⓸Negligible Unforced Historical Pattern Effect on Climate Feedback Strength Found in HadISST-Based AMIP Simulations

NICHOLAS LEWIS

Bath, United Kingdom

THORSTEN MAURITSEN

Department of Meteorology, Stockholm University, Stockholm, Sweden

ABSTRACT: Recently it has been suggested that natural variability in sea surface temperature (SST) patterns over the historical period causes a low bias in estimates of climate sensitivity based on instrumental records, in addition to that suggested by time variation of the climate feedback parameter in atmospheric general circulation models (GCMs) coupled to dynamic oceans. This excess, unforced, historical "pattern effect" (the effect of evolving surface temperature patterns on climate feedback strength) has been found in simulations performed using GCMs driven by AMIPII SST and sea ice changes (amipPiForcing). Here we show, in both amipPiForcing experiments with one GCM and by using Green's functions derived from another GCM, that whether such an unforced historical pattern effect is found depends on the underlying SST dataset used. When replacing the usual AMIPII SSTs with those from the HadISST1 dataset in amipPiForcing experiments, with sea ice changes unaltered, the first GCM indicates pattern effects that are indistinguishable from the forced pattern effect of the corresponding coupled GCM. Diagnosis of pattern effects using Green's functions derived from the second GCM supports this result for five out of six non-AMIPII SST reconstruction datasets. Moreover, internal variability in coupled GCMs is rarely sufficient to account for an unforced historical pattern effect of even one-quarter the strength previously reported. The presented evidence indicates that, if unforced pattern effects have been as small over the historical record as our findings suggest, they are unlikely to significantly bias climate sensitivity estimates that are based on long-term instrumental observations and account for forced pattern effects obtained from GCMs.

KEYWORDS: Sea surface temperature; Warm pool; Climate sensitivity; Feedback; Regression analysis

1. Introduction

It has become clear that, in general circulation climate models (GCMs) at least, the spatial pattern of the planet's surface warming, which is largely determined by that over the ocean, is a key factor controlling the global-mean radiative response to surface warming, or the global climate feedback parameter, creating so-called pattern effects (Stevens et al. 2016). Inference of Earth's climate sensitivity, on all time scales, based on centennial historical warming and a linear energy balance framework is sensitive to assumptions about the strength of pattern effects over the historical period and in the future, since these control the ratio of future warming, both transient and in equilibrium, to that estimated in a fixed-feedback response framework.

On physical grounds, warming in tropical (30°S–30°N) ascent regions relative to elsewhere is expected to produce a strongly cooling global radiative response, in the form of an increase in outgoing radiation at top of atmosphere, while warming in tropical descent regions alone should produce a warming global radiative response. That is because surface temperature in convective areas, of which the most important

is the Indo-Pacific warm pool, controls temperature in the tropical free troposphere, which spatially is fairly uniform, and influences temperature in the extratropics, while low clouds are a key determinant of outgoing shortwave radiation in tropical descent regions such as the eastern Pacific (Sobel et al. 2002). In addition to directly increasing outgoing longwave radiation, much of which is emitted from the troposphere, an increase in free tropospheric temperature relative to surface temperature in descent regions strengthens the boundary layer inversion, which is known to increase low cloud cover (Wood and Bretherton 2006; Ceppi and Gregory 2017). Moreover, ocean dynamics are thought to cause delayed warming in the eastern tropical Pacific, the extratropical Southern Ocean, and to an extent the North Atlantic, where either the mixed layer is deep and the thermohaline circulation has sinking branches or cool upwelling water influences SST (Winton et al. 2010; Andrews et al. 2015, 2018). Delayed warming increases inversion strength and hence low cloud cover, transiently causing a greater radiative response. Thus, following an increase in greenhouse gas concentrations or other imposition of a radiative forcing, the spatial pattern of surface warming changes over time, and a pattern effect arises. The climate system is expected to display a stronger radiative response to warming—a larger

---

Corresponding author: Nicholas Lewis, nhlewis@btinternet.com

climate feedback parameter λ, the change in top-of-atmosphere (TOA) global-mean outgoing radiative flux R caused by a unit increase in global-mean surface temperature T[1]—initially than subsequently. This constitutes a time-related forced pattern effect, attributable to a reduction in deep-ocean heat uptake over time.

Few AOGCMs (atmospheric GCMs coupled to a dynamic ocean model) have been run to equilibrium (Rugenstein et al. 2020). However, their equilibrium climate sensitivity (ECS)—the eventual global-mean surface warming following a doubling of preindustrial CO_2 concentration—and climate feedback can be estimated from their behavior following an initial step quadrupling of CO_2 concentration, in 150-yr-long abrupt4xCO2 simulations (Gregory et al. 2004; Andrews et al. 2012). Climate feedback can be estimated from abrupt4xCO2 simulation data, over any selected period, as the absolute slope of top-of-atmosphere radiative imbalance when regressed on T. The AOGCM response in these abrupt4xCO2 simulations can be well emulated as the sum of a fast response pattern that dominates over the first 20 years, but is almost complete by then, and thereafter a slow response pattern (initially suppressed by deep-ocean heat uptake) with an e-folding time scale of one to several hundred years (Held et al. 2010; Geoffroy and Saint-Martin 2014). The main differences between the patterns are that, relative to the global mean, the slow pattern shows much more warming poleward of 45°S, somewhat more warming in the eastern tropical/southeastern subtropical Pacific Ocean, and less warming in the tropical west Pacific and Indian Ocean, than does the fast pattern (Andrews et al. 2015). These changes appear to account, in the vast majority of AOGCMs, for climate feedback decreasing noticeably 20 years or so after a step in forcing is applied. As a result, deriving ECS by extrapolating an AOGCM's top-of-atmosphere radiative imbalance versus T relationship estimated by regression over years 1–150 of abrupt4xCO2 simulations, as is often done (e.g., Flato et al. 2013), tends to underestimate it. When we refer to ECS_{AOGCM}, we mean ECS estimated by regressing instead over years 21–150 of abrupt4xCO2 simulations. Doing so better captures the AOGCM slow-pattern response and usually gives slightly higher ECS_{AOGCM} estimates (Geoffroy et al. 2013; Lewis and Curry 2018; Geoffroy and Saint-Martin 2020). Andrews et al. (2015) estimated that, for the CMIP5 AOGCM mean, the CO_2-forced fast-to-slow pattern λ ratio is 1.65:1, with about 60% of the difference in global climate feedback between fast and slow responses coming from the tropics. As a result, in AOGCMs ECS_{AOGCM} generally exceeds effective climate sensitivity (EffCS), a proxy for ECS derived by dividing λ estimated from transient changes in R and T using a linear energy balance framework into the effective radiative forcing from a doubling of atmospheric CO_2 concentration (e.g., Senior and Mitchell 2000; Armour et al. 2013; Armour 2017; Lewis and Curry 2018).

In recent years, EffCS estimates for the real world based on observed warming over the historical period ($EffCS_{hist-obs}$)

have generally been in the range from 1.5 to 2.0 K (Otto et al. 2013; Lewis and Curry 2015, 2018; Mauritsen and Pincus 2017; Skeie et al. 2018). In AOGCMs λ and hence EffCS varies with time elapsed since forcing is imposed. Therefore, $EffCS_{hist-obs}$ estimates based on observed real-world warming should be compared with EffCS in AOGCMs over a period with a comparable forcing duration to that over the historical period [being from the third quarter of the nineteenth century to recently (between 2009 and 2016 in the studies cited)]. Forcing was generally not diagnosed in CMIP5 historical simulations, but comparable EffCS estimates for them ($EffCS_{hist-AOGCM}$) can be derived from appropriate subperiods of CMIP5 CO_2-forced simulations. Armour (2017) estimated $EffCS_{hist-AOGCM}$ from T and R responses at year 100 of 1% yr^{-1} CO_2 ramping (1pctCO2) simulations. Such $EffCS_{hist-AOGCM}$ estimates range from 1.9 to 4.1 K (Lewis and Curry 2018). The median excess in CMIP5 models of ECS_{AOGCM} over thus-estimated $EffCS_{hist-AOGCM}$ is approximately 10%. Armour (2017) derived a significantly higher than 10% excess of ECS_{AOGCM} over $EffCS_{hist-AOGCM}$. However, Lewis and Curry (2018) showed that Armour's estimate was biased high, mainly due to use of CO_2 forcing estimates reflecting neither the nonlogarithmic element of the forcing–concentration relationship nor all rapid adjustments. The ratio of ECS_{AOGCM} to $EffCS_{hist-AOGCM}$ estimates is much smaller than their ratio of fast to slow warming pattern climate feedback because much of the surface warming and radiative response to a forcing change occurs while the fast response pattern dominates, with the weaker slow warming pattern climate feedback only applying to the remaining response. Although the time-related forced pattern effect causes a ~10% excess of ECS_{AOGCM} over $EffCS_{hist-AOGCM}$, projected twenty-first-century warming constrained by the historical record is only ~3% lower if ECS_{AOGCM} is set equal to $EffCS_{hist-AOGCM}$ (Mauritsen and Pincus 2017; Geoffroy and Saint-Martin 2020).

EffCS applicable to the historical period mixture of non-volcanic forcings does not appear to differ from that to purely CO_2 forcing over a comparable period (Hansen et al. 2005; Lewis and Curry 2018, 2020; Richardson et al. 2019). That is consistent with the similarity of the surface temperature response patterns to historical forcings both individually and in combination being similar to that to CO_2 forcing (Hansen et al. 2005; Richardson et al. 2019). Moreover, over the historical period the linear trend in volcanic forcing, for which the applicable EffCS does seem different (Lewis and Curry 2015; Gregory et al. 2016, 2019), was negligible. An implication is that any forced pattern effect over the historical period arising from the mixture of forcings is likely to be small.

In addition to forced pattern effects, internal variability can produce unforced pattern effects (Gregory and Andrews 2016; Zhou et al. 2016; Mauritsen 2016; Dessler et al. 2018; Dessler 2020), in the sense of global-mean R varying with surface temperature patterns without any change in global-mean T. We use the term "unforced historical pattern effect" to refer to such internal variability caused deviation of the ratio of changes in R and T over the entire historical period (at a minimum, 1871–2005) from its underlying forced ratio, with a positive pattern effect corresponding to a more positive change in R than the underlying forced change. Unlike forced pattern

---

[1] We adopt thus adopt a sign convention under which a climate stabilizing λ is positive. The shorter terms "climate feedback" and "feedback" refer to the climate feedback parameter λ, unless the context requires otherwise.

effects, unforced pattern effects could, if changing negatively, lead to rapid near-term warming, and could bias estimation of the transient climate response and EffCS as well as ECS. Strictly, unforced pattern effects contribute to a random variability/error term rather than a change in climate feedback, since that relates to the causative effect of T on R. However, since measured changes in actual or simulated R include random variability, estimation of λ is affected by unforced pattern effects, particularly over relatively short periods (Marvel et al. 2018). A positive unforced historical pattern effect has been proposed as the explanation for λ in GCM AMIP simulations, in which they are driven over the historical period by evolving historical sea surface temperature (SST) and sea ice patterns from an observationally based dataset rather than by changing forcing (amipPiForcing simulations), being surprisingly large (Gregory and Andrews 2016; Andrews et al. 2018; Gregory et al. 2019). The large amipPiForcing estimated λ arises mainly from a strong increase in simulated R over the last four decades, during which period CO_2 forcing increases have dominated changes in non-CO_2 forcing (in particular, aerosol forcing) considerably more strongly than previously. If a positive unforced historical pattern effect did occur, EffCS estimates based on observed warming over the historical period would be biased low, even if forced pattern effects are accounted for.

Here we investigate a simpler alternative explanation, namely that the apparent positive unforced historical pattern effect may largely or wholly be an artifact arising from use in amipPiForcing simulations of a particular melded, observationally based infilled SST reconstruction (AMIPII; Hurrell et al. 2008), the warming patterns of which may be questionable (Flannaghan et al. 2014; Fueglistaler et al. 2015). As further discussed in section 4, the AMIPII dataset merges two SST reconstructions that employ different bias correction and interpolation methods, and its post-1981 interpolation method may be suboptimal for the study of pattern effects.

We show that using alternative, internally consistent SST reconstructions, which show weaker relative warming in the tropical Indian Ocean and the west Pacific than the AMIPII reconstruction, yields substantially lower λ estimates in amipPiForcing simulations by the ECHAM6.3 GCM. These λ estimates are close to those from forced simulations by the parent AOGCM and hence do not indicate any unforced historical pattern effect. We find similar results using a CAM5.3-based Green's function approach (Zhou et al. 2017), and investigate other SST reconstructions.

2. Methods and data

a. ECHAM6.3 simulations

We investigated historical pattern effects in two ways. First, we carried out an ensemble of GCM amipPiForcing experiments over 1871–2010, using ECHAM6.3 (Mauritsen et al. 2019). ECHAM6.3 is an advanced atmosphere-only GCM, with improved representation of clouds, convection, tropospheric aerosol, and radiative transfer, compared to its predecessor CMIP5 version, and was run with a T63 spectral truncation and 47 vertical levels. Our amipPiForcing SST boundary conditions were based on the HadISST1 dataset. The

same sea ice boundary conditions as in the AMIPII dataset were used, so that differences between the results using AMIPII and HadISST1 SST data are unrelated to changes in sea ice. Natural and anthropogenic forcings were held constant at preindustrial levels. Accordingly, nonrandom variations in radiative fluxes reflect only changes in SST and sea ice boundary conditions. That enables changes in R to be obtained directly from top-of-atmosphere radiation fields, and thus climate feedback to be diagnosed. Five simulation runs with slightly different initial conditions were executed; except where stated otherwise results are based on their ensemble-mean values. Five simulation runs identical to the current simulations apart from using AMIPII SST boundary conditions have already been carried out using ECHAM6.3 (Andrews et al. 2018).

Using ECHAM6.3 has the major advantage that the effective radiative forcing (ERF) during historical simulations by its parent AOGCM, MPI-ESM1.1, can be accurately estimated from diagnostic fixed-SST simulations. Moreover, data from very large ensembles of MPI-ESM1.1 historical and 1pctCO2 simulations are available. In this model it is therefore, uniquely, possible to accurately estimate climate feedback strength in response both to composite historical period ERF and to a broadly similar time-profile of CO_2-only ramped ERF, and to compare the radiative response in the amipPiForcing simulations with that expected on the basis of those climate feedback estimates. A disadvantage of using a single GCM is that we cannot sample model uncertainties. The model exhibited slightly less than average historical pattern effect in the model intercomparison of Andrews et al. (2018), and results should be interpreted in light of this.

As in Zhou et al. (2017), we measure T as surface (skin) temperature globally; over ice-free ocean this is treated in this GCM as equaling SST. This measure (denoted T_s), as well as enabling use of the Green's functions derived by Zhou et al. (2017), is preferred to using near-surface air temperature (denoted T) for several other reasons. First, because doing so reduces temperature noise in SST-driven GCM experiments, improving feedback estimation. Second, because near-surface air temperature is not well measured over the oceans in the real climate system. Third, because the radiative response arises primarily from changes in surface temperature not in near-surface air temperature. Finally, because near-surface air temperature, unlike surface temperature, is a diagnostic rather than prognostic variable in GCMs (Jiménez-de-la-Cuesta and Mauritsen 2019).

We estimate λ from MPI-ESM1.1 simulation data, as the ordinary least squares (OLS) slope coefficient when regressing annual-mean R on T_s, similarly to in previous work (e.g., Gregory et al. 2004; Gregory and Andrews 2016; Andrews et al. 2018). We use ensemble-mean data from 100 historical simulation runs and 68 1pctCO2 simulation runs.

For the MPI-ESM1.1 historical and 1pctCO2 experiments R is obtained from the relationship $R = F - N$, with F being the ERF and N the net downward TOA radiative imbalance. We obtain ERF estimates for the 1850–2005 historical simulation from ensemble-mean N in three runs with 1850–2008 historical atmospheric composition and land use changes but fixed SST.

We augment the N values to account for the small increase in T due to land surface warming (Hansen et al. 2005). The applied adjustment factor of 0.07 represents the ratio of changes in mean T and N between 1860–82 and 1999–2008, being periods unaffected by volcanism, in that simulation multiplied by a λ value of 1.36 W m^{-2} K^{-1} estimated by regression over years 1–150 of the MPI-ESM1.1 abrupt4xCO2 simulation. Volcanic ERF exhibits a low (equilibrium) efficacy in MPI-ESM1.1: it responds as if volcanic ERF were less than that included in the diagnosed historical ERF (Lewis and Curry 2020). This may be due to tropical volcanic eruptions weakening the zonal Pacific tropical SST gradient (Clement et al. 1996; Gregory et al. 2016; Miao et al. 2018). To eliminate any resulting bias in feedback estimation when regressing MPI-ESM1.1 historical simulation R on T_s, volcanic forcing (per the IPCC AR5 estimated time series) is included as a separate regressor. Using data over 1852–2005, omitting 1850 and 1851 as their forcing estimates were affected by spinup issues, the regression fit is excellent ($r =$ 0.96), with a λ estimate of 1.56 $\pm$ 0.04 W m^{-2} K^{-1} (1 standard error regression uncertainty). Estimating over 1871–2005 gave essentially identical results: $\lambda = 1.56 \pm 0.04$ W m^{-2} K^{-1}. So did regressing using pentadal-mean data, with an even closer fit ($r =$ 0.994) due to suppression of interannual variability.

We estimate F in the MPI-ESM1.1 1pctCO2 experiment using a quadratic fit to the TOA radiative imbalances between two MPI-ESM1.1 fixed SST runs, one with CO$_2$ increasing at 1% per year and the other with CO$_2$ fixed (Adams and Dessler 2019), adjusting for land surface warming in those periods on the same basis as aforesaid. We obtain $F_{2\times CO2}$ and $F_{4\times CO2}$ estimates of respectively 4.00 and 8.50 W m^{-2}. This $F_{4\times CO2}$ value is within 1% of the average of estimates derived by regression over years 1 or 2 to years 10, 20, or 50 of the MPI-ESM1.1 abrupt4xCO2 simulation, all of which are very similar. Using the fitted F, we estimate λ over years 1–70 of the 1pctCO2 experiment as 1.59 W m^{-2} K^{-1}; regression uncertainty in the estimate is negligible. The λ estimate from regressing over years 1–100 is 1.57 W m^{-2} K^{-1}, while the estimate from regression over years 2–50 of the abrupt2xCO2 simulation—involving the same span of average forcing age—by the nearly identical MPI-ESM1.2 model is 1.59 W m^{-2} K^{-1}; MPI-ESM1.1 has not performed this simulation.

We likewise use OLS regression to estimate λ in amipPiForcing simulations, but here we regress pentadal means, for reasons explained subsequently (section 3c). For amipPiForcing experiments, where forcing is fixed at preindustrial levels, $R = -N$.

We also use OLS regression to estimate trends in annual-mean SST data, comparing area-weighted linear trends in the Indo-Pacific warm pool with those both over the tropics (30°S–30°N) as a whole and over the tropics and midlatitudes combined (50°S–50°N). SST comparisons over 50°S–50°N are preferred to those over the global mean, or over 60°S–60°N, which are affected by the quantification and treatment of sea ice coverage. Using instead comparisons over 60°S–60°N would have a negligible effect on our findings. We define the Indo-Pacific warm pool (IPWP) as area-mean SST within the region 15°S–15°N, 45°–195°E, as SST is very warm over all but a small part of it. Dong et al. (2019) focused on a region with a similar longitudinal boundaries (50°–200°E) but spanning 30°S–30°N.

Earlier studies defined warm pool regions that extended over 15°S–15°N (Andrews and Webb 2018), 20°S–20°N (Visser et al. 2003), or 10°S–10°N (Barlow et al. 2002). We refer, for any stated period, to the percentage excess IPWP SST linear trend over the period relative to that over 30°S–30°N or 50°S–50°N, as the "excess warm pool SST trend" in relation to the relevant zone.

b. Green's functions derived from CAM5.3

Second, we use a Green's function approach to explore climate feedback in response to historical SST warming patterns. Using Green's functions both provides results derived from another GCM and makes it practical to explore a wider variety of SST datasets, including those that only provide temperature anomalies. The Green's function approach exploits the apparent linear superpositionality in space of GCM responses to warming (Barsugli and Sardeshmukh 2002). In particular, the global changes ΔT_s and ΔR resulting from an imposed SST change pattern can be approximated as the sum of their global responses to SST changes in individual locations weighted by time-invariant Green's function values for each location (Dong et al. 2019). Climate feedback λ can then be estimated as $\Delta R/\Delta T_s$. We use Green's functions for ΔR and ΔT_s derived from 74 pairs of patch experiments using the CAM5.3 GCM (Zhou et al. 2017). A 6-yr control run with SST, sea ice and forcings fixed at year 2000 levels was first performed. Each patch experiment involved then imposing a centered 1 K average cosine-squared humped SST warming or cooling spanning a 20° latitude × 80° longitude patch. Sea ice was held fixed. Distorting effects of the associated changes in SST gradients were reduced by differencing the warming and cooling responses when computing the Green's functions. Between them, the 74 patches, which are at 10° latitude and 40° longitude spacings and hence overlap, cover the ice-free ocean. The Green's functions for individual grid cells (each 1.9° latitude × 2.5° longitude) were generated by first allocating the simulated global ΔR and ΔT_s responses to SST change in each patch experiment between grid cells falling within the patch proportionately to their ocean area. For each grid cell the average of those allocated global responses across all patches within which the grid cell falls, weighted by the SST changes at that grid cell imposed in each of the patch experiments involved, was then taken. Fuller details are given in Zhou et al. (2017). The Green's function approach thus includes nonlocal T_s and R responses to local SST change, but does not incorporate the effects of changes in sea ice.

Zhou et al. derived global ΔR_{cloud} responses to local SST change by applying radiative kernels to the Green's function total (all-sky) ΔR and clear-sky ΔR responses. We use the total ΔR responses directly. Figure 1 shows the CAM5.3 Green's function global-mean T_s and R responses at all gridcell locations, as well as the climate feedback λ implied by their ratio. The T_s response is positive everywhere and particularly strong in the west Pacific warm pool, while λ is strongly positive over most of the tropical west and central Pacific, the Indian, and the North Atlantic Oceans. Elsewhere λ is more commonly negative. Outside the tropics the global R response to local SST change is generally small.
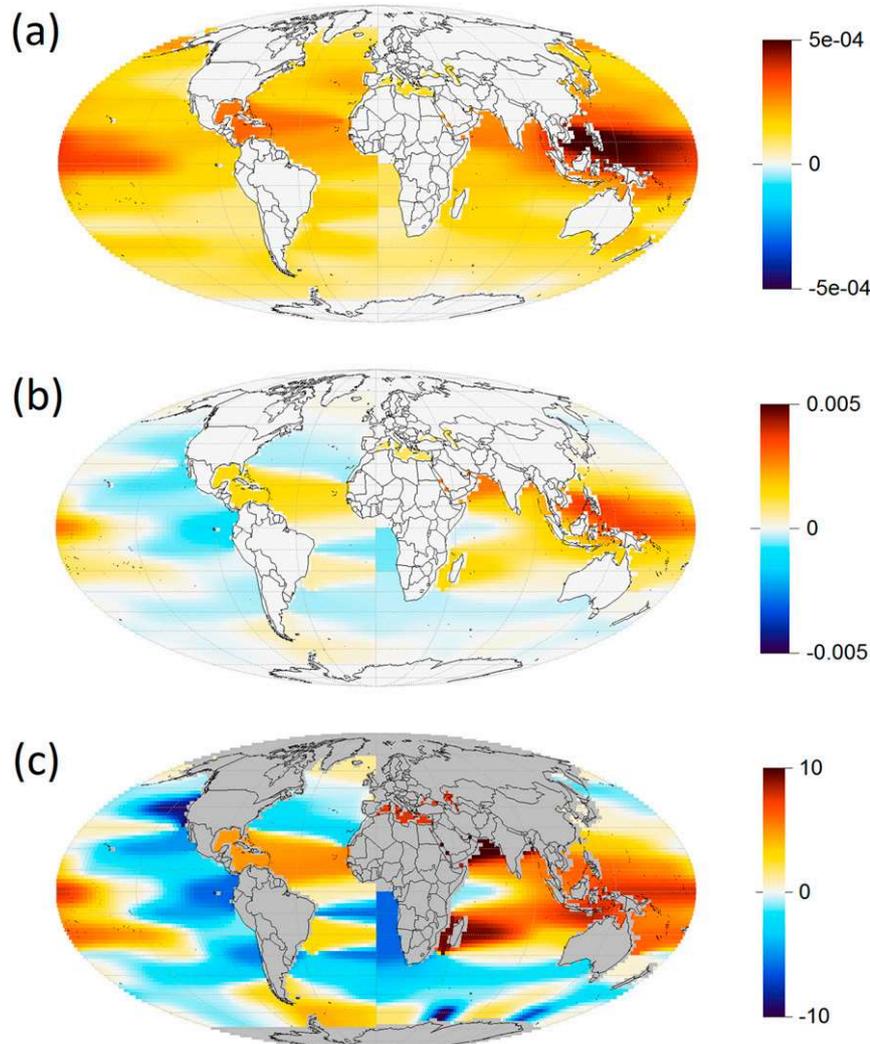
FIG. 1. CAM5.3 Green's functions. (a),(b) The change, respectively, in global-mean T_s (K) and in global-mean R (W m^{-2}) per 1 K increase in local gridcell SST. (c) The global climate feedback parameter λ (W m^{-2} K^{-1}) for a change in local gridcell SST [the ratio of the values plotted in (b) to those plotted in (a)].

Although the Dong et al. (2019) Green's function experiments, which used CAM4, involve more patches and varying rather than fixed sea ice, we prefer the Zhou et al. (2017) Green's functions since CAM5.3 appears to have a more a realistic representation of clouds than CAM4. Cloud behavior is key to the realism of the radiative response to SST warming patterns. Zhou et al. (2017) showed that use of their Green's functions provided good emulation of cloud response for a variety of SST warming patterns, notwithstanding nonnegligible radiative kernel error. Our results show that direct use of their Green's functions likewise provide good emulation of λ, estimated by regression, in the CAM5.3 amipPiForcing simulation. For the Zhou et al. (2017) pattern experiments based on CMIP5 mean SST warming over years 1–150 and 1–20 of abrupt4xCO2 simulations, λ estimated from the CAM5.3 Green's functions (as $\Delta R/\Delta T_s$) is respectively fairly accurate (2.3% high) and modestly

underestimated (by 6.1%). The CAM5.3 Green's functions also reasonably emulate λ for the Zhou et al. (2017) AMO pattern experiment (underestimating λ by 7.6%), despite ΔT_s being quite small (0.37 K). The other Zhou et al. (2017) CAM5.3 pattern experiments (PDO and ENSO) produced very small global-mean ΔT_s and ΔR values, preventing meaningful estimation of λ. Their CAM5.3 Green's functions typically somewhat underestimate the actual simulated ΔT_s and ΔR, but with only minor resulting bias in the estimated λ, which is diagnosed from the relationship between changes in ΔR and in ΔT_s. One might expect that the absence of sea ice variation in the formulation of the Green's functions would cause misestimation of λ; however, there presumably are compensating errors in the Green's-function-emulated ΔR and in ΔT_s values that counteract much of the sea ice feedback-related bias. It is unfortunately not possible to test the accuracy of the Green's function

λ estimates on SST patterns from CAM5.3-based coupled model abrupt4xCO2 or 1pctCO2 simulations, as none have been carried out—hence Andrews et al. (2018) did not estimate a historical pattern effect for CAM5.3.

c. SST datasets

We use the following infilled SST datasets, all of which provide complete coverage over the ice-free ocean for the periods stated.

1) HADISST1 (1870 ONWARD)

The HadISST1 dataset (Rayner et al. 2003) provides SST and sea ice data at $1° \times 1°$ resolution. The SST data are based on ship and buoy data taken from the Met Office Marine Data Bank; SSTs for 1871–1995 from the Comprehensive Ocean–Atmosphere Dataset (COADS: Woodruff et al. 1987) were also used. After 1981 surface skin-temperature estimates from the Advanced Very High Resolution Radiometer (AVHRR) satellite instrument are used in conjunction with the in situ data. The AVHRR data provide nearly complete observational coverage, but require time-varying bias correction, as do the in situ ship SST data. The sea ice data were taken from a variety of sources. HadISST1 temperatures are reconstructed using a two-stage reduced-space optimal interpolation procedure, followed by superposition of quality-improved gridded observations onto the reconstructions to restore local detail. SSTs near sea ice are estimated using statistical relationships between SST and sea ice concentration.

2) AMIPII (1870–2017)

AMIPII (Hurrell et al. 2008) uses HadISST1 SST fields before November 1981 and thereafter uses Optimum Interpolation v2 (OIv2; Reynolds et al. 2002) SST fields, which are based on essentially the same sources of ship and buoy data as HadISST1 uses. OIv2 uses the same AVHRR data source as HadISST1, from November 1981 onward, but employs different methods for assimilation and bias correction. OIv2 applies optimum interpolation directly rather than in a reduced space, achieving greater spatial resolution (Flannaghan et al. 2014) but possibly preserving climate signals at subregional and larger scales less well (Kaplan et al. 1997). Prior to merging the two $1° \times 1°$ resolution datasets, HadISST1 SST anomalies are rebased so that each grid cell's mean SST over 1971–2000 is the same as in OIv2. Doing so alters relative gridcell temperatures prior to 1981. HadISST1 sea ice data are used throughout, with some adjustments. Monthly data are adjusted so as to preserve the seasonal cycle amplitude when interpolated to daily resolution. Use of the AMIPII dataset is standard for historical AMIP experiments, including amipPiForcing.

3) HADISST2 (1850–2010)

HadISST2 is intended to improve on HadISST1 in a number of key areas. Compared with HadISST1, its treatment of sea ice uses new data sources, applies new bias adjustments, and improves the method of estimating concentrations where only information about the sea ice edge is known, with the aim of providing a more consistent records of sea ice concentrations (Titchner and Rayner 2014). The SST component of HadISST2, which has $0.5° \times 0.5°$

resolution, assimilates Along-Track Scanning Radiometer (ATSR) as well as AVHRR satellite data. Greater reliance is placed on the lower-coverage but higher-quality ATSR data. Only the sea ice component of HadISST2 is currently fully documented and regularly updated. However, an ensemble of 10 realizations of the SST component spanning 1850–2010 is publicly available (at https://www.metoffice.gov.uk/hadobs/hadisst2/data/HadISST.2.1.0.0/index.html), is documented in outline (at https://www.eeo.ed.ac.uk/earthtemp/themes/1_in_situ_satellite/Rayner_EarthTemp_Edinburgh_2012_Poster.pdf), and has been used elsewhere (Andrews et al. 2018). We use ensemble-mean HadISST2 SST data.

4) COWTAN AND WAY (HAD4_KRIG_V2_0_0 AND HAD4SST4_KRIG_V2_0_0: 1850 ONWARD)

The had4_krig_v2_0_0 dataset (Cowtan and Way 2014a,b,c) is a version of HadCRUT4v6 (Morice et al. 2012), infilled using kriging. Its SST data are a kriged version of HadSST3 (Kennedy et al. 2011a,b), while the SST data in had4sst4_krig_v2_0_0 are a kriged version of HadSST4 (Kennedy et al. 2019). HadSST3 and HadSST4 are produced, at $5° \times 5°$ resolution, from in situ SST measurements from ships and buoys. Both employ detailed bias correction methods, but the actual adjustments applied differ somewhat.

5) COBE-SST2 (1850–2017)

The COBE-SST2 dataset (Hirahara et al. 2014a,b) is produced from in situ SST measurements from ships and buoys. Bias adjustments applied to ship SST measurements are derived somewhat differently from those applied in constructing HadSST3, HadSST4, HadISST1, and HadISST2. COBE-SST2 is infilled at $1° \times 1°$ resolution using multi-time-scale analysis, as the sum of a time-varying secular trend with a fixed pattern, and spatially varying interannual variations and daily changes. Satellite observations are used only in producing empirical orthogonal functions for an optimal interpolation scheme used in reconstructing interannual-to-decadal fluctuations. The authors find their infilling method to be superior to direct use of optimal interpolation, at least when data are sparse.

6) ERSSTv5 (1854 ONWARD)

The Extended Reconstructed Sea Surface Temperature, version 5, SST dataset (Huang et al. 2017) uses in situ SST measurements from ships and buoys. However, up to 2010 its ship SST values are effectively replaced, on decadal and longer time scales, with HadNMAT2 nighttime marine air temperature data (Kent et al. 2013), which are less widely sampled than SST data. ERSSTv5 uses the same OIv2 dataset for infilling as does the AMIPII SST dataset, but indirectly (for reduced-space interpolation of high-frequency SST components) as part of a complex infilling procedure, with low-frequency SST components instead being infilled by a nearest-neighbor method and then smoothed.

d. Preprocessing and regridding SST data

We mark as NA (not available) all gridcell values in each SST dataset that appear to represent only land and/or sea ice. Where necessary we deduce full sea ice coverage from

near-freezing gridcell SSTs. We form annual-mean SSTs for each grid cell, marked as NA if SST is NA in any month. We then regrid all the observational datasets to the same 2.5° × 1.875° resolution grid as CAM5.3, for use with the CAM5.3 Green's functions, using bilinear interpolation (to a subdivision of the CAM5.3 grid, followed by aggregation, where the observational dataset has significantly finer resolution than CAM5.3). We give NA values to any NA valued grid cell in the regridded HadISST1 dataset in the same year (which indicates that some original grid cell contributing to it represents entirely land and/ or sea ice), thus eliminating possible inconsistencies in regridded data at the boundaries of land and sea ice covered areas and also ocean-masking the combined land and ocean Cowtan and Way data. When carrying out regressions, we exclude grid cells that were marked as NA in any year during the analysis period, to avoid influence from changes in sea ice differing between SST datasets. We also regrid HadISST1 data to match the ECHAM6.3 grid for our amipPiForcing simulation experiments, using with it the HadISST1-based AMIPII sea ice boundary condition dataset.

3. Results

a. ECHAM6.3 simulations

We first present the ECHAM6.3 amipPiForcing simulation results, using the standard OLS regression method (Andrews et al. 2018; Gregory et al. 2019). Figures 2a and 2b show the 1871–2010 ensemble-mean T_s and R time series for the AMIPII- and HadISST1-based experiments, while the scatterplots in Fig. 2c show their relationships in each experiment, on a pentadal-mean basis, along with the best-fit lines, the slope of which represents the estimated λ value ($\lambda_{hist-amip}$). Until November 1981, AMIPII SST was based on HadISST1 SST data, although with the pattern of absolute temperatures altered by the AMIPII method of merging HadISST1 and OIv2 SST data. The post-1980 differences between the AMIPII and HadISST1 T_s time series are relatively small and well within the uncertainty ranges attributable to surface temperature observation datasets. All major differences between AMIPII and HadISST1 R anomalies arise after 1980. The AMIPII– HadISST1 R difference spiked in 1982; T_s was much less affected. It is possible that the first year or so of OIv2 data had some issues, or that the effects of the El Chichón eruption differentially affected HadISST1 and OIv2 SST patterns. The R differences rose again in 1997, and remained high for another decade.

Table 1 shows excess warm pool SST trends and feedback strengths for the AMIPII- and HadISST1-based amipPiForcing experiments and, for comparison, for the parent MPI-ESM1.1 coupled model's 1pctCO2 and historical forced experiments, over both the full 1871–2010 amipPiForcing period and 1871–2005 (for comparison with the historical simulation ending then). The λ estimates for the HadISST1-based experiment are significantly lower than for the AMIPII-based experiment. Applying Welch's t test (Welch 1947) to the λ estimates over 1871–2010 from both sets of five individual runs gives the probability that the HadISST1-based and AMIPII-based single-run λ estimates came from populations with the same mean λ as 0.01%. For the estimates over 1871–2005 the probability was 0.03%.
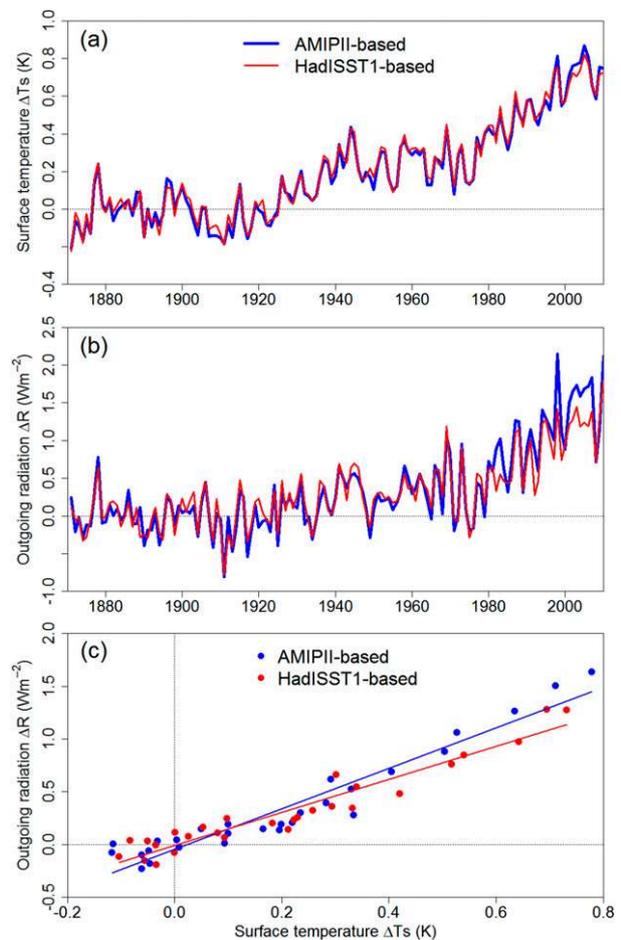

FIG. 2. Comparison of changes over 1871–2010 in annual ensemble-mean (a) surface temperature (ΔT_s) and (b) TOA outgoing radiation flux (ΔR) in the two ECHAM6.3 amipPiForcing experiments, and (c) the relationship in the two experiments between pentadal ensemble-mean ΔT_s and ΔR, with the best-fit lines to those pentadal-mean points. Changes are relative to 1871–1900 means.

Comparing each of the sets of five amipPiForcing 1871–2010 λ estimates with the 68 λ estimates over years 1–70 of individual 1pctCO2 runs using the same test gives a 39% probability that the HadISST1-based single-run λ estimates came from populations with the same mean λ as the 1pctCO2 run λ estimates, whereas for AMIPII the corresponding probability is merely 0.07%. Accordingly, the 1871–2010 HadISST1-based amipPiForcing ensemble-mean λ estimate, but not the AMIPII-based estimate, is statistically indistinguishable from that over years 1–70 of the 1pctCO2 experiment. Similarly, we found that the 1871–2005 HadISST1-based amipPiForcing ensemble-mean λ estimate, but not the AMIPII-based estimate, is statistically indistinguishable from that over the same period of the historical experiment (the relevant probabilities being respectively 12% and 0.2%).

Figure 3 maps SST 1871–2010 spatial warming trends in the AMIPII and HadISST1 datasets, and their differences.

TABLE 1. Excess Indo-Pacific warm pool SST trends and climate feedback, in ECHAM6.3 amipPiForcing simulations and in MPI-ESM1.1 coupled 1pctCO2 and historical simulations. All values are based on ensemble-mean T_s and R data (except for AMIPII and HadISST1 SST trends and standard deviations of individual run feedback estimates). Feedback estimates are from OLS regression of pentadal-mean data for amipPiForcing simulations (see section 3c). Values in parentheses are standard errors of the OLS regression feedback estimates, which reflect underlying deviations from a linear relationship as well as internal variability.

ECHAM6.3/MPI-ESM1.1 simulation data and SST data source	Excess warm pool vs 30°S–30°N SST trend (%)	Excess warm pool vs 50°S–50°N SST trend (%)	Climate feedback on T_s in simulation (W m^{-2} K^{-1})	Standard deviation of individual run feedback estimates (W m^{-2} K^{-1})
AMIPII amipPiForcing simulation, 1871–2010	7	0	1.92 (0.11)	0.08
HadISST1 amipPiForcing simulation, 1871–2010	2	−7	1.56 (0.09)	0.06
1pctCO2 coupled simulation, years 1–70	5	7	1.59 (0.01)	0.06
AMIPII amipPiForcing simulation, 1871–2005	5	−3	1.85 (0.11)	0.10
HadISST1 amipPiForcing simulation, 1871–2005	0	−10	1.49 (0.09)	0.09
Historical coupled simulation, 1871–2005	7	12	1.56 (0.04)	0.09

AMIPII warms more than HadISST1 in both the west Pacific and deep tropical Indian Ocean, and also in the tropical southeastern Pacific, but less in the tropical northeastern Pacific and extratropical North Pacific. There are mixed differences in the Atlantic Ocean, while in the southern extratropics AMIPII warms less apart from over a narrow band centered on 55°–60°S, where it warms considerably more. Since sea ice boundary conditions are identical in the two experiments, there are no differences in areas covered by sea ice.

Averaged across the two periods involved, feedback in the amipPiForcing simulation was 19% lower with HadISST1 rather than AMIPII SST boundary conditions. Feedback in the amipPiForcing simulation when using HadISST1 was also slightly lower than in the historical simulation, and almost the same as that over years 1–70 of the 1pctCO2 simulations, by the parent AOGCM. Thus, in this model there is little evidence for any unforced historical pattern effect when using HadISST1 SST boundary conditions. Moreover, the similarity of feedback in the 1pctCO2 and historical coupled simulations implies that a forced historical pattern effect due to non-CO$_2$ changes, as found by Shindell (2014) in a subset of CMIP5 models, is absent in this model.

The Indo-Pacific warm pool SST trend was lower, relative to both the whole tropics and to 50°S–50°N, in the amipPiForcing simulation with HadISST1 boundary conditions than in the amipPiForcing simulation with AMIPII boundary conditions or in either MPI-ESM1.1 coupled simulation (Table 1). However, there were no clear relationships between the Indo-Pacific warm-pool-relative SST trends and feedback differences of either amipPiForcing simulation or the two coupled simulations. As is evident from Table 1, the relationship between the Indo-Pacific warm pool SST trend relative to 50°S–50°N and the differences between feedback in the amipPiForcing simulation with AMIPII boundary conditions and in the two coupled simulations was negative. These results indicate that there are also other factors, such as sea ice variation, determining feedback besides relative SST trends in the warm pool.

We estimate the unforced historical pattern effect in ECHAM6.3 by differencing ensemble-mean feedback estimates in respectively the ECHAM6.3 AMIPII and HadISST1 amipPiForcing simulations from those in the 1pctCO2 and historical coupled simulations, and quantify one standard deviation uncertainty in them by adding in quadrature our standard error estimates for the feedback estimates being differenced. The resulting unforced historical pattern effect estimates over 1871–2005, relative to feedback in historical coupled simulations, are 0.30 ± 0.12 and -0.07 ± 0.10 W m^{-2} K^{-1} for respectively AMIPII- and HadISST1-based amipPiForcing simulations. The resulting unforced historical pattern effect estimates over 1871–2010, relative to feedback over years 1–70 in 1pctCO2 coupled simulations, are 0.33 ± 0.11 and -0.02 ± 0.09 W m^{-2} K^{-1} for respectively AMIPII- and HadISST1-based amipPiForcing simulations.

Dessler (2020) quantified the possible magnitude of the unforced historical pattern effect based on internal variability in the MPI-ESM1.1 ensemble of 100 historical coupled simulations, measuring climate feedback relative to the ensemble mean (thus implying zero effect on average). We use the same data to evaluate whether our unforced historical pattern effect estimates are consistent with AMIPII and/or HadISST1 representing single realizations of possible SST trajectories that might be generated by such internal variability, allowing for the uncertainty in the unforced historical pattern effect estimates. The standard deviation of 1871–2005 regression-based feedback estimates from individual historical simulations, being 0.09 W m^{-2} K^{-1}, provides the appropriate estimate of variability of the unforced historical pattern effect in MPI-ESM1.1. Making the assumption that error distributions are approximately normal, the estimated unforced historical pattern effect in ECHAM6.3 for the HadISST1 amipPiForcing experiment (of -0.07 ± 0.10 or -0.02 ± 0.09 W m^{-2} K^{-1} relative to feedback in, respectively, the historical experiment and the 1pctCO2 experiment) are statistically consistent with internal variability ($p = 0.63$ and $p = 0.87$ respectively). On the
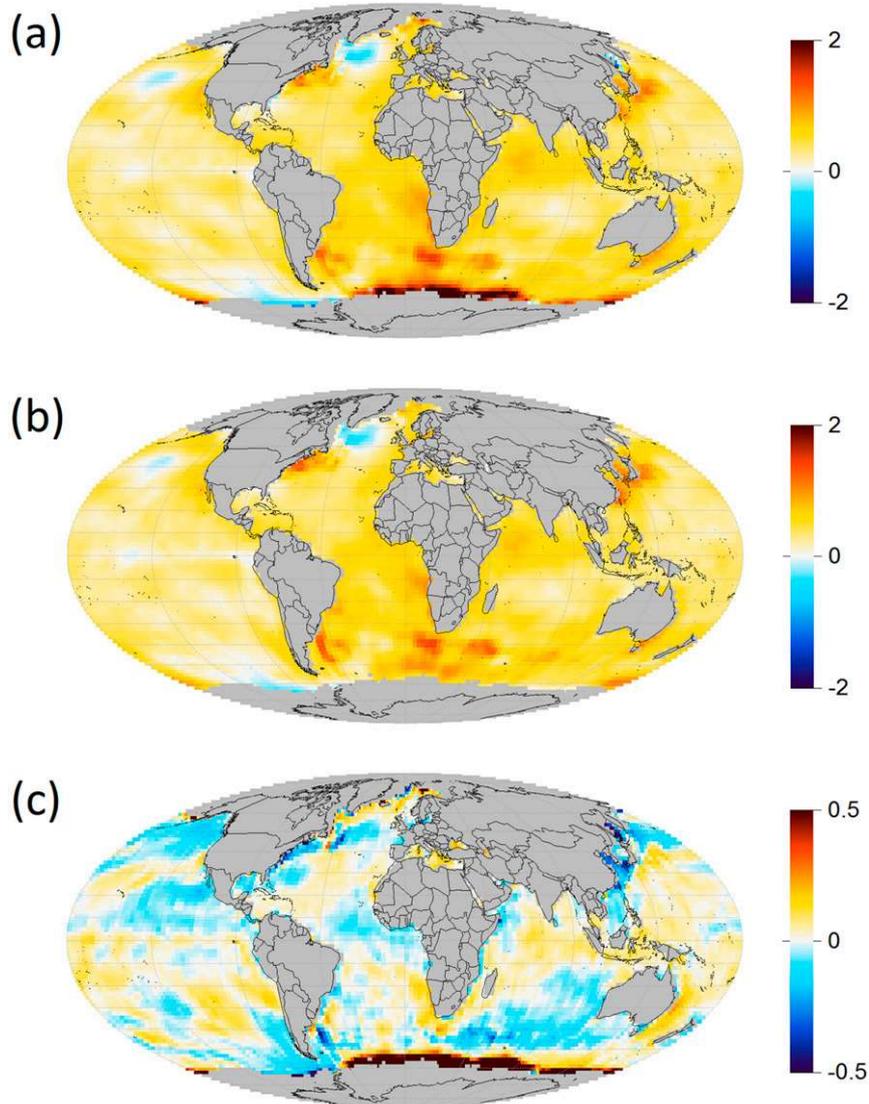
Fig. 3. Local SST linear warming trends over 1871–2010 (K century^{-1}) (a) in the AMIPII dataset, (b) in the HadISST1 dataset, and (c) for AMIPII minus HadISST1 data. Note that (c) has a 4-times-finer SST trend scale. No values are shown for areas where HadISST1 lacks data for any month during 1871–2010, indicating coverage by land or sea ice.

other hand, the estimated unforced historical pattern effect for the AMIPII amipPiForcing experiment (of 0.30 ± 0.12 or 0.33 ± 0.11 W m^{-2} K^{-1} relative to feedback in, respectively, the historical experiment and the 1pctCO2 experiment) is statistically inconsistent with internal variability ($p = 0.05$ and $p = 0.02$, respectively). We caution that these probability estimates are dependent inter alia on the realism of internal variability in MPI-ESM1.1.

b. Investigating historical period feedback using Green's functions derived from CAM5.3

Using the Green's functions, we are able to emulate, in a computationally inexpensive way, time series for the T_s and R responses to historical warming patterns in a wide range of observational SST datasets, and hence produce associated feedback estimates by the same regression method as for the simulation data discussed above. Table 2 shows that using the CAM5.3 Green's functions provides an accurate estimate of feedback in the AMIPII-based amipPiForcing simulation, although that does not imply that the Green's function feedback estimates are necessarily similarly accurate in the other cases. The Green's function feedback estimates based on regressing, using pentadal means, emulated global-mean T_s and R responses to evolving 1871–2010 warming patterns in the CESM1-CAM5 1pctCO2 and historical simulations, and per HadISST1 and five other observationally based SST datasets, are also shown.

TABLE 2. Excess Indo-Pacific warm pool SST trends and Green's function derived estimates of climate feedback in CAM5.3 AMIPII-based amipPiForcing simulations, in CESM1-CAM5 coupled 1pctCO2 and historical/RCP8.5 simulations, and for warming in six observational SST datasets, along with feedback estimated from the actual CAM5.3 AMIPII-based amipPiForcing simulation data. Feedback estimates are from OLS regression of pentadal-mean R and T_s values derived from the evolving SST warming patterns in the relevant simulation or observationally based dataset. Data over 1871–2010, the amipPiForcing experiment period, is used, with data from the historical experiment extended using RCP8.5 experiment data, except in the 1pctCO2 simulation case where years 1–70 data are used. Small differences in AMIPII and HadISST1 excess warm pool SST trends from those in Table 1 reflect regridding to different grids.

Model or SST data source involved	Excess warm pool vs 30°S–30°N SST trend (%)	Excess warm pool vs 50°S–50°N SST trend (%)	Green's function feedback on T_s estimate (W m^{-2} K^{-1})	Feedback on T_s in simulation (W m^{-2} K^{-1})
AMIPII amipPiForcing simulation (CAM5.3)	8	0	1.71	1.70
1pctCO2 coupled simulation (CESM1-CAM5)	7	11	1.68	—
Historical coupled simulation (CESM1-CAM5)	10	5	1.65	—
HadISST1 warming pattern	2	−8	1.56	—
HadISST2 warming pattern	1	−6	1.59	—
Had4_krig_v2 warming pattern	−2	−9	1.56	—
HadSST4 krig_v2 warming pattern	−1	−5	1.64	—
COBE-SST2 warming pattern	0	−3	1.69	—
ERSSTv5 warming pattern	−3	0	1.84	—

CESM1-CAM5 is the most closely related coupled model to CAM5.3 for which such simulation data were available.[2] Indo-Pacific warm-pool-relative SST trends are also given. The use of SST linear trends derived by OLS regression matches the method of estimating feedback. Using instead differences in mean SST between the first and last decades would show higher excess warm pool SST trends, but with the differences between the various observational datasets showing a similar pattern to that for 1871–2010 linear trends.

Two points are noteworthy. First, the excess warm pool SST trends are lower in all the non-AMIPII observationally based SST datasets than in the AMIPII-based amipPiForcing simulation and the 1pctCO2 and historical CESM1-CAM5 coupled simulations, apart from ERSSTv5 matching the AMIPII zero excess warming over 50°S–50°N. The Green's function feedback estimates for the seven observationally based SST datasets are strongly correlated with warm pool SST trends relative to those over the tropics and midlatitudes ($r = 0.90$), but not relative to those over the tropics alone ($r = -0.10$) (Fig. 4). Second, there is no systematic tendency for Green's function feedback estimates from all the non-AMIPII observationally based SST datasets to be above those for the SST warming

patterns in the CESM1-CAM5 historical and 1pctCO2 simulations; in a majority of cases they are lower. Moreover, feedback estimated from the warming pattern in the AMIPII amipPiForcing simulation is only slightly higher than that for the CESM1-CAM5 1pctCO2 and historical coupled simulations. These findings appear to imply that unforced historical pattern effects are relatively weak in this model. Based on CAM5.3 Green's function SST warming pattern-based feedback estimates, the seven observationally based SST datasets exhibit almost zero average unforced historical pattern effect relative to the feedback estimates for the CESM1-CAM5 1pctCO2 simulation pattern. ERSSTv5 produces a +10% unforced historical pattern effect; in other cases it ranges between −7% (HadISST1, Had4_krig_v2) and +2% (AMIPII). It is not possible to enumerate statistical significances for these effects as the accuracy of the Green's function–derived λ estimates is not adequately quantified. However, the magnitude of the largest fractional error in the Green's function–derived λ estimates in cases where it has been quantified, being 7.6% (section 2) could be taken as providing a crude uncertainty bound. On that basis, the pattern effect would be regarded as indistinguishable from zero for all datasets except ERSSTv5.

We have also used the CAM5.3 Green's functions to identify the fractional contributions of differences between AMIPII and HadISST1 warming trends in different zones to differences in the resulting trends in global R and T_s (Fig. 5). SST trend differences over 15°S–15°N, primarily in the IPWP, contributed

---

[2] No forcing estimates are available for CESM1-CAM5 1pctCO2 and historical simulations, so feedback in them cannot be estimated directly from simulation data.
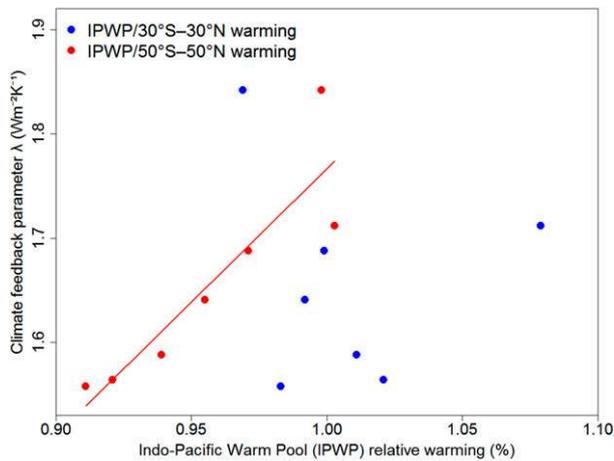
Fig. 4. The relationship between climate feedback strength, estimated using the CAM5.3 Green's functions and pentadal regression, and the warming trend in the Indo-Pacific warm pool relative to that over either 30°S–30°N (blue circles) or 50°S–50°N (red circles), both over 1871–2010, for SST per seven observational datasets (AMIPII, HadISST1, HadISST2, Had4_krig_v2, HadSST4_krig_v2, COBE-SST2, and ERSSTv5). The red line shows a linear fit between the warming trend in the IPWP relative to that over 50°S–50°N and estimated climate feedback strength ($r = 0.90$). No equivalent fit is shown for the warming trend in the IPWP relative to that over 30°S–30°N, as the relationship is very weak ($r = -0.10$).

Fig. 5. The difference in 1871–2010 linear trend in (a) global-mean T_s (K century^{-1}) and (b) global-mean R (W m^{-2} century^{-1}), as estimated using the CAM5.3 Green's functions, caused by the local gridcell SST evolving as per the AMIPII dataset rather than as per the HadISST1 dataset. Values have been divided by each cell's area weight, so that they reflect the magnitude of the difference per unit area. Note that trends in (a) include the estimated global land T_s response as well as the local SST change and hence are generally higher than those in Fig. 3c, which includes only the local SST change and has a different scale.

slightly over two-thirds of the total difference in the global R/T_s trend ratio (a proxy for the difference in λ) between the two SST datasets. Additional contributions for SST trend differences between latitudes 15° and 30°, mainly from east of Australia, led to 30°S–30°N contributing five-sixths of the total difference in the global R/T_s trend ratio. Differences over 50°S–50°N contributed one-fifth more than the global difference, with AMIPII producing less warming, but a stronger increase in R, than HadISST1 over much of 30°–50°N, where λ is generally negative for local SST changes. Poleward of 50° the overall difference in global R trend between the AMIPII and HadISST1 datasets was positive but small, while the difference in global T_s trend was large and positive, due to much stronger warming for AMIPII over 60°S–50°S, mainly over 45°W–45°E. Therefore, SST trend differences poleward of 50° reduced the overall excess R/T_s ratio per AMIPII over that per HadISST1.

As Fig. 2c shows, it is the last 15 years (1996–2010) that show the largest difference in AMIPII- and HadISST1-based feedback estimates in ECHAM6.3 simulations. The same is true for the CAM5.3 Green's function–based feedback estimates. Based on regressing 1871–1995 pentadal data, the difference between the AMIPII and HadISST1 based Green's function feedback estimates is reduced from 0.15 W m^{-2} K^{-1} per Table 2 to 0.03 W m^{-2} K^{-1}. Since the two SST datasets are essentially identical until late 1981, and the post-1995 period has the strongest signal, one would expect excluding that period to greatly reduce the difference between feedback estimates from the two datasets. The 1998 El Niño event appears to be responsible for at most a small part of the difference. When excluding 1997–2000, the period affected by the buildup of El
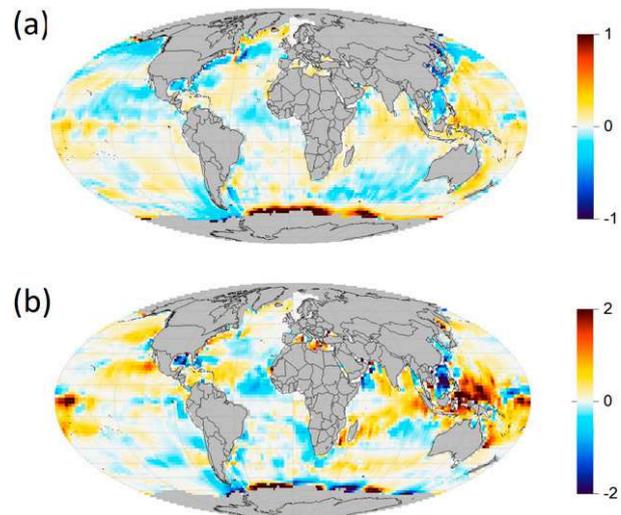
Niño, its peak and the subsequent La Niña, or the 1996–2000 pentad, the difference in AMIPII and HadISST2 based feedback estimates is little changed from when regressing over the full 1871–2010 period.

It follows that it is SST differences over the 2001–10 period that account for the majority of the difference in the Green's function–derived full-period feedback estimates. Figure 6 repeats the Fig. 5 analysis but based on changes in 2001–10 mean R and T_s anomalies. The patterns of R and T_s differences are similar to the full-period regression-derived patterns in Fig. 5.

We computed feedback estimates when regressing annual data over 1871–2010 but with each year from 2001 to 2010 excluded in turn. The resulting AMIPII versus HadISST1 feedback differences are on average marginally smaller than when no year is excluded, but they have a standard deviation of only 0.005 W m^{-2} K^{-1}, with no years standing out as having unusually large effects.

c. Regression issues when estimating feedback in amipPiForcing experiments

We are cautious of feedback estimates based on the usual method of regressing annual-mean data, instead preferring estimates from regressing pentadal-mean data. Using pentadal-mean data substantially reduces noise in the regressor variable, which through regression dilution causes a downward bias in the slope coefficient, and also greatly diminishes the effect of responses to interannual fluctuations (Gregory et al. 2019), thus providing more robust estimation. When estimating λ from MPI-ESM1.1 historical and 1pctCO2 simulation large
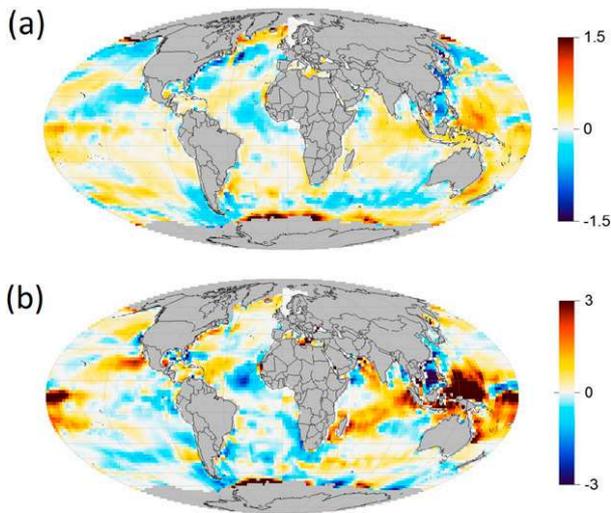
FIG. 6. The difference between average 1871–1900 and 2001–10 (a) global-mean T_s (K) and (b) global-mean R (W m^{-2}), as estimated using the CAM5.3 Green's functions, caused by the local gridcell SST evolving as per the AMIPII dataset rather than as per the HadISST1 dataset. Values have been divided by each cell's area weight, so that they reflect the magnitude of the difference per unit area.

ensemble-mean data, which contain little interannual fluctuation noise, results from regressing annual and pentadal-mean data were essentially identical. However, for the amipPiForcing simulations, feedback estimates tended to be slightly lower when regressing pentadal rather than annual-mean data, contrary to what regression dilution alone would cause.

Table 3 shows, for the eight models analyzed in Andrews et al. (2018), feedback estimates (here based on 2-m air temperature T, not T_s, to match their results) when each performed, for the number of runs stated, an AMIPII-based amipPiForcing experiment over 1871–2010 (over 1871–2004 for GFDL-AM2.1 and GFDL-AM3). The feedback estimates, per Andrews et al. in column 3, are essentially identical to means of estimates from regressing annual-mean R on T data for each run separately. Columns 4 and 6 show feedback estimates from regressing ensemble-mean data using respectively annual and pentadal means, with associated standard errors. The standard deviations of individual run feedback estimates are also given. In these simulations, where identical SST and sea ice boundary conditions are imposed in each run, regression dilution is minimal and it makes very little difference whether ensemble means are taken before or after regression. Feedback estimates over years 1–50 of abrupt4xCO2 simulation runs ($\lambda_{4\times CO2\_1-50}$) are shown for comparison, this period providing a broadly comparable average forcing duration to the historical period.[3] EffCS estimated over the first 50 years of abrupt4xCO2 simulations is very similar to that estimated over

the first 100 years of 1pctCO2 simulations, which has the same average forcing duration as a 50-yr constant forcing (Lewis and Curry 2018). EffCS estimated over the first 100 years of 1pctCO2 simulations in turn corresponds closely to that estimated from historical forcing (Armour 2017).

Feedback estimates from regressing annual ensemble-mean data are on average 0.06 W m^{-2} K^{-1} (4%) higher than the more robust estimates from regressing pentadal ensemble-mean data, and 7%–9% higher for two of the models. The within-model standard deviations of all the individual run differences in λ estimated by regressing annual and pentadal means were small: 0.02 W m^{-2} K^{-1} on average. Investigation indicated that these ensemble-mean differences were due to responses to interannual fluctuations generally being considerably stronger than those to longer-term climate change. Feedback estimates from regressing year-to-year *changes* in annual-mean R and T were 8%–55% higher than those from regressing annual-mean R on T, and those excesses correlated strongly ($r = 0.92$) with the excesses of estimates from regressing annual-mean R on T over pentadal-mean R on T. We obtained very similar results when using smoothing splines with 32 degrees of freedom to remove low-frequency variability in annual-mean R and T, rather than taking year-to-year changes. With AMIP simulations, interannual fluctuations in T mainly arise from fluctuations in the prescribed boundary conditions and hence do not average out when ensemble-mean values are used, unlike with coupled simulations. However, interannual fluctuations have less influence on pentadal means, which therefore provide more robust regression-based feedback estimates. Moreover, the standard deviation of feedback estimates from separate amipPiForcing runs was slightly lower when using pentadal rather than annual-mean data. Given the varying and in some cases material bias when estimating climate feedback in amipPiForcing experiments using annual-mean data, it seems clearly preferable to regress instead pentadal-mean data. The AMIPII-based amipPiForcing feedback estimates when doing so are still significantly higher than comparable feedback estimates from the first 50 years of idealized CO_2 forced experiments by the parent coupled models, except for GFDL-AM2.1, thus still indicating positive unforced historical pattern effects when using the AMIPII dataset.

d. Inconsistency of AMIPII-based historical pattern effects with AOGCM internal variability

Another way of appraising whether the AMIPII-based amipPiForcing feedback estimates reported in Andrews et al. (2018) are likely to represent an unforced historical pattern effect is to investigate whether the magnitude of their deviation from $\lambda_{hist-AOGCM}$ can be reproduced by internal variability in AOGCM piControl simulations. We have tested this for the five GCMs listed in Table 3 for which $\lambda_{hist-AOGCM}$ can be estimated (as equal to $\lambda_{4\times CO2\_1-50}$) and a significant unforced historical pattern effect was detected (CAM4, ECHAM6.3, GFDL-AM3, HadAM3, and HadGEM2). It should be noted that this test will provide purely statistical evidence, with no recourse to a physical principle that would enable one SST dataset to be preferred over another.

---

[3] Temperature changes are much larger in this case so it makes little difference whether annual or pentadal-mean data are regressed.

TABLE 3. Climate feedback parameters λ (for 2 m air temperature) in amipPiForcing simulations with AMIPII boundary conditions as estimated from 1871–2010 data (1871–2004 for GFDL-AM2.1 and AM3 models) using different OLS regression approaches. Feedback for the parent AOGCM of GFDL-AM2 is the mean of those for GFDL-ESM2G and GFDL-ESM2M. All climate feedback values are in W m^{-2} K^{-1}. Values in parentheses are standard errors of the OLS regression feedback estimates, which reflect underlying deviations from a linear relationship as well as internal variability.

Model	No. of runs in ensemble	Feedback per Andrews et al. (2018): mean of each run's annual regression estimate	Feedback by regressing annual ensemble-mean data	Standard deviation of individual run annual regression feedback estimates	Feedback by regressing pentadal ensemble-mean data	Standard deviation of individual run pentadal regression feedback estimates	Feedback from regressing parent AOGCM years 1–50 abrupt4xCO2 run annual data
CAM4	1	2.27	2.27 (0.13)	—	2.27 (0.17)	—	1.52 (0.06)
CAM5.3	1	1.71	1.71 (0.11)	—	1.70 (0.16)	—	—
ECHAM6.3	5	1.90	1.92 (0.08)	0.08	1.88 (0.10)	0.08	1.46 (0.04)
GFDL-AM2.1	6	1.67	1.70 (0.09)	0.05	1.56 (0.10)	0.05	1.57 (0.09)
GFDL-AM3	6	1.40	1.44 (0.10)	0.04	1.34 (0.16)	0.04	0.97 (0.06)
GFDL-AM4.0	5	1.91	1.93 (0.10)	0.04	1.86 (0.14)	0.04	—
HadAM3	4	1.65	1.65 (0.09)	0.02	1.57 (0.11)	0.01	1.19 (0.03)
HadGEM2	4	1.37	1.39 (0.09)	0.06	1.33 (0.11)	0.05	0.83 (0.04)
Model mean		1.74	1.75	—	1.69	—	—

We took all 140-yr-long piControl simulation segments obtainable (allowing overlap between different segments) from the control runs of 43 CMIP5 models. For each of the resulting 18 391 segments we calculated the average T and R anomalies over the last 15 years, $\Delta T_{\text{piControl}}$ and $\Delta R_{\text{piControl}}$, relative to their average over years 1–30. For each of the aforementioned five GCMs (Table 3) we then calculated as follows. For each of the 18 391 piControl segments we subtracted their $\Delta T_{\text{piControl}}$ from the ensemble-mean 1871–1900 to 1996–2010 average ΔT in their AMIPII-based amipPiForcing simulations, ΔT_{amip}. We then multiplied the resulting internal variability affected ΔT values by the relevant $\lambda_{\text{hist-AOGCM}}$ (thus providing an estimate of the forced ΔR response). Since internal variability affects R as well as T, we then added the corresponding $\Delta R_{\text{piControl}}$ anomalies from the same piControl segment to the estimated forced radiative response. The logic here is that if internal variability increased warming over the historical period, then the forced warming—and thus also the forced radiative response, based on $\lambda_{\text{hist-AOGCM}}$, in the absence of an unforced pattern effect—was smaller, than that in the AMIPII-based simulation, and vice versa. We thus created, for each of the five GCMs, a set of 18 391 samples of 1871–1900 to 1996–2010 ΔR values that reflected the estimated forced response of each model to global warming equal to that over its 140-yr AMIPII-based simulation but that were affected by temporally matching 140-yr duration ΔT and ΔR internal variability from different CMIP5 models. We then compared each GCM's set of samples with the corresponding ensemble-mean ΔR in its AMIPII-based amipPiForcing simulations, ΔR_{amip}.

We found that in only 0.06% of the nearly 92 000 cases tested was simulated internal variability sufficient to increase the

estimated forced ΔR response enough to reach ΔR in the AMIPII-based amipPiForcing simulations. Mathematically, that equates to cases where

$$\lambda_{\text{hist-AOGCM}}(\Delta T_{\text{amip}} - \Delta T_{\text{amip}}) \geq (\Delta R_{\text{amip}} - \Delta R_{\text{amip}}). \quad (1)$$

The probability was slightly lower still when using anomalies over the last 20 rather than 15 years of the 140-yr periods. These findings suggest that, if both multidecadal internal variability and the radiative response to patterned warming are realistically estimated in CMIP5 models, the historical warming patterns in the AMIPII (and, a fortiori, ERSSTv5) SST datasets are unlikely to be correct. We issue a caveat about this finding in that the realism of multidecadal internal variability in AOGCMs—although regularly relied upon in detection and attribution studies—is unproven. However, the multidecadal internal variability would have to be unrealistic in almost all CMIP5 models for our conclusion to be unwarranted.

We also investigated how likely it was that internal variability could account for weaker unforced historical pattern effects than those in the AMIPII-based amipPiForcing simulations by the aforementioned five GCMs listed in Table 3. That is, for any chosen pattern effect strength β, by replacing ΔR_{amip} by $\{\beta \Delta R_{\text{amip}} + (1 - \beta)\lambda_{\text{hist-AOGCM}}\Delta T_{\text{amip}}\}$ in (1), giving

$$\lambda_{\text{hist-AOGCM}}(\beta \Delta T_{\text{amip}} - \Delta T_{\text{piControl}}) \geq (\beta \Delta R_{\text{amip}} - \Delta R_{\text{piControl}}). \quad (2)$$

The probability of internal variability being sufficient to account for a weakened, β-strength, unforced pattern effect reached 1.6% at $\beta = 0.5$, and 10% at $\beta = 0.25$.

4. Discussion and conclusions

In this study we have found no evidence for a substantial unforced pattern effect over the historical period, arising from internal variability, in the available sea surface temperature datasets, except for when the AMIPII and ERSSTv5 datasets are used. Our results imply that the evidence suggesting existing constraints on EffCS from historical-period energy budget considerations are biased low due to unusual internal variability in SST warming patterns is too weak to support such a conclusion, and suggest that any such bias is likely to be small and of uncertain sign. This should not be mistaken for a finding relating to a forced pattern effect that acts to temporarily dampen global warming in AOGCM simulations on decadal to centennial time scales.

It is worth noting that none of the datasets inspected here provides a perfectly homogenized temperature record, which is a source of concern when looking at changes over extended periods. In all cases time-varying bias corrections must be applied due to the evolving observing system, and observational data with partial coverage must be interpolated to provide a globally complete reconstruction. Although all SST reconstructions involve making compromises, an additional concern with the AMIPII dataset is that it merges two SST reconstructions that employ different bias correction and interpolation methods, and in doing so alters pre-merger SST patterns. The various datasets try, in different ways, to take advantage of the satellite observations from when they become available around 1980. The post-1981 AMIPII dataset interpolation method, however, does so in a way that emphasizes small-scale features at the expense of the large-scale patterns central to the study of pattern effects (Hurrell et al. 2008). Perhaps as a result, AMIPII warms more in the western tropical ocean basins and less in the eastern subsidence regions when compared to HadISST1. Earlier studies have in other contexts pointed to issues with the patterns of tropical warming in AMIPII (Rayner et al. 2003: Fig. 16f; Flannaghan et al. 2014). These potential issues with the AMIPII dataset are particularly problematic since the ongoing CFMIP protocol contains amipPiForcing experiments (Webb et al. 2017). On a separate point, in relation to ERSSTv5 it may be relevant that over most of its record gradual changes are actually determined by measurements of nighttime marine air temperatures, which are arguably poorer than SST data (Rayner et al. 2003).

Although only indirect evidence, we find that in only 0.06% of the cases is internal variability as generated in preindustrial control simulations with CMIP5 coupled climate models able to capture the strong unforced pattern effects estimated in amipPiForcing experiments based on the AMIPII dataset (Andrews et al. 2018), and in only 10% of cases is it sufficient to capture unforced pattern effects of one-quarter their strength. Therefore, if internal variability in at least some CMIP5 AOGCMs is realistic, it seems highly probable that either the AMIPII SST dataset is flawed or at least part of the historical pattern effect detected when using AMIPII SST data is forced. Supporting this argument, Zhou et al. (2016) found that if decadal time scale internal variability in CMIP5 piControl simulations is realistic then at least part of the 1980–2005 AMIPII SST trend pattern was likely forced. Moreover, if there

were strong unforced pattern effects associated with internal variability one would expect the rate of warming relative to the rate of forcing to vary substantially over time. However, such variations appear surprisingly small. Taking non-overlapping 15-yr means to average out shorter-term variability and adjusting for the low efficacy of volcanic forcing, since 1941 that ratio has remained remarkably constant, being unusually low only over 1972–86 (Lewis and Curry 2018).

It is unclear from our results to what extent there is a robust relationship between stronger climate feedback and higher SST trends in the Indo-Pacific warm pool compared with elsewhere, at least where the comparison is limited to the tropics.

We caution that care is needed when using regression to estimate feedback in AMIP simulations, with nonnegligible bias toward overly strong estimates possible when regressing annual-mean data.

Sea ice variation is an important factor for climate feedback in AOGCM simulations. A limitation of this study, and those with which it compares and contrasts results, is that AMIP experiments are used in which sea ice is prescribed, generally using AMIPII sea ice (essentially HadISST1) data. There are large uncertainties in sea ice data prior to the satellite era, particularly around Antarctica. Nevertheless, Gregory and Andrews (2016) showed that even when sea ice is fixed at climatological 1871–1900 levels, much the same SST-driven pattern effect arises. They found that feedback for the AMIPII SST pattern with fixed climatological sea ice does not differ greatly from that when sea ice varies per the AMIPII dataset, and feedback for the years 1–20 abrupt4xCO2 SST pattern with fixed climatological sea ice is little different from that in the AOGCM abrupt4xCO2 experiment. However, Andrews et al. (2018) found that climate feedback in amipPiForcing simulations by two Met Office GCMs was much weaker when the HadISST2 rather than the AMIPII sea ice dataset was used, in conjunction with HadISST2 SST data, mainly due to the change in sea ice data rather than in SST data, and corresponded to a negative unforced historical pattern effect.[4]

Although sea ice uncertainty represents a further, unquantified, source of uncertainty in estimates of the absolute level of the unforced historical pattern effect, it is unlikely to greatly affect our estimates of the differences in that effect between SST datasets. The main focus of our Green's function based investigations, which suffer from greater limitations in relation to sea ice (since they incorporate no variation in it), is on the differences in estimated feedbacks between various SST datasets. Moreover, the accurate estimation of climate feedback in the AMIPII driven amipPiForcing simulation provided by the CAM5.3 Green's functions suggests that the lack of sea ice variation is unlikely to significantly bias the Green's function–based feedback estimates for other SST datasets.

A further limitation of this study is that it is based on simulations by a single GCM, combined with estimates using

---

[4] Using feedback estimated by regression over years 1–50 of the parent AOGCMs' abrupt4xCO2 simulations as a proxy for their forced historical feedback over 1871–2010.

Green's functions derived from a different GCM. It would therefore be useful if simulations employing alternative SST datasets were run with more models such that the feedback parameter can be compared with that from the corresponding coupled AOGCMs in historical and purely CO_2 forced simulations. The necessary forcing estimates, which were only available to us for ECHAM6.3, could become available from a range of models through experiments in the RFMIP protocol (Pincus et al. 2016).

The potential presence of a strong unforced pattern effect, as suggested by studies based on the AMIPII dataset, is particularly worrying since such internal variability could change in unpredictable ways over short periods of time. More so, since these patterns were thought to dampen global warming one might assert that rapid global warming could lie ahead. On the contrary, if it turns out that the historical record is not substantially influenced by unforced pattern effects—as suggested here—then global warming could continue in a more predictable fashion in line with anthropogenic and natural forcing over this century.

REFERENCES

Adams, B. K., and A. E. Dessler, 2019: Estimating transient climate response in a large-ensemble global climate model simulation. *Geophys. Res. Lett.*, **46**, 311–317, https://doi.org/10.1029/2018GL080714.

Andrews, T., and M. J. Webb, 2018: The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical Pacific warming. *J. Climate*, **31**, 641–654, https://doi.org/10.1175/JCLI-D-17-0087.1.

——, J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models. *Geophys .Res. Lett.*, **39**, L09712, https://doi.org/101029/2012GL051607.

——, ——, and ——, 2015: The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Climate*, **28**, 1630–1648, https://doi.org/10.1175/JCLI-D-14-00545.1.

——, and Coauthors, 2018: Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophys. Res. Lett.*, **45**, 8490–8499, https://doi.org/10.1029/2018GL078887.

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

——, C. M. Bitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *J. Climate*, **26**, 4518–4534, https://doi.org/10.1175/JCLI-D-12-00544.1.

Barlow, M., H. Cullen, and B. Lyon, 2002: Drought in central and Southwest Asia: La Niña, the warm pool, and Indian Ocean precipitation. *J. Climate*, **15**, 697–700, https://doi.org/10.1175/1520-0442(2002)015<0697:DICASA>2.0.CO;2.

Barsugli, J. J., and P. D. Sardeshmukh, 2002: Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. *J. Climate*, **15**, 3427–3442, https://doi.org/10.1175/1520-0442(2002)015<3427:GASTTS>2.0.CO;2.

Ceppi, P., and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proc. Natl. Acad. Sci. USA*, **114**, 13 126–13 131, https://doi.org/10.1073/pnas.1714308114.

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

Cowtan, K., and R. G. Way, 2014a: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, https://doi.org/10.1002/qj.2297.

——, and ——, 2014b: Coverage bias in the HadCrut4 temperature record and its impact on recent temperature trends. Update: Temperature reconstruction by domain: Version 2.0 temperature series. 9 pp., https://www.webcitation.org/6t09bN8vM.

——, and ——, 2014c: Coverage bias in the HadCrut4 temperature record and its impact on recent temperature trends. University of York Department of Chemistry, accessed 20 March 2019 and 8 October 2019, https://www-users.york.ac.uk/%7Ekdc3/papers/coverage2013//had4sst4_krig_v2_0_0.nc.gz.

Dessler, A. E., 2020: Potential problems measuring climate sensitivity from the historical record. *J. Climate*, **33**, 2237–2248, https://doi.org/10.1175/JCLI-D-19-0476.1.

——, T. Mauritsen, and B. Stevens, 2018: The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity. *Atmos. Chem. Phys.*, **18**, 5147–5155, https://doi.org/10.5194/acp-18-5147-2018.

Dong, Y., C. Proistosescu, K. C. Armour, and D. S. Battisti, 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. *J. Climate*, **32**, 5471–5491, https://doi.org/10.1175/JCLI-D-18-0843.1

Flannaghan, T. J., S. Fueglistaler, I. M. Held, S. Po-Chedley, B. Wyman, and M. Zhao, 2014: Tropical temperature trends in atmospheric general circulation model simulations and the impact of uncertainties in observed SSTs. *J. Geophys. Res. Atmos.*, **119**, 13 327–13 337, https://doi.org/10.1002/2014JD022365.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Fueglistaler, S., C. Radley, and I. M. Held, 2015: The distribution of precipitation and the spread in tropical upper tropospheric temperature trends in CMIP5/AMIP simulations. *Geophys. Res. Lett.*, **42**, 6000–6007, https://doi.org/10.1002/2015GL064966.

Geoffroy, O., and D. Saint-Martin, 2014: Pattern decomposition of the transient climate response. *Tellus*, **66A**, 23393, https://doi.org/10.3402/tellusa.v66.23393.

——, and ——, 2020: Equilibrium- and transient-state dependencies of the radiative feedback: Are they important for climate projections? *J. Climate*, **33**, 1863–1879, https://doi.org/10.1175/JCLI-D-19-0248.1.

——, ——, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéca, 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. *J. Climate*, **26**, 1859–1876, https://doi.org/10.1175/JCLI-D-12-00196.1.

Gregory, J. M., and T. Andrews, 2016: Variation in climate sensitivity and feedback parameters during the historical period. *Geophys. Res. Lett.*, **43**, 3911–3920, https://doi.org/10.1002/2016GL068406.

——, and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, https://doi.org/10.1029/2003GL018747.

——, T. Andrews, P. Good, T. Mauritsen, and P. M. Forster, 2016: Small global-mean cooling due to volcanic radiative forcing. *Climate Dyn.*, **47**, 3979–3991, https://doi.org/10.1007/s00382-016-3055-1.

——, ——, P. Ceppi, T. Mauritsen, and M. J. Webb, 2019: How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? *Climate Dyn.*, **54**, 129–157, https://doi.org/10.1007/S00382-019-04991-Y.

Hansen, J., and Coauthor, 2005: Efficacy of climate forcings. *J. Geophys Res.*, **110**, D18104, https://doi.org/10.1029/2005JD005776.

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

Hirahara, S., M. Ishii, and Y. Fukuda, 2014a: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57–75, https://doi.org/10.1175/JCLI-D-12-00837.1.

——, and ——, 2014b: COBE-SST2 data provided by the NOAA/OAR/ESRL PSD, accessed 8 November 2019, ftp://ftp.cdc.noaa.gov/Datasets/COBE2/sst.mon.mean.nc.

Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153, https://doi.org/10.1175/2008JCLI2292.1.

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

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res. Oceans*, **102**, 27 835–27 860, https://doi.org/10.1029/97JC01734.

Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker, 2011a: Reassessing biases and other uncertainties in sea-surface temperature observations since 1850: 1. Measurement and sampling errors. *J. Geophys. Res.*, **116**, D14103, https://doi.org/10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea-surface temperature observations since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, https://doi.org/10.1029/2010JD015220.

——, ——, C. P. Atkinson, and R. E. Killick, 2019: An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set. *J. Geophys. Res. Atmos.*, **124**, 7719–7763, https://doi.org/10.1029/2018JD029867.

Kent, E. C., and Coauthors, 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *J. Geophys. Res. Atmos.*, **118**, 1281–1298, https://doi.org/10.1002/JGRD.50152.

Lewis, N., and J. A. Curry, 2015: The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dyn.*, **45**, 1009–1023, https://doi.org/10.1007/s00382-014-2342-y.

——, and ——, 2018: The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *J. Climate*, **31**, 6051–6071, https://doi.org/10.1175/JCLI-D-17-0667.1.

——, and ——, 2020: Reply to "Comment on 'The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity.'" *J. Climate*, **33**, 397–404, https://doi.org/10.1175/JCLI-D-18-0669.1.

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

Mauritsen, T., 2016: Clouds cooled the earth. *Nat. Geosci.*, **9**, 865–867, https://doi.org/10.1038/ngeo2838.

——, and R. Pincus, 2017: Committed warming inferred from observations. *Nat. Climate Change*, **7**, 625–655, https://doi.org/10.1038/NCLIMATE3357.

——, and Coauthors, 2019: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *J. Adv. Model. Earth Syst.*, **11**, 998–1038, https://doi.org/10.1029/2018MS001400.

Miao, J., T. Wang, H. Wang, and J. Sun, 2018: Interannual weakening of the tropical Pacific Walker circulation due to strong tropical volcanism. *Adv. Atmos. Sci.*, **35**, 645–658, https://doi.org/10.1007/s00376-017-7134-y.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nat. Geosci.*, **6**, 415–416, https://doi.org/10.1038/ngeo1836.

Pincus, R., P. M. Forster, and B. Stevens, 2016: The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental protocol for CMIP6. *Geosci. Model Dev.*, **9**, 3447–3460, https://doi.org/10.5194/GMD-9-3447-2016.

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

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Richardson, T. B., and Coauthors, 2019: Efficacy of climate forcings in PDRMIP models. *J. Geophys. Res. Atmos.*, **124**, 12 824–12 844, https://doi.org/10.1029/2019JD030581.

Rugenstein, M., and Coauthors, 2020: Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophys. Res. Lett.*, **47**, e2019GL083898, https://doi.org/10.1029/2019GL083898.

Senior, C. A., and J. F. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophys. Res. Lett.*, **27**, 2685–2688, https://doi.org/10.1029/2000GL011373.

Shindell, D. T., 2014: Inhomogeneous forcing and transient climate sensitivity. *Nat. Climate Change*, **4**, 274–277, https://doi.org/10.1038/nclimate2136.

Skeie, R. B., T. Berntsen, M. Aldrin, M. Holden, and G. Myhre, 2018: Climate sensitivity estimates—Sensitivity to radiative forcing time series and observational data. *Earth Syst. Dyn.*, **9**, 879–894, https://doi.org/10.5194/esd-9-879-2018.

Sobel, A. H., I. M. Held, and C. S. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature. *J. Climate*, **15**, 2702–2706, https://doi.org/10.1175/1520-0442(2002)015<2702: TESITT>2.0.CO;2.

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

Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, https://doi.org/10.1002/2013JD020316.

Visser, K., R. Thunell, and L. Stott, 2003: Magnitude and timing of temperature change in the Indo-Pacific warm pool during deglaciation. *Nature*, **421**, 152–155, https://doi.org/10.1038/nature01297.

Webb, M. J., and Coauthors, 2017: The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geosci. Model Dev.*, **2017**, 359–384, https://doi.org/10.5194/gmd-10-359-2017.

Welch, B. L., 1947: The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, **34**, 28–35, https://doi.org/10.1093/BIOMET/34.1-2.28.

Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, **23**, 2333–2344, https://doi.org/10.1175/2009JCLI3139.1.

Wood, R., and C. S. Bretherton, 2006: On the relationship between stratiform low cloud cover and lower-tropospheric stability. *J. Climate*, **19**, 6425–6432, https://doi.org/10.1175/JCLI3988.1.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer, 1987: A comprehensive ocean–atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**, 1239–1250, https://doi.org/10.1175/1520-0477(1987) 068<1239:ACOADS>2.0.CO;2.

Zhou, C., M. D. Zelinka, and S. A. Klein, 2016: Impact of decadal cloud variations on the Earth's energy budget. *Nat. Geosci.*, **9**, 871–874, https://doi.org/10.1038/ngeo2828.

——, ——, and ——, 2017: Analyzing the dependence of global cloud feedback one the spatial pattern of sea surface temperature change with a Green's function approach. *J. Adv. Model. Earth Syst.*, **9**, 2174–2189, https://doi.org/10.1002/2017MS001096.