Change in Climate Sensitivity and Its Dependence on the Lapse-Rate Feedback in 4 × CO_2 Climate Model Experiments

KAI-UWE EISELT[a] AND RUNE GRAND GRAVERSEN[a,b]

[a] *Institute for Physics and Technology, University of Tromsø, Tromsø, Norway*
[b] *Norwegian Meteorological Institute, Tromsø, Norway*

(Manuscript received 13 August 2021, in final form 18 January 2022)

ABSTRACT: Robust estimates of climate sensitivity are important for decision-making on mitigation of climate change. However, climate sensitivity and its governing processes are still subject to large uncertainty. Recently it has been established that climate sensitivity changes over time in numerical climate model experiments with abrupt quadrupling of the CO_2 concentration. Here we conduct an analysis of such experiments from a range of climate models from phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP). Climate feedbacks associated with clouds, lapse rate, Planck radiation, surface albedo, and water vapor and their changes over time are diagnosed based on a radiative kernel method. We find two clearly distinct model groups, one with weak and one with strong lapse-rate feedback change. The Arctic is the region showing the largest differences between these two model groups, with respect to both warming change and individual feedback changes. We retrace this change to the development over time of the Arctic sea ice, which impacts both the surface-albedo and lapse-rate feedbacks. Generally, models that warm quickly, both globally and in the Arctic, also quickly lose their Arctic sea ice and change their total global-mean climate feedback only little, and vice versa. However, it remains unclear if the Arctic changes are a cause or rather a by-product of the total global-mean feedback change. Finally, we find support for the results of previous studies finding that the relative warming in the tropical Indo-Pacific region may control the change of total climate feedback over time.

KEYWORDS: Climate change; Climate sensitivity; Feedback; Radiative fluxes; Regional effects; Climate models; Coupled models; Model comparison

1. Introduction

The response of Earth's climate system to a forcing due to an increase in atmospheric greenhouse gas concentration has been of interest for more than 100 years (Arrhenius 1896). Equilibrium climate sensitivity (ECS), which is the equilibrium global-mean temperature increase due to a doubling of the CO_2 concentration in the atmosphere, is a traditional measure for the climate system's response to such a forcing (Charney et al. 1979; Sherwood et al. 2020). It is often estimated based on numerical climate model experiments. A straightforward model-based method for deriving ECS is to start from a model state in radiative equilibrium, then implementing a forcing corresponding to a doubling of the CO_2 concentration in the atmosphere, and finally running the model until radiative equilibrium is restored. However, in complex Earth system models (ESMs), which realistically represent the climate system, more than 1000 simulation years are necessary for the model to reach a new equilibrium because the deep ocean has a large heat capacity and thus warms only slowly. Hence, large computational resources are required in order to equilibrate a perturbed ESM (e.g., Paynter et al. 2018; Rugenstein et al. 2020). Gregory et al. (2004) introduced an alternative but indirect approach (known as the Gregory method) to estimate ECS from a much smaller amount of simulation years in instantaneous CO_2 forcing experiments, hence without the need of running the model to a new equilibrium. The global-mean top-of-atmosphere (TOA) radiative imbalance is expressed in terms of global-mean surface-air temperature (SAT) change as follows:

$$N = F + \alpha \Delta T_s, \tag{1}$$

where N is the TOA radiative imbalance, F is the forcing (e.g., due to an increase in CO_2), α is the climate feedback parameter, and ΔT_s is the SAT change. The climate feedback parameter α represents the total feedback of the climate system and is usually estimated by regressing global-mean TOA imbalance on global-mean SAT change, and climate sensitivity is estimated by extrapolating this relationship to zero imbalance [i.e., where N in Eq. (1) is zero]. Sherwood et al. (2020) call the climate sensitivity thus estimated the *effective climate sensitivity* (S), to separate it from ECS derived from equilibrating an ESM. The climate system's response and hence its climate sensitivity is controlled by its total feedback, which is negative provided that the climate system is stable.

As mentioned by Gregory et al. (2004), and later confirmed for several other ESMs (e.g., Andrews et al. 2015), the estimates for S and α depend on the interval of simulation years included in the ΔT_s–N regression. In most simulations, S is smaller when based on the years directly following the forcing than when based on years later in the simulation, meaning

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0623.s1.

Corresponding author: Kai-Uwe Eiselt, kai-uwe.eiselt@uit.no

DOI: 10.1175/JCLI-D-21-0623.1

that S increases over time. Correspondingly, the total feedback estimate becomes less negative over time.

The time dependence of climate sensitivity and feedback has been extensively researched but is still not well understood. Understanding the mechanism leading to a weakening of total feedback and hence an increase in S will help to gauge the likelihood of the possible occurrence of this shift in ongoing and future climate change. These model-based results are of great importance since they call into question the reliability of climate sensitivity estimates based on historical data, which, by necessity, cannot take into account future feedback changes. The estimates based on observations and proxy data typically fall on the lower end of expected climate sensitivity (Sherwood et al. 2020, and references therein).

Senior and Mitchell (2000) report a change over time of climate sensitivity in their CO_2-doubling experiment with an early coupled climate model, and argue that this is due to a change of cloud feedback over time that is associated with stability changes resulting from the delayed warming of the Southern Hemisphere compared to the Northern Hemisphere. Williams et al. (2008) take an alternative approach and introduce the concept of *effective forcing*, which is the greenhouse gas forcing adjusted for processes that occur on time scales that are short compared to those of climate stabilization. They argue that after the initial adjustment process, the ΔT_s–N relation is linear on centennial time scales and hence that the change over time of climate sensitivity is an artifact of not accounting for the relatively fast forcing adjustments. However, Winton et al. (2010) challenge this notion, since the adjustments occur on time scales on the order of decades where the "oceanic adjustment" is an important factor as well. They propose a time-varying *ocean heat uptake efficacy*, considering ocean heat uptake as a forcing that accounts for the change in the ΔT_s–N relationship over time. Armour et al. (2013) show that the evolution of global total feedback can be explained by a change over time in spatial weighting of time-invariant local feedbacks due to evolving patterns of surface warming. However, Ceppi and Gregory (2017) find that this cannot explain the global total feedback changes in a suite of ESMs that they analyze.

The perhaps most recent hypothesis for explaining the time dependence of climate feedback and S as presented by Ceppi and Gregory (2017) is the following: Upon CO_2 quadrupling, sea surface temperature (SST) warming patterns change over time to increasingly favor stably stratified regions where surface warming remains trapped near the surface and is not readily communicated vertically and horizontally, thus inducing weaker global cooling efficiency over time. In the following we refer to this as the stability hypothesis. The southeastern tropical Pacific (EP) is a region of descent (stably stratified) and the tropical western Pacific (WP) a region of ascent (close to neutrally stratified). Across members of phase 5 of the Coupled Model Intercomparison Project (CMIP5) a *delayed* warming in the EP and an *early* warming in the WP in response to an instantaneous forcing are observed (e.g., Andrews et al. 2015). A physical mechanism is suggested by Zhou et al. (2016) and further developed by Andrews and Webb (2018) that explains how this change in warming pattern causes a weakening of total feedback over time: The earlier warming in the WP readily affects higher levels in the atmosphere whereby energy is efficiently radiated out to space. Over time the warming shifts more to the EP and because this region is stably stratified, the heat remains trapped close to the surface, implying a less negative lapse-rate feedback and a reduction of the cooling efficiency of the Earth system. Moreover, the surface warming in the stably stratified regions decreases the stability there, reducing the low cloud cover (see also Wood and Bretherton 2006). Since low clouds mainly act to reflect incoming solar radiation back to space, a reduction in low cloud cover enhances the absorption of solar radiation, implying a positive cloud feedback. This means that as the warming shifts from WP to EP over time, the total feedback becomes less negative, due to a change in both lapse-rate and cloud feedback. Andrews and Webb (2018) confirm this mechanism for one atmosphere-only model (HadGEM2-A; Martin et al. 2011) by performing experiments with prescribed surface warmings in the dedicated regions. Ceppi and Gregory (2017) show that changes over time in climate feedbacks (especially cloud and lapse rate) are consistent with this mechanism and the corresponding stability reduction over a range of CMIP5 members. Andrews and Webb (2018) construct the tropical Pacific warming pattern index (TPI) where the mean warming change in the WP is subtracted from that in the EP. If the mechanism proposed by Zhou et al. (2016) and Andrews and Webb (2018) is a main driver of feedback change over time across models we would expect a positive correlation of the TPI change with the change of total feedback across models. However, Andrews and Webb (2018) do not find such a correlation, but suggest that this lack of linkage might be because the index is not refined well enough, or that other regions are drivers of the total feedback change. Zhou et al. (2017) introduced a Green's function approach to investigate the influence of regional warming on global cloud feedback evolution and Dong et al. (2019) expand this to global total feedback. They find that the influence of the EP as well as of northern and southern polar regions on total feedback is small and argue that this is because these regions are stably stratified and local changes are confined there. However, their analysis points to the Indo-Pacific warm pool (IPWP) region, including the WP and large parts of the Indian Ocean, as having the largest impact on the change over time of total feedback. In a follow-up study, Dong et al. (2020) succeed to reconstruct the change of total feedback from the IPWP surface warming reasonably well for members of CMIP5, but not of CMIP6. Dong et al. (2020) suggest that the failure of reconstruction for CMIP6 results from strong EP and Southern Hemisphere midlatitude cloud feedbacks.

Like Dong et al. (2020), we perform a feedback analysis of instantaneous CO_2 forcing experiments in a suite of ESMs from CMIP5 and CMIP6. We adopt their terminology in assigning the years 1–20 to the *early* period and the years 21–150 to the *late* period (see also Ceppi and Gregory 2017). "Change over time" in the following refers to the change from early to late period. Unlike in previous studies, which

either use individual models (e.g., Andrews and Webb 2018; Dong et al. 2019) or compare CMIP5 and CMIP6 (e.g., Dong et al. 2020), we group models according to their feedback changes in order to find prominent similarities and differences that potentially can explain the differences in total feedback change. Given our model grouping, we examine the influence of surface warming, stability, and feedback changes in various specific regions on total feedback change and we investigate correlations across models.

The remainder of this study is structured as follows. Section 2 gives a short overview over the model simulation data we consider and section 3 describes the radiative kernel method used to estimate individual climate feedbacks. In section 4 we present the results and offer some discussion. Section 5 concludes with a summary.

2. Models and experiments

We use data from the CMIP5 and CMIP6 abrupt4xCO2 experiments where the CO_2 concentration is abruptly increased by a factor of 4 in comparison to a preindustrial equilibrium condition. The minimum length of this experiment in CMIP5 and CMIP6 is 150 years although there exist longer simulations for some models (Taylor et al. 2009; Eyring et al. 2016). Hence, we only analyze the first 150 years for each available model run. All abrupt4xCO2 experiments have a corresponding preindustrial control experiment (piControl) that is in radiative equilibrium. For each of the quantities, following Zelinka et al. (2020), we take a 21-yr running mean over this control run and subtract from the abrupt4xCO2 run to derive the changes due to the CO_2 forcing and to remove a possible model drift. All models used in the present study are listed and referenced in Tables S1 and S2 in the online supplemental material.

3. Methods

Radiative kernel method

The radiative kernel method (Soden et al. 2008) can be used to derive the magnitude of individual radiative feedbacks activated due to a temperature change associated with a forcing of the climate system. It is based on the assumptions that 1) the TOA net radiation (N) can be described as a function of independent climate system variables (i.e., these variables do not interact or interact only little) and that 2) the radiative flux change due to a small change in one of these variables is linear. As an example, the surface albedo (SA) kernel can hence be defined as

$$\begin{aligned} &N(a + \delta a, T, w, c) - N(a, T, w, c) \\ &\quad = N(\delta a) = \frac{\partial N}{\partial a}(a, T, w, c)\delta a \equiv K^a \delta a, \end{aligned} \tag{2}$$

where a is the SA, T the temperature, w the water vapor (WV) mixing ratio, c a set of cloud properties, and δa a small SA perturbation; K^a represents the SA kernel.

Radiative kernels for a specific climate system variable are derived from a climate model or a reanalysis. In a given climate state with TOA imbalance $N(a, T, w, c)$, the variable in question is perturbed by a specific amount (e.g., 1% for SA per grid cell) and then only the radiation code is executed. This yields the TOA imbalance of the given climate state with perturbed SA, $N(a + \delta a, T, w, c)$, where δa is 1%. The radiative kernel, K^a, can then be calculated from Eq. (2). Since radiative transfer schemes in climate models are well tested and fairly similar across models, a set of kernels based on a given model is expected to be applicable across models (e.g., Shell et al. 2008). This is supported by the fact that feedbacks calculated with radiative kernels derived from different climate models yield mostly similar results (e.g., Soden et al. 2008), although results for SA feedback can vary considerably across kernels (Donohoe et al. 2020; Hahn et al. 2021). Radiative kernels can be generated for surface and air temperature, WV mixing ration, and SA. These are all either direct outputs of the here-used CMIP model simulations or can be calculated from them.

Since the effects of clouds on radiation are strongly nonlinear, the radiative kernels described here are not appropriate for cloud feedback [for radiative kernels for cloud feedback using cloud-top pressure and optical thickness that are not available in the simulations used here, see Zelinka et al. (2012, 2016)]. However, the radiative flux change due to cloud changes, $N(\delta c)$, can still be estimated by adjusting the change in cloud radiative effect (CRE; calculated as TOA all-sky minus TOA clear-sky radiation, for TOA radiation being positive downward) by the cloud masking of the other radiative feedbacks. This procedure is described in Soden et al. (2008) and can be performed by adding the difference of kernel-derived clear-sky minus all-sky radiative flux changes to the CRE as calculated from model output:

$$\begin{aligned} N(\delta c) = \mathrm{CRE} &+ (K^a_{\mathrm{cs}} - K^a_{\mathrm{as}})\delta a + \left(K^T_{\mathrm{cs}} - K^T_{\mathrm{as}}\right)\delta T \\ &+ (K^w_{\mathrm{cs}} - K^w_{\mathrm{as}})\delta w + \left(F^{4x}_{\mathrm{cs}} - F^{4x}_{\mathrm{as}}\right). \end{aligned} \tag{3}$$

The subscripts "as" and "cs" denote all-sky and clear-sky, respectively, and F^{4x} represents the forcing due to a quadrupling of the CO_2 concentration. For a derivation of Eq. (3) see appendix A.

The above-given assumptions of the kernel method allow a decomposition of the total feedback parameter α in Eq. (1) into individual feedback parameters related to physical processes and we can rewrite the equation as

$$N = F + (\alpha_a + \alpha_T + \alpha_w + \alpha_c)\Delta T_s, \tag{4}$$

where α_a, α_T, α_w, and α_c are the feedback parameters due to a change in SA, surface and air temperature, WV mixing ratio, and cloud properties, respectively. This decomposition together with the kernel-derived radiative responses for the individual variables yields the possibility of performing the Gregory method for each individual feedback and the feedback parameters themselves can be derived via a linear regression, similar to the total feedback parameter (e.g., Block and Mauritsen 2013). As an example we can write for the radiative effect of the SA feedback:

$$N(\delta a) = \alpha_a \Delta T_s, \tag{5}$$

where $N(\delta a)$ is the TOA radiative flux change due to the change in SA, which is derived with the radiative kernels according to Eq. (2). Note that the forcing terms in Eqs. (3) and (4) are constant in abrupt forcing experiments. Thus, they do not influence the feedback which is calculated as the linear regression slope. The temperature feedback (α_T) may be split into the Planck feedback and the lapse-rate (LR) feedback. The Planck feedback can be derived by assuming a uniform temperature change equal to the surface temperature change throughout the atmospheric column. It can hence be calculated by using the surface (skin) temperature change on all atmospheric levels and multiplying it on the air temperature kernel. Subtracting the Planck feedback from the air temperature feedback gives the LR feedback. This feedback is strongly linked to tropospheric stability because it originates from the stratification of warming in the atmosphere. In a stably stratified region a surface warming is confined at lower levels, which implies that, excluding other influences, the warming at higher levels in that region is lower (i.e., the stability decreases). This results in a positive LR feedback, since the lapse rate reduces the cooling efficiency. Hence, a reduction in tropospheric stability is typically connected with an increase in LR feedback. On the other hand, if there is an upper-level warming, while the surface warming is comparatively weaker, meaning an increase in stability, a negative LR feedback ensues since the cooling to space is hereby stronger than it would have been under uniform warming. An increase in stability therefore tends to be associated with a decrease in LR feedback.

If the assumptions of the kernel method are appropriate, the sum of the individual feedback parameters should be approximately equal to the total feedback as derived from model output radiative fluxes using the Gregory method. To test if these assumptions are sufficiently fulfilled in a given model simulation, we employ the clear-sky linearity test (CSLT; Shell et al. 2008). This test consists of comparing the sum of the kernel-derived clear-sky feedbacks (i.e., $\alpha_a^{cs} + \alpha_T^{cs} + \alpha_w^{cs}$) with the total clear-sky feedback derived from the Gregory method using model output clear-sky radiative fluxes. In appendix B we show that in the kernel decomposition used here it does not matter if all-sky or clear-sky feedbacks are compared. Following Ceppi and Gregory (2017), we exclude models whose kernel-derived clear-sky feedback sum deviates more than 15% from the total clear-sky feedback in *both* the early and late period calculated independently.

We calculated the feedbacks for 66 members of CMIP5 and CMIP6 listed in Tables S1 and S2 using six different sets of radiative kernels (Soden et al. 2008; Shell et al. 2008; Block and Mauritsen 2013; Huang et al. 2017; Pendergrass et al. 2018; Smith et al. 2018). A maximum of 37 members pass the CSLT for the Shell et al. (2008) kernels and we focus our analysis on these results. There is, however, only a small variation in feedbacks derived from different kernels. It is found that three outliers (MIROC-ES2L, GISS-E2-R, and GISS-E2.2-G) lie on opposite sides of the distribution of models, especially for the WV and LR feedbacks. We exclude these models since they may artificially enhance correlations across models of different aspects that are investigated in the present study. Where in the following we mention "all models" we refer to the remaining 34 models. The whole analysis was repeated with a CSLT error threshold of 20% for which 50 models pass using the Shell et al. (2008) kernels. Generally, the results are little different from the original analysis and the conclusions do not change. The same is true if the complete analysis is conducted using the Pendergrass et al. (2018) kernels for which 36 (CSLT threshold 15%) and 44 models (CSLT threshold 20%) pass the CSLT. Note that SA feedback and its change over time calculated with the Shell et al. (2008) kernels exhibit the smallest standard deviation across models of all tested kernels while for the Pendergrass et al. (2018) kernels they exhibit the largest standard deviation. Hence, we conclude that the above-mentioned dependence of SA feedback on the chosen set of radiative kernels does not significantly impact our results.

We apply a simple mask for the stratosphere by excluding values on pressure levels exceeding a lower threshold that varies with the cosine of the latitude between 300 hPa at the poles and 100 hPa at the equator. Thus, our feedbacks do not include stratospheric adjustments.

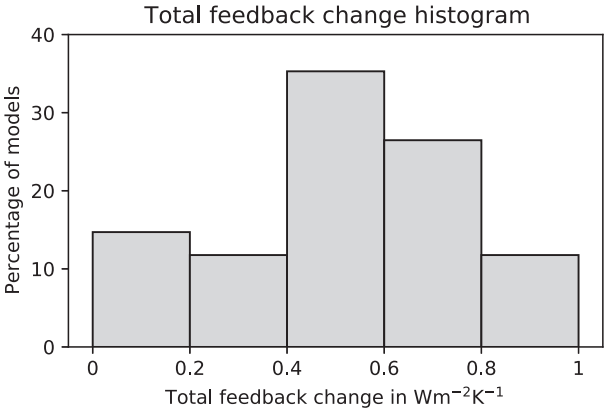

FIG. 1. Distribution of total feedback change from early (years 1–20) to late period (year 21–150) across models ("all models"; refer to section 3 for details). For the distribution of all 66 models with available data see Fig. S1 in the online supplemental material.

4. Results

We now first present the results of the kernel feedback analysis and then proceed to compare groups of models distinguished by their feedback changes. Informed by this analysis we investigate the impact of regional warming and stability changes on total feedback change.

a. Feedback analysis

Figure 1 shows the distribution of total feedback change from early to late period derived from the Gregory method across all models. Most models concentrate

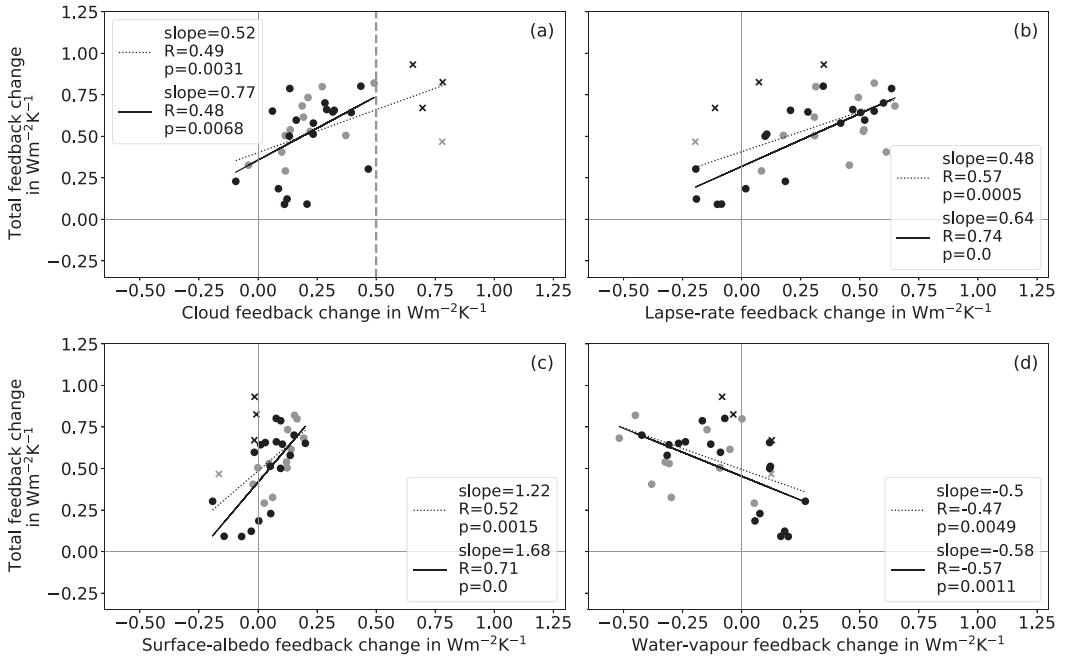

FIG. 2. Total vs individual kernel-derived feedback changes from early (years 1–20) to late period (years 21–150). The dotted black line represents the linear regression for all models and the solid black line represents the regression if the strong cloud feedback change models are excluded (see text for discussion). These models are indicated by × markers. In (a) the dashed gray vertical line shows the threshold value (0.5 W m^{-2} K^{-1}) for "strong cloud feedback change". Members of CMIP5 are depicted in gray and members of CMIP6 in black. The R and p values are the correlation coefficient and significance based on a two-sided p value calculated from a Wald test with a t distribution, respectively.

around 0.5 W m^{-2} K^{-1} change but there is considerable spread with some models showing little feedback change while others show considerable change of close to 1 W m^{-2} K^{-1}. Notably, none of the investigated models exhibits a negative change. The few models that do in fact exhibit negative change fail the CSLT or are considered outliers (see section 3 as well as Tables S1 and S2 and Fig. S1).

We now proceed with the kernel feedback analysis, and decompose the total feedback into individual components to investigate the cause for the total feedback change. Figure 2 shows the intermodel linkage between individual feedback changes and the total (i.e., kernel sum) feedback change. For some models with a large feedback change, the change of the cloud feedback is particularly strong ($\Delta\alpha_c > 0.5$ W m^{-2} K^{-1}; Fig. 2a and Table S3). From comparison with the other feedbacks it is clear that for these models, the change in cloud feedback is the dominating mechanism. These models are shown with × markers in Fig. 2 and correlations are shown both for all models (black dotted line) and for models excluding the strong cloud feedback change models (black solid line). In the latter set of models, a large total feedback change mostly results from a large LR feedback change (Fig. 2b), although the cloud feedback change seems important as well. For this set of models there appears to be a strong and significant correlation between the LR feedback change and the total feedback change. The SA (Fig. 2c) and Planck (not shown) feedback changes are generally smaller but their correlation with the total feedback change is strong provided that the models with large cloud feedback change are excluded. The WV feedback change is anticorrelated with the total feedback change (Fig. 2d), and this anticorrelation is again strong if the strong cloud feedback change models are excluded. We conclude that the set of models with strong cloud feedback change responds somewhat differently to a forcing compared to the remainder of models in that the cloud feedback change is exceptionally large and explains most of the total feedback change while the other feedback changes are less important for these models. Most models in this group (this also holds in the case of a CSLT threshold of 20%) have particularly large positive cloud feedback change over the Southern Ocean (not shown). This region has been noted for positive cloud feedback change due to cloud solid–liquid phase change due to recent updates of cloud-physics parameterization (i.e., feedback changes not directly related to stability changes; Bjordal et al. 2020). Hence, we disregard the models with strong cloud feedback change from our analysis. For the other models there appears to be a strong dependence of the total feedback change on the change of the LR feedback indicating that the change of this feedback explains a large part of the total feedback change. In the following section we thus investigate this group more extensively.

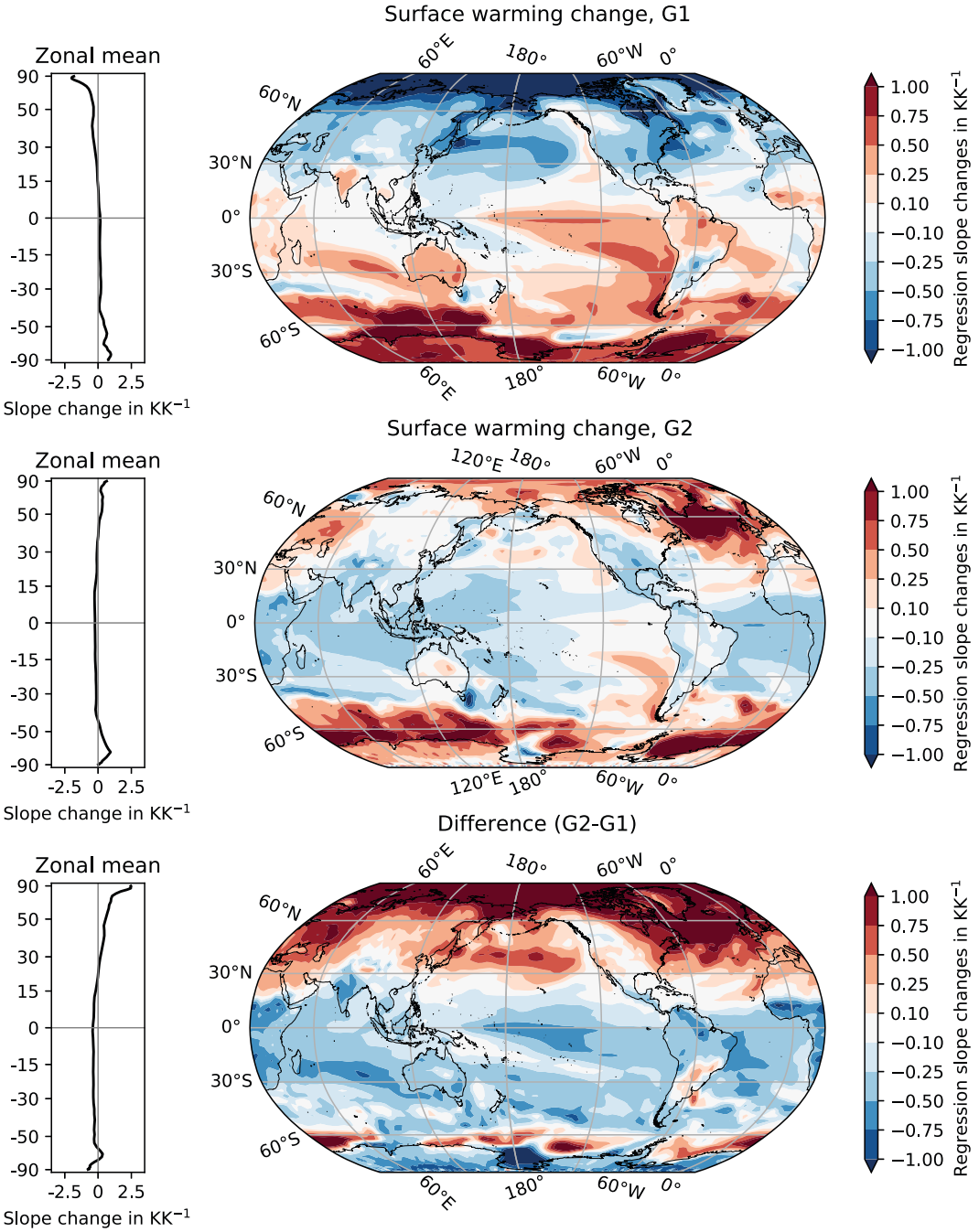

FIG. 3. (right) Maps and (left) zonal means of surface-warming change relative to the global-mean change time averaged over models with (top) weak (G1) and (middle) strong (G2) lapse-rate feedback change, and (bottom) their difference (G2 − G1). The surface-warming change is calculated by first regressing local annual-mean surface temperature on global annual-mean surface temperature for the periods of years 1–20 (early) and 21–150 (late), and then subtracting the results of the early from the late period (cf. Ceppi and Gregory 2017). Note that the *y* axis in the zonal-mean plot is scaled by the cosine of the latitude to obtain an equal-area perspective. See Figs. S2 and S3 for the maps for the early and late period, respectively.

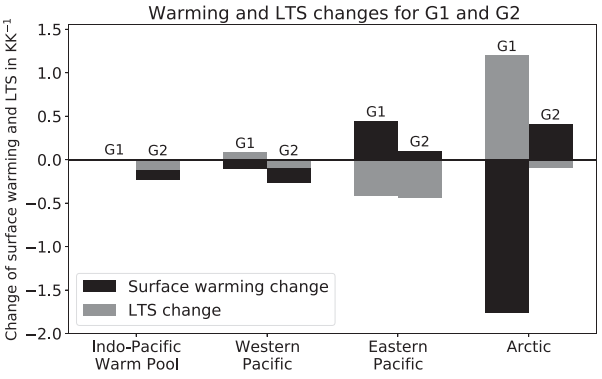

FIG. 4. Indo-Pacific warm pool, western Pacific, eastern Pacific, and Arctic surface-warming change (black) as well as lower-tropospheric stability (LTS; see text for details) change (gray) averaged over models with weak (G1) and strong (G2) lapse-rate feedback change (refer to the text for details). As in Fig. 3, the change of surface warming and LTS is calculated similarly by first regressing local surface temperature and LTS on global-mean surface temperature for both the early (years 1–20) and late period (years 21–150) and then subtracting the results of the early from the late period (cf. Ceppi and Gregory 2017). Note that while the Arctic, eastern Pacific, and western Pacific regions are of similar size, the Indo-Pacific warm pool region is much larger (see text for discussion and Table S6).

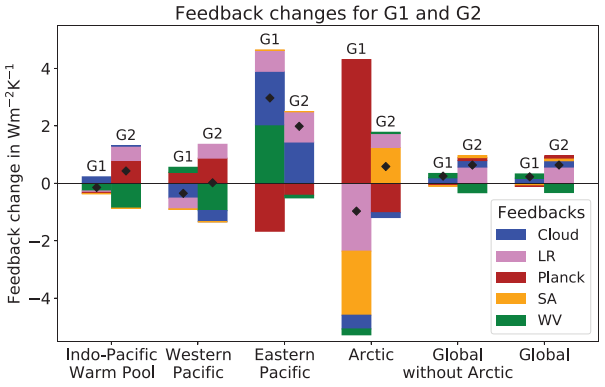

FIG. 5. Indo-Pacific warm pool, western Pacific, eastern Pacific, Arctic, global mean without Arctic, and global-mean feedback changes from the early (years 1–20) to the late period (years 21–150) decomposed into parts indicated by the color coding averaged over models with weak (G1) and strong (G2) lapse-rate feedback change (refer to the text for details). Feedbacks are added on top of each other for each sign and ordered according to their individual magnitudes. The black diamonds represent the total local feedback change.

b. *Comparing two model groups*

Excluding the models with strong cloud feedback change, we now more closely examine the remainder for which a strong coupling between the LR feedback change and the total feedback change is found. To investigate general differences between models, we divide this group into two subgroups, models with a weak LR feedback change (G1; $\Delta\alpha_{LR} > 0.1$ W m^{-2} K^{-1}; see Table S4) and models with a strong LR feedback change (G2; $\Delta\alpha_{LR} > 0.5$ Wm^{-2} K^{-1}; see Table S5). We conduct a sensitivity analysis by varying the thresholds by + and −0.05 W m^{-2} K^{-1}, respectively, and find that the overall conclusions do not change. Notably, since for these models the LR feedback change dominates the total feedback change, G1 tends to include models with weak total feedback change whereas G2 models with strong total feedback change. The slight positive feedback change in G1 is due to small positive changes in cloud and WV feedback that are counteracted by even smaller negative changes in Planck, LR, and SA feedback.

Similar to Ceppi and Gregory (2017) and Andrews and Webb (2018), we focus first on surface temperature. Figure 3 shows the change from early to late period of local surface temperatures regressed on global-mean surface temperature averaged over G1 (top panel) and G2 (middle panel), as well as the difference between the two groups (G2 minus G1, bottom panel). It is clear that both groups exhibit positive warming change relative to the global-mean change in the eastern subtropical Pacific as well as over the Southern Ocean and Antarctica. Furthermore, G1 exhibits stronger positive changes relative to the global average across almost the whole Southern Hemisphere, while G2 shows mostly negative changes between 30°S and 30°N. Accordingly, the difference between G1 and G2 is mostly negative south of 30°N with the exception of some parts north of Antarctica. The IPWP is noticeably different between the two groups, with little change in G1 but negative change in G2. However, the largest difference between the groups occurs in the Arctic. Here G1 exhibits negative change while the G2 change is positive. This means that the models showing little LR feedback change tend to quickly warm the Arctic in response to the abrupt greenhouse gas forcing while the Arctic warming relative to global mean during the later period is weaker. In contrast, there is a delay in relative Arctic warming in the models having large LR feedback change and this relative warming increases over time.

We now proceed with an investigation of individual regions with respect to their warming and stability change. As a stability metric we use the lower tropospheric stability (LTS), which is taken as the difference between surface and 700-hPa potential temperature (Klein and Hartmann 1993). First, we examine the Arctic since this region exhibits the largest difference between the two model groups. We also consider the regions pointed out in Andrews and Webb (2018) and Dong et al. (2019), namely the EP and the WP as well as the IPWP (see section 1). The warming and LTS changes averaged over the four regions are shown in Fig. 4 for both model groups. Note that the Arctic, the EP, and the WP are of roughly similar size, while the IPWP is considerably larger (~20 times as large; see Table S6).

In the Arctic (which we define as 75°–90°N) the largest surface-warming and LTS changes as compared to the other regions are found. Moreover, also here the biggest difference between the two model groups is apparent. G1 exhibits a strong negative warming change over time, while in G2 the change is positive but much weaker. The changes in LTS are opposite those of the surface warming. In accordance with accelerated early relative Arctic surface warming, the Arctic

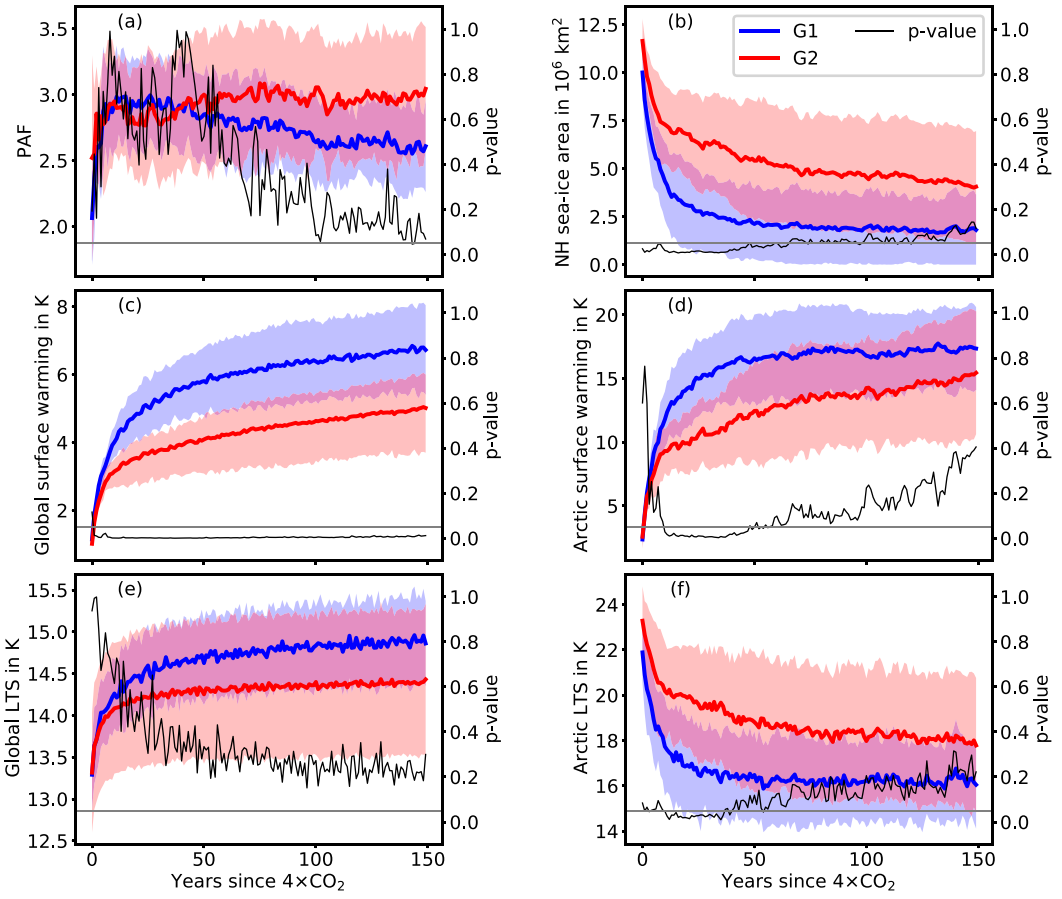

FIG. 6. (a) Polar amplification factor (PAF), (b) Northern Hemisphere (NH) sea ice area, (c) global mean, and (d) Arctic surface temperature anomaly, and (e) global mean and (f) Arctic lower-tropospheric stability (LTS; see text) averaged over models with weak (G1; blue) and strong (G2; red) lapse-rate feedback change (refer to the text for details). The lines indicate the multimodel means and the shading denotes the ±1-sigma spread. The PAF is calculated as the Arctic (75°–90°N) surface temperature anomaly divided by that of the global mean. The Northern Hemisphere sea ice area is calculated by multiplying the sea ice fraction in a grid cell by the grid cell area and then integrating over the NH. The black line shows the p value of a two-sided Welch's t test for the difference in group mean and the gray horizontal line indicates a p value of 0.05. No sea ice data were available for one member of G2 (BCC-CSM2-MR) so this model is excluded in (b).

LTS in G1 decreases strongly in the early period but as the relative Arctic surface warming subsequently slows down (cf. Fig. 6a), the decrease of LTS is much weaker in the late period, leading to the positive change of LTS seen in Fig. 4. In G2 the relative Arctic surface warming does not slow down (cf. also Fig. 6a) and the Arctic LTS decreases similarly in both early and late period, leading to the small negative change seen in Fig. 4.

The EP exhibits a positive surface warming change in G1, which is much larger than in G2. However, the magnitude of the change of G1 in the EP is generally much smaller than in the Arctic. Regarding the LTS change in the EP, G1 and G2 show smaller difference as compared to the surface warming. Note that the change is slightly more negative in G2. Hence, even though the surface in the EP warms less strongly in G2, the EP stability in G2 decreases more than in G1. Since total global-mean feedback changes more in G2 than in G1, this may indicate that the delayed EP surface warming, although a robust feature across ESMs in response to an abrupt greenhouse gas concentration increase, does not significantly influence total feedback change. This supports the findings of Dong et al. (2019) that it is the IPWP surface warming rather than the tropical east–west SST gradient that is important for total global feedback change.

As expected for regions of ascent, the LTS changes only little in the WP and the IPWP in both groups. In G1, both the WP and IPWP exhibit little or no surface warming change, while there is a negative change in both regions in G2. Since G1 consists mostly of models with weak total feedback change and G2 mostly of models with strong total feedback change this lends support to the findings of Dong et al. (2019, 2020) that the IPWP surface warming largely controls the total feedback change. Since the IPWP is a region of ascent, a surface warming there is quickly spread vertically

and in upper levels also horizontally through the atmosphere and hence increases the global cooling efficiency. In contrast, energy associated with a surface warming in other, generally more stable regions is not radiated to space as efficiently. The surface warming change in the IPWP thus determines the total global feedback change to a significant degree. Since G2 experiences a reduction in IPWP surface warming relative to the global average while in G1 it is almost constant, this implicates a reduction of the magnitude of the negative total feedback in G2 while the G1 total feedback should only change little, which is what we see.

We further investigate the changes of the kernel-derived feedbacks in the respective regions. Figure 5 shows the individual feedback changes averaged over the regions IPWP, WP, EP, and Arctic as well as the globe without the Arctic, and the entire globe, for G1 and G2.

For both groups the most important feedback changes in the Arctic are the Planck, SA, and LR feedbacks. However, the groups differ in the magnitude of the individual changes and, more importantly, the sign of each individual feedback change is different between groups. As expected from the surface warming change (Fig. 4), the Planck feedback change is positive in G1 and negative in G2. In contrast, the LR and SA feedback changes are both negative in G1 while they are both positive in G2. The sum of the LR and SA feedback change is stronger than the Planck feedback change in both groups so that for G1 the Arctic feedback change is in total negative, while for G2 it is positive.

The LR and SA feedbacks in the Arctic are both strongly connected to the melting of sea ice and they are known to interact considerably with each other (e.g., Graversen et al. 2014). As Earth warms, the sea ice melts and exposes the darker ocean surface, reducing the albedo and hence causing a positive feedback. Furthermore, the sea ice acts as an insulating layer impairing heat exchange between the atmosphere and the much warmer ocean so that above the sea ice low atmospheric temperatures far below the melting point of sea ice may prevail especially during the dark seasons. The melting of sea ice therefore strongly enhances the heat exchange between atmosphere and ocean, leading to a strong increase in Arctic surface temperatures. Because the Arctic is stably stratified, this warming is trapped at the surface and a positive LR feedback is induced. Hence, the melting of sea ice controls the evolution of both the LR and SA feedbacks. Thus, a plausible explanation for the large difference in Arctic feedback changes in G1 and G2 is that as a result of the stronger early-period Arctic warming in G1 (Fig. 6d), the sea ice quickly melts (Fig. 6b), causing strong positive LR and SA feedbacks. In the later period most of the ice has already melted, resulting in a weakening of both the LR and SA feedback in the Arctic. Accordingly, these feedbacks are much less positive in the late period, inducing the strong negative change of these feedbacks in G1 seen in Fig. 5. Furthermore, the strong early warming leads to a considerable reduction in atmospheric stability, causing an increase of the local cooling efficiency from the early to the late period, which hampers further warming (Fig. 6f). In G2, on the other hand, the Arctic warming is weaker in the early period, inducing much less sea ice melt and weaker positive LR and SA feedbacks. Also, the Arctic remains more stable in G2 than in G1. As the Arctic in the late period continues to warm more strongly in G2 than in G1 (Figs. 6a,d), more sea ice melts (Fig. 6b), hereby strengthening the LR and SA feedbacks and causing the positive change as shown in Fig. 5. As the Arctic surface warming is weaker in G2 than in G1 (Fig. 6d), the stratification remains more stable and the local cooling efficiency weaker (Fig. 6f), enhancing the surface warming also in the late period.

In the EP the two model groups exhibit differences especially in WV and Planck feedback change but in both groups the total feedback change is strongly positive. Notably, the difference in total feedback change in the EP between G1 and G2 is the opposite of the global total feedback change. The feedback changes in the WP and the IPWP are similar in G2, while in G1 the individual feedback changes have opposite sign between the two regions. However, the total feedback change for both regions in G1 and G2 is small compared to the EP and the Arctic. The differences in total feedback between G1 and G2 in the regions IPWP, WP, and Arctic are similar to the difference in global-mean total feedback change between G1 and G2 and hence appear important for the difference between the two model groups. Conversely, the difference in the EP is opposite that of the global mean, indicating that this region is less important for the general between-group difference.

By construction, in the global mean the positive LR feedback change dominates for G2 with some smaller positive changes in cloud and SA feedback. The WV feedback change is negative and somewhat compensates the other feedbacks but there still remains a considerable positive total feedback change. In contrast, the total G1 feedback change is only slightly positive. The SA and LR feedback changes are small but negative and compensated for by the positive cloud and WV feedback changes. In spite of the large changes in the Arctic, excluding this region from the global mean has little effect on the feedback decomposition, owing to the fact that the Arctic (>75°N) covers less than 2% of Earth's surface area. Thus, in this simple global-mean perspective, the Arctic changes, though large, have only little influence. However, previous research has pointed to remote influences of Arctic changes, especially on the development of the Hadley circulation (Feldl and Bordoni 2016; Feldl et al. 2017). Generally, positive high-latitude feedbacks reduce the meridional temperature gradient, hereby weakening the meridional heat transport, which at low latitudes is partly accomplished by the thermally direct cell constituting the Hadley circulation. Stronger positive Arctic SA and LR feedbacks are thus associated with a stronger weakening of the Hadley circulation. We find evidence for this coupling since, on the one hand, in G1 the meridional overturning streamfunction in the Northern Hemisphere tropics weakens more strongly in the early period but then remains constant, and on the other hand, in G2 this streamfunction weakens less in the early period but continues to weaken further in the late period (Fig. S4).

Other research highlights the influence of low-latitude changes on the Arctic. For example, Yoo et al. (2012) show that convective heating in the IPWP enhances poleward

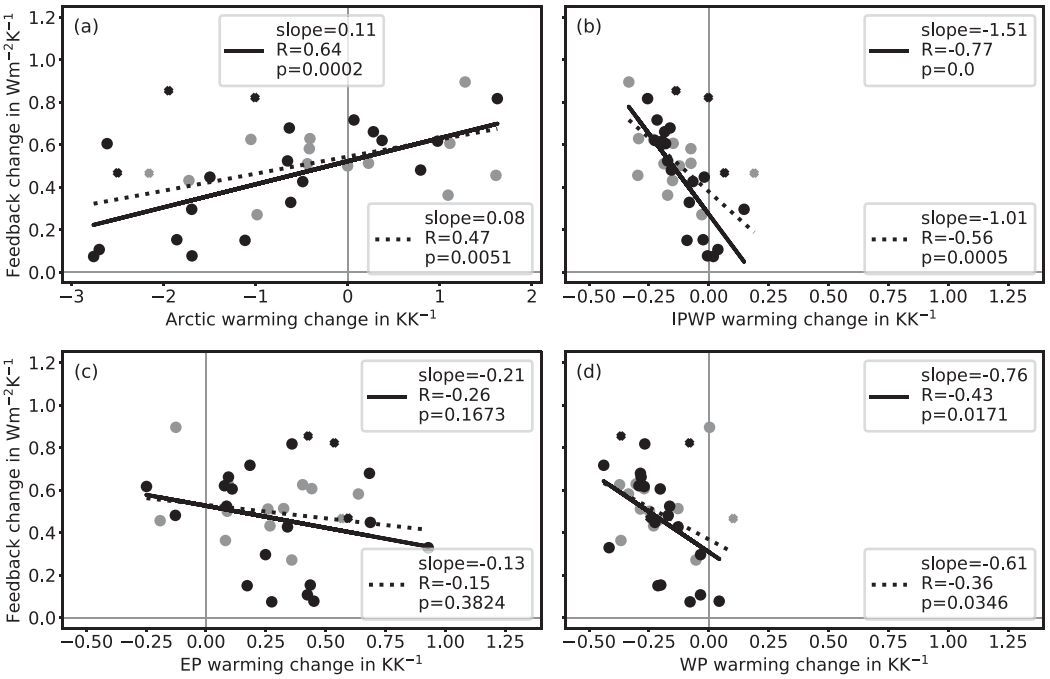

FIG. 7. As in Fig. 2, but for total feedback change derived with the Gregory method vs surface-warming change in (a) the Arctic, (b) the Indo-Pacific warm pool (IPWP), (c) the eastern Pacific (EP), and (d) the western Pacific (WP). Note that the *x*-axis scale in (a) differs from that in (b)–(d).

propagating Rossby waves leading to increased Arctic surface air temperatures. Lee (2014) presents a theory for this, according to which poleward heat transport intensifies and increases Arctic temperatures given that a greenhouse gas forcing leads to enhanced localized convection in the tropics.

In summary, previous literature highlights both the influence of the Arctic on lower latitudes and vice versa. The effects of the coupling of these regions on our results have not been disentangled in the present study but are left for future investigations.

We now move away from the two-model-group frame and consider correlations across all models.

c. *Regional warming and feedbacks across models*

Similar to the approach by Andrews and Webb (2018) using the TPI, we investigate correlations of the change over time of the surface warming of the four regions described above with the change over time of total feedback across models (Fig. 7). If the surface warming in a specific region has a major influence on the global cooling efficiency, as the stability hypothesis states, and as shown for the IPWP in Dong et al. (2019), this should be supported by a correlation of the change of the surface warming in that region relative to global-mean surface warming with the change of total feedback over time. As Fig. 7 (black solid lines) shows, the

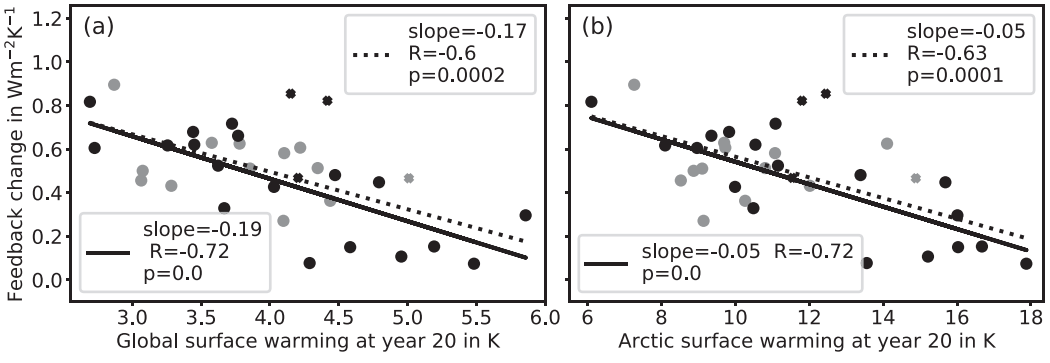

FIG. 8. As in Fig. 2, but for total feedback change derived with the Gregory method vs (a) global- and (b) Arctic-mean surface warming at year 20 (averaged over years 18–22) of the abrupt4xCO2 simulation. Note the different *x*-axis scales.

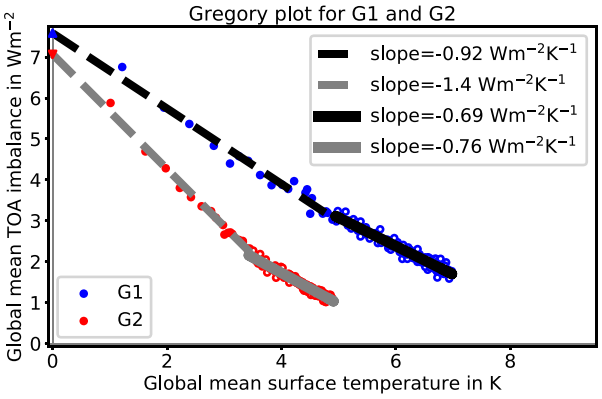

FIG. 9. Gregory plot for models with weak (G1; blue) and strong (G2; red) lapse-rate feedback change. The filled circles represent the early period (years 1–20) and the unfilled circles the late period (years 21–150). The dashed and solid lines show the early- and late-period linear regressions, respectively. The regression slope indicates the total feedback (see text for details).

strongest correlation is found for the IPWP and there is a slightly weaker correlation for the Arctic. There is a moderate correlation for the WP but close to no correlation for the EP, supporting the suggestion (see section 4b) that the latter region has little influence on the global-mean total feedback change. Note that the correlations weaken considerably if the models with a strong cloud feedback change are included (Fig. 7, black dotted lines). However, if the cloud feedback is excluded from the kernel-derived feedback sum, the correlations for the Arctic, WP, and IPWP, but not the EP, are strong regardless of these models being included or not (Fig. S5). This provides support for the stability hypothesis in general and the findings of Dong et al. (2019) and Dong et al. (2020) in particular (i.e., the IPWP surface-warming evolution largely controls total climate feedback change). Here CMIP5 and CMIP6 are not investigated separately as is done by Dong et al. (2020). They point out that while they can reconstruct the total feedback evolution over time from the IPWP surface warming evolution for CMIP5, this fails for CMIP6. With the caveat that we use a somewhat different set of models here, our results support the suggestion in Dong et al. (2020) that the failure to reconstruct total feedback evolution from IPWP warming in CMIP6 is because of a particularly large cloud feedback change in a number of members of CMIP6. If these members are excluded, the total feedback evolution may still be reasonably well reconstructed from the IPWP surface warming. Considering that due to a likely missing negative cloud lifetime feedback component (Mülmenstädt et al. 2021) some members of CMIP6 may overestimate the cloud feedback change, adjusting the cloud physics parameterization to include this feedback component might make a reconstruction of the total feedback evolution from IPWP surface warming as suggested by Dong et al. (2019, 2020) successful for CMIP6 as well.

As pointed out above, the surface warming change in the Arctic is significantly positively correlated with the change in total feedback. This means that the more a model increases its Arctic warming relative to the global mean over time, the more it changes its total feedback over time. Since the Arctic is a stably stratified region, this fits the stability hypothesis, since the surface warming over time shifts to more stable regions, decreasing Earth's cooling efficiency and hence weakening the total feedback. However, as pointed out in section 4b, it is difficult to establish causality between Arctic and global changes and the large differences in Arctic warming could be a by-product of the following: Models that warm strongly in the global mean in the early period (such as the members of G1) generally exhibit an even stronger early warming in the Arctic, while in the later period the global warming has a larger pace than the Arctic warming (Figs. 6c,d). Notably, these tend to be the models that have a weak total feedback change (Fig. 8). On the other hand, models with strong total feedback change tend to warm less quickly, and while they exhibit similar Arctic amplification their *absolute* Arctic warming is weaker (see Figs. 6a,c,d for the G1–G2 comparison). As explained in section 4b, the Arctic warming is strongly connected to sea ice changes and the resulting SA and LR feedbacks, and due to this double feedback loop, small differences in surface warming are quickly enhanced locally. Hence, the difference in early warming across models (the stronger the early warming, the weaker the total feedback change; Fig. 8) explains the relationship of local Arctic warming and feedback evolution with total feedback change across models. However, as discussed in section 4b, Arctic changes have been found to have remote influences (Feldl and Bordoni 2016; Feldl et al. 2017) and hence may have important indirect influences on feedback change. Further research is needed to establish the relevance of the large changes in the Arctic for the total global feedback change.

5. Discussion and conclusions

This study investigates why different numerical climate models change their total feedback differently over time using members of the CMIP5 and CMIP6 archives. We perform a radiative kernel analysis to decompose the total feedback into individual parts associated with different feedback processes and group the models according to similarities in individual feedbacks.

We investigate the differences between a group of models with weak (G1) and strong (G2) lapse-rate feedback change. It is revealed that the Arctic is the region with the largest difference between these groups and a region with large warming, stability, and individual feedback changes over time. These changes as well as their differences between the groups are strongly linked to Arctic sea ice changes. It is found that members of G1 warm much more quickly and exhibit faster Arctic sea ice melt, triggering stronger positive early-period Arctic surface-albedo and lapse-rate feedbacks than those of G2. Since the Arctic sea ice after a few decades has mostly vanished in the G1 models, the surface-albedo and lapse-rate feedbacks are much weaker in the late period for those models. Conversely, in the members of G2, more Arctic sea ice remains and the surface-albedo and lapse-rate feedbacks

become more positive in the late period. We furthermore find evidence for influence of Arctic changes on lower-latitude circulation, in accordance with previous studies (Feldl and Bordoni 2016; Feldl et al. 2017). However, it is difficult to determine if the Arctic changes and their low-latitude influence have causal relevance for the total feedback change.

As the members of G1 warm up much faster than the members of G2, they also experience much more rapid changes. Hence, the G1 members quickly reach a new warmer climate state with almost constant total feedback on the time scales considered here. In contrast, the members of G2, due to their slower warming, experience changes over climate states on longer time scales that become apparent in the present study. Evidence for this is provided by the fact that whereas the early-period feedback in G2 is much more negative than in G1 and the early-period warming of G2 is smaller than in G1, the late-period feedbacks of both groups are similar (Fig. 9).

Additionally, we generally find support for the results of previous studies (Dong et al. 2019, 2020) that the relative surface warming in the tropical Indo-Pacific is well correlated with total feedback change across models, which fits the previous finding that the surface warming in this region may control the total feedback change via the mechanism explained by the stability hypothesis.

Future research should focus on disentangling the cause-and-effect relationships. Questions of interest include the following: Why do some models warm much more quickly than others? Why do models warming slower have a large feedback change over time? How much does the sensitivity to warming of Arctic sea ice vary across models? How strong is the influence of Arctic changes on the Hadley circulation and how do changes in the Hadley circulation affect global feedbacks?

The change of climate feedback in numerical climate models as investigated here is of relevance, since it affects estimates of Earth's sensitivity to a forcing (e.g., due to anthropogenic greenhouse gas concentration changes in the atmosphere). Understanding the reasons behind climate feedback change will make it possible to compare the climate model results with real-world changes under ongoing climate change and hence improve the robustness of climate sensitivity estimates.

Acknowledgments. We thank Hege-Beate Fredriksen and Martin Rypdal for useful discussions and suggestions. We acknowledge the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling as well as the Earth System Grid Federation (ESGF) for making available and archiving the model output in the CMIP5 and CMIP6 archives. The data downloaded and generated during this research is stored at the Nird storage facilities provided by the Norwegian e-infrastructure for research and education, UNINETT Sigma2, under the project NS9063K. The work is part of the project "UiT–Climate Initiative, Ice–ocean–atmosphere interactions in the Arctic—From the past to the future," funded by the Faculty of Science and Technology, University of Tromsø. We thank three anonymous reviewers and the editor, Isaac Held, for their help in improving the work.

Data availability statement. The original model output is available on the WCRP's CMIP5 (https://esgf-node.llnl.gov/search/cmip5/) and CMIP6 archives (https://esgf-node.llnl.gov/search/cmip6/). The procedure for generating the climate feedbacks using the radiative kernels is described in section 3. Scripts and data generated during this research are stored at Nird storage facilities provided by UNINETT Sigma2, and are available on request from the corresponding author.

APPENDIX A

Derivation of Adjusted Cloud Radiative Effect

Here we present a brief derivation of Eq. (3). We start by splitting up the clear-sky and all-sky top-of-atmosphere (TOA) imbalances (N_{cs} and N_{as}, respectively) into contributions of individual climate state variables:

$$N_{cs} = N_{cs}(\delta a) + N_{cs}(\delta T) + N_{cs}(\delta w) + F_{cs}^{4x}, \tag{A1}$$

$$N_{as} = N_{as}(\delta a) + N_{as}(\delta T) + N_{as}(\delta w) + N(\delta c) + F_{as}^{4x}. \tag{A2}$$

Note that to fully represent the TOA imbalance, we need to include the forcing due to the quadrupling of the CO_2 concentration in both cases (i.e., F_{cs}^{4x} and F_{as}^{4x}). Since, as described in section 3, the cloud radiative effect (CRE) is defined as $N_{as} - N_{cs}$ we now subtract Eq. (A1) from Eq. (A2) to derive this quantity:

$$\begin{aligned}\mathrm{CRE} = N_{as} - N_{cs} &= N_{as}(\delta a) + N_{as}(\delta T) + N_{as}(\delta w) + N(\delta c)\\ &\quad + F_{as}^{4x} - N_{cs}(\delta a) - N_{cs}(\delta T) - N_{cs}(\delta w) - F_{cs}^{4x}.\end{aligned} \tag{A3}$$

Solving for $N(\delta c)$, substituting CRE, and rearranging yields

$$\begin{aligned}N(\delta c) = \mathrm{CRE} &+ N_{cs}(\delta a) - N_{as}(\delta a) + N_{cs}(\delta T) - N_{as}(\delta T)\\ &+ N_{cs}(\delta w) - N_{as}(\delta w) + F_{cs}^{4x} - F_{as}^{4x}.\end{aligned} \tag{A4}$$

Using Eq. (2) for the surface-albedo flux and its equivalents for the other climate state variables with both clear-sky and all-sky kernels, Eq. (3) is obtained.

APPENDIX B

Why Clear-Sky and Not All-Sky Linearity Test?

We here briefly show that in the kernel decomposition used in the present study the difference between total kernel-derived feedback and total feedback derived from model output radiative fluxes is the same for both clear-sky and all-sky conditions. We start from Eq. (A2), but now indicate the quantities involving radiative kernels by the superscript K:

$$N_{as}^{K} = N_{as}^{K}(\delta a) + N_{as}^{K}(\delta T) + N_{as}^{K}(\delta w) + N^{K}(\delta c) + F_{as}^{4x}. \tag{B1}$$

Note that $N^K(\delta c)$ is not solely based on radiative kernels since it involves CRE, which is calculated from model

output radiative fluxes. Substituting Eq. (2) for the surface-albedo flux and its equivalents for temperature and water vapor as well as Eq. (3) for the cloud flux, this expression reduces to

$$N_{\text{as}}^{K} = \text{CRE} + K_{\text{cs}}^{a}\delta a + K_{\text{cs}}^{w}\delta w + K_{\text{cs}}^{T}\delta T + F_{\text{cs}}^{4x}. \quad \text{(B2)}$$

On the right-hand side we can now substitute N_{cs}^{K} as well as the definition of CRE (i.e., $N_{\text{as}}^{M} - N_{\text{cs}}^{M}$) and obtain

$$N_{\text{as}}^{K} - N_{\text{as}}^{M} = N_{\text{cs}}^{K} - N_{\text{cs}}^{M}, \quad \text{(B3)}$$

where the superscript M denotes a model output radiative flux quantity. The CSLT is chosen here since this test is the direct comparison between the sum of the kernel-derived feedbacks and the total feedback derived from model output radiative fluxes with the Gregory method.

REFERENCES

Andrews, T., and M. J. Webb, 2018: The dependence of global cloud and lapse rate feedback 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, and M. J. Webb, 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.

Armour, K. C., C. M. Blitz, 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.

Arrhenius, S., 1896: On the influence of carbonic acid in the air upon the temperature of the ground. *London Edinburgh Dublin Philos. Mag. J. Sci.*, **41**, 237–276, https://doi.org/10.1080/14786449608620846.

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

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

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.

Charney, J. G., and Coauthors, 1979: Carbon dioxide and climate: A scientific assessment. Report of an Ad Hoc Study Group on Carbon Dioxide and Climate, National Academy of Sciences, 20 pp.

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.

——, K. C. Armour, M. D. Zelinka, C. Proistosescu, D. S. Battisti, C. Zhou, and T. Andrews, 2020: Intermodel spread in the pattern effect and its contribution of climate sensitivity in CMIP5 and CMIP6 models. *J. Climate*, **33**, 7755–7775, https://doi.org/10.1175/JCLI-D-19-1011.1.

Donohoe, A., E. Blanchard-Wrigglesworth, A. Schweiger, and P. J. Rasch, 2020: The effect of atmospheric transmissivity on model and observational estimates of the sea ice albedo feedback. *J. Climate*, **33**, 5743–5765, https://doi.org/10.1175/JCLI-D-19-0674.1.

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

Feldl, N., and S. Bordoni, 2016: Characterizing the Hadley circulation response through regional climate feedbacks. *J. Climate*, **29**, 613–622, https://doi.org/10.1175/JCLI-D-15-0424.1.

——, ——, and T. M. Merlis, 2017: Coupled high-latitude climate feedbacks and their impact on atmospheric heat transport. *J. Climate*, **30**, 189–201, https://doi.org/10.1175/JCLI-D-16-0324.1.

Graversen, R. G., P. L. Langen, and T. Mauritsen, 2014: Polar amplification in CCSM4: Contributions from the lapse rate and surface albedo feedbacks. *J. Climate*, **27**, 4433–4450, https://doi.org/10.1175/JCLI-D-13-00551.1.

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

Hahn, L. C., K. C. Armour, M. D. Zelinka, C. M. Bitz, and A. Donohoe, 2021: Contributions to polar amplification in CMIP5 and CMIP6 models. *Front. Earth Sci.*, **9**, 710036, https://doi.org/10.3389/feart.2021.710036.

Huang, Y., Y. Xia, and X. Tan, 2017: On the pattern of CO_2 radiative forcing and poleward energy transport. *J. Geophys. Res. Atmos.*, **122**, 10 578–10 593, https://doi.org/10.1002/2017JD027221.

Klein, S. A., and D. L. Hartmann, 1993: The seasonal cycle of low stratiform clouds. *J. Climate*, **6**, 1587–1606, https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2.

Lee, S., 2014: A theory for polar amplification from a general circulation perspective. *Asia-Pac. J. Atmos. Sci.*, **50**, 31–43, https://doi.org/10.1007/s13143-014-0024-7.

Martin, G. M., and Coauthors, 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.

Mülmenstädt, J., and Coauthors, 2021: An underestimated negative cloud feedback from cloud life-time changes. *Nat. Climate Change*, **11**, 508–513, https://doi.org/10.1038/s41558-021-01038-1.

Paynter, D., T. L. Frölicher, L. W. Horowitz, and L. G. Silvers, 2018: Equilibrium climate sensitivity obtained from multimillennial runs of two GFDL climate models. *J. Geophys. Res. Atmos.*, **123**, 1921–1941, https://doi.org/10.1002/2017JD027885.

Pendergrass, A. M., A. Conley, and F. M. Vitt, 2018: Surface and top-of-atmosphere radiative feedback kernels for CESM-CAM5. *Earth Syst. Sci. Data*, **10**, 317–324, https://doi.org/10.5194/essd-10-317-2018.

Rugenstein, M. A. A., 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. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophys. Res. Lett.*, **27**, 2685–2688, https://doi.org/10.1029/2000GL011373.

Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the radiative kernel technique to calculate climate feedbacks in

NCAR's Community Atmospheric Model. *J. Climate*, **21**, 2269–2282, https://doi.org/10.1175/2007JCLI2044.1.

Sherwood, S. C., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, **58**, e2019RG000, https://doi.org/10.1029/2019RG000678.

Smith, C. J., and Coauthors, 2018: Understanding rapid adjustments to diverse forcing agents. *Geophys. Res. Lett.*, **45**, 12 023–12 031, https://doi.org/10.1029/2018GL079826.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *J. Climate*, **21**, 3504–3520, https://doi.org/10.1175/2007JCLI2110.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. Tech. Rep. 33 pp., https://pcmdi.llnl.gov/mips/cmip5/docs/Taylor_CMIP5_design.pdf?id=98.

Williams, K. D., W. J. Ingram, and J. M. Gregory, 2008: Time variation of effective climate sensitivity in GCMs. *J. Climate*, **21**, 5076–5090, https://doi.org/10.1175/2008JCLI2371.1.

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.

Yoo, C., S. Lee, and S. B. Feldstein, 2012: Arctic response to an MJO-like tropical heating in an idealized GCM. *J. Atmos. Sci.*, **69**, 2379–2393, https://doi.org/10.1175/JAS-D-11-0261.1.

Zelinka, M. D., S. A. Klein, and D. L. Hartmann, 2012: Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative kernels. *J. Climate*, **25**, 3715–3735, https://doi.org/10.1175/JCLI-D-11-00248.1.

——, C. Zhou, and S. A. Klein, 2016: Insights from a refined decomposition of cloud feedbacks. *Geophys. Res. Lett.*, **43**, 9259–9269, https://doi.org/10.1002/2016GL069917.

——, T. A. Meyers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

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 on 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.