in such a configuration to reduce its contribution to the surface layer, then the surface/mixed-layer taps into the deep-ocean to compensate for this loss, and vice versa. We speculate that such a configuration of TOA radiative flux is potentially consistent with a large negative feedback, since in this configuration of atmospheric feedbacks the surface efficiently radiates heat back to space. This again suggests a potential anti-correlation between the ocean heat uptake efficiency and λ during unforced decadal variability timescales as discussed previously.

Going forward, a critical question for future research is to understand what caused the particular configuration of SST patterns over recent decades (e.g. strong warming in the western Pacific while cooling in the eastern Pacific and Southern Ocean, despite temperature increasing in the global-mean; Figure 4 and 9), and how might this pattern evolve in the future. For example, various hypotheses have been put forward:

1. It could represent a mode of unforced coupled atmosphere-ocean variability (e.g. Xie et al., 2016; Watanabe et al. 2021), albeit an unusual one is that is rarely simulated by AOGCMs (Fueglistaler and Silvers, 2021). In this scenario, we might expect the pattern effect to reduce in the near-future as the configuration of tropical SST patterns shift to more warming in the east than the west. There is some evidence (Loeb et al. 2020; 2021) this has already begun to happen in the most recent years, as we have also shown. We might therefore expect an acceleration of warming trends, unless the additional heat at the surface from the reduced pattern effect is tempered by compensating heat exchanges with the deep-ocean (Hedemann et al. 2017).
2. Spatiotemporal variations in anthropogenic forcings such as aerosols (e.g., Smith et al., 2015; Takahashi & Watanabe, 2016; Moseid et al., 2020; Heede and Fedorov, 2021) or explosive volcanic eruptions (Smith et al. 2015; Gregory et al. 2020) have been implicated in driving tropical Pacific SST patterns. In these scenarios, the pattern effect may decline with the reduction in aerosol emissions in the future, or continue to have decadal variations associated with future volcanism. Whether changes in deep-ocean fluxes will be accompanied with such forced changes in the pattern effect is unclear.
3. While not explaining the eastern Pacific cooling per se, a delayed warming in the eastern Pacific relative to the west is an expected transient response to forcing due to the upwelling of (as yet) unperturbed waters from below (Clement et al., 1993; Held et al. 2010; Heede and Fedorov, 2021). The implication of this is that eventually the eastern Pacific will warm, and hence we might expect the pattern effect to reduce and the Earth to warm with stronger (positive) cloud feedbacks (e.g. Dessler, 2020).
4. In contrast, AOGCMs may overstate the expected warming in the eastern Pacific (e.g. Seager et al., 2020). Under this scenario, we might expect the pattern effect to reduce after the eastern Pacific stops cooling, but the full pattern effect according to AOGCMs may never materialise if they incorrectly simulate a strong 'ENSO-like' pattern in their long-term response to CO_2. However, a lack of eastern Pacific warming in the long-term seems unlikely according to paleoclimate records (Tierney et al. 2019; 2020).
5. Teleconnections from either the Atlantic Ocean (McGregor et al. 2018) or Southern Ocean (Hwang et al. 2017) have potentially driven the tropical Pacific SST patterns. Under the scenario of an Atlantic influence, we might expect the pattern effect to

reduce as Atlantic SST trends evolve over the next few decades. Under the scenario of a Southern Ocean influence, we might expect the pattern effect to reduce as the Southern Ocean surface warms; this could take years to decades if the Southern Ocean temperature trends have been largely mediated by internal variability (e.g., Zhang et al. 2019) but could take centuries or longer if Southern Ocean cooling continues due, for instance, to freshwater input from ongoing Antarctic ice shelf melt (e.g., Sadai et al. 2020).

These are merely some of the proposed hypotheses, and not meant to be an exhaustive list. But whatever the reason, the fact that AOGCMs rarely simulate this pattern (e.g. Watanbe et al., 2021; Fueglistaler and Silvers, 2021; Dong et al., 2021) is a concern, suggesting either that their unforced decadal variability is deficient, or that their forced response is biased, and in either case there is a serious systematic error which affects all AOGCMs. Moreover, each of the above interpretations imply different futures, and therefore untangling them is critical for informing both near-term and long-term climate projections. This is time critical because satellite evidence suggests the Pacific SST pattern that has dominated recent decades is currently shifting (Loeb et al., 2020) and indeed the Earth's energy balance is rapidly changing with it (Loeb et al. 2021; Raghuraman et al., 2021). Predicting the near future therefore depends on maintaining the continuity of the satellite record and untangling the above mechanisms.

Acknowledgements

TA and JMG thank Richard Wood and Mark Ringer for useful discussions. We thank Bosong Zhang, Zhihong Tan and Knut von Salzen for useful comments on an early draft version of the manuscript. We thank Maria Rugenstein and an anonymous reviewer for positive and constructive comments. TA, JMG and ABS were supported by the Met Office Hadley Centre Climate Programme funded by BEIS. TA, TM, AM and RR have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 820829. JMG's work is also supported by the European Research Council (ERC, grant agreement No 786427, project "Couplet"). TM and AM received funding from the European Research Council grant 770765. The contribution of JB to this work was funded by the U. S. Department of Energy's Regional and Global Modeling Analysis program area. TO was supported by the Integrated Research Program for Advancing Climate Models (TOUGOU) Grant Number JPMXD0717935457 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. CL was supported by the National Natural Science Foundation of China (42075036). KCA and YD acknowledge support from the National Science Foundation (Grant AGS-1752796) and from the National Oceanic and Atmospheric Administration MAPP Program (Award NA20OAR4310391). BM acknowledges support by the U.S. Department of Energy under Award Number DE-SC0022070 and National Science Foundation (NSF) IA 1947282; the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977; and the National Oceanic and Atmospheric Administration under award NA20OAR4310392. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

Data Availability

Global-annual-ensemble-mean dT and dN data from all *amip-piForcing*, *hadSST-piForcing* and *abrupt-4xCO2* simulations used in this study are provided at https://doi.org/10.5281/zenodo.6799004 (Andrews et al. 2022). Raw data from CMIP6 *amip-piForcing* simulations (indicated in Table 1) are available at https://pcmdi.llnl.gov/CMIP6/ (Eyring et al., 2016). *abrupt-4xCO2* raw data for most models is available at CMIP5 (https://esgf-node.llnl.gov/projects/cmip5/) (Taylor et al., 2012) or CMIP6 (https://pcmdi.llnl.gov/CMIP6/) (Eyring et al., 2016). The HadCRUT5 analysis dataset is available at https://www.metoffice.gov.uk/hadobs/hadcrut5/ (Morice et al., 2021). IPCC AR6 ERF timeseries is available at https://doi.org/10.5281/zenodo.5211358 (Smith et al., 2021). DEEP-C v5 dN radiative fluxes can be obtained from https://doi.org/10.17864/1947.000347 (Lui and Allan, 2022) and previous versions described at http://www.met.reading.ac.uk/~sgs02rpa/research/DEEP-C/GRL/. The HadISST1 SSTs used to force the *hadSST-piForcing* simulations are available at https://www.metoffice.gov.uk/hadobs/hadisst/ (Rayner et al. 2003).

References

Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., and Vidale, P.-L. (2014), Changes in global net radiative imbalance 1985–2012, *Geophys. Res. Lett.*, 41, 5588– 5597, doi:10.1002/2014GL060962.

Andrews, Timothy, Bodas-Salcedo, Alejandro, Gregory, Jonathan, Dong, Yue, Armour, Kyle, Paynter, David, Lin, Pu, Modak, Angshuman, Mauritsen, Thorsten, Cole, Jason, Medeiros, Brian, Benedict, James, Douville, Herve, Roehrig, Romain, Koshiro, Tsuyoshi, Kawai, Hideaki, Ogura, Tomoo, Dufresne, Jean-Loius, Allan, Richard, & Liu, Chunlei. (2022). amip-piForcing and HadSST-piForcing simulation data (v1.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.6799004

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., Webb, M. J., Armour, K. C., Forster, P. M., & Titchner, H. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, 45, 8490– 8499. https://doi.org/10.1029/2018GL078887

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, 1630–1648.

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., & Ringer, M. A. (2014). Cloud feedbacks, rapid adjustments, and the forcing–response relationship in a transient CO_2 reversibility scenario. *Journal of Climate*, 27(4), 1799– 1818. https://doi.org/10.1175/JCLI-D-13-00421.1

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. https://doi.org/10.1175/JCLI-D-17-0087.1

Annan, J. D., & Hargreaves, J. C. (2017). On the meaning of independence in climate science. *Earth System Dynamics*, 8, 221– 224.

Annan, J. D., & Hargreaves, J. C. (2011). Understanding the CMIP3 Multimodel Ensemble, *Journal of Climate*, *24*(16), 4529-4538. https://doi.org/10.1175/2011JCLI3873.1.

Armour, K. C. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Natural Climate Change*, 7, 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, 4518– 4534.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., & Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002010. https://doi.org/10.1029/2019MS002010

Bloch-Johnson, J., Rugenstein, M., Stolpe, M. B., Rohrschneider, T., Zheng, Y., & Gregory, J. M. (2021). Climate sensitivity increases under higher CO_2 levels due to feedback temperature dependence. *Geophysical Research Letters*, 48, e2020GL089074. https://doi.org/10.1029/2020GL089074

Block, K., and Mauritsen, T. (2013), Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO_2, *J. Adv. Model. Earth Syst.*, 5, 676– 691, doi:10.1002/jame.20041.

Bjordal, J., Storelvmo, T., Alterskjær, K. *et al.*, 2022: Equilibrium climate sensitivity above 5 °C plausible due to state-dependent cloud feedback. *Nat. Geosci.* **13,** 718–721 (2020).doi, 10.1038/s41561-020-00649-1.

Caballero, R., & Huber, M. (2013). State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 14,162– 14,167. https://doi.org/10.1073/pnas.1303365110

Ceppi, P., & Fueglistaler, S. (2021). The El Niño–Southern Oscillation pattern effect. *Geophysical Research Letters*, 48, e2021GL095261. https://doi.org/10.1029/2021GL095261.

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, 13,126– 13,131. https://doi.org/10.1073/pnas.1714308114

Chalmers, J., Kay, J. E., Middlemas, E. A., Maroon, E. A., & DiNezio, P., 2022: Does disabling cloud radiative feedbacks change spatial patterns of surface greenhouse warming and cooling?, Journal of Climate (published online ahead of print 2022; https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-21-0391.1)

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.

Clement, A. C., Seager, R., Cane, M. A., & Zebiak, S. E. (1996). An ocean dynamical thermostat. *Journal of Climate*, 9, 2190– 2196. https://doi.org/10.1175/1520-0442(1996)009<2,190:AODT>2.0.CO2

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001916. https://doi.org/10.1029/2019MS001916

Dessler, A. E. (2020). Potential Problems Measuring Climate Sensitivity from the Historical Record, *Journal of Climate*, *33*(6), 2237-2248.

Dong, Y., Armour, K. C., Proistosescu, C., Andrews, T., Battisti, D. S., Forster, P. M., et al. (2021). Biased estimates of equilibrium climate sensitivity and transient climate response derived from historical CMIP6 simulations. *Geophysical Research Letters*, 48, e2021GL095778. https://doi.org/10.1029/2021GL095778

Dong, Y., Armour, K. C., Zelinka, M. D., Proistosescu, C., Battisti, D. S., Zhou, C., & Andrews, T. (2020). Inter-model spread in the sea-surface temperature pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. *Journal of Climate*, 33(18), 7755– 7775. https://doi.org/10.1175/JCLI-D-19-1011.1

Dong, Y., Proistosescu, C., Armour, K. C., & Battisti, D. S. (2019). Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a Green's Function approach: The preeminence of the western Pacific. *Journal of Climate*, 32(17), 5471– 5491. https://doi.org/10.1175/JCLI-D-18-0843.1

Donner, L. et al.. (2011). The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3, *Journal of Climate*, *24*(13), 3484-3519.

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

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

Fueglistaler, S., & Silvers, L.G. (2021). The peculiar trajectory of global warming. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033629. https://doi.org/10.1029/2020JD033629

Forster, P., T. Storelvmo, K. Armour, W. Collins, J. L. Dufresne, D. Frame, D. J. Lunt, T. Mauritsen, M. D.Palmer, M. Watanabe, M. Wild, H. Zhang, 2021, The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 923–1054, doi:10.1017/9781009157896.009.

Flynn, C. M. and Mauritsen, T.: On the climate sensitivity and historical warming evolution in recent coupled model ensembles, Atmos. Chem. Phys., 20, 7829–7842, https://doi.org/10.5194/acp-20-7829-2020, 2020.

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

Geoffroy, O., et al. (2013). Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *Journal of Climate*, 26, 1859– 1876.

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., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO_2. *Philosophical Transactions of the Royal Society A*, 373, 140– 417. http://doi.org/10.1098/rsta.2014.0417

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., & Webb, M. J. (2020). How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? *Climate Dynamics*, 54(1–2), 129– 157. https://doi.org/10.1007/s00382-019-04991-y

Gregory, J. M., Stouffer, R. J., Raper, S. C. B., 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

Gregory, J. M., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, 31, L03205. https://doi.org/10.1029/2003GL018747

Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254.

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

Hansen, J., Sato, M. K. I., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S. & Bell, N. (2005). Efficacy of climate forcings, mathematical physical and engineering sciences 365, 1925–54. *Journal of Geophysical Research*, 110, D18104. https://doi.org/10.1029/2005JD005776

Heede, U.K., Fedorov, A.V. Eastern equatorial Pacific warming delayed by aerosols and thermostat response to CO_2 increase. *Nat. Clim. Chang.* **11,** 696–703 (2021). https://doi.org/10.1038/s41558-021-01101-x

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, 11, 3691– 3727. https://doi.org/10.1029/2019MS001829

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

Hedemann, C., Mauritsen, T., Jungclaus, J. *et al.* The subtle origins of surface-warming hiatuses. *Nature Clim Change* **7,** 336–339 (2017). https://doi.org/10.1038/nclimate3274

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

Hwang, Y.-T., Xie, S.-P., Deser, C., and Kang, S. M. (2017), Connecting tropical climate change with Southern Ocean heat uptake, *Geophys. Res. Lett.*, 44, 9449– 9457, doi:10.1002/2017GL074972.

Jiménez-de-la-Cuesta, D., & Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post–1970s global warming. *Nature Geoscience*, 12, 902–905. https://doi.org/10.1038/s41561-019-0463-y

Kawai, H., S. Yukimoto, T. Koshiro, N. Oshima, T. Tanaka, H. Yoshimura, and R. Nagasawa, 2019: Significant Improvement of Cloud Representation in Global Climate Model MRI-ESM2. Geosci. Model Dev., 12, 2875-2897.

Lewis, N., & Mauritsen, T. (2021). Negligible Unforced Historical Pattern Effect on Climate Feedback Strength Found in HadISST-Based AMIP Simulations, *Journal of Climate*, *34*(1), 39-55. https://doi.org/10.1175/JCLI-D-19-0941.1.

Lewis, N., & Curry, J. A. (2018). The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *Journal of Climate*, 31, 6051– 6071.

Li, C., von Storch, J.-S., & Marotzke, J. (2013). Deep-ocean heat uptake and equilibrium climate response. *Climate Dynamics*, 40, 1071– 1086.

Liu, C., Allan, R. P., Berrisford, P., Mayer, M., Hyder, P., Loeb, N., Smith, D., Vidale, P.-L., and Edwards, J. M. (2015), Combining satellite observations and reanalysis energy transports to estimate global net surface energy fluxes 1985–2012, *J. Geophys. Res. Atmos.*, 120, 9374– 9389, doi:10.1002/2015JD023264.

Liu, C., Allan, R. P., Mayer, M., Hyder, P., Loeb, N. G., Roberts, C. D., Valdivieso, M., Edwards, J. M., and Vidale, P.-L. (2017), Evaluation of satellite and reanalysis-based global net surface energy flux and uncertainty estimates, *J. Geophys. Res. Atmos.*, 122, 6250– 6272, doi:10.1002/2017JD026616.

Liu, C., Allan, R.P., Mayer, M. *et al.* Variability in the global energy budget and transports 1985–2017. *Clim Dyn* **55,** 3381–3396 (2020). https://doi.org/10.1007/s00382-020-05451-8

Liu, C. and R. Allan (2022): Reconstructions of the radiation fluxes at the top of atmosphere and net surface energy flux: DEEP-C Version 5.0 [Dataset]. University of Reading. https://doi.org/10.17864/1947.000347.

Loeb, N. G., Johnson, G. C., Thorsen, T. J., Lyman, J. M., Rose, F. G., & Kato, S. (2021). Satellite and ocean data reveal marked increase in Earth's heating rate. *Geophysical Research Letters*, 48, e2021GL093047. https://doi.org/10.1029/2021GL093047.

Loeb, N. G., Wang, H., Allan, R., Andrews, T., Armour, K., Cole, J. N. S., et al. (2020). New generation of climate models track recent unprecedented changes in earth's radiation budget observed by CERES. *Geophysical Research Letters*, 47, e2019GL086705. https://doi.org/10.1029/2019GL086705.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, 11, 998– 1038. https://doi.org/10.1029/2018MS001400

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

Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. (2016). Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change*, 6, 386– 389). https://doi.org/10.1038/nclimate2888.

Martin, G.M., et al., 2011: The HadGEM2 family of Met Office Unified Model climate configurations, Geosci. Model Dev., 4, 723–757, https://doi.org/10.5194/gmd-4-723-2011, 2011.

McGregor, S., Stuecker, M.F., Kajtar, J.B. *et al.* Model tropical Atlantic biases underpin diminished Pacific decadal variability. *Nature Clim Change* **8,** 493–498 (2018). https://doi.org/10.1038/s41558-018-0163-4

Meehl G A, Senior C A, Eyring V, Flato G, Lamarque J-F, Stouffer R J, Taylor K E and Schlund M, (2020), Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. Sci. Adv. 6, 26, https://doi.org/10.1126/sciadv.aba1981.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., et al. (2021). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, 126, e2019JD032361. https://doi.org/10.1029/2019JD032361

Moseid, K. O., Schulz, M., Storelvmo, T., Julsrud, I. R., Olivié, D., Nabat, P., Wild, M., Cole, J. N. S., Takemura, T., Oshima, N., Bauer, S. E., and Gastineau, G.: Bias in CMIP6 models as compared to observed regional dimming and brightening, Atmos. Chem. Phys., 20, 16023–16040, https://doi.org/10.5194/acp-20-16023-2020, 2020.

Neale, R. B., Richter, J., Park, S., Lauritzen, P. H., Vavrus, S. J., Rasch, P. J., & Zhang, M. (2013). The Mean Climate of the Community Atmosphere Model (CAM4) in Forced SST and Fully Coupled Experiments, *Journal of Climate*, *26*(14), 5150-5168. https://doi.org/10.1175/JCLI-D-12-00236.1.

Newsom, E., Zanna, L., Khatiwala, S., & Gregory, J. M. (2020). The influence of warming patterns on passive ocean heat uptake. *Geophysical Research Letters*, 47, e2020GL088429. https://doi.org/10.1029/2020GL088429

Nijsse, F. J. M. M., Cox, P. M., & Williamson, M. S. (2020). An emergent constraint on Transient Climate Response from simulated historical warming in CMIP6 models. *Earth System Dynamics*. https://doi.org/10.5194/esd-2019-86.

Olonscheck, D., Rugenstein, M., & Marotzke, J. (2020). Broad consistency between observed and simulated trends in sea surface temperature patterns. *Geophysical Research Letters*, 47, e2019GL086773. https://doi.org/10.1029/2019GL086773

Otto, A., Otto, F. E. L., Boucher, O., Church, J., Hegerl, G., Forster, P. M., Gillett, N. P., Gregory, J., Johnson, G. C., Knutti, R., Lewis, N., Lohmann, U., Marotzke, J., Myhre, G., Shindell, D., Stevens, B., & Allen, M. R. (2013). Energy budget constraints on climate response. *Nature Geoscience*, 6(6), 415–416. https://doi.org/10.1038/ngeo1836

Pope, D. V., M. Gallani, R. Rowntree, and A. Stratton (2000), The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3, Clim. Dyn., 16(2–3), 123–146.

Power, S., et al., (2021). Decadal climate variability in the tropical Pacific: Characteristics, causes, predictability, and prospects. Science. 374. eaay9165. 10.1126/science.aay9165.

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, 3, 1– 7. https://doi.org/10.1126/sciadv.1602821

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A. (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108, 4407, doi:10.1029/2002JD002670, D14.

Raghuraman, S.P., Paynter, D. & Ramaswamy, V. Anthropogenic forcing and response yield observed positive trend in Earth's energy imbalance. *Nat Commun* **12,** 4577 (2021). https://doi.org/10.1038/s41467-021-24544-4

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An Improved In Situ and Satellite SST Analysis for Climate, *Journal of Climate*, *15*(13), 1609-1625.

Richardson, T. B., Forster, P. M., Smith, C. J., Maycock, A. C., Wood, T., Andrews, T., et al. (2019). Efficacy of climate forcings in PDRMIP models. *Journal of Geophysical Research: Atmospheres*, 124, 12824– 12844. https://doi.org/10.1029/2019JD030581

Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., & Koll, D. D. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, 41, 1– 8.

Rugenstein, M. A. A., Caldiera, K., & Knutti, R. (2016). Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophysical Research Letters*, 43, 9877– 9885. https://doi.org/10.1002/2016GL070907

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, 47, e2019GL083898. https://doi.org/10.1029/2019GL083898

Rugenstein, M. A. A., & Armour, K. C. (2021). Three flavors of radiative feedbacks and their implications for estimating equilibrium climate sensitivity. *Geophysical Research Letters*, 48, e2021GL092983. https://doi.org/10.1029/2021GL092983

Sadai, S., Condron, A., DeConto, R., & Pollard, D. (2020). Future climate response to Antarctic Ice Sheet melt caused by anthropogenic warming. Science advances, 6(39), eaaz1169.

Sanderson, B. M., and Knutti, R. (2012), On the interpretation of constrained climate model ensembles, *Geophys. Res. Lett.*, 39, L16708, doi:10.1029/2012GL052665.

Schneider, A., Flanner, M. & Perket, J. Multidecadal variability in surface albedo feedback across CMIP5 models. *Geophys. Res. Lett.* **45**, 1972–1980 (2018).

Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019). Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, 9, 517– 522.

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

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58, e2019RG000678. https://doi.org/10.1029/2019RG000678.

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, C., Piers Forster, Matt Palmer, Bill Collins, Nick Leach, Masa Watanabe, Sophie Berger, & Brad Hall, (2021). IPCC-WG1/Chapter-7: IPCC WGI AR6 Chapter 7 (v.1.0) [dataset]. Zenodo. https://doi.org/10.5281/zenodo.5211358.

Smith, D. M., and Coauthors, 2015: Earth's energy imbalance since 1960 in observations and CMIP5 models. *Geophys. Res. Lett.*, **42**, 1205–1213, https://doi.org/10.1002/2014GL062669.

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, 512– 522. https://doi.org/10.1002/2016EF000376.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019.

Takahashi, C., and M. Watanabe, 2016: Pacific trade winds accelerated by aerosol forcing over the past two decades. *Nature Climate Change*, 6, 768-772, doi: 10.1038/nclimate2996.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: 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, 2019.

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, 485– 498.

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.

Tierney, J. E., Haywood, A. M., Feng, R., Bhattacharya, T., & Otto-Bliesner, B. L. (2019). Pliocene warmth consistent with greenhouse gas forcing. *Geophysical Research Letters*, 46, 9136– 9144. https://doi.org/10.1029/2019GL083802.

Tierney, J. E., Zhu, J., King, J., Malevich, S.B., Hakim, G.J., & Poulsen, C.J. (2020). Global cooling and climate sensitivity revisited. https://doi.org/10.31223/osf.io/me5uj.

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

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advance*, 6, eaaz9549. https://doi.org/10.1126/sciadv.aaz9549.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, 11, 2177– 2213. https://doi.org/10.1029/2019MS001683

Watanabe, M., Dufresne, JL., Kosaka, Y. *et al.*, (2021): Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient. *Nat. Clim. Chang.* **11,** 33–37 (2021). https://doi.org/10.1038/s41558-020-00933-3

Watanabe, M., Kamae, Y., Yoshimori, M., Oka, A., Sato, M., Ishii, M., Mochizuki, T., and Kimoto, M. (2013), Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus, *Geophys. Res. Lett.*, 40, 3175– 3179, doi:10.1002/grl.50541.

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J. E., Klein, S. A., Marchand, R., Medeiros, B., Siebesma, A. P., Skinner, C. B., Stevens, B., Tselioudis, G., Tsushima, Y., and Watanabe, M.: The Cloud Feedback Model

Intercomparison Project (CFMIP) contribution to CMIP6, Geosci. Model Dev., 10, 359–384, https://doi.org/10.5194/gmd-10-359-2017, 2017.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., … Xavier, P. K. (2017). The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *Journal of Advances in Modeling Earth Systems*, 10, 357– 380. https://doi.org/10.1002/2017MS001115

Yukimoto, S., H. Kawai, T. Koshiro, N. Oshima, K. Yoshida, S. Urakawa, H. Tsujino, M. Deushi, T. Tanaka, M. Hosaka, S. Yabu, H. Yoshimura, E. Shindo, R. Mizuta, A. Obata, Y. Adachi, and M. Ishii, 2019: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. J. Meteor. Soc. Japan, 97, 931-965.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, 47, e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhang, L., Delworth, T.L., Cooke, W. *et al.* Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Clim Change* **9,** 59–65 (2019). https://doi.org/10.1038/s41558-018-0350-3

Zhou, C., Zelinka, M. D., & Klein, S. A. (2016). Impact of decadal cloud variations on the Earth's energy budget. *Nature Geoscience*, 9, 871– 875.

Zhou, C., Zelinka, M.D., Dessler, A.E. *et al.* Greater committed warming after accounting for the pattern effect. *Nat. Clim. Chang.* **11,** 132–136 (2021). https://doi.org/10.1038/s41558-020-00955-x

Zhu, J., B.L. Otto-Bliesner, E.C. Brady, C. Poulson, J.K. Shaw, J.E. Kay (2022), LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2. Earth and Space Science Open Archive, https://doi.org/10.1002/essoar.10507790.1.

Table1: Summary of the Atmospheric General Circulation Model (AGM) simulations used in this study. a*mip-piForcing* refers to an AGCM simulation forced with time-varying observed monthly SSTs and sea-ice using the AMIP II boundary condition SST and sea-ice dataset, forcing agents such greenhouse gases, aerosol emission etc. are kept at pre-industrial levels. *hadSST-piForcing* is identical in all aspects except SSTs are taken from the HadISST1 database (sea-ice remains the same as *amip-piForcing*). The ensemble size and time-periods covered for each experiment and AGCM is indicated. *amip-piForcing* simulations included in the CFMIP3 (Webb et al. 2017) contribution to CMIP6 are indicated by a y/n. The corresponding name of each AGCMs parent AOGCM is indicated. Global-annual-ensemble-mean dT and dN timeseries data are available for all *amip-piForcing* and *hadSST-piForcing* AGCM simulations (see Data Availability Statement).

AGCM	**Corresponding AOGCM name**	**Model description**	***amip-piForcing***			***hadSST-piForcing***	
			CMIP6? (y/n)	**Ensemble size**	**Time-period covered**	**Ensemble size**	**Time-period covered**
CAM4	CCSM4	Neale et al. (2013)	n	3	1870 – 2014	3	1870 – 2014
CESM2	unchanged	Danabasoglu et al. (2020)	y	1	1870 – 2014	1	1870 - 2015
CNRM-CM6-1	unchanged	Voldoire et al. (2019)	y	1	1870 – 2014	-	-
CanESM5	unchanged	Swart et al. (2019)	y	3	1870 – 2014	-	-
ECHAM6.3	MPI-ESM1.1	Mauritsen et al. (2019)	n	5	1871 – 2010	5	1871 – 2015
GFDL-AM3	GFDL-CM3	Donner et al. (2011)	n	1	1870 – 2014	1	1870 – 2014
GFDL-AM4	GFDL-CM4	Held et al. (2019)	n	1	1870 – 2016	1	1870 – 2016
HadAM3	HadCM3	Pope et al. (2000)	n	4	1871 – 2012	4	1871 – 2012
HadGEM2	HadGEM2-ES	Martin et al. (2011)	n	4	1871 – 2012	1	1871 – 2012
HadGEM3-GC31-LL	unchanged	Williams et al. (2017)	y	1	1870 – 2014	1	1871 – 2016
IPSL-CM6A-LR	unchanged	Boucher et al. (2020)	y	1	1870 – 2014	-	-
MIROC6	unchanged	Tatebe et al. (2019)	y	1	1870 – 2014	-	-
MRI-ESM2-0	unchanged	Yukimoto et al. (2019), Kawai et al. (2019)	y	1	1870 – 2014	-	-
MPI-ESM1-2-LR	unchanged	Mauritsen et al. (2019)	n	3	1871 – 2017	3	1871 – 2017

Table 2: Feedback parameter in *amip-piForcing* and *hadSST-piForcing* simulations over various historical time-periods, as well as *abrupt-4xCO2* sensitivity parameters. λ values from *amip-piForcing* and *hadSST-piForcing* are calculated from OLS regression ($\lambda = dN/dT$) over the relevant time-periods using global-annual-mean timeseries data. F_{2xCO2} is calculated as $F_{4xCO2}/2$ and ECS=$-F_{2x}/\lambda_{4xCO2}$ from 150 years of *abrupt-4xCO2* experiments (λ_{4xCO2} calculated over years 1-20 and 21-150 is also shown) (see Andrews et al., 2012; 2015).

	abrupt-4xCO2					$\lambda_{1871-2010}$ (W m^{-2} K^{-1})		$\lambda_{1871-1980}$ (W m^{-2} K^{-1})		$\lambda_{1981-2010}$ (W m^{-2} K^{-1})	
	ECS (K)	**F_{2x} (W m^{-2})**	**λ_{4xCO2} (W m^{-2} K^{-1})**	**λ_{4xCO2_1-20} (W m^{-2} K^{-1})**	**λ_{4xCO2_21-150} (W m^{-2} K^{-1})**	**AMIP**	**HadISST1**	**AMIP**	**HadISST1**	**AMIP**	**HadISST1**
CAM4	2.95	3.64	-1.23	-1.52	-0.94	-2.14	-1.77	-1.22	-1.45	-2.84	-2.70
CESM2	5.16	3.39	-0.66	-1.17	-0.49	-1.93	-1.49	-0.87	-0.95	-3.08	-2.92
CNRM-CM6-1	4.88	3.66	-0.75	-0.93	-0.87	-1.23	-	-1.10	-	-1.64	-
CanESM5	5.61	3.64	-0.65	-0.70	-0.59	-1.44	-	-0.93	-	-1.83	-
ECHAM6_3	3.01	4.10	-1.36	-1.47	-1.08	-1.92	-1.57	-1.43	-1.38	-2.69	-2.42
GFDL-AM3	3.99	2.97	-0.74	-1.13	-0.61	-1.44	-1.35	-0.72	-0.99	-1.90	-1.41
GFDL-AM4	3.84	3.32	-0.86	-1.54	-0.60	-1.84	-1.66	-1.33	-1.40	-2.57	-2.93
HadAM3	3.37	3.52	-1.04	-1.25	-0.75	-1.65	-1.44	-1.35	-1.40	-2.19	-1.86
HadGEM2	4.62	2.90	-0.63	-0.81	-0.33	-1.39	-1.04	-1.12	-1.08	-2.26	-1.54
HadGEM3-GC31-LL	5.54	3.49	-0.63	-0.81	-0.60	-1.28	-1.01	-0.95	-0.84	-1.87	-1.55
IPSL-CM6A-LR	4.56	3.41	-0.75	-0.98	-0.61	-1.59	-	-1.17	-	-2.50	-
MIROC6	2.58	3.72	-1.44	-1.61	-1.60	-1.42	-	-1.21	-	-1.87	-
MRI-ESM2-0	3.13	3.44	-1.10	-1.68	-0.78	-1.93	-	-1.23	-	-2.79	-
MPI-ESM1-2-LR	3.02	4.21	-1.39	-1.61	-1.34	-1.88	-1.58	-1.30	-1.45	-2.55	-2.42
MEAN	4.02	3.53	-0.95	-1.23	-0.80	-1.65	-1.43	-1.14	-1.21	-2.33	-2.19
1.645σ	1.64	0.57	0.49	0.54	0.55	0.46	0.41	0.33	0.38	0.72	0.95

Table 3: The pattern effect ($\Delta\lambda = \lambda_{4xCO2} - \lambda_{hist}$, with λ_{4xCO2} from years 1-150 of a*brupt-4xCO2*) between *abrupt-4xCO2* radiative feedback and radiative feedback calculated over different historical periods (i.e. λ_{hist} from 1871-2010, and its separation into 1871-1980 and 1981-2010) in *amip-piForcing* and *hadSST-piForcing*, as well as their difference.

	1871 – 2010 ($W\ m^{-2}\ K^{-1}$)			1871 – 1980 ($W\ m^{-2}\ K^{-1}$)			1981 – 2010 ($W\ m^{-2}\ K^{-1}$)		
	AMIP	**HadSST**	**Diff**	**AMIP**	**HadSST**	**Diff**	**AMIP**	**HadSST**	**Diff**
CAM4	0.90	0.53	0.37	-0.01	0.22	-0.23	1.60	1.47	0.13
CESM2	1.27	0.84	0.43	0.21	0.29	-0.08	2.43	2.26	0.17
CNRM-CM6-1	0.48			0.35			0.89		
CanESM5	0.80			0.28			1.19		
ECHAM6_3	0.56	0.21	0.35	0.07	0.02	0.05	1.32	1.06	0.26
GFDL-AM3	0.69	0.61	0.08	-0.03	0.24	-0.27	1.15	0.67	0.48
GFDL-AM4	0.97	0.80	0.17	0.47	0.53	-0.06	1.70	2.07	-0.37
HadAM3	0.61	0.40	0.21	0.31	0.35	-0.04	1.15	0.82	0.33
HadGEM2	0.76	0.41	0.35	0.49	0.45	0.04	1.63	0.91	0.72
HadGEM3-GC31-LL	0.65	0.38	0.27	0.32	0.21	0.11	1.24	0.92	0.32
IPSL-CM6A-LR	0.84			0.43			1.76		
MIROC6	-0.02			-0.23			0.42		
MRI-ESM2-0	0.83			0.14			1.69		
MPI-ESM1-2-LR	0.49	0.19	0.30	-0.09	0.06	-0.15	1.16	1.03	0.13
MEAN	0.70	0.48	0.28	0.19	0.26	-0.07	1.38	1.24	0.24
1.645σ	0.47	0.36	0.17	0.35	0.26	0.20	0.75	0.88	0.46

Table 4: Comparison of the 1985-2014 climate resistance (ρ = dF/dT), feedback parameter (-λ = -d(N – F)/dT and ocean heat uptake efficiency (κ = dN/dT) using different versions of the DEEP-C (Allan et al., 2014) satellite based reconstruction of dN (see Section 2.4). The lower half of the table shows how ρ, λ and κ estimates change as the 30 year moving window advances to 1990-2019. In all calculations HadCRUT5 analysis dT (Morice et al. 2021) and IPCC AR6 dF (Forster et al., 2021; Smith et al., 2021) are used. Years 1991-2 are excluded from the calculation as these years are identified as being strongly impacted by the volcanic forcing from the Pinatubo eruption (Section 4).

dN dataset version	Start year	End year	ρ ($W\ m^{-2}\ K^{-1}$)	-λ ($W\ m^{-2}\ K^{-1}$)	κ ($W\ m^{-2}\ K^{-1}$)
DEEP-C v2G			2.38	2.24	0.14
DEEP-C v3			2.38	2.24	0.14
DEEP-C v3G	1985	2014	2.38	2.24	0.14
DEEP-C v4			2.38	1.98	0.41
DEEP-C v5			2.38	1.98	0.41
DEEP-C v5	1986	2015	2.38	1.75	0.63
DEEP-C v5	1987	2016	2.25	1.55	0.70
DEEP-C v5	1988	2017	2.21	1.62	0.59
DEEP-C v5	1989	2018	2.23	1.66	0.57
DEEP-C v5	1990	2019	2.30	1.44	0.86

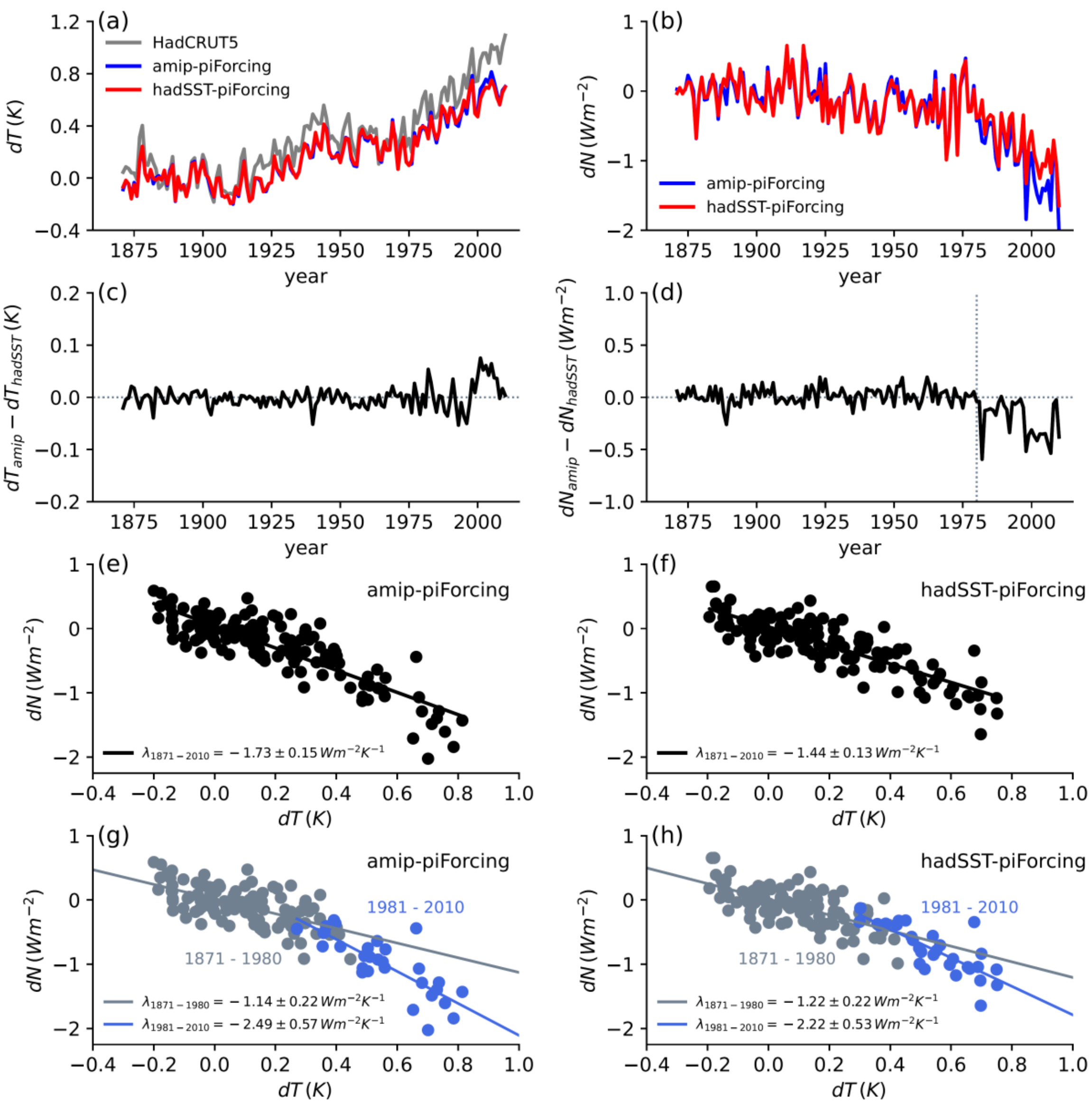

Figure 1: Comparison of multi-model ensemble-annual-mean (a) d*T* and (b) d*N* in the amip-*piForcing* and *hadSST-piForcing* simulations. (c) and (d) shows the difference in d*T* and d*N* respectively, highlighting 1980 as a key year where the d*N* response diverges according to the SST dataset. In (a) the HadCRUT5 observed d*T* evolution is shown for comparison. (e) and (f) show the relationship between global-annual-mean d*T* and d*N* in *amip-piForcing* and *hadSST-piForcing* respectively, where λ=d*N*/d*T* is calculated from OLS regression on the global-annual-mean data points. The stated 5-95% uncertainty is ±1.645σ from the standard error of the linear fit. (g) and (h) show the d*T* and d*N* relationship separated into two time-periods: years 1871-1980 (grey) and years 1981-2010 (blue). The multi-model ensemble-means are restricted to the nine AGCMs that performed both simulations (see Table 1).

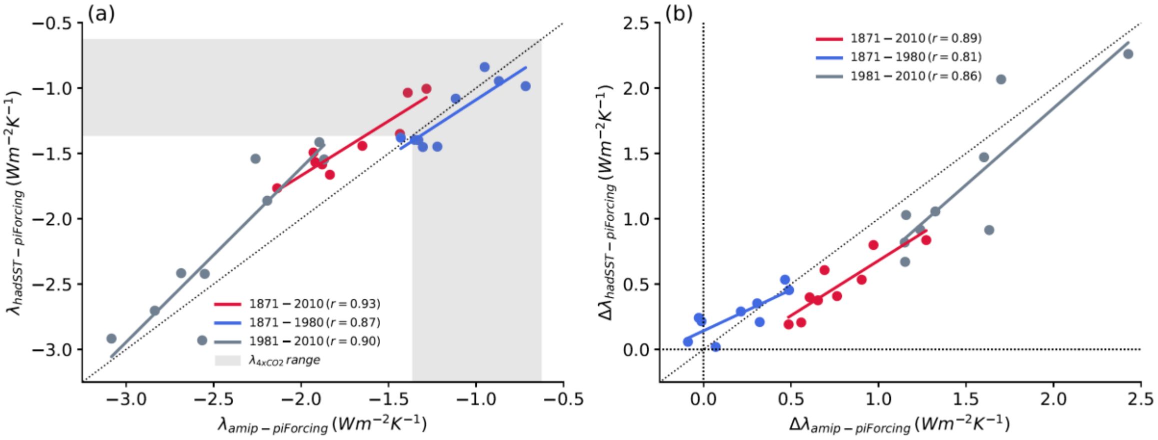

Figure 2: (a) Relationship between the feedback parameter, λ, in the *amip-piForcing* and *hadSST-piForcing* simulations over various historical time-periods. Each point is a single AGCM. The shaded grey region shows the range of λ_{4xCO2} from the AGCMs corresponding parent AOGCM *abrupt-4xCO2* simulation. The one-to-one line (dotted) is shown. (b) Relationship between the pattern effect, $\Delta\lambda = \lambda_{4xCO2} - \lambda_{hist}$, diagnosed from the *amip-piForcing* and *hadSST-piForcing* simulations over various historical time-periods.

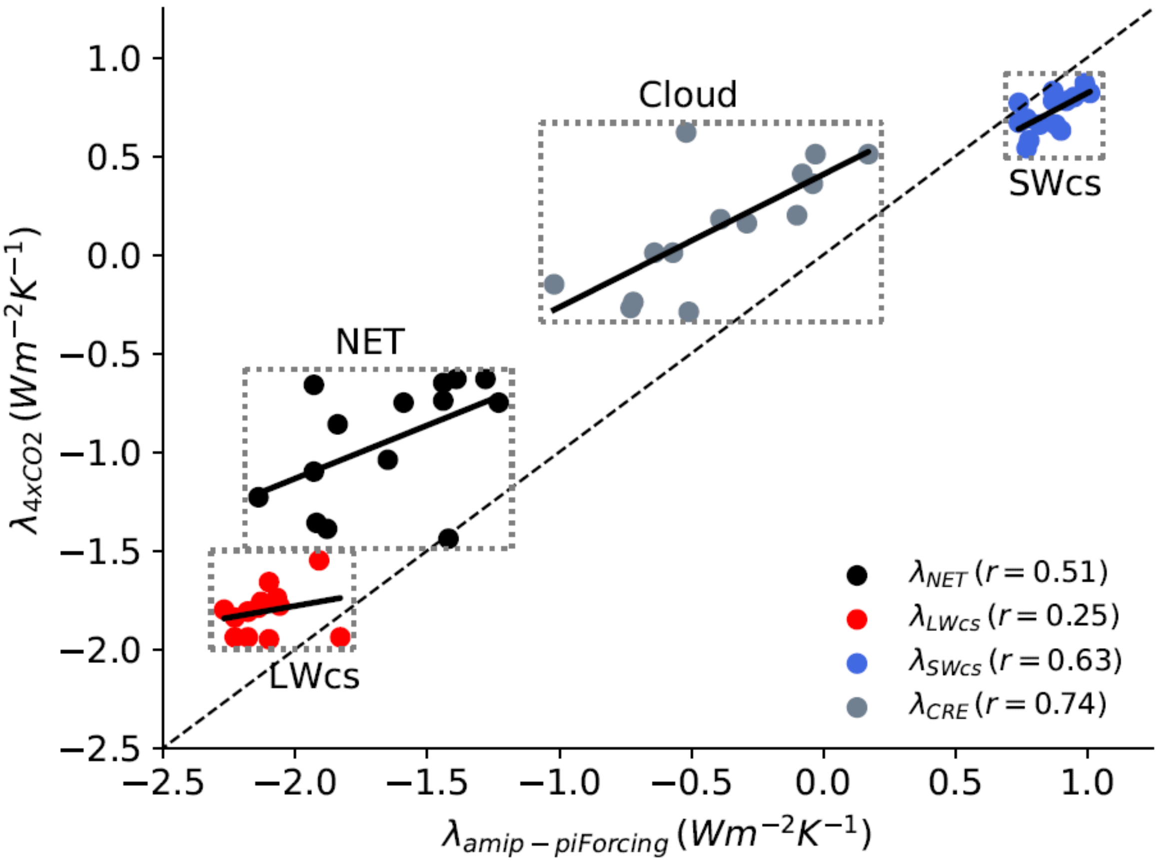

Figure 3: Relationship across models (dots) between the feedback parameter in *amip-piForcing* (calculated over years 1871-2010) and *abrupt-4xCO2* simulation (calculated over years 1-150). The net feedback parameter is decomposed into its longwave clear-sky, SW clear-sky and cloud radiative effect components.

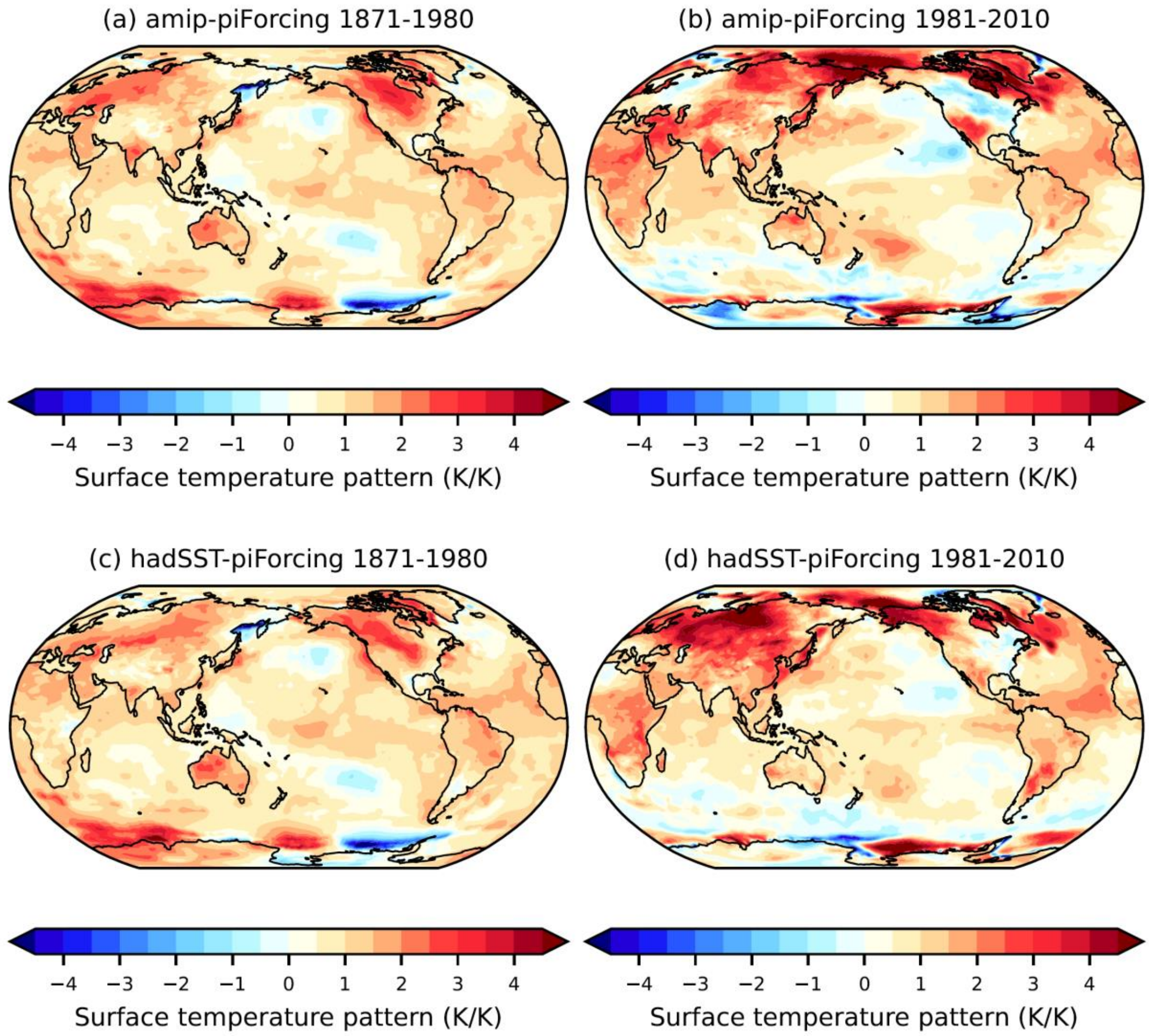

Figure 4: Pattern of near-surface temperature change (local d*T* per global-mean d*T*) for the time-periods 1870-1980 and 1981-2010 in (a) and (b) *amip-piForcing* and (c) and (d) *hadSST-piForcing*. Patterns are calculated from the slope of the linear regression of local temperature change against global-mean temperature change using annual-mean data points. Note that by definition the global-means are unity. Data from HadGEM3-GC31-LL simulations have been used for this illustration.

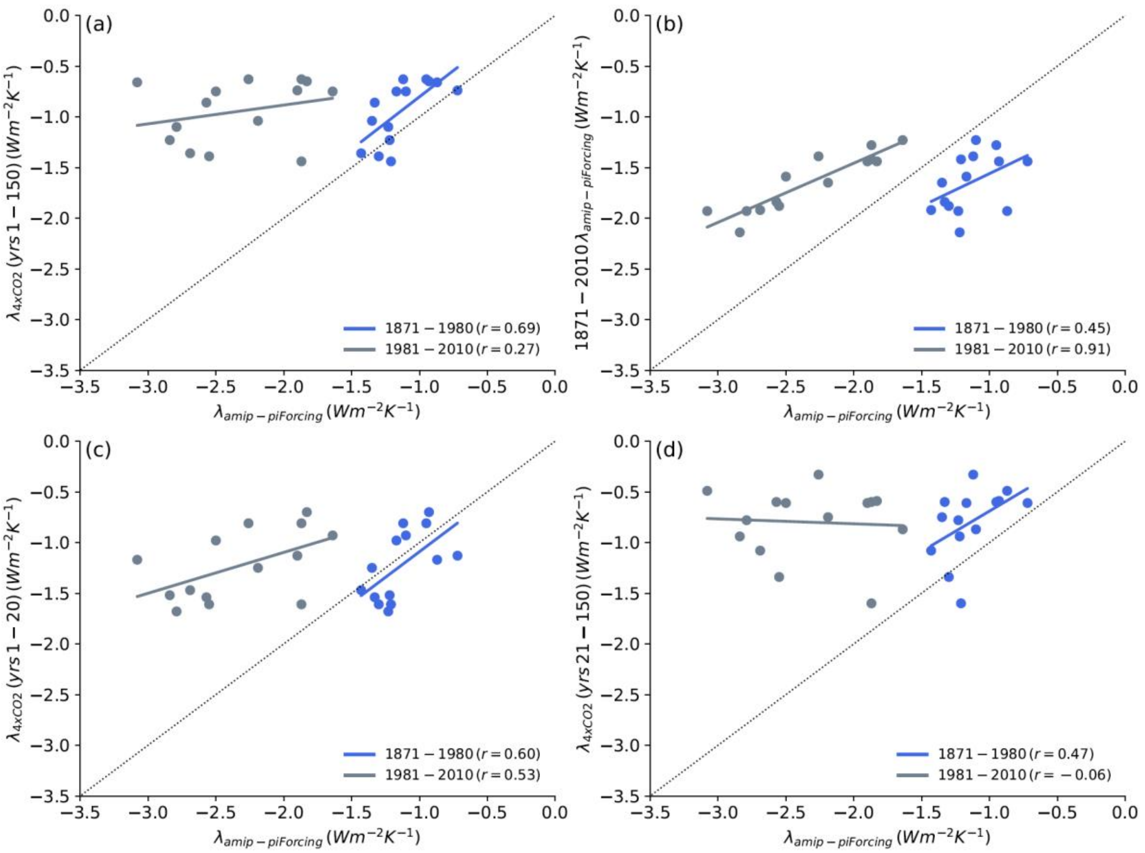

Figure 5: Relationships between model simulated feedbacks in *amip-piForcing* over years 1871-1980 (blue) or 1981-2010 (grey) and (a) λ_{4xCO2} from *abrupt-4xCO2*, (b) λ_{hist} over the entire historical record (1871-2010), (c) λ_{4xCO2} from *abrupt-4xCO2* over years 1-20 and (d) years 21-150.

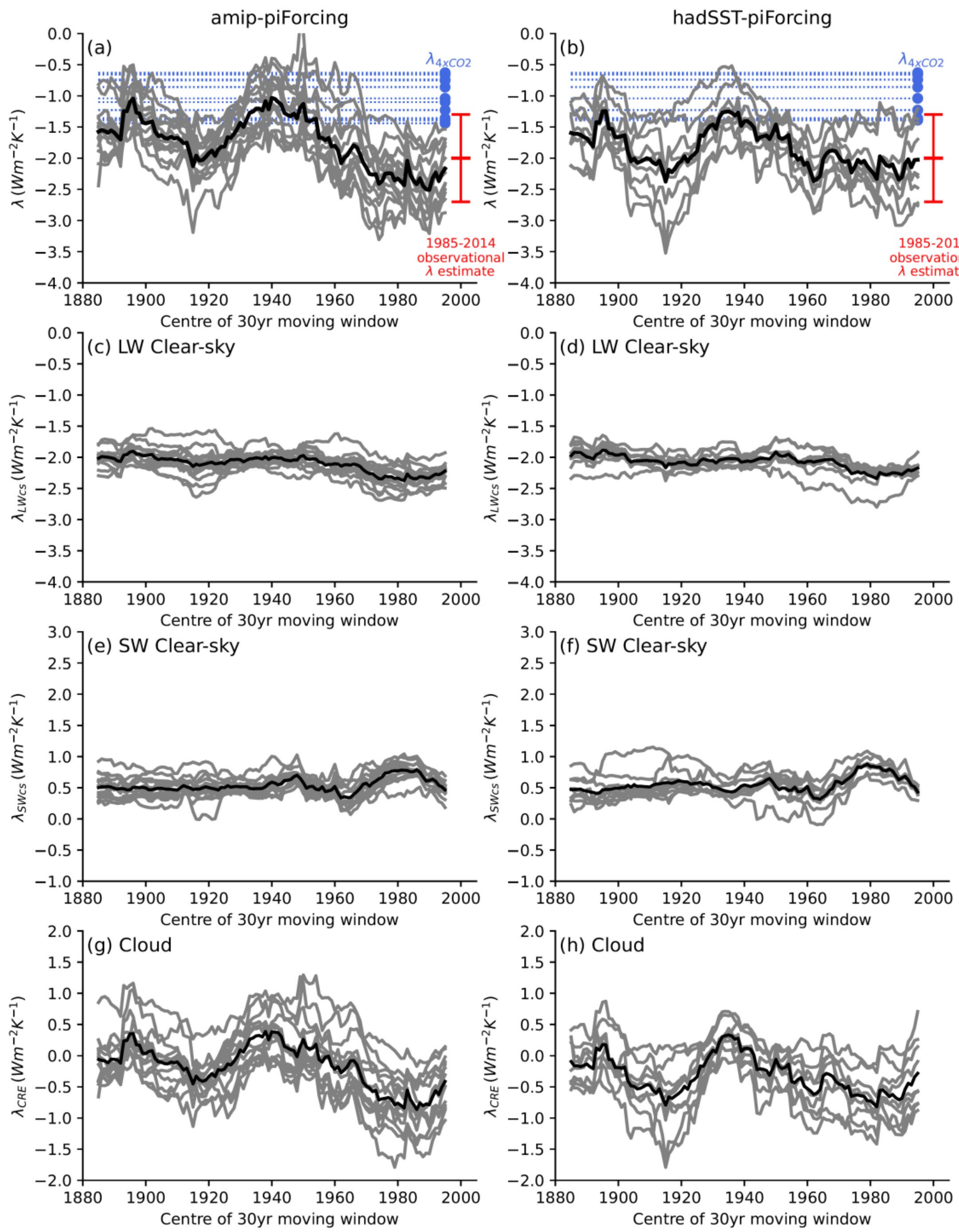

Figure 6: Decadal variation in the feedback parameter λ from 1871 to 2010. Left column shows results from *amip-piForcing* and right column shows results from *hadSST-piForcing*. Each grey line represents a single AGCM (see Table 1). Thick black is the ensemble-mean of the results. X-axis represents the centre of a 30 year moving window in which λ=d*N*/d*T* is calculated from OLS regression on annual-mean data, i.e. λ at 1980.5 represents the feedback parameter over years 1966 to 1995. Shown in (a) and (b) is the net feedback parameter. Blue dots and lines represent the corresponding λ_{4xCO2} values from AOGCM *abrupt-4xCO2* simulations (Table 2). Red shows an observational estimate and 5-95% uncertainty of λ=d(*N* – *F*)/d*T* ~ -2.0 ± 0.7 W m^{-2} K^{-1} over years 1985-2014 (see Section 4). (c) – (h) shows the corresponding LW clear-sky, SW clear-sky and cloud radiative effect (CRE) components of λ.

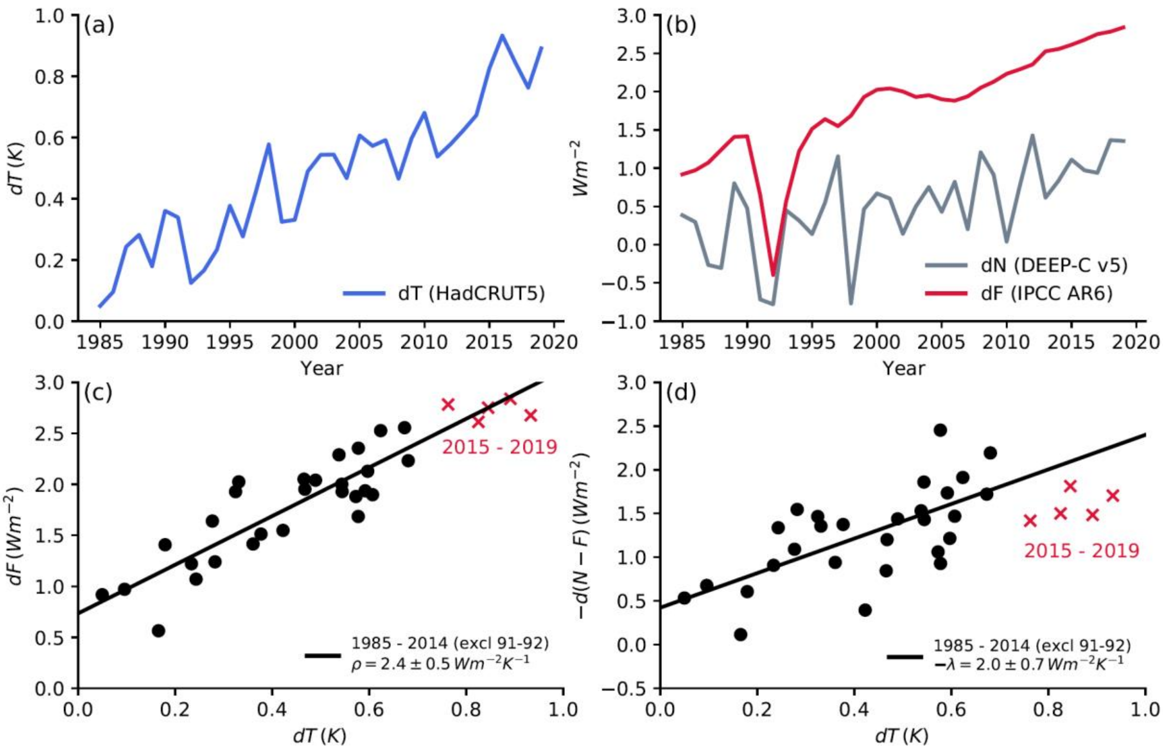

Figure 7: Observational estimate of the Earth's 1985-2019 energy balance. All points are global-annual-means. (a) d*T* (HadCRUT5 analysis dataset; Morice et al., 2021), (b) d*N* (DEEP-C v5; Allan et al., 2014; Liu and Allan, 2022) and d*F* (IPCC AR6; Forster et al., 2021; Smith et al., 2021). (c) ρ = d*F*/d*T* relationship and (d) -λ_{hist}=-d(*N* – *F*)/d*T* relationship over years 1985-2014. Black dots are global-annual means over years 1985-2014 excluding years 1991-2 which are strongly influenced by the Pinatubo explosive volcanic eruption (see red line panel b). Red points in (c) and (d) are years 2015-2019. The stated 5-95% uncertainties are ±1.645σ from the standard error of the linear fit.

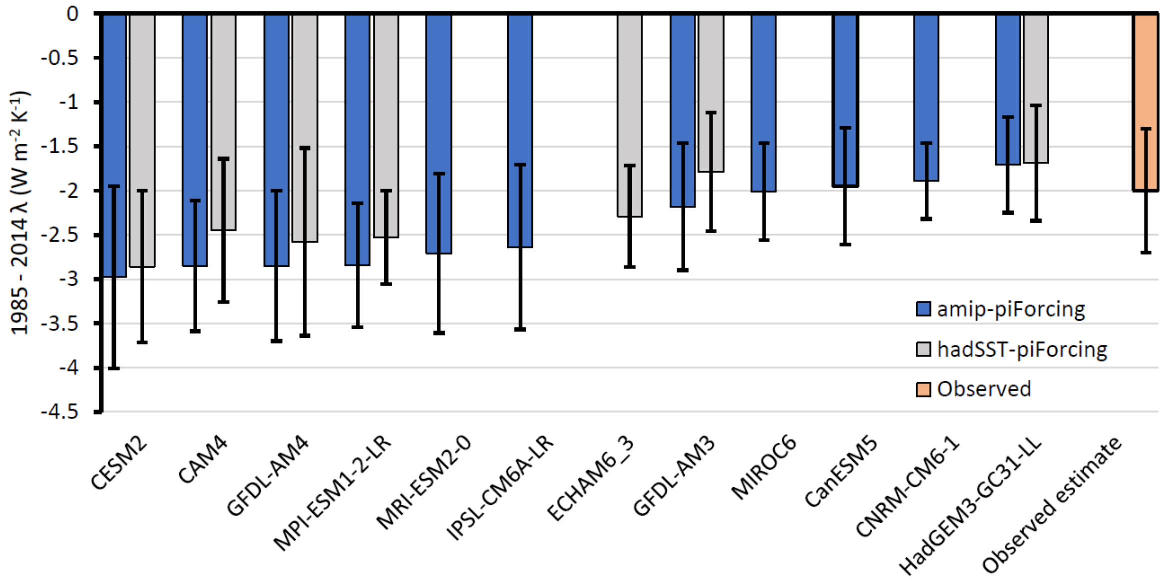

Figure 8: Comparison of the 1985-2014 feedback parameter, $\lambda_{hist} = d(N - F)/dT$, in *amip-piForcing* and *hadSST-piForcing* simulations to an observed estimate based on DEEP-C V5 d*N* (Allan et al., 2014; Liu and Allan, 2022), HadCRUT5 analysis d*T* (Morice et al. 2021) and IPCC AR6 d*F* (Forster et al., 2021; Smith et al., 2021). The 5-95% uncertainty is simply 1.645σ from the standard error of the linear fit, with no allowance for systematic uncertainties. Note also that years 1991-2 are excluded from the calculation as these years are identified as being strongly impacted by the volcanic forcing from the Pinatubo eruption (Figure 7b).

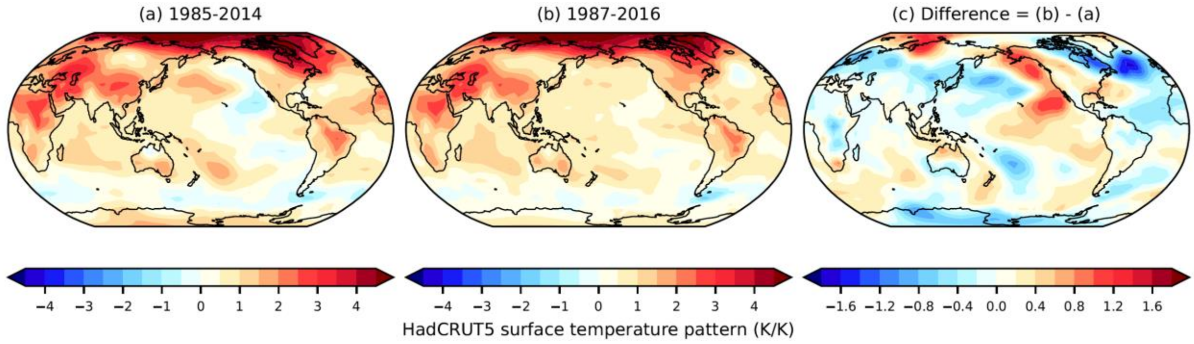

Figure 9: Pattern of near-surface temperature change (local d*T* per global-mean d*T*) for the time-periods (a) 1985-2014 and (b) 1987-2016, and (c) shows the difference (b minus a). Data is the HadCRUT5 analysis dataset (Morice et al. 2021). Patterns are calculated from the slope of the linear regression of local temperature change against global-mean temperature change using annual-mean data points. Note that by definition the global-means of panels (a) and (b) are unity.

On the effect of historical SST patterns on radiative feedback

Timothy Andrews et al.

Supplementary Tables

Contents:

Table S1: Longwave clear-sky feedback parameters in *amip-piForcing* and *hadSST-piForcing* simulations over various historical time-periods, as well as *abrupt-4xCO2* sensitivity parameters.

	abrupt-4xCO2			$\lambda_{1871-2010}$ (W m^{-2} K^{-1})		$\lambda_{1871-1980}$ (W m^{-2} K^{-1})		$\lambda_{1981-2010}$ (W m^{-2} K^{-1})	
	λ_{4xCO2} (W m^{-2} K^{-1})	λ_{4xCO2_1-20} (W m^{-2} K^{-1})	λ_{4xCO2_21-150} (W m^{-2} K^{-1})	**AMIP**	**HadISST1**	**AMIP**	**HadISST1**	**AMIP**	**HadISST1**
CAM4	-1.95	-1.99	-1.90	-2.10	-2.07	-1.99	-2.03	-2.19	-2.20
CESM2	-1.81	-1.88	-1.74	-2.18	-2.07	-2.01	-1.95	-2.53	-2.26
CNRM-CM6-1	-1.76	-1.81	-1.74	-2.13		-1.91		-2.23	
CanESM5	-1.84	-1.89	-1.81	-2.23		-2.17		-2.33	
ECHAM6_3	-1.74	-1.75	-1.68	-2.07	-2.03	-1.94	-1.93	-2.20	-2.19
GFDL-AM3	-1.94	-2.03	-1.93	-2.18	-2.20	-1.88	-1.97	-2.34	-2.28
GFDL-AM4	-1.81	-1.90	-1.78	-2.18	-2.14	-2.03	-2.07	-2.23	-2.32
HadAM3	-1.79	-1.84	-1.71	-2.14	-2.08	-2.04	-2.02	-2.22	-2.16
HadGEM2	-1.66	-1.81	-1.64	-2.10	-2.08	-1.96	-1.97	-2.16	-1.94
HadGEM3-GC31-LL	-1.80	-1.88	-1.78	-2.27	-2.17	-2.08	-2.07	-2.28	-2.24
IPSL-CM6A-LR	-1.55	-1.58	-1.54	-1.91		-1.81		-1.95	
MIROC6	-1.94	-1.99	-1.91	-1.83		-1.78		-2.15	
MRI-ESM2-0	-1.94	-2.04	-1.86	-2.23		-1.94		-2.47	
MPI-ESM1-2-LR	-1.78	-1.81	-1.78	-2.06	-2.00	-1.89	-1.91	-2.13	-2.16
MEAN	-1.81	-1.87	-1.77	-2.12	-2.09	-1.96	-1.99	-2.24	-2.20
1.645*sigma	0.18	0.19	0.17	0.19	0.10	0.17	0.09	0.23	0.17

Table S2: Shortwave clear-sky feedback parameters in *amip-piForcing* and *hadSST-piForcing* simulations over various historical time-periods, as well as *abrupt-4xCO2* sensitivity parameters.

	abrupt-4xCO2				$\lambda_{1871\text{-}2010}$ (W m^{-2} K^{-1})			$\lambda_{1871\text{-}1980}$ (W m^{-2} K^{-1})			$\lambda_{1981\text{-}2010}$ (W m^{-2} K^{-1})	
	λ_{4xCO2} (W m^{-2} K^{-1})	$\lambda_{4xCO2\_1\text{-}20}$ (W m^{-2} K^{-1})	$\lambda_{4xCO2\_21\text{-}150}$ (W m^{-2} K^{-1})		**AMIP**	**HadISST1**		**AMIP**	**HadISST1**		**AMIP**	**HadISST1**
CAM4	0.87	0.84	0.89		0.99	0.98		0.77	0.73		0.50	0.39
CESM2	0.54	0.72	0.44		0.77	0.88		0.74	0.83		0.40	0.29
CNRM-CM6-1	0.82	0.84	0.60		1.01			0.72			0.47	
CanESM5	0.78	0.82	0.74		0.87			0.75			0.58	
ECHAM6_3	0.66	0.67	0.69		0.88	0.90		0.61	0.63		0.42	0.41
GFDL-AM3	0.69	0.65	0.67		0.77	0.76		0.65	0.64		0.63	0.43
GFDL-AM4	0.77	0.79	0.67		0.74	0.75		0.59	0.58		0.26	0.36
HadAM3	0.58	0.58	0.58		0.78	0.79		0.57	0.55		0.43	0.46
HadGEM2	0.67	1.05	0.77		0.74	0.99		0.56	0.68		0.15	0.33
HadGEM3-GC31-LL	0.66	0.74	0.56		0.82	0.90		0.70	0.75		0.33	0.48
IPSL-CM6A-LR	0.80	0.78	0.81		0.95			0.72			0.46	
MIROC6	0.78	0.75	0.63		0.92			0.91			0.41	
MRI-ESM2-0	0.83	0.97	0.81		0.87			0.68			0.35	
MPI-ESM1-2-LR	0.63	0.52	0.61		0.90	0.91		0.63	0.63		0.39	0.33
MEAN	0.72	0.76	0.68		0.86	0.87		0.69	0.67		0.41	0.39
1.645*sigma	0.16	0.22	0.19		0.14	0.14		0.15	0.14		0.19	0.10

Table S3: Cloud radiative effect feedback parameters in *amip-piForcing* and *hadSST-piForcing* simulations over various historical time-periods, as well as *abrupt-4xCO2* sensitivity parameters.

	abrupt-4xCO2				**$\lambda_{1871\text{-}2010}$ (W m^{-2} K^{-1})**			**$\lambda_{1871\text{-}1980}$ (W m^{-2} K^{-1})**			**$\lambda_{1981\text{-}2010}$ (W m^{-2} K^{-1})**	
	λ_{4xCO2} (W m^{-2} K^{-1})	**$\lambda_{4xCO2_1\text{-}20}$ (W m^{-2} K^{-1})**	**$\lambda_{4xCO2_21\text{-}150}$ (W m^{-2} K^{-1})**		**AMIP**	**HadISST1**		**AMIP**	**HadISST1**		**AMIP**	**HadISST1**
CAM4	-0.15	-0.37	0.08		-1.02	-0.67		0.00	-0.15		-1.15	-0.89
CESM2	0.62	-0.01	0.81		-0.52	-0.30		0.40	0.18		-0.96	-0.95
CNRM-CM6-1	0.20	0.03	0.27		-0.10			0.10			0.12	
CanESM5	0.41	0.37	0.48		-0.08			0.49			-0.09	
ECHAM6_3	-0.27	-0.39	-0.08		-0.73	-0.45		-0.10	-0.08		-0.91	-0.64
GFDL-AM3	0.51	0.25	0.65		-0.03	0.09		0.51	0.34		-0.18	0.43
GFDL-AM4	0.18	-0.43	0.51		-0.39	-0.27		0.10	0.09		-0.60	-0.97
HadAM3	0.16	0.01	0.38		-0.29	-0.15		0.11	0.07		-0.41	-0.16
HadGEM2	0.36	-0.05	0.54		-0.04	0.05		0.28	0.21		-0.26	0.07
HadGEM3-GC31-LL	0.51	0.33	0.61		0.17	0.26		0.43	0.48		0.08	0.21
IPSL-CM6A-LR	0.01	-0.17	0.13		-0.64			-0.08			-1.01	
MIROC6	-0.29	-0.36	-0.32		-0.51			-0.34			-0.12	
MRI-ESM2-0	0.01	-0.60	0.27		-0.57			0.02			-0.68	
MPI-ESM1-2-LR	-0.24	-0.32	-0.17		-0.72	-0.49		-0.05	-0.17		-0.82	-0.59
MEAN	0.14	-0.12	0.30		-0.39	-0.21		0.13	0.11		-0.50	-0.39
1.645*sigma	0.49	0.48	0.53		0.54	0.47		0.40	0.34		0.68	0.83

Table S4: Growth of the historical feedback parameter, λ_{hist}, from 2010 to 2014 in *amip-piForcing* and *hadSST-piForcing*. Shown is λ_{hist} calculated over 1871-2010 and 1871-2014, and their difference.

	AMIP λ_{hist} (W m^{-2} K^{-1})				HadSST λ_{hist} (W m^{-2} K^{-1})		
	1871-2010	**1871-2014**	**change**		**1871-2010**	**1871-2014**	**change**
CAM4	-2.14	-2.24	-0.10		-1.77	-1.81	-0.05
CESM2	-1.93	-2.09	-0.16		-1.49	-1.59	-0.10
CNRM-CM6-1	-1.23	-1.27	-0.04		-	-	-
CanESM5	-1.44	-1.48	-0.04		-	-	-
GFDL-AM3	-1.44	-1.48	-0.04		-1.35	-1.38	-0.03
GFDL-AM4	-1.84	-1.90	-0.07		-1.66	-1.68	-0.01
HadGEM3-GC31-LL	-1.28	-1.33	-0.04		-1.01	-1.09	-0.08
IPSL-CM6A-LR	-1.59	-1.65	-0.06		-	-	-
MIROC6	-1.42	-1.50	-0.08		-	-	-
MRI-ESM2-0	-1.93	-1.97	-0.05		-	-	-
MPI-ESM1-2-LR	-1.88	-1.92	-0.04		-1.58	-1.64	-0.06
MEAN	-1.65	-1.71	-0.07		-1.48	-1.53	-0.05
1.645*sigma	0.48	0.51	0.06		0.41	0.39	0.05