On the Correspondence between Atmosphere-Only and Coupled Simulations for Radiative Feedbacks and Forcing from CO2

Yi Qin[1], Mark D. Zelinka[2], and Stephen A. Klein[2]

[1]Lawrence Livermore National Laboratory
[2]Lawrence Livermore National Laboratory (DOE)

November 24, 2022

Abstract

Atmosphere-only experiments are widely used to investigate climate feedbacks simulated in more computationally expensive fully-coupled global climate model simulations. We confirm that this remains a valid approach by comparing the radiative feedbacks and forcing between coupled and atmosphere-only simulations for the latest models taking part in the 6th phase of the Coupled Model Intercomparison Project (CMIP6). For cloud feedbacks, we find a better than previously known correspondence between these experiments, which applies even to the response of individual cloud properties (amount, altitude and optical depth), is present at nearly every geographic location, and holds even when considering atmosphere-only simulations of only 1 year duration. In the tropics, the correspondence between the two experiments is better revealed when considering feedbacks stratified by vertical motion rather than by geography, owing to the non-uniform warming pattern in the coupled experiment. For the lapse rate and surface albedo feedbacks, the correspondence between the two experiments is weaker due to the lack of sea-ice changes in the atmosphere-only experiment. For the across-model relationship between 4xCO_2 radiative forcing and feedback, we find a different behavior across experiments in CMIP6 than in CMIP5, casting doubt on the physical significance of previous results that highlighted an anti-correlation between the two quantities. Overall, these results confirm the utility of atmosphere-only experiments particularly to study cloud feedbacks, which are the dominant source of inter-model spread in climate sensitivity.

On the Correspondence between Atmosphere-Only and Coupled Simulations for Radiative Feedbacks and Forcing from CO_2

Yi Qin[1], Mark D. Zelinka[1], Stephen A. Klein[1]

[1]PCMDI/Lawrence Livermore National Laboratory, Livermore, CA, USA.

Corresponding author: Yi Qin (qin4@llnl.gov)

Key Points:

- Cloud radiative feedbacks in atmosphere-only and coupled simulations are highly correlated across both CMIP5 and CMIP6 models
- This correlation extends to cloud property feedbacks and the regional distribution of cloud feedbacks
- Atmosphere-only experiments need only be run for 1 year to capture the inter-model spread of global-mean coupled cloud feedbacks.

Abstract

Atmosphere-only experiments are widely used to investigate climate feedbacks simulated in more computationally expensive fully-coupled global climate model simulations. We confirm that this remains a valid approach by comparing the radiative feedbacks and forcing between coupled and atmosphere-only simulations for the latest models taking part in the 6th phase of the Coupled Model Intercomparison Project (CMIP6). For cloud feedbacks, we find a better than previously known correspondence between these experiments, which applies even to the response of individual cloud properties (amount, altitude and optical depth), is present at nearly every geographic location, and holds even when considering atmosphere-only simulations of only 1 year duration. In the tropics, the correspondence between the two experiments is better revealed when considering feedbacks stratified by vertical motion rather than by geography, owing to the non-uniform warming pattern in the coupled experiment. For the lapse rate and surface albedo feedbacks, the correspondence between the two experiments is weaker due to the lack of sea-ice changes in the atmosphere-only experiment. For the across-model relationship between 4xCO_2 radiative forcing and feedback, we find a different behavior across experiments in CMIP6 than in CMIP5, casting doubt on the physical significance of previous results that highlighted an anti-correlation between the two quantities. Overall, these results confirm the utility of atmosphere-only experiments particularly to study cloud feedbacks, which are the dominant source of inter-model spread in climate sensitivity.

1 Introduction

Radiative feedback and forcing are generally calculated from fully-coupled simulations of global climate models (GCMs) forced by abruptly quadrupled CO_2 concentration run for 150 years or longer. However, conducting sensitivity experiments to better understand the physical mechanisms driving feedbacks is generally not feasible with fully coupled simulations, which are computationally expensive. Hence, simplified atmosphere-only experiments, including Atmospheric Model Intercomparison Project (AMIP) and aquaplanet experiments with globally uniform increases in sea surface temperature (SST), are more commonly used to understand inter-model differences in radiative feedbacks and forcing (Bony & Dufresne, 2005; del Genio et al., 2007; Ringer et al., 2006; Medeiros et al., 2015) or to investigate physical mechanisms involved in feedbacks in individual models (Bretherton et al., 2014; Brient & Bony, 2012, 2013; Ceppi et al., 2016; Demoto et al., 2013; Gettelman et al., 2012, 2013, 2019; Kamae et al., 2016; Webb et al., 2015; Xu & Cheng, 2016). An additional benefit of atmosphere-only simulations is that radiative feedbacks and forcing can be estimated in a straightforward manner via experiments forced by imposed changes in SSTs or CO_2 concentration as described in Cloud Feedback Model Intercomparison Project (CFMIP) protocols (Bony et al., 2011; Taylor et al., 2012; Webb et al 2017) rather than from the estimate by the Gregory method (Gregory et al., 2004) for fully-coupled experiments. This helps disentangle the separate contributions of radiative feedbacks and forcing to the diversity of equilibrium climate sensitivities (ECS) across models.

But this raises an important question: to what extent can AMIP simulations reproduce the climate feedbacks, especially the uncertain cloud feedback, in coupled simulations? Ringer et al. (2014) found global-mean feedbacks from AMIP experiments agree well with those from coupled experiments using a set of CMIP5 models. In this study, we assess whether this correspondence continues to hold in the latest generation of models that are part of CMIP6. Additionally, we will determine whether CMIP6 models exhibit the across-model anti-correlation between radiative feedback and forcing which was found to be stronger in simpler experiments in CMIP5 models (Andrews et al., 2012; Webb et al., 2012; Ringer et al., 2014; Caldwell et al., 2016; Chung & Soden, 2018).

The complexity of feedback processes, especially those related to clouds, hinders our understanding of mechanisms causing the large uncertainty of climate feedbacks. It is informative to decompose the total feedback into components (Bony & Dufresne, 2005; Shell et al., 2008; Soden et al., 2008; Soden & Held, 2006; Webb et al., 2006). So doing reveals that cloud feedbacks are particularly uncertain and drive inter-model spread in climate sensitivity. The cloud feedback itself comprises several cloud property feedbacks, which have been elucidated using the cloud radiative kernel method and shown to be widely-varying across models (Zelinka et al., 2012, 2016). Thus, combining these different diagnostic methods provides a more comprehensive evaluation of the consistency between atmosphere-only and coupled feedbacks not only for the global average but also for spatial patterns and individual cloud components.

Given the correspondence of radiative feedback and forcing between AMIP and coupled experiments, it is useful to know whether we can use AMIP experiments to estimate the ECS of the corresponding coupled model in advance of performing the coupled model simulation. This would be helpful in the case of a new atmosphere model, which might be a very expensive stormresolving model (e.g., DYAMOND models, Stevens et al., 2019), or one of a multitude of perturbed parameter versions of a given model, or a candidate version of the next version of the GCM for which a coupled model is not yet available. This further motivates the topic from two perspectives -- first, what combinations of AMIP experiments are in the best agreement with the ECS of the coupled model, and second, how long an AMIP simulation must be performed in order for its feedbacks to be representative of that from its corresponding coupled simulation. Previous studies generally use as few as 5-year AMIP experiments to investigate the radiative feedback in low-resolution (100~200 km), super-parameterized, and even global cloud-resolving climate models (Bretherton et al., 2014; Gettelman et al., 2012, 2019; Noda et al., 2019; Parishani et al., 2018; Zhang et al., 2018). However, CFMIP protocols (Bony et al., 2011; Taylor et al., 2012; Webb et al., 2017) require longer AMIP simulations (e.g., ~20 years in CFMIP2, ~36 years in CFMIP3). It would be useful to know the duration of atmosphere-only simulations necessary to get robust radiative feedbacks that are comparable to those from coupled experiments, especially given the rapid development of global cloud-resolving models (Stevens et al., 2019), whose huge computational expense may not permit AMIP-style simulations of more than a few months or years (Miura et al. 2005; Satoh et al., 2012; Tsushima et al., 2015).

The paper is organized as follows. Section 2 presents the used model data and methods. Detailed examination of the correspondence between AMIP and coupled radiative feedbacks and forcing from 4xCO_2 will be shown in Sections 3.1 and 3.2, respectively. In Section 3.3, the relationship between radiative feedback and forcing will be also examined in a hierarchy of models to check whether simpler experiments better capture this relationship as was found in CMIP5. Section 3.4 will discuss what combination of AMIP experiments gives the best estimate of ECS from coupled experiments and Section 3.5 will further discuss the minimum duration of AMIP simulation needed to represent the coupled feedback and the inter-model spread. Conclusions and discussion are in Section 4.

2 Materials and Methods

2.1 Data

We use output from:

1) fully coupled GCM experiments in which CO_2 concentrations are abruptly quadrupled from preindustrial concentrations and held fixed (abrupt4xCO_2) and their control experiments (piControl);

2) atmosphere-only experiments in which their CO_2 concentrations are abruptly quadrupled (sstClim4xCO_2) and their control experiments with preindustrial SST (sstClim);

3) atmosphere-only experiments in which SST is uniformly increased by 4K (amip4K) or a composite SST warming pattern derived from CMIP3 coupled simulations of idealized 1% per year increase in atmospheric CO_2, scaled to an ice-free ocean mean of 4K, is imposed (amipFuture) or CO_2 concentration is abruptly quadrupled (amip4xCO_2) and their control experiments with prescribed observed monthly sea surface temperature and sea ice concentrations starting from 1979 (amip);

4) aqua-planet experiments in which SST is increased by 4K (aqua4K) or CO_2 concentration is abruptly quadrupled (aqua4xCO_2) and their control experiments with a prescribed SST profile (aquaControl).

Please see Taylor et al. (2012) or Webb et al. (2017) for a more detailed definition for these experiments. All anomalies are computed relative to their corresponding period in their control experiments.

To simplify the experiment descriptions used hereafter, we define the following annotations: feedbacks calculated from abrupt4xCO_2 and piControl, amip4K and amip, amipFuture and amip, aqua4K and aquaControl are referred to as coupled, amip4K, amipFuture and aqua4K feedbacks, respectively. Similarly, forcing calculated from abrupt4xCO_2 and piControl, amip4xCO_2 and amip, sstClim4xCO_2 and sstClim, aqua4xCO_2 and aquaControl are referred to as coupled, amip4xCO_2, sstClim4xCO_2 and aqua4xCO_2 forcing, respectively.

2.2 Methods to calculate radiative feedbacks and forcing

For coupled simulations (abrupt4xCO_2 and piControl), regression of global- and annual-mean surface air temperature anomalies (ΔT_s) against global- and annual-mean TOA net downward radiation anomalies is used to derive the 4xCO_2 radiative forcing (Y-intercept) and feedback (slope) following the Gregory method (Gregory et al., 2004). This method is also applied to cloud radiative effect (CRE) anomalies to obtain CRE adjustments (Y-intercept) and feedbacks (slope).

For AMIP and aquaplanet simulations, the feedback is derived from global-mean and climatological net TOA downward radiative flux anomalies (amip4K minus amip; amipFuture minus amip; aqua4K minus aquaControl) divided by ΔT_s. 4xCO_2 radiative forcing is the global- and annual-mean net TOA downward radiative flux anomalies between amip4xCO_2/aqua4xCO_2 and amip/aquaControl. Correspondingly, the CRE adjustments are calculated from the related CRE anomalies between amip4xCO_2/aqua4xCO_2 and amip/aquaControl.

To decompose the total feedback into individual components, the radiative kernel method is used to quantify the sensitivity of TOA net radiative flux anomalies to surface temperature (Planck feedback; PL), atmospheric temperature (lapse rate feedback; LR), water vapor (water vapor feedback; WV) and surface albedo (albedo feedback; ALB) (Shell et al., 2008; Soden et al., 2008). First, the monthly temperature, water vapor, and albedo anomalies are multiplied by the corresponding radiative kernels and in the case of atmospheric temperature and water vapor integrated from the surface to a varying tropopause (Reichler et al., 2003). Finally, the annual-mean TOA radiative anomalies due to each field are regressed on ΔT_s to get individual feedback components for coupled experiments. For AMIP and aquaplanet experiments, the individual feedback components are calculated by dividing the annual-mean TOA net radiative anomalies by ΔT_s. We also implement an alternative decomposition method, which avoids the large compensation between LR feedback and WV feedback by using relative humidity as the state variable (Held & Shell, 2012).

The cloud feedback is computed by adjusting the change in cloud radiative effect (CRE; clear- minus all-sky upwelling radiation) for non-cloud influences (Shell et al., 2008; Soden et al., 2008). We use Huang et al. (2017) kernels as more models passed the clear-sky linearity test (Zelinka et al., 2020). To get more insights about different cloud types on the total cloud feedback, cloud radiative kernel analysis (Zelinka et al., 2012, 2016) is applied to those models with ISCCP simulator output to estimate the cloud feedback due to the changed cloud amount, altitude, and optical depth for low (cloud top pressure > 680 hPa) and non-low (cloud top pressure < 680 hPa) clouds.

In this study, all available models for radiative feedback/forcing calculations, radiative kernel analysis, and cloud radiative kernel analysis are respectively labeled ‘O’, ‘R’ and ‘C’ in Table 1 and 2. Those models with both available AMIP and coupled simulations for radiative kernel analysis are further labelled by numbers.

The correspondence between AMIP and coupled feedbacks/forcing is evaluated by two main metrics: Pearson correlation coefficient (R) with Student’s *t*-test and revised coefficient of determination (γ). For the correlation, the statistical significance uses a 95% significance level. The γ is defined as:

$$\gamma(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $\hat{y_i}$ is the AMIP feedback/forcing of the i-th model and y_i is the corresponding coupled feedback/forcing for total n samples. Overbars denote the average of all coupled feedback/forcing. So γ describes the percentage of the coupled radiative feedback/forcing variation (y_i) that is explained by AMIP radiative feedback/forcing ($\hat{y}_i$). For example, a γ of 0.8 means AMIP feedback/forcing can explain 80% of the variation of coupled feedback/forcing. It is a statistical measure of how closely the AMIP and coupled data fit to the 1-1 line. The higher the γ, the better fit to the 1-1 line. The maximum value of γ is 1.0 which occurs when all of the data lies on the 1-1 line.

Table 1. CMIP5 models. ‘O’ denotes models used in radiative feedback calculation. ‘R’ denotes models used in radiative kernel analysis, and ‘C’ denotes models used in cloud-radiative kernel analysis. Models with both abrupt4xCO_2 and amip4K experiments are labelled by numbers in the first column.

Label	MODEL	RIPF	abrupt4xCO2	amip4K	amipFuture	amip4xCO2	sstClim4xCO2	aqua4K	aqua4xCO2
	ACCESS1-0	r1i1p1	OR						
	ACCESS1-3	r1i1p1	OR						
	BNU-ESM	r1i1p1	OR				O		
0	CCSM4	r1i1p1	OCR	OCR	OCR		O	O	O
1	CNRM-CM5	r1i1p1	OR	OCR	OCR	O		O	O
	CNRM-CM5-2	r1i1p1	OR						
	CSIRO-Mk3-6-0	r1i1p1	OR				O		
2	CanESM2	r1i1p1	OCR	OCR	OCR	O	O		
3	FGOALS-g2	r1i1p1	OR	OR		O		O	O
	FGOALS-s2	r1i1p1	OR				O	O	O
	GFDL-CM3	r1i1p1	OR						
	GFDL-ESM2G	r1i1p1	OR						
	GFDL-ESM2M	r1i1p1	OR						
	GISS-E2-H	r1i1p1	OR						
	GISS-E2-R	r1i1p1	OR						
4	HadGEM2-ES	r1i1p1	OCR	OCR	OCR	O		O	O
	IPSL-CM5A-LR	r2i1p1		R		O		O	O
5	IPSL-CM5A-LR	r1i1p1	OR	OR	OR	O	O	O	O
	IPSL-CM5A-MR	r1i1p1	OR						
6	IPSL-CM5B-LR	r1i1p1	OR	OR	OR	O			
	MIROC-ESM	r1i1p1	OCR						
7	MIROC5	r1i1p1	OCR	OCR	OCR	O	O	O	O

8	MPI-ESM-LR	r1i1p1	OCR	OCR	OCR	O	O	O	O
9	MPI-ESM-MR	r1i1p1	OR	OR	OR	O	O	O	O
	MPI-ESM-P	r1i1p1	OR				O		
10	MRI-CGCM3	r1i1p1	OCR	OCR	OCR	O	O	O	O
	NorESM1-M	r1i1p1	OR			O	O		
11	bcc-csm1-1	r1i1p1	OR	OR	OR	O	O		
	bcc-csm1-1-m	r1i1p1	OR						
	inmcm4	r1i1p1	OR				O		

Table 2. As in Table 1, but for CMIP6 models.

Label	MODEL	RIPF	abrupt-4xCO2	amip-p4K	amip-future4K	amip-4xCO2	piClim-4xCO2	aqua-p4K	aqua-4xCO2
	ACCESS-CM2	r1i1p1f1	OR				O		
	ACCESS-ESM1-5	r1i1p1f1	OR						
	AWI-CM-1-1-MR	r1i1p1f1	OR						
12	BCC-CSM2-MR	r1i1p1f1	OR	OCR	OCR	O			
	BCC-ESM1	r1i1p1f1	OR						
	CAMS-CSM1-0	r1i1p1f1	OR						
	CAS-ESM2-0	r1i1p1f1	O						
13	CESM2	r1i1p1f1	OR	OCR	OR	O	O	O	O
	CESM2-FV2	r1i1p1f1	OR						
	CESM2-WACCM	r1i1p1f1	OR						
	CESM2-WACCM-FV2	r1i1p1f1	OR						
	CIESM	r1i1p1f1	OR						
	CMCC-CM2-SR5	r1i1p1f1	OR						
	CMCC-ESM2	r1i1p1f1	OR						

14	CNRM-CM6-1	r1i1p1f2	OR	OCR	OCR	O	O	O	O
	CNRM-CM6-1-HR	r1i1p1f2	O						
	CNRM-ESM2-1	r1i1p1f2	OR				O		
15	CanESM5	r1i1p2f1	OCR	OCR	OCR	O	O		
16	E3SM-1-0	r1i1p1f1	OCR	OCR	OCR	O			
	EC-Earth3	r1i1p1f1	O				O		
	EC-Earth3-AerChem	r1i1p1f1	OR						
	EC-Earth3-Veg	r1i1p1f1	OR						
	FGOALS-f3-L	r1i1p1f1	OR						
	FGOALS-g3	r1i1p1f1	OR						
17	GFDL-CM4	r1i1p1f1	OCR	OCR	O	O	O	O	O
	GFDL-ESM4	r1i1p1f1	OR						
18	GISS-E2-1-G	r1i1p1f1	OR	OR		O	O		
	GISS-E2-1-H	r1i1p1f1	OR						
	GISS-E2-2-G	r1i1p1f1	OR						
19	HadGEM3-GC31-LL	r1i1p1f3	OCR	OCR	OCR	O	O	O	O
	HadGEM3-GC31-MM	r1i1p1f3	O						
	IITM-ESM	r1i1p1f1	OR						
	INM-CM4-8	r1i1p1f1	OR						
	INM-CM5-0	r1i1p1f1	OR						
	IPSL-CM5A2-INCA	r1i1p1f1	OR						
20	IPSL-CM6A-LR	r1i1p1f1	OCR	OCR	OCR	O	O	O	O
	KACE-1-0-G	r1i1p1f1	O						
	KIOST-ESM	r1i1p1f1	OR						
	MIROC-ES2L	r1i1p1f2	OCR						

21	MIROC6	r1i1p1f1	OCR	OCR	OCR	O	O		
	MPI-ESM-1-2-HAM	r1i1p1f1	OR						
	MPI-ESM1-2-HR	r1i1p1f1	OR						
	MPI-ESM1-2-LR	r1i1p1f1	OR				O		
22	MRI-ESM2-0	r1i1p1f1	OCR	OCR	OCR	O	O		
	NESM3	r1i1p1f1	OR						
	NorESM2-LM	r1i1p1f1	OR				O		
	NorESM2-MM	r1i1p1f1	OR				O		
	SAM0-UNICON	r1i1p1f1	OR						
	TaiESM1	r1i1p1f1	OR						
	UKESM1-0-LL	r1i1p1f2	OCR						

3 Results

3.1 Relationships between radiative feedbacks in AMIP and coupled experiments

3.1.1 Global-mean radiative feedbacks

Figure 1 examines the relationship between amip4K and coupled radiative feedbacks for CMIP5 and CMIP6 models. Both clear-sky SW (SWCLR) and clear-sky LW (LWCLR) feedbacks lie to the left of the 1-1 line, indicating weaker positive SWCLR feedback and more negative LWCLR feedback in amip4K experiments compared with coupled experiments, as found in Ringer et al. (2014). The weaker positive SWCLR feedback from amip4K experiments is because their SST and sea ice are fixed and there is no strong sea ice reduction in response to the warming as that in coupled experiments (Figure 2g). The more negative LWCLR feedback in amip4K experiments is partly related to the greater atmospheric LW transmissivity in the absence of increased CO_2 concentrations (Good et al., 2012). This is confirmed by comparing the radiative kernel-derived, instead of model-calculated, clear-sky LW feedbacks between amip4K and coupled experiments (Figure S1). Because the radiative kernels are computed with respect to present-day rather than quadrupled CO_2 concentrations, radiative-kernel derived clear-sky LW feedbacks in abrupt4xCO_2 experiments are more negative than those derived from direct model output and hence in better agreement with those from amip4K. We find the model spread is reduced and models lie much closer to the 1-1 line with γ increasing from -1.45 to -0.43.

The large spread of SW, LW and net CRE feedbacks in coupled experiments is well captured by amip4K, with significant correlations of 0.84, 0.96 and 0.82, respectively (Figure 1d-f). However, their γ suggests there is a systematic bias for both LW and SW CRE feedbacks: most models exhibit slightly stronger SWCRE feedbacks and weaker LWCRE feedbacks in amip4K experiments (Figure 1d and 1e), which can also be seen in Figure 2 of Ringer et al. (2014). However, if we compare the adjusted SW and LW CRE feedbacks derived from radiative kernel methods between amip4K and coupled experiments, we find the systematic biases for LW and SW CRE feedbacks are largely alleviated. The γ is increased from 0.55 to 0.73 for SW CRE feedbacks (Figure 1g) and 0.37 to 0.80 for LW CRE feedbacks (Figure 1h). Although the unadjusted net CRE feedback bias is much weaker (Figure 1f) due to the 'bias' compensation between SW and LW CRE feedbacks, the γ is also improved from 0.67 to 0.74 for net CRE feedbacks (Figure 1i). These results suggest that the systematic biases in unadjusted CRE feedbacks between amip4K and coupled experiments are mostly an artifact of not correcting for cloud masking. For simplicity, the adjusted CRE feedbacks are called cloud feedbacks hereafter.

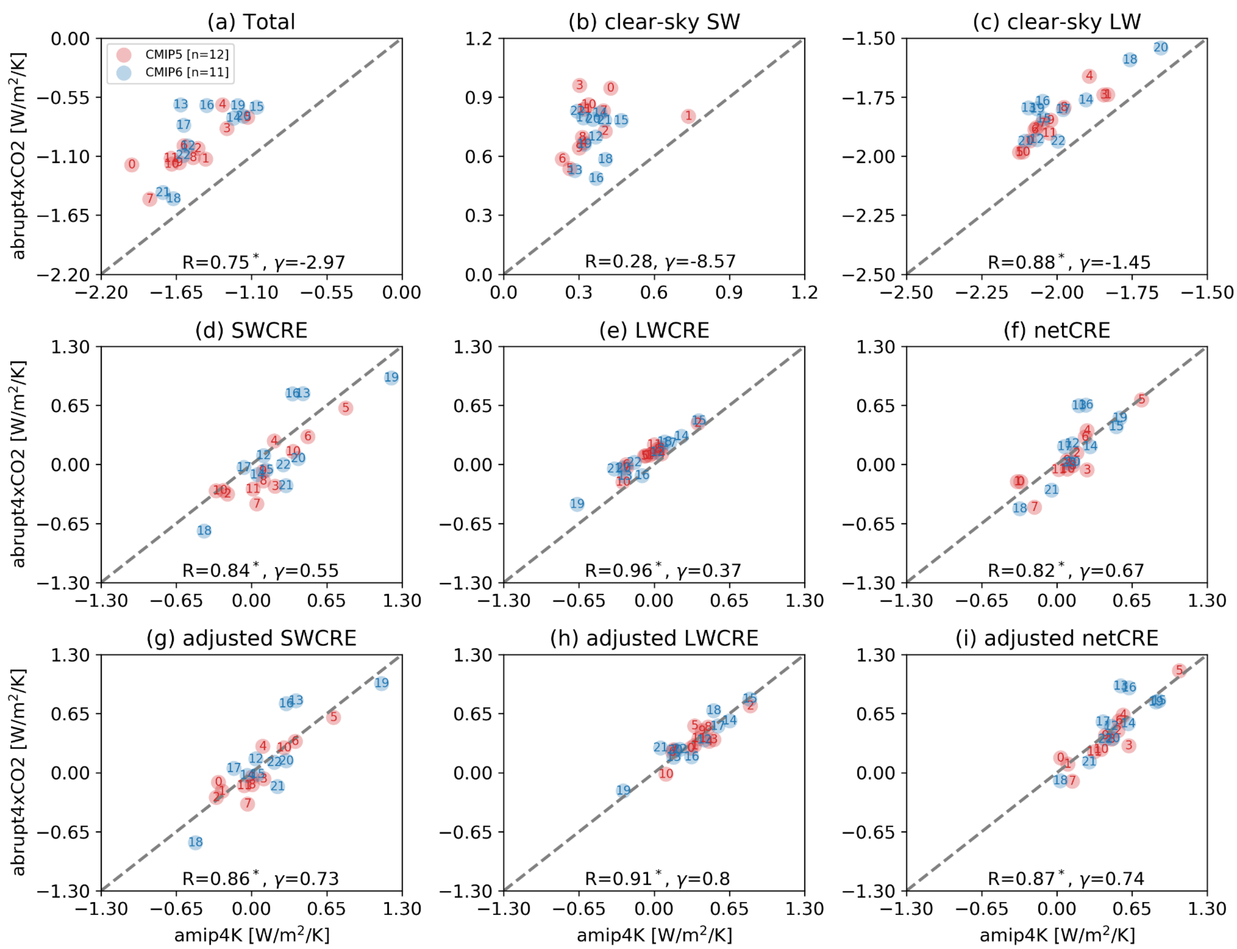

Figure 1. Global-mean radiative feedbacks (W/m^2/K) compared between amip4K and fully coupled abrupt4xCO_2 experiments. (a) total climate feedback; (b) clear-sky SW feedback; (c) clear-sky LW feedback; (d) unadjusted SWCRE, LWCRE and netCRE feedbacks; (g-i) adjusted SWCRE, LWCRE and netCRE feedbacks derived from the radiative kernel method. Red and blue dots denote CMIP5 and CMIP6 models respectively. Models used in later radiative kernel and cloud radiative kernel analysis are labelled by numbers as denoted in Table 1 and 2. R denotes the correlation coefficient with single asterisk indicating significance at the 95% level and γ denotes the fraction of the variation in the value of abrupt4xCO_2 feedback (Y) that is explained by the Y=X regression line where X is the amip4K feedback.

The close agreement between coupled and amip4K cloud feedbacks means that the stronger negative total climate feedback in amip4K than in coupled experiments (Figure 1a) comes solely from the combination of weaker positive SWCLR and stronger negative LWCLR feedbacks in amip4K (Figure 1b and 1c). The good agreement of cloud feedbacks also implies that the evolving surface temperature pattern (‘pattern effect’) in coupled experiments is not the first-order impact on the model diversity in those experiments, in agreement with Dong et al. (2020). Two models, CESM2 (#13) and E3SM-1-0 (#16), diverge from the other models in having much stronger positive SW and net cloud feedbacks in their coupled than amip4K experiments (Figure 1g and 1i). This behavior will be elucidated in more detail in Section 3.1.2.

The radiative kernel analysis allows us to further separate the total feedback parameter into terms corresponding to the effects of different climate components as we described in Section 2.2. We present the comparison of radiative kernel-derived non-cloud feedbacks between amip4K and coupled experiments in Figure 2. Compared with amip4K feedbacks, both the negative Planck feedback (Figure 2a) and the positive water vapor feedback are weaker (Figure 2c) in coupled experiments. Given that the Planck feedback goes as roughly $4\sigma T^3$, where σ is the Stefan-Boltzmann constant and T is the global mean temperature, the weaker negative Planck feedback in coupled than amip4K experiments arises in part because amip4K feedbacks are computed with respect to the warmer present-day state (amip) than the piControl climate that coupled feedbacks are computed with respect to. Indeed, the relative warming between present day and pre-industrial of about 1 K implies a more negative Planck feedback in the present-day of about 0.06 $W/m^2/K$ in present-day, which is close to the multi-model mean difference between amip4K and coupled experiments. The less negative coupled lapse rate feedback (Figure 2b and 2e) is mainly due to polar amplification of surface warming in the coupled experiments, which leads to a stronger positive lapse rate feedback in polar regions (where the warming is confined to the lower troposphere) that compensates the negative lapse rate feedback in the tropics (where warming is amplified with height). Feedbacks derived from the fixed relative humidity (RH) framework (Figure 2d-f), exhibit much smaller inter-model spread than do the traditional Planck, water vapor and lapse rate feedbacks (Figure 2a-c), consistent with previous studies (Held and Shell, 2008; Zelinka et al., 2020). Moreover, models lie closer to the 1-1 line for constant-RH Planck and relative humidity feedbacks (Figure 2d and 2f). The stronger positive surface albedo feedback (Figure 2g) is due to the sea-ice reduction in coupled experiments, which is not present in amip4K experiments. The inter-model spread of these non-cloud feedbacks in coupled experiments, though narrower than for cloud feedbacks, is not negligible and might be related to model differences in the pattern of surface warming (Po-Chedley et al., 2018).

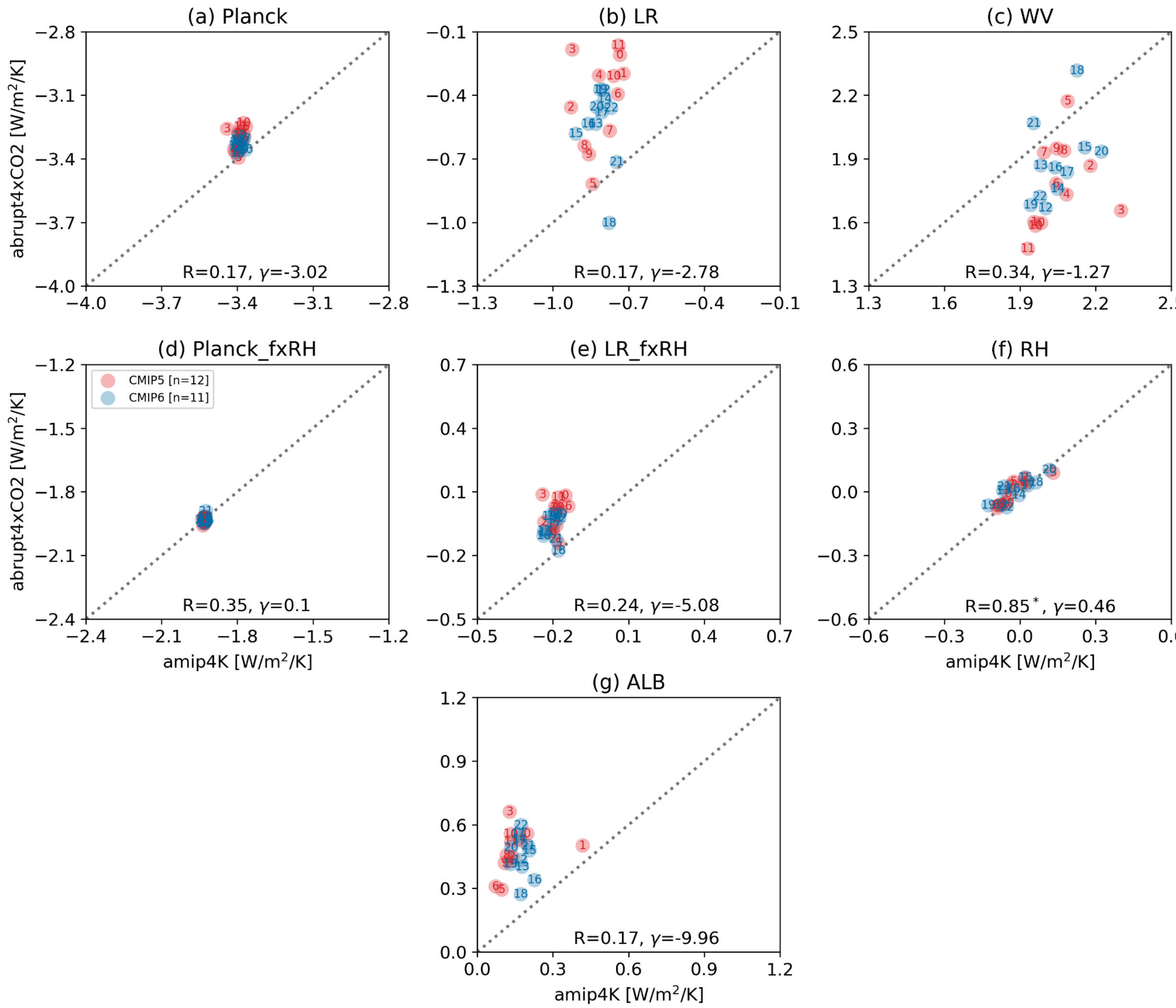

Figure 2. Global mean feedbacks (W/m^2/K) compared between amip4K and abrupt4xCO_2 experiments. (a) Planck and (b) lapse rate (LR) feedback computed holding absolute humidity fixed, (c) water vapor (WV) feedback, (d) Planck and (e) LR feedback computed holding relative humidity fixed (Held and Shell, 2012), (f) relative humidity (RH) feedback, and (g) surface albedo feedback. The sum of (a-c) is identical to the sum of (d-f). Red and blue dots denote CMIP5 and CMIP6 models, respectively.

From this, we conclude that the more negative total feedback in amip4K relative to coupled experiments comes from a stronger negative clear-sky LW feedback and weaker positive clear-sky SW feedback. The former is due to a stronger negative lapse rate feedback in amip4K experiments, where polar amplification and its attendant locally positive lapse rate feedback is strongly muted. The latter is due to the lack of sea ice reduction with warming in amip4K experiments. The strong correlation between amip4K and coupled cloud feedbacks after correcting for the cloud masking effect motivates an even more detailed examination of the correspondence of individual cloud types and in different regions in the section below.

3.1.2 Cloud decomposition

The cloud feedbacks are decomposed into the components due to changes in the individual cloud properties of amount, altitude and optical depth and for clouds at different vertical levels using the cloud radiative kernel method as described in Section 2.2 (Zelinka et al. 2012, 2016). The correspondence between amip4K and coupled cloud feedback components is shown in Figure 3. The significant correlation between amip4K and coupled feedbacks for all cloud feedback components indicates that the close correspondence between amip4K and coupled total cloud feedbacks identified above extends to the individual cloud responses composing the total cloud feedback, and amip4K simulations can largely capture the diversity of individual cloud feedback components in coupled experiments. For each component, most models also lie closely to the 1-1 line with relatively high γ. For example, the non-low cloud altitude and low cloud amount feedbacks (two large and important terms) agree fairly well between amip4K and coupled experiments, with γ around 0.8 for both LW, SW and net components (Figure 3g and 3j). With the near-zero altitude and optical depth feedbacks for low clouds, the good correspondence for total low cloud feedbacks is dominated by the low cloud amount feedback (Figure 3i). Slightly weaker consistency (γ is smaller) is shown for non-low cloud amount and optical depth feedbacks (Figure 3f and 3h). Coupled SW non-low cloud optical depth feedbacks tend to be more positive than those in amip4K, and vice versa for the LW (Figure 3h). Therefore, most intermodel spread of coupled cloud feedback components can be well captured by amip4K cloud feedback components. This decomposition is also helpful to identify the source of differences between coupled and amip4K feedbacks related to individual cloud properties for individual models. For example, the stronger positive SW cloud feedback in coupled experiments than that in AMIP experiments for E3SM-1-0 (model #16) can be further traced to the non-low cloud optical depth feedback (Figure 3h).

An assumption of the decomposition used in Figure 3 is that the cloud radiative kernel analysis using ISCCP simulator output (which has some methodological limitations) is able to reconstruct the total cloud feedback calculated from the radiative kernel method (which has fewer methodological limitations). Therefore, we also verified that total global-mean LW, SW and net cloud feedbacks estimated from the radiative kernel method agree well (with correlations 0.95, 0.96, and 0.93 for LW, SW, and net cloud feedbacks respectively) with those computed using the above cloud radiative kernel analysis for 13 models (6 CMIP5 models + 7 CMIP6 models) for which ISCCP simulator output is available. However, owing to the limited model samples for the cloud radiative kernel analysis, we use adjusted CRE feedbacks in later analyses to ensure a larger sample size for more robust comparison and evaluation.

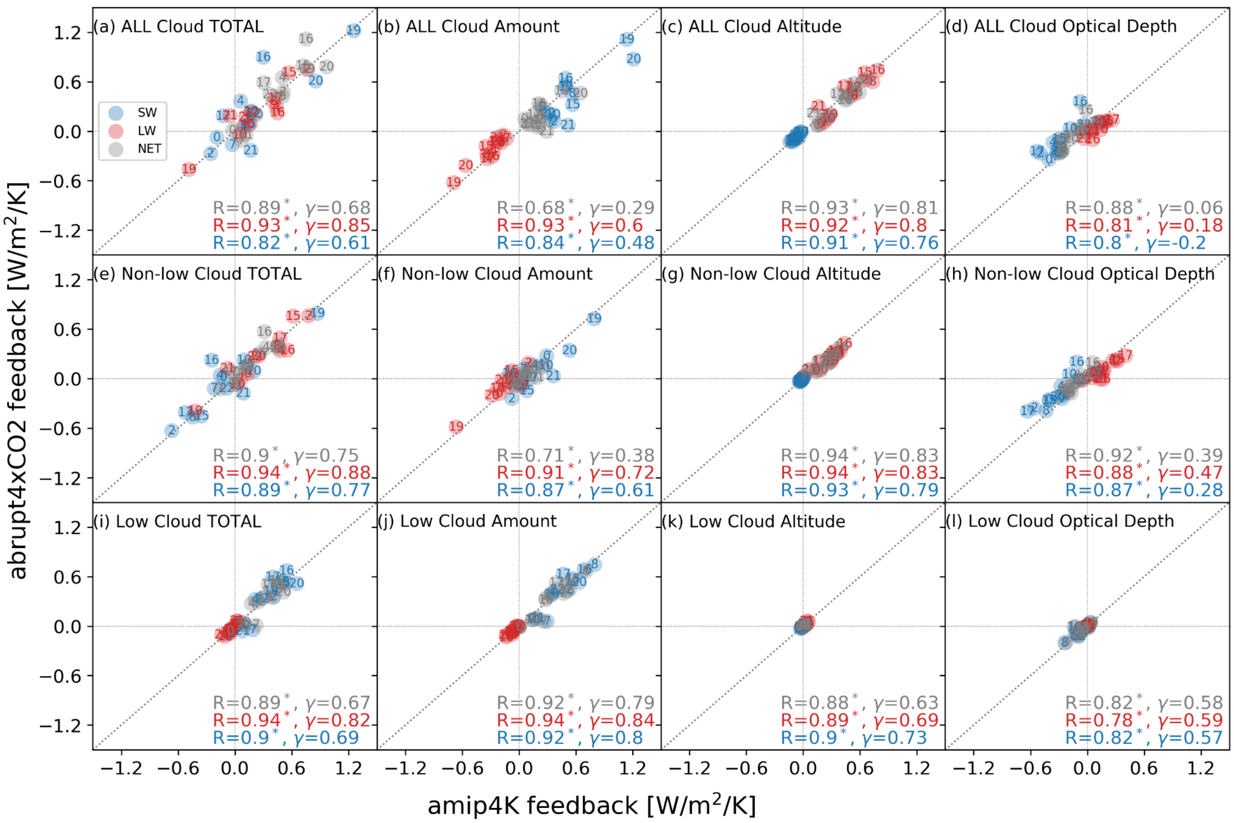

Figure 3. Global mean SW (blue), LW (red), and net (grey) cloud feedbacks (W/m^2/K) compared between amip4K and abrupt4xCO_2 experiments. (a, e, i) Total cloud feedbacks are decomposed into (b, f, j) amount, (c, g, k) altitude, (d, h, l) and optical depth components for (a-d) all clouds, (e-h) non-low clouds only (cloud top pressures less than 680 hPa), and (i-l) low clouds only (cloud top pressures greater than 680 hPa). Decomposition residuals are very small in all models and are not shown for clarity.

3.1.3 Spatial distribution

Given that global-mean cloud feedbacks agree well between amip4K and coupled experiments, the next question is whether the agreement is maintained for the spatial distribution. Maps of across-model correlation indicate that LW, SW, and net cloud feedbacks in amip4K experiments significantly correlate with those in coupled experiments (Figure 4). A notable exception is in the tropical Pacific, where the correspondence for LW and SW cloud feedbacks is much weaker. This could be understood as follows: High cloud changes -- which strongly affect both LW and SW radiation without strongly affecting net radiation -- are closely tied to large scale circulation changes. Therefore, in regions where the circulation regime changes substantially in coupled but not in amip4K experiments, the across-model correlation of LW and SW feedbacks will be degraded. Indeed, this interpretation is supported by maps of the response of 500 hPa vertical velocity (ω_{500}), shown in Figure 5. In coupled models, deep convection moves towards the central Pacific where SST anomalies are much greater than in amip4K (the "El-Nino like response"), hence there are much larger ascent anomalies in this region compared to that in amip4K experiments (Figure 5c). To demonstrate this more quantitatively, in Figure S2, we sort the control and warming CRE by the corresponding ω_{500} first, and then get the CRE anomalies in each dynamic regime for both amip and coupled experiments. Consistent with the interpretation above, the amip-coupled correlation turns out to be significant in each dynamic regime. This indicates the inconsistency in tropical SW and LW cloud feedbacks between amip4K and coupled experiments is mainly due to the ascent/descent regions moving around with different surface warming patterns in amip4K and coupled experiments.

The across-model correlation of net cloud feedbacks between amip4K and coupled experiments is significant near-globally including most tropical regions (Figure 4c). The good agreement of tropical net cloud feedbacks suggests most models have a compensation between LW and SW components, tied to large-scale circulation and cloud responses, which do not strongly affect the change of net CRE. However, even for the net cloud feedback, some regions exhibit less consistency, like India, western Pacific Ocean, North Atlantic Ocean, and high-latitude oceans. The different warming pattern in Indian and Pacific Ocean between amip4K and coupled experiments might lead to different cloud feedbacks over India because monsoon simulation is very sensitive to the air-sea coupling and land-sea temperature contrast (Wang et al., 2005; Endo et al. 2018; Singh et al., 2019; Geen et al., 2020). A "warming hole" is commonly simulated by coupled models in the North Atlantic, which could cause a locally different cloud feedback compared to that occurring when SSTs are warmed uniformly. The lack of correspondence of net cloud feedback over the high-latitude oceans near Antarctica and in the far north Atlantic and Arctic oceans is tied to cloud responses near the sea-ice edge, which retreats poleward with warming in coupled but not in amip4K experiments. Notwithstanding these differences, the above investigation shows that the amip4K experiments can be widely used to infer the model diversity of 150-year coupled cloud feedbacks, not only for the global average, but also for the spatial distribution.

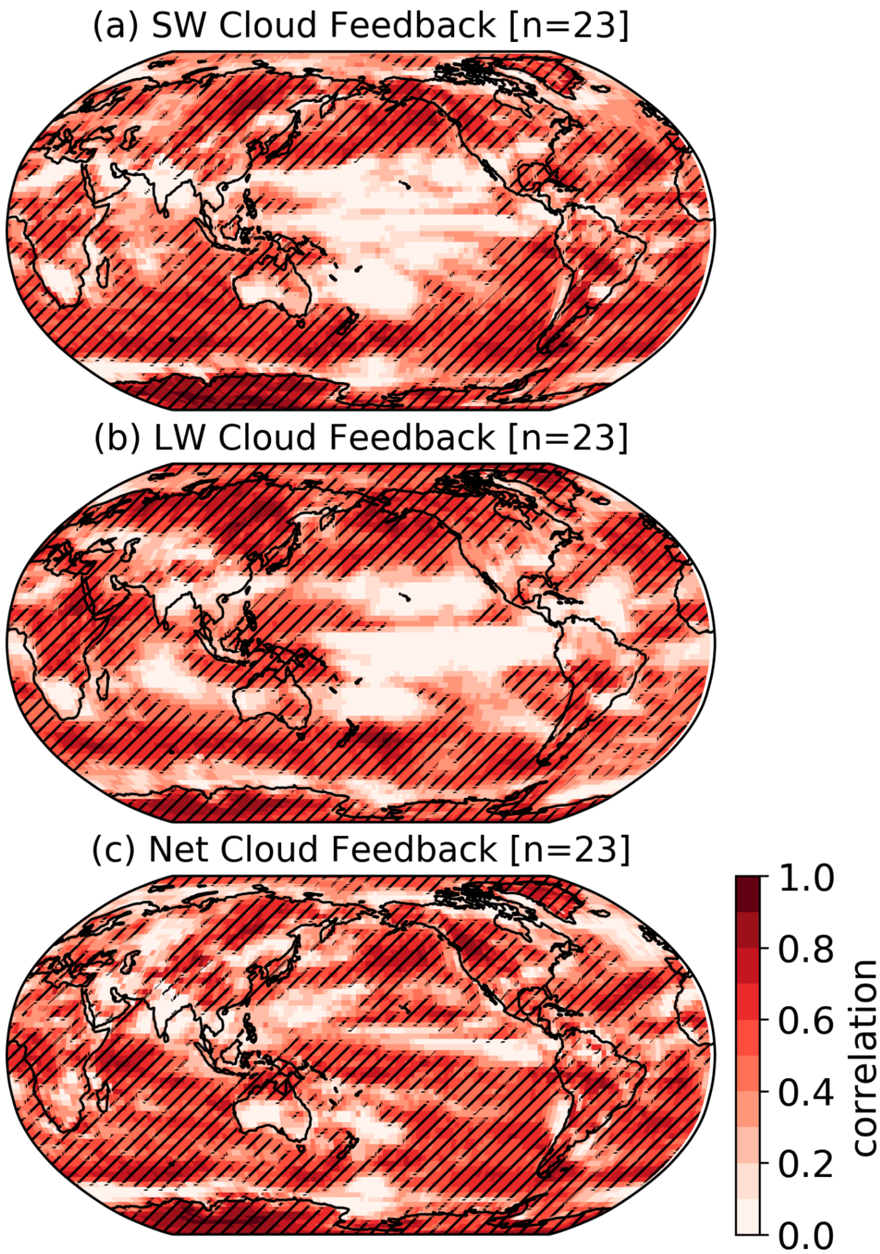

Figure 4. Across-model correlations of adjusted (a) SW, (b) LW and (c) net CRE feedbacks between amip4K and coupled experiments. Correlation coefficients significant at the 95% confidence level are indicated with hatching.

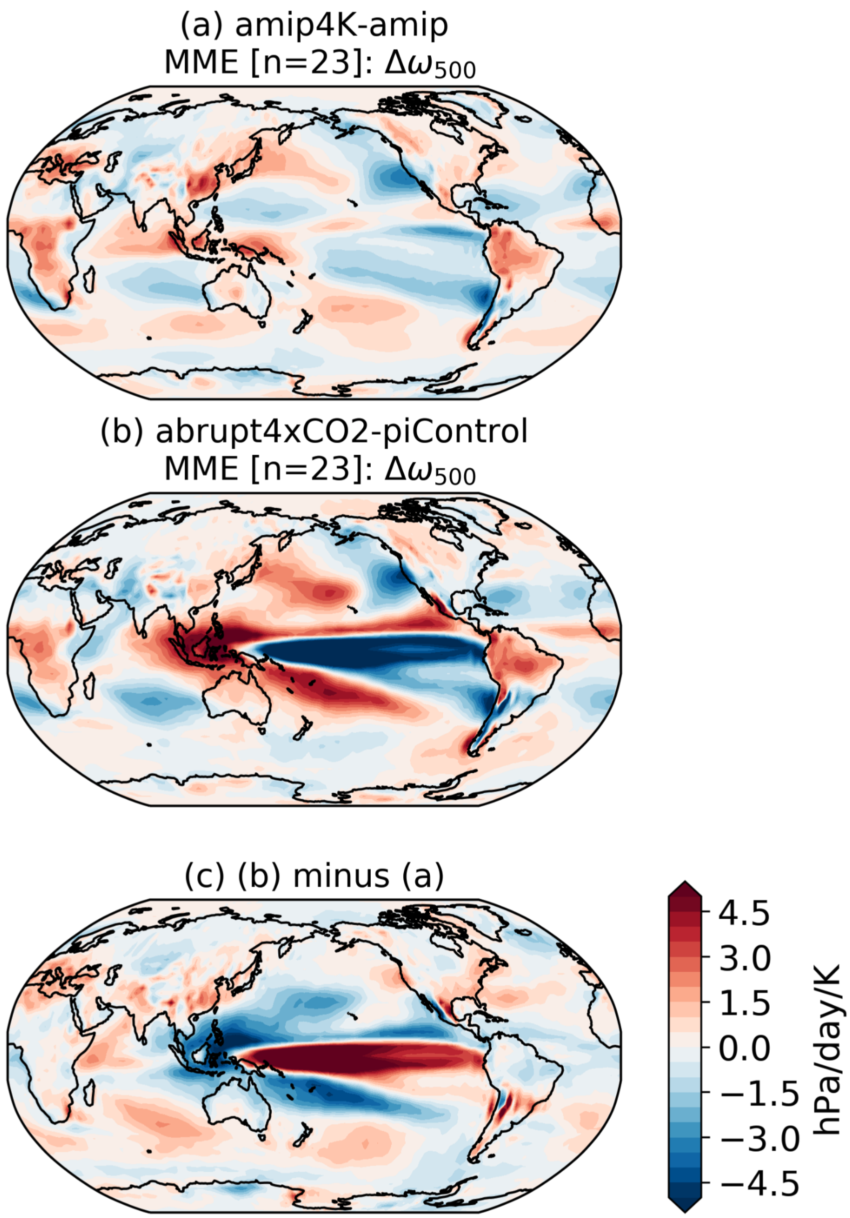

Figure 5. The multi-model ensemble-mean (MME) change in 500 hPa vertical pressure velocity (ω_{500}) per degree global warming (hPa/day/K); positive values are downward. The change is computed by differencing the (a) amip4K and amip simulations or (b) abrupt4xCO_2 and piControl simulations and normalizing this difference by the change in global mean temperature for each model, and then averaging the result across all models. (b) minus (a) is shown in (c).

3.2 Relationships between 4xCO_2 radiative forcing in AMIP and coupled experiments

In this section, we use the Gregory method (Gregory et al., 2004) and Hansen method (Hansen et al., 2005) to estimate the 4xCO_2 effective radiative forcing (ERF) for coupled and amip4xCO_2/sstClim4xCO_2 experiments, respectively. Previous studies have evaluated the strengths and weaknesses of various methods of calculating ERF (e.g., Forster et al., 2016; Smith et al., 2020; Chung and Soden, 2015; Andrew et al., 2012). The Gregory method derives the ERF (Y-intercept) by linearly regressing the TOA radiative anomalies against the global-mean surface temperature anomalies. It is generally applied to coupled experiments with its simple split of radiative forcing and feedback in one framework (Zelinka et al., 2020). However, a simple linear regression over the full 150-year experiment does not capture the time-evolving response, so the derived ERF is sensitive to the selected years (Andrew et al., 2012). The Hansen method estimates the ERF by differencing the radiation between a fixed SST simulation with the forcing agent imposed and one without the forcing agent imposed. Compared with the Gregory method, Hansen method is more computationally efficient and less sensitive to the selected simulation years (Forster et al., 2016). However, the land surface temperature in fixed-SST experiments is allowed to change and it could contribute to the change of global-mean surface temperature (Andrews et al., 2021). Whereas their definitions are different, comparing these two types of ERF across models can help understand the model diversity of ERF and the correspondence between AMIP and coupled ERF among CMIP models.

Because the estimate of 4xCO_2 ERF using Gregory method is sensitive to the starting and ending years considered when computing the regression, we calculate coupled ERFs using all possible windows of 5- to 50-year duration with starting years ranging from 1 to 10. We also consider the radiation anomaly from the first year of the coupled simulation as an additional ERF estimate. We then diagnose the correlation and γ between every coupled ERF value and those derived from amip4xCO_2/sstClim4xCO_2 experiments (Figure S3). We find the best correspondence with amip4xCO_2 when deriving ERF using the first 10 years of the coupled simulation (γ = 0.28; Figure 6a) and the best correspondence with sstClim4xCO_2 when deriving ERF using the first 36 years of the coupled simulation (γ = 0.18; Figure 6b). However, if we use those models with both sstClim4xCO_2 and amip4xCO_2 experiments available, the best correspondence (i.e., largest γ) occurs when deriving ERF using the first 14 or 15 years of the coupled simulation (Figure S4). This implies that the best segment of the coupled simulation to match the sstClim4xCO_2/amip4xCO_2 ERF is sensitive to the selected model samples. However, we find that the correlation of ERF between coupled and amip4xCO_2/sstClim4xCO_2 experiments is best (0.78; 0.75) when simply taking the first year of the coupled simulation and is less sensitive to the selected model samples (Figure S3a and c; Figure S4a and c). This suggests that the TOA radiation anomaly in the first year of the coupled simulation can largely capture the inter-model spread of ERF derived from amip4xCO_2/sstClim4xCO_2 simulations, although the former is generally smaller than the latter.

Because the sstClim4xCO_2 experiment has more consistent base state and radiatively active constituents (aerosols, ozone, etc) with abrupt4xCO_2 (Webb et al., 2017), the correlation with coupled ERF improves when using sstClim4xCO_2 rather than amip4xCO_2, regardless of what simulation period of coupled experiments is used to derive the coupled ERF (Figure 6 and Figure S3a). Figure 6c further shows ERFs from amip4xCO_2 and sstClim4xCO_2 are highly correlated (R: 0.85; γ: 0.70), indicating the quadrupled CO_2 is still the dominant factor in affecting the net TOA radiation anomalies although the difference of other forcing agencies and initial conditions can affect the ERF. GISS-E2-1-G (#18) diverges from other models in having a stronger forcing from sstClim4xCO_2 than that from amip4xCO_2 experiments, which needs further investigation. Chung and Soden (2015) found small differences of ERF exist among sstClim4xCO_2, amip4xCO_2 and aqua4xCO_2 experiments in CMIP5 models owing to differences in base states, consistent with our results. In the multi-model space, it is plausible to use the global-mean amip4xCO_2 ERF to represent its sstClim4xCO_2 ERF.

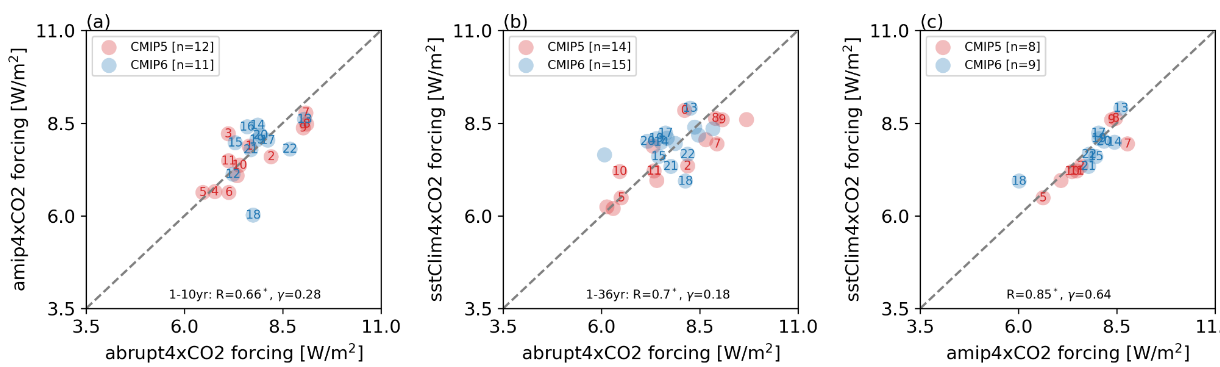

Figure 6. Global-mean effective radiative forcing (W/m^2) compared between (a) amip4xCO_2 and abrupt4xCO_2 experiments, (b) sstClim4xCO_2 and abrupt4xCO_2 experiments, and (c) amip4xCO_2 and sstClim4xCO_2 experiments. The first 10 years of abrupt4xCO_2 data is used in (a) and the first 36 years of abrupt4xCO_2 data is used in (b) to derive the coupled ERF (see section 3.2 for explanation of these choices). Red and blue dots denote CMIP5 and CMIP6 models respectively.

3.3 Relationships between radiative forcing and feedback

Previous studies identified that 4xCO_2 radiative forcing and feedbacks are anti-correlated across models, which damps inter-model spread in ECS (Andrews et al., 2012; Ringer et al., 2014; Caldwell et al., 2016). Understanding whether there is any physical basis for such a relationship between radiative forcing and feedback is an important topic (Sherwood et al., 2020). From CMIP5 models, Ringer et al. (2014) found that the increased complexity of model configuration blurs the relationship between radiative forcing and feedback in coupled simulations, and that AMIP and aquaplanet simulations are simpler configurations for studying this relationship. In this section we re-examine this relationship using CMIP6 models.

Table 3 summarizes the across-model correlation between radiative forcing and feedback in different model configurations. From fully coupled to AMIP and aquaplanet experiments, the correlation between radiative forcing and total feedback is indeed increased in CMIP5 models as Ringer et al. (2014) found, but this feature does not exist in CMIP6 models. Whereas the correlation strength increases with decreasing model complexity in CMIP5 models from -0.46 in abrupt4xCO_2 to -0.46 in amip4K to -0.87 in aqua4K, it varies non-monotonically from -0.52 in abrupt4xCO_2 to +0.40 in amip4K to -0.27 in aqua4K. This relation is quite consistent for net CRE for CMIP5 and CMIP6 models, for which only 5 models are currently available in aquaplanet experiments of CMIP6. When considering all CMIP5 and CMIP6 models together, the strengthening of the anti-correlation as experiments become simpler (Ringer et al., 2014) is no longer present due to the non-monotonic relation with model complexity.

Different model samples are used to calculate the forcing-feedback relationship in different experiments (labeled model numbers in Table 3), and coupled experiments generally have larger model samples than AMIP and aquaplanet experiments. To eliminate the potential systematic bias on correlation due to using different model samples across different experiments, the across-model correlation is recalculated using those models with both AMIP and coupled experiments (Table 4), reducing the model sample size to 11 CMIP5 and 11 CMIP6 models. The anti-correlation between forcing and feedback no longer increases from coupled (-0.66) to AMIP (-0.46) experiments in CMIP5. This suggests that the stronger anti-correlation with decreased model complexity is not robust and is sensitive to the selected model samples.

Due to the limited model samples for aqua4K/aqua4xCO_2 from CMIP6, results from CMIP6 models should be viewed with caution. Nevertheless, based on the combination of all CMIP5 and CMIP6 models, our results are inconsistent with Ringer et al. (2014). Furthermore, the lack of anti-correlation between forcing and feedback in the AMIP experiments when using all models suggests that there is no physical basis relating forcing to feedback.

Table 3. Cross-model correlation between 4xCO_2 radiative forcing/cloud adjustments and total radiative feedback/unadjusted CRE feedbacks. Labels '1-150' indicate that first 150 years of coupled experiments are used to calculate the radiative feedback/forcing. The number of model samples used for each correlation entry is listed in parentheses. Single asterisk indicates correlations significant at 95% level.

	TOTAL			**netCRE**			**SWCRE**			**LWCRE**		
Experiments used to derive feedback/ forcing	ALL	CMIP5	CMIP6	ALL	CMIP5	CMIP6	ALL	CMIP5	CMIP6	ALL	CMIP5	CMIP6
abrupt4xCO2 (1-150) /abrupt4xCO2 (1-150)	-0.48* (79)	-0.46* (29)	-0.53* (50)	-0.40* (79)	-0.46* (29)	-0.42* (50)	-0.38* (79)	-0.54* (29)	-0.38* (50)	0.21 (79)	0.42* (29)	0.19 (50)
amip4K/ amip4xCO2	0.01 (22)	-0.46 (11)	0.40 (11)	-0.01 (22)	-0.59 (11)	0.66* (11)	0.06 (22)	-0.56 (11)	0.54 (11)	-0.10 (22)	0.07 (11)	-0.19 (11)
aqua4K/ aqua4xCO2	-0.54* (16)	-0.87* (11)	-0.27 (5)	-0.61* (16)	-0.96* (11)	-0.03 (5)	-0.65* (16)	-0.97* (11)	-0.25 (5)	0.60* (16)	0.91* (11)	0.45 (5)

Table 4. As in Table 3, but for those models with both coupled and amip experiments.

	TOTAL			**netCRE**			**SWCRE**			**LWCRE**		
Experiments used to derive feedback/ forcing	ALL	CMIP5	CMIP6	ALL	CMIP5	CMIP6	ALL	CMIP5	CMIP6	ALL	CMIP5	CMIP6
abrupt4xCO2 (1-150)/ abrupt4xCO2 (1-150)	-0.51* (22)	-0.66* (11)	-0.54 (11)	-0.49* (22)	-0.53 (11)	-0.61* (11)	-0.48* (22)	-0.61* (11)	-0.57 (11)	0.20 (22)	0.23 (11)	0.26 (11)
amip4K/ amip4xCO2	0.01 (22)	-0.46 (11)	0.40 (11)	-0.01 (22)	-0.59 (11)	0.66* (11)	0.06 (22)	-0.56 (11)	0.54 (11)	-0.10 (22)	0.07 (11)	-0.19 (11)

3.4 What combination of AMIP experiments gives ECS estimates in best agreement with coupled experiments?

Many studies use atmosphere-only models to infer the feedbacks and ECS for fully coupled GCMs owing to the much lower computational expense for atmosphere-only experiments. However, different AMIP experiments with different configurations are available to estimate feedbacks and forcing as shown in previous sections. Thus, it is useful to know what combination of radiative forcing and feedbacks from AMIP experiments is most predictive of the coupled models' feedbacks and ECS.

We consider three options for radiative forcing: ERF = 4 W/m^2 (Sherwood et al. 2020), ERF derived from sstClim4xCO_2, and ERF derived from amip4xCO_2, and two options for total radiative feedback: amip4K feedback and amipFuture feedback. To compensate for the lack of polar warming and sea ice reduction on the total radiative feedback in AMIP experiments (Figure 1a), an estimate of 0.50 W/m^2/K from Figure 1a is added to all total feedbacks from AMIP experiments. The ECS values from fully-coupled experiments are derived using the ordinary Gregory method (i.e., the x-intercept of the regression of radiative imbalance on surface temperature) and are obtained from the analysis of Zelinka et al. (2020, https://github.com/mzelinka/cmip56_forcing_feedback_ecs).

Figure 7 shows the across-model correlation, root mean squared error (RMSE) and γ between the ECS predicted from AMIP experiments and the actual coupled models' ECS. To avoid model sampling problems, we make the comparison only for the same 15 models which have performed all the necessary experiments. Predicting ECS with the combination of sstClim4xCO_2 forcing and amip4K feedback gives the best agreement with the coupled ECS, with a correlation of 0.88, RMSE of 0.69, and γ of 0.55. The combination of 4 W/m^2 forcing and amipFuture feedback gives the worst correspondence. Two models, CESM2 and IPSL-CM5A-LR, show a weaker agreement between amip-predicted and coupled ECS due to the larger difference of cloud feedback between amip4K and coupled experiments and a relatively much weaker sea ice feedback in coupled experiments, respectively. As discussed in Section 3.2, compared to amip4xCO_2, the sstClim4xCO_2 radiative forcing is generally closer to the coupled forcing because its base state and emissions are similar to the coupled experiments. Hence, it is reasonable that the derived ECS using sstClim4xCO_2 forcing agrees better with the coupled ECS than using the amip4xCO_2 forcing. An issue with the sstClim4xCO_2 experiment is that one would need to know the climatology of the corresponding coupled model in order to perform the simulation. In the case of atmospheric models without a corresponding coupled model this would be unavailable. For those models, one could only perform the amip4xCO_2 experiment to derive the forcing since it does not need a corresponding coupled model. Fortunately, Figure 7 indicates that using the forcing from the amip4xCO_2 experiment together with the feedbacks from the amip4K experiment is almost as predictive as when using the forcing from the sstClim4xCO_2 experiment with the feedbacks from the amip4K experiment.

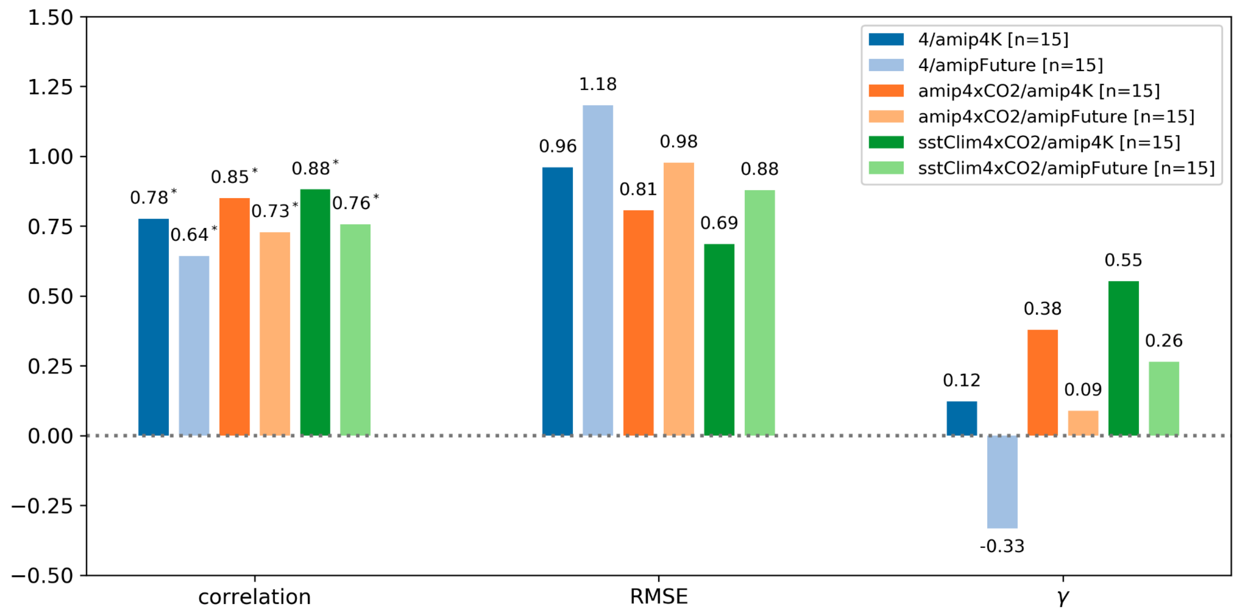

Figure 7. Pearson correlation coefficient, root mean square error (RMSE) and γ between ECS derived from atmosphere-only experiments and ECS derived from fully coupled experiments. All combinations between atmosphere-only ERF and feedback are shown in the legend (ERF/feedback). Three options for ERF (W/m^2) are: 4, amip4xCO_2 and sstClim4xCO_2, and two options for total feedback (W/m^2/K) are: amip4K and amipFuture. Single asterisks indicate correlations significant at the 95% level. Used model samples are shown in brackets.

It is interesting that the amipFuture feedback does not exhibit a better agreement than the amip4K feedback. This is likely because the imposed SST warming pattern in amipFuture experiments comes from the ensemble mean sea surface temperature anomaly pattern in coupled

CMIP3 experiments with 1% per year increase in atmospheric CO_2, which differs from the warming trend in the latest CMIP models. In particular, CMIP6 models show a stronger warming in the Southern Hemisphere (Dong et al., 2020) than is present in the pattern imposed for amipFuture experiments. This difference leads to models that are sensitive to the warming pattern (e.g., CESM2 #16) getting a closer correspondence to the coupled feedback with a uniform warming pattern instead of the amipFuture warming pattern.

3.5 What is the minimum duration required for AMIP simulations to capture the coupled cloud feedbacks?

Given the good agreement described in the previous section between ECS derived from sstClim4xCO_2/amip4K (or amip4xCO_2/amip4K) experiments and coupled experiments, it is useful to know how long one needs to run AMIP experiments to capture forcing and feedbacks in coupled experiments. Since the radiative forcing from AMIP experiments is more direct and stable than that from the coupled experiments, the minimum duration required for AMIP simulations to capture the coupled forcing will not be discussed. For the radiative forcing from AMIP experiments, Forster et al. (2016) found that 30-year duration is sufficient to keep the global mean ERF to 0.1 W/m^2 in the 5%-95% confidence interval. Therefore, the following discussion will focus on the feedback.

We calculate the amip4K feedbacks using different simulation lengths of amip4K experiments at both yearly and monthly timescales. For example, for a full 27-year amip4K experiment, we consider 27 samples of yearly feedbacks for each model, derived using data from every 1-year period in turn. In this way, we can also calculate other N-year feedbacks. Every N-year feedback will have a total 27-N+1 overlapping samples. For example, 26-year feedback includes two samples, which calculates feedback using 1-26 years and 2-27 years. respectively. Similarly, we consider 27*12 samples of monthly feedbacks for each model, derived using data from every 1-month period in turn. This method is helpful to increase the sample size for further statistical evaluation and quantify the uncertainty brought in due to the varying selected duration. We use the same 23 models in calculating diagnostic variables below as those used in radiative kernel analysis (Section 3.1.1). For each N-year/N-month feedback and each diagnostic variable, all available samples are used to calculate the corresponding standard deviation.

To determine the minimum duration necessary for amip4K feedbacks to capture the inter-model spread of the coupled model feedbacks, we first examine the ratio of amip4K across-model standard deviation relative to the coupled (std_ratio), as well as the correlation and γ between coupled and amip4K feedbacks as a function of amip4K simulation duration (Figure S5). These diagnostic variables are nearly invariant with increased AMIP duration for the total feedback and each component (SWCLR, LWCLR, SWCRE and LWCRE as in Figure 1). This suggests that the feedback difference between AMIP and coupled experiments is hardly reduced with increased simulation length. Considering (1) the better correspondence of cloud feedbacks between AMIP and coupled experiments (Figure 1g-i) than that of other feedback components (Figure 1a-f) and (2) the larger uncertainty of cloud feedbacks, it is more useful to get the minimum duration of amip4K experiments to capture the coupled cloud feedbacks. It is also important to know whether amip4K vs coupled cloud feedback differences tend to decrease with increased amip4K experiment length or asymptote quickly to some systematic bias, like the bias exhibited by models #13 (CESM2) and #16 (E3SM-1-0) in Figure 1g and i.

Figure 8 shows the evolution of the global mean cloud feedback difference between amip4K and coupled experiments ($\Delta\lambda_c$) as a function of simulation duration from amip4K experiments for all available models. First, the multi-model mean $\Delta\lambda_c$ is quite close to zero for both LW, SW and net cloud feedbacks (Figure 8) and the spread of LW feedback is weaker than that of SW and net cloud feedbacks in both monthly and yearly timescales (Figure 8b). The inter-model spread reduces with increased simulation months and becomes quite stable with further increased simulation years. Furthermore, $\Delta\lambda_c$ for each model is also stable with increased years with reduced uncertainty (Figure S6-S8). Different models tend to get different systematic biases for $\Delta\lambda_c$but increasing the amip4K simulation length does not reduce the magnitude of $\Delta\lambda_c$. Two models, CESM2 (model #13) and E3SM-1-0 (model #16), have larger biases for SW and net cloud feedbacks than other models as shown in Figure 1. Nonetheless, the systematic differences between amip4K and coupled feedbacks for these two models can also be estimated from 1-year feedback without running longer amip4K experiments (Figure S6-S8).

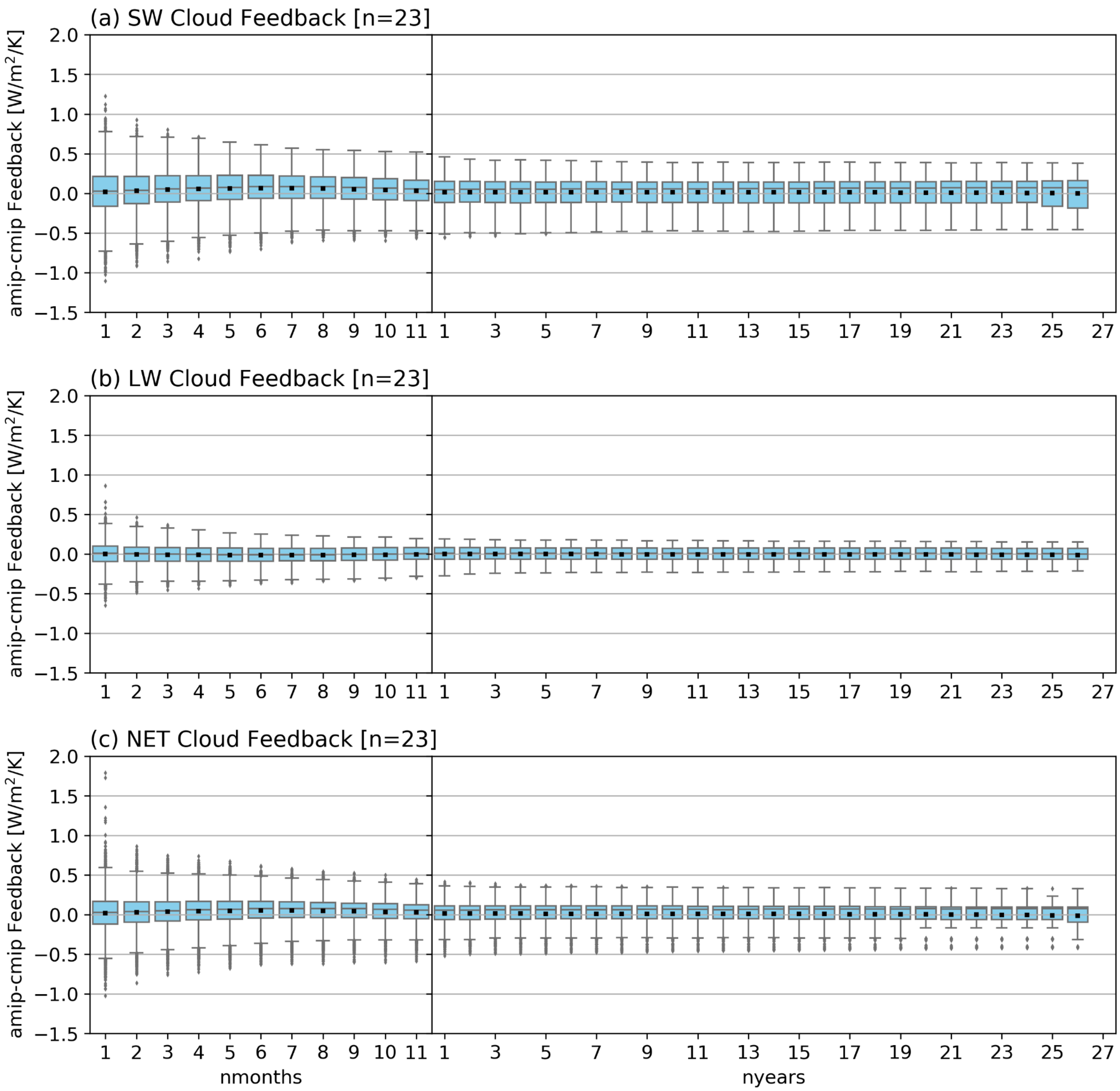

Figure 8. Adjusted (a) SWCRE, (b) LWCRE and (c) netCRE feedback ($W/m^2/K$) difference between amip4K and abrupt4xCO_2 as a function of years used in computing amip4K feedbacks. The box extends from the 25th percentile (Q1) and 75th percentile (Q3) with the horizontal line at the median (Q2) and the square at the mean. The whiskers indicate the range of the nonoutliers [outliers are either > (Q3 + 1.5 * IQR) or < (Q1 - 1.5*IQR); IQR=Q3-Q1]. Outliers are plotted as separate dots.

Figure 9 further presents the std_ratio, the correlations and γ between coupled and amip4K feedbacks as a function of amip4K simulation duration. The std_ratio for SW and net cloud feedbacks decreases from around 1.0 to 0.8 with increased months and stabilizes at around 0.8-0.9 in the yearly scale (Figure 9a), indicating that inter-model spread of amip4K feedbacks is slightly reduced compared to that in coupled. In contrast, although the std_ratio also decreases with increased months, the stabilized std_ratio exceeds 1 for the LW cloud feedback (Figure 9a), indicating slightly greater spread in amip4K than in coupled. The std_ratio is nearly invariant with further increased amip4K duration for each component in the yearly scale. A similar conclusion is reached from considering the Pearson and Spearman correlation coefficients and γ between amip4K and coupled feedbacks, which show little variation with amip4K simulation duration (Figure 9b-c). Overall, amip4K feedbacks derived from the first year would be sufficient to capture the inter-model spread of coupled feedbacks. Further investigation on those 1-year cloud feedbacks indicates that the exact year chosen does not matter much (Figure S9), although for some models, it is slightly better if one avoids ENSO/volcano years (not shown). If taking the monthly feedback into account, correlations and γ are smaller and the variation of diagnostic variables is larger than that from the yearly feedback (Figure 9). Nonetheless, we find that solstice months should be avoided if only one month of atmosphere-only simulation is to be run (Figure S10d-i), as they show systematically less agreement with the coupled feedbacks.

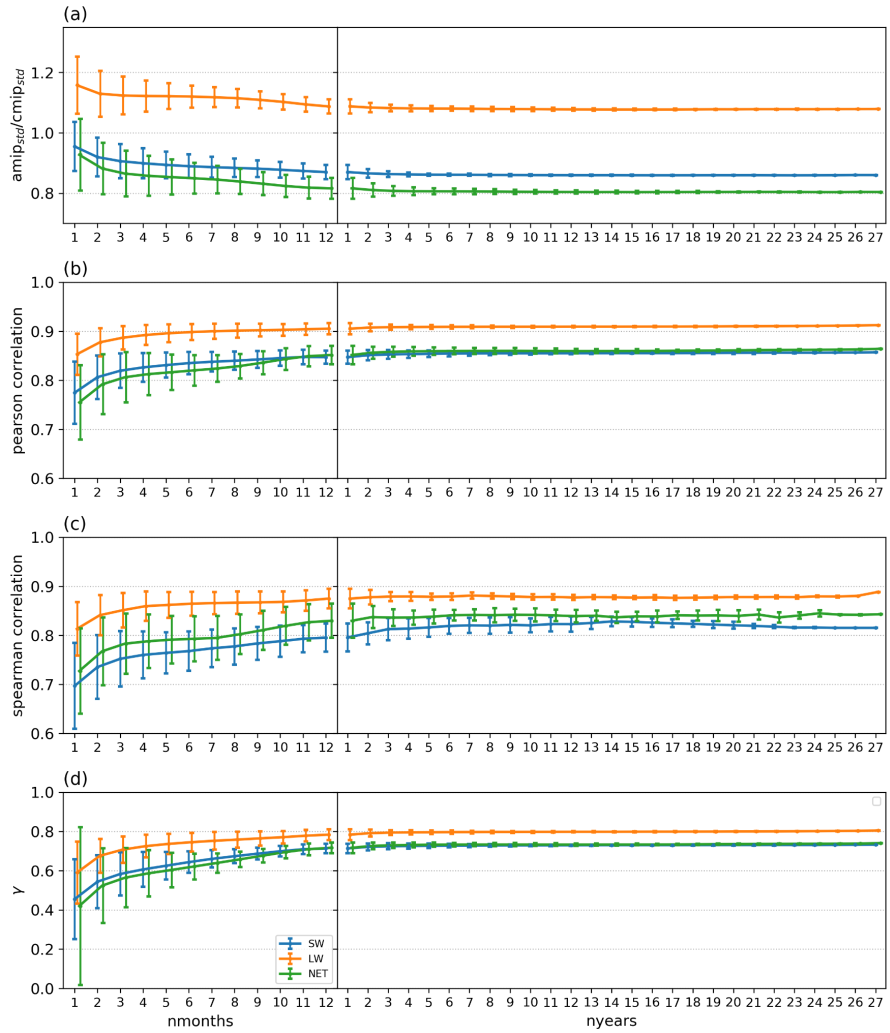

Figure 9. (a) the ratio of amip4K cross-model standard deviation (std) to abrupt-4xCO_2 cross-model standard deviation, (b) Pearson correlation coefficient, (c) Spearman correlation coefficient and (d) γ for adjusted (blue) SW, (orange) LW and (green) net CRE feedbacks between abrupt4xCO_2 and amip4K experiments as a function of months/years used in computing amip4K cloud feedbacks. The error bar denotes the standard deviation of each variable due to the variation of selected time slices.

Section 3.1.3 shows a near global correspondence between amip and coupled cloud feedbacks, especially for net cloud feedback (Figure 4). Given that 1-year global-mean amip4K cloud feedbacks would be enough to capture the inter-model spread of coupled feedbacks, it would be useful to know whether the good correspondence holds regionally using 1-year amip4K cloud feedbacks or whether some regions need more amip4K simulation duration to capture the coupled cloud feedbacks (if correspondence is possible). Figure 10 shows the spatial distribution of required minimum simulation years and the corresponding fraction area of the planet with significant correlations (p-value smaller than 0.05) with increased amip4K years for SW, LW and net cloud feedbacks. For each grid point, the minimum simulation year is defined as the simulation duration which (1) first exhibits a p-value smaller 0.05 and (2) the significance is held for the following 5 simulation duration. For example, if the 1-year duration for one grid point first exhibits significance and the following 2-, 3-, 4-, 5- and 6-year durations are also significant, then we regard 1-year duration as the minimum simulation year for this grid point. The 1-year amip4K simulation can largely capture the inter-model spread of local coupled feedbacks in many regions including but not limited to the southern Indian and Atlantic Oceans (Figure 10a, c, e). The signal is slightly more complicated in the Pacific Ocean where 2 or more years are often needed to get a significant correlation. Over some land regions in the northern hemisphere, longer than 5 years are necessary. Of the spatial area of the planet in which a statistically significant correspondence between coupled and amip4K cloud feedbacks occurs, about half is achieved with a single year of the amip4K simulation, and about 90% is achieved with 5 years. Note that the choice of p-threshold (0.05 or 0.01) does not fundamentally affect this result (not shown).

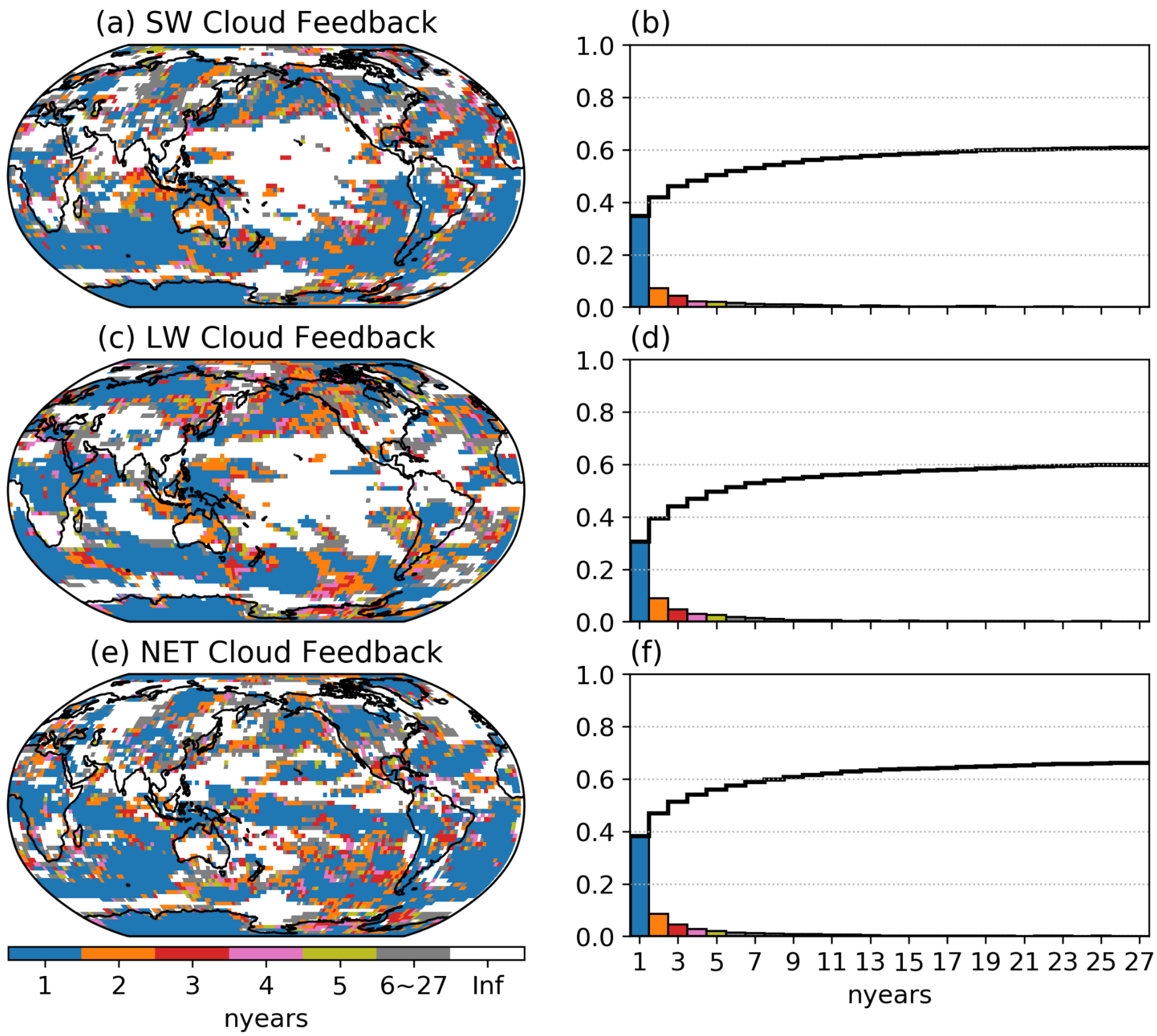

Figure 10. (a, c, e) The spatial distribution of the required minimum years of amip4K simulations to capture coupled cloud feedbacks (see the text for a description of the criteria applied), and (b, d, f) the fractional area of the planet with significant inter-model correlations as a function of years used in computing the amip4K feedbacks for the adjusted (a, b) SW, (c, d) LW and (e, f) net cloud feedbacks. White regions in (a, c, e) indicate locations where the correlation is not significant even using the full 27 years amip4K experiments. The black line in (b, d, f) denotes the cumulative curve.

In summary, we conclude that cloud feedbacks computed from amip4K experiments of only 1 year duration can closely capture the inter-model spread of global mean coupled feedbacks. Increased simulation duration does not improve this agreement materially. Furthermore, for each model, increasing the simulation years does not reduce the cloud feedback difference between

AMIP and coupled experiments. For most models, the systematic bias between amip4K and coupled cloud feedbacks is apparent with only a single year of output and does not change much with increased simulation duration. For regional feedbacks, 1-year experiments can capture around half of the significant regions and 5-year experiments are sufficient to capture almost all the regions shown to have a significant correlation when using the full 27 years of amip4K simulations. This is reassuring evidence that short duration atmosphere-only experiments, such as those often performed while developing new atmosphere model versions, provide highly valuable information about the cloud feedbacks operating in the corresponding fully coupled model.

4 Conclusions

We have compared radiative feedbacks between amip4K and coupled experiments in CMIP5 and CMIP6 models, including their global-mean values, spatial distribution, and breakdown into individual cloud feedback components. Consistent with previous studies (Ringer et al., 2014), the total negative radiative feedback is weaker in coupled experiments, which arises solely from differences in clear-sky feedback strengths. Weaker positive global-mean clear-sky SW radiative feedbacks are related to the weaker surface albedo feedbacks in amip4K experiments, which lack sea ice reduction. Stronger negative global-mean clear-sky LW radiative feedbacks arise from stronger negative lapse rate feedbacks in amip4K experiments, which lack polar-amplified surface warming. In contrast to clear-sky feedbacks, global-mean cloud feedbacks are highly correlated between amip4K and coupled experiments. This correspondence is better than previously reported in the literature because we have accounted for non-cloud influences that alias onto raw changes in cloud radiative effect. This good correspondence also extends to the cloud feedbacks resulting from individual cloud property changes, as we showed that amip4K experiments successfully capture most of the coupled model diversity in global-mean cloud amount, altitude, and optical depth feedback components for all, low, and non-low clouds.

The close correspondence between amip4K and coupled cloud feedback extends beyond the global mean to the spatial distribution with around ⅔ of the planet exhibiting significant local correlations. Poor correspondence is present in the tropical Pacific for LW and SW cloud feedbacks. This arises because of a disparate response of high clouds between the two experiments, which have very different patterns of surface warming and therefore very different large-scale circulation responses. Tropical cloud feedbacks segregated into vertical motion regimes are, however, well-correlated between the two experiments.

Radiative forcing derived from the first 10 years and 36 years coupled experiments agrees best with the forcing from amip4xCO_2 and sstClim4xCO_2 experiments, respectively. The best time segment of coupled experiments to match the amip4xCO_2 or sstClim4xCO_2 radiative forcing is sensitive to the used model samples though. The higher similarity (control climate state, emissions, et al.) between sstClim and coupled experiments leads to a stronger correlation (relative to amip4xCO_2) between sstClim4xCO_2 and coupled forcing. However, the good correspondence between amip4xCO_2 and sstClim4xCO_2 forcing suggests the difference of model setup for amip and sstClim experiments play a second order role in the inter-model spread of forcing, consistent with Forster et al. (2016).

Ringer et al. (2014) found an anti-correlation between radiative forcing and feedback across CMIP5 models that becomes monotonically stronger with reduced complexity of experiments (from coupled to AMIP to aquaplanet). This is no longer the case in CMIP6 because the correlation between amip4xCO_2 forcing and amip4K feedback is now positive. The strong anti-correlation between cloud feedbacks and rapid cloud adjustments that drove the forcing-feedback relationship across CMIP5 models has also become weaker in CMIP6, for reasons that remain to be investigated. The lack of anti-correlation between forcing and feedback in the AMIP experiments when using all models suggests that there is no physical basis relating forcing to feedback.

In all possible options for forcing and feedback, the estimated ECS using the sstClim4xCO_2 forcing and amip4K feedback agrees best with the coupled ECS, with the values from amip4xCO_2 forcing and amip4K feedback close behind. Furthermore, we find that cloud feedbacks derived from 1-year atmosphere-only simulations can largely capture the inter-model spread of the coupled feedbacks. The feedback difference between amip4K and coupled experiments asymptotes quickly to a small systematic bias for most models. Further examination of the correspondence of regional feedbacks shows that 1-year amip4K simulation can capture about half of the regions with significant correlation, and 5 years get a very similar correspondence as that using the full 27-year amip4K experiments. The good agreement of cloud feedbacks in both global-mean and spatial distribution justifies using amip4K experiments to further understand coupled cloud feedbacks not only for global-mean, but also for the ⅔ of the planet with significant local correlations and all tropical vertical motion regimes.

Acknowledgments, Samples, and Data

This work was supported by the U.S. Department of Energy (DOE) Regional and Global Modeling Analysis program area and was performed under the auspices of the DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP5 and CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP and ESGF. We are grateful for insightful discussions with Peter Caldwell, Tim Myers, Chris Golaz.

References

Andrews, T., Gregory, J. M., Forster, P. M., & Webb, M. J. (2012). Cloud Adjustment and its Role in CO2 Radiative Forcing and Climate Sensitivity: A Review. Surveys in Geophysics, 33(3-4), 619–635. doi: 10.1007/ s10712-011-9152-0

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forc- ing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. Geophysical Research Letters, 39(9), 1–7. doi: 10.1029/2012GL051607

Andrews, T., Smith, C. J., Myhre, G., Forster, P. M., Chadwick, R., & Ackerley, D. (2021). Effective Radiative Forcing in a GCM With Fixed Surface Temperatures. Journal of Geophysical Research: Atmospheres, 126(4). doi: 10.1029/2020JD033880

Bony, S., & Dufresne, J. L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. Geophysical Research Letters, 32(20). doi: 10.1029/2005GL023851

Bony, S., Webb, M., Bretherton, C., Klein, S. A., Siebesma, P., Tselioudis, G., & Zhang, M. (2011). CFMIP: Towards a better evaluation and understand- ing of clouds and cloud feedbacks in CMIP5 models. Clivar Exchanges, 56(2), 20–22. Retrieved from http://www.euclipse.eu/downloads/ CFMIP CMIP5 Exchanges May2011.pdf

Bretherton, C. S., Blossey, P. N., & Stan, C. (2014, 12). Cloud feedbacks on greenhouse warming in the superparameterized climate model SP-CCSM4. Journal of Advances in Modeling Earth Systems, 6(4), 1185–1204. Retrieved from http://doi.wiley.com/10.1002/2014MS000355 doi: 10.1002/2014MS000355

Brient, F., & Bony, S. (2012). How may low-cloud radiative properties simulated in the current climate influence low-cloud feedbacks under global warming? Geophysical Research Letters, 39(20), 1–6. doi: 10.1029/2012GL053265

Brient, F., & Bony, S. (2013). Interpretation of the positive low-cloud feedback pre- dicted by a climate model under global warming. Climate Dynamics, 40(9-10), 2415–2431. doi: 10.1007/s00382-011-1279-7

Caldwell, P. M., Zelinka, M. D., Taylor, K. E., & Marvel, K. (2016). Quantifying the sources of intermodel spread in equilibrium climate sensitivity. Journal of Climate, 29(2), 513–524. doi: 10.1175/JCLI-D-15-0352.1

Ceppi, P., Hartmann, D. L., & Webb, M. J. (2016). Mechanisms of the negative shortwave cloud feedback in middle to high latitudes. Journal of Climate, 29(1), 139–157. doi: 10.1175/JCLI-D-15-0327.1

Chung, E. S., & Soden, B. J. (2015). An assessment of methods for computing ra- diative forcing in climate models. Environmental Research Letters, 10(7). doi: 10.1088/1748-9326/10/7/074004

Chung, E. S., & Soden, B. J. (2018). On the compensation between cloud feedback and cloud adjustment in climate models. Climate Dynamics, 50(3-4). doi: 10 .1007/s00382-017-3682-1

Del Genio, A. D., Yao, M. S., & Jonas, J. (2007). Will moist convection be stronger in a warmer climate? Geophysical Research Letters, 34(16), 7790–7795. doi: 10 .1029/2007GL030525

Demoto, S., Watanabe, M., & Kamae, Y. (2013, 5). Mechanism of tropical low-cloud response to surface warming using weather and climate simulations. Geophys- ical Research Letters, 40(10), 2427–2432. Retrieved from http://doi.wiley .com/10.1002/grl.50474 doi: 10.1002/grl.50474

Dong, Y., Armour, K. C., Zelinka, M. D., Proistosescu, C., Battisti, D. S., Zhou, C., & Andrews, T. (2020). Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. Journal of Climate, 33(18). doi: 10.1175/JCLI-D-19-1011.1

Endo, H., Kitoh, A., & Ueda, H. (2018). A unique feature of the Asian summer monsoon response to global warming: The role of different land-sea thermal contrast change between the lower and upper troposphere. Scientific Online Letters on the Atmosphere, 14. doi: 10.2151/SOLA.2018-010

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., . . . Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. Journal of Geophysical Research, 121(20). doi: 10.1002/2016JD025320

Geen, R., Bordoni, S., Battisti, D. S., & Hui, K. (2020). Monsoons, ITCZs, and the Concept of the Global Monsoon (Vol. 58) (No. 4). doi: 10.1029/ 2020RG000700

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., . . . Mills, M. J. (2019). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). Geophysical Research Letters, 2, 8329–8337. doi: 10.1029/2019gl083978

Gettelman, A., Kay, J. E., & Fasullo, J. T. (2013, 6). Spatial decomposition of climate feedbacks in the community earth system model. Journal of Climate, 26(11), 3544–3561. Retrieved from http://journals.ametsoc.org/doi/abs/ 10.1175/JCLI-D-12-00497.1 doi: 10.1175/JCLI-D-12-00497.1

Gettelman, A., Kay, J. E., & Shell, K. M. (2012). The evolution of climate sensitiv- ity and climate feedbacks in the community atmosphere model. Journal of Cli- mate, 25(5), 1453–1469. doi: 10.1175/JCLI-D-11-00197.1

Good, P., Ingram, W., Lambert, F. H., Lowe, J. A., Gregory, J. M., Webb, M. J.,. . . Wu, P. (2012). A step-response approach for predicting and under- standing non-linear precipitation changes. Climate Dynamics, 39(12). doi: 10.1007/s00382-012-1571-1

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., . . . Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. Geophysical Research Letters, 31(3), 2–5. doi: 10.1029/2003GL018747

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., . . . Zhang, S. (2005). Efficacy of climate forcings. Journal of Geophysical Re- search D: Atmospheres, 110(18), 1–45. doi: 10.1029/2005JD005776

Held, I. M., & Shell, K. M. (2012). Using relative humidity as a state variable in cli- mate feedback analysis. Journal of Climate, 25(8). doi: 10.1175/JCLI-D-11 -00721.1

Huang, Y., Xia, Y., & Tan, X. (2017). On the pattern of CO2 radiative forcing and poleward energy transport. Journal of Geophysical Research: Atmospheres, 122(20). doi: 10.1002/2017JD027221

J. Smith, C., J. Kramer, R., Myhre, G., Alterskjr, K., Collins, W., Sima, A., . . . M. Forster, P. (2020). Effective radiative forcing and adjustments in CMIP6 models. Atmospheric Chemistry and Physics, 20(16). doi: 10.5194/acp-20-9591-2020

Kamae, Y., Shiogama, H., Watanabe, M., Ogura, T., Yokohata, T., & Kimoto, M. (2016, 9). Lower-tropospheric mixing as a constraint on cloud feedback in a multiparameter multiphysics ensemble. Journal of Climate, 29(17), 6259–6275. Retrieved from http://journals.ametsoc.org/doi/10.1175/ JCLI-D-16-0042.1 doi: 10.1175/JCLI-D-16-0042.1

Medeiros, B., Stevens, B., & Bony, S. (2015). Using aquaplanets to understand the robust responses of comprehensive climate models to forcing. Climate Dynam- ics, 44(7-8), 1957–1977. doi: 10.1007/s00382-014-2138-0

Miura, H., Tomita, H., Nasuno, T., Iga, S. I., Satoh, M., & Matsuno, T. A climate sensitivity test using a global cloud resolving model under an aqua planet condition. Geophysical Research Letters, 32(19). 10.1029/2005GL023672 (2005). doi: 10.1029/2005GL023672

Noda, A. T., Kodama, C., Yamada, Y., Satoh, M., Ogura, T., & Ohno, T. (2019). Responses of Clouds and Large-Scale Circulation to Global Warming Eval- uated From Multidecadal Simulations Using a Global Nonhydrostatic Model. Journal of Advances in Modeling Earth Systems, 11(9). doi: 10.1029/2019MS001658

Parishani, H., Pritchard, M. S., Bretherton, C. S., Terai, C. R., Wyant, M. C., Khairoutdinov, M., & Singh, B. (2018). Insensitivity of the Cloud Response to Surface Warming Under Radical Changes to Boundary Layer Turbulence and Cloud Microphysics: Results From the Ultraparameterized CAM. Journal of Advances in Modeling Earth Systems, 10(12). doi: 10.1029/2018MS001409

Po-Chedley, S., Armour, K. C., Bitz, C. M., Zelinka, M. D., & Santer, B. D. (2018). Sources of intermodel spread in the lapse rate and water vapor feedbacks. Journal of Climate, 31(8). doi: 10.1175/JCLI-D-17-0674.1

Reichler, T., Dameris, M., & Sausen, R. (2003). Determining the tropopause height from gridded data. Geophysical Research Letters, 30(20). doi: 10.1029/2003GL018240

Ringer, M. A., Andrews, T., & Webb, M. J. (2014, 6). Global-mean radiative feedbacks and forcing in atmosphere-only and coupled atmosphere-ocean climate change experiments. Geophysical Research Letters, 41(11), 4035–4042. Retrieved from http://doi.wiley.com/10.1002/2014GL060347 doi: 10.1002/2014GL060347

Satoh, M., Iga, S. I., Tomita, H., Tsushima, Y., & Noda, A. T. (2012). Response of upper clouds in global warming experiments obtained using a global nonhy- drostatic model with explicit cloud processes. Journal of Climate, 25(6). doi: 10.1175/JCLI-D-11-00152.1

Shell, K. M., Kiehl, J. T., & Shields, C. A. (2008). Using the radiative kernel technique to calculate climate feedbacks in NCAR's Community Atmospheric Model. Journal of Climate, 21(10), 2269–2282. doi: 10.1175/2007JCLI2044.1

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Har- greaves, J. C., . . . Zelinka, M. D. (2020). An Assessment of Earth's Cli- mate Sensitivity Using Multiple Lines of Evidence (Vol. 58) (No. 4). doi: 10.1029/2019RG000678

Singh, D., Ghosh, S., Roxy, M. K., & McDermid, S. (2019). Indian summer monsoon: Extreme events, historical changes, and role of anthropogenic forcings (Vol. 10) (No. 2). doi: 10.1002/wcc.571

Soden, B. J., & Held, I. M. (2006, 7). An assessment of climate feedbacks in coupled ocean-atmosphere models. Journal of Climate, 19(14), 3354–3360. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/JCLI3799.1 doi: 10 .1175/JCLI3799.1

Soden, B. J., Held, I. M., Colman, R. C., Shell, K. M., Kiehl, J. T., & Shields, C. A. (2008). Quantifying climate feedbacks using radiative kernels. Journal of Climate, 21(14), 3504–3520. doi: 10.1175/2007JCLI2110.1

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., . . . Zhou, L. (2019). DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (Vol. 6) (No. 1). doi: 10.1186/s40645-019-0304-z

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012, 4). An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485–498. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/ BAMS-D-11-00094.1 doi: 10.1175/BAMS-D-11-00094.1

Tsushima, Y., Iga, S. I., Tomita, H., Satoh, M., Noda, A. T., & Webb, M. J. (2015). High cloud increase in a perturbed SST experiment with a global nonhydro- static model including explicit convective processes. Journal of Advances in Modeling Earth Systems, 6(3). doi: 10.1002/2013MS000301

Wang, B., Ding, Q., Fu, X., Kang, I. S., Jin, K., Shukla, J., & Doblas-Reyes, F. (2005). Fundamental challenge in simulation and prediction of summer monsoon rainfall. Geophysical Research Letters, 32(15). doi: 10.1029/ 2005GL022734

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chad- wick, R., . . . Watanabe, M. (2017, 1). The Cloud Feedback Model Inter- comparison Project (CFMIP) contribution to CMIP6. Geoscientific Model Development, 10(1), 359–384. Retrieved from https://www.geosci-model -dev.net/10/359/2017/ doi: 10.5194/gmd-10-359-2017

Webb, M. J., & Lock, A. P. (2013). Coupling between subtropical cloud feedback and the local hydrological cycle in a climate model. Climate Dynamics, 41(7- 8), 1923–1939. doi: 10.1007/s00382-012-1608-5

Webb, M. J., Lock, A. P., Bretherton, C. S., Bony, S., Cole, J. N. S., Idelkadi, A., . . . Zhao, M. (2015, 11). The impact of parametrized convection on cloud feedback. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373(2054), 20140414. Retrieved from https://royalsocietypublishing.org/doi/10.1098/rsta.2014.0414 doi: 10.1098/rsta.2014.0414

Webb, M. J., Senior, C. A., Sexton, D. M., Ingram, W. J., Williams, K. D., Ringer, M. A., . . . Taylor, K. E. (2006). On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles (Vol. 27) (No. 1). doi: 10.1007/s00382-006-0111-2

Xu, K. M., & Cheng, A. (2016). Understanding the tropical cloud feedback from an analysis of the circulation and stability regimes simulated from an upgraded multiscale modeling framework. Journal of Advances in Modeling Earth Sys- tems, 8(4). doi: 10.1002/2016MS000767

Zelinka, M. D., Klein, S. A., & Hartmann, D. L. (2012). Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative ker- nels. Journal of Climate, 25(11), 3715–3735. doi: 10.1175/JCLI-D-11-00248.1

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., . . . Taylor, K. E. (2020, 1). Causes of higher climate sensitivity in CMIP6 models. Geophysical Research Letters, n/a(n/a), 2019GL085782. Retrieved from https://doi.org/10.1029/2019GL085782https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085782 doi: 10.1029/2019GL085782

Zelinka, M. D., Zhou, C., & Klein, S. A. (2016). Insights from a refined decomposi- tion of cloud feedbacks. Geophysical Research Letters, 43(17), 9259–9269. doi: 10.1002/2016GL069917

Zhang, H., Wang, M., Guo, Z., Zhou, C., Zhou, T., Qian, Y., . . . Gettelman, A. (2018, 11). Low-Cloud Feedback in CAM5-CLUBB: Physical Mechanisms and Parameter Sensitivity Analysis. Journal of Advances in Modeling Earth Sys- tems, 10(11), 2844–2864. Retrieved from http://doi.wiley.com/10.1029/ 2018MS001423https://onlinelibrary.wiley.com/doi/abs/10.1029/ 2018MS001423 doi: 10.1029/2018MS001423

Journal of Geophysical Research: Atmospheres

Supporting Information for

On the Correspondence between Atmosphere-Only and Coupled Simulations for Radiative Feedbacks and Forcing from CO_2

Yi Qin[1], Mark D. Zelinka[1], Stephen A. Klein[1]

[1]PCMDI/Lawrence Livermore National Laboratory, Livermore, CA, USA

Contents of this file

Figures S1 to S10

Introduction

In this Supporting Information, we provide additional figures that support the results in the main text (Figure S1-S10).

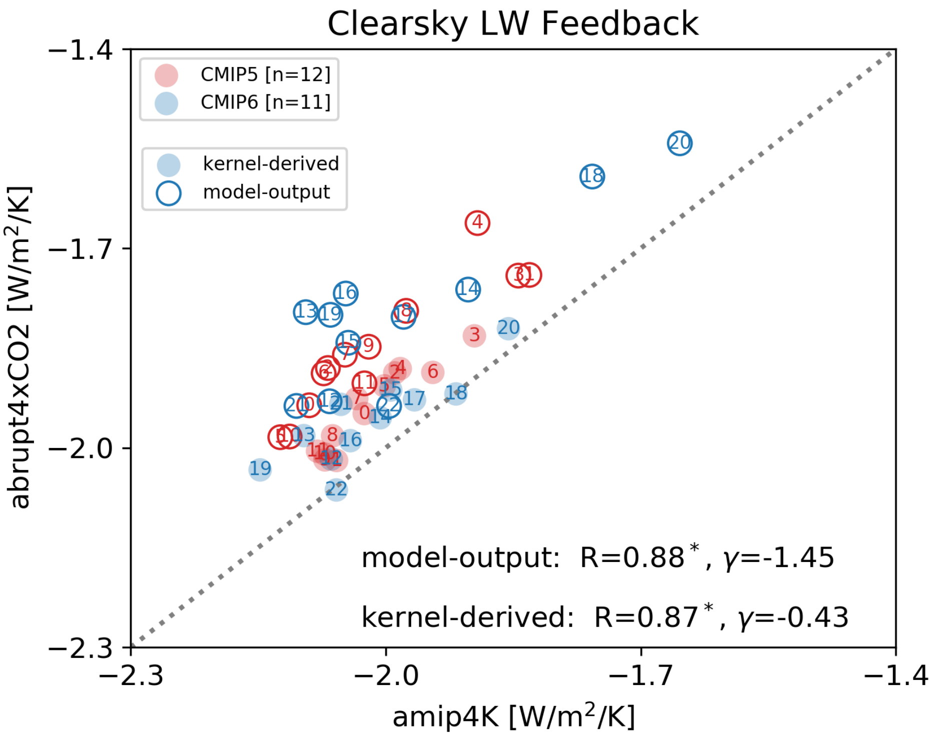

Figure S1. Global-mean net clear-sky LW radiative feedbacks (W/m^2/K) calculated by adding contributions from primary contributors (water vapor, surface temperature, atmospheric temperature) determined using the radiative kernel method (filled markers) and directly calculated by the model output (unfilled markers) compared between amip4K and fully coupled abrupt4xCO$_2$ experiments. Red and blue dots denote CMIP5 and CMIP6 models respectively. Models used in later radiative kernel and cloud radiative kernel analysis are labelled by numbers as denoted in Table 1 and 2.

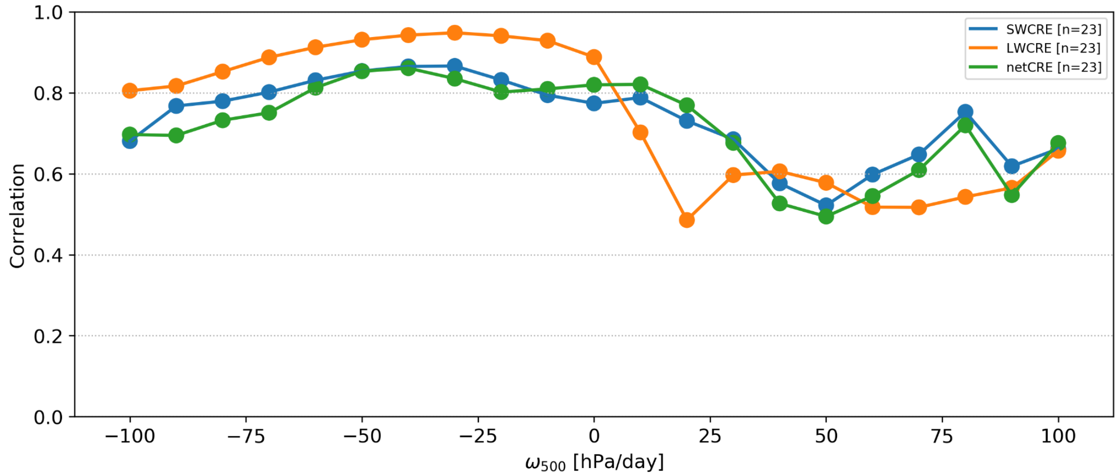

Figure S2. Across-model correlation of tropical (30°S-30°N) ocean unadjusted SW (blue), LW (orange) and net (green) CRE feedbacks (amip4K minus amip; abrupt4xCO_2 minus piControl) between amip4K and coupled experiments in each 500 hPa vertical velocity regime. Correlation coefficients significant at the 95% confidence level are marked by solid dots.

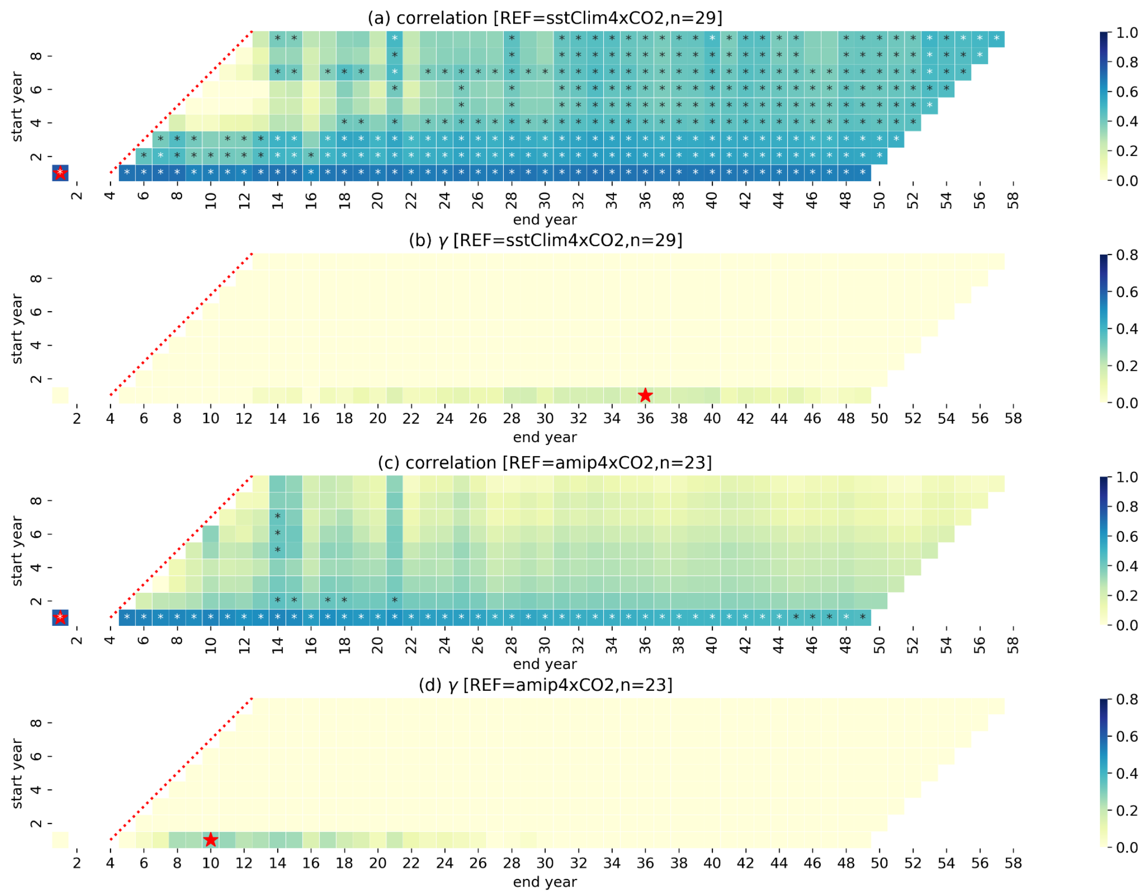

Figure S3. The (a, c) correlation and (b, d) γ matrix between (a-b) sstClim4xCO_2/ (c-d) amip4xCO_2 forcing and coupled forcing derived using varying starting and end years. The selected minimum length of coupled experiments is 5 years (red dotted lines). Single asterisk indicates correlations significant at the 95% level. Red stars denote the largest correlation/γ. The starting year ranges from 1 to 10 and the total duration considered ranges from 5 to 50. The year-1 coupled radiation anomaly is also considered.

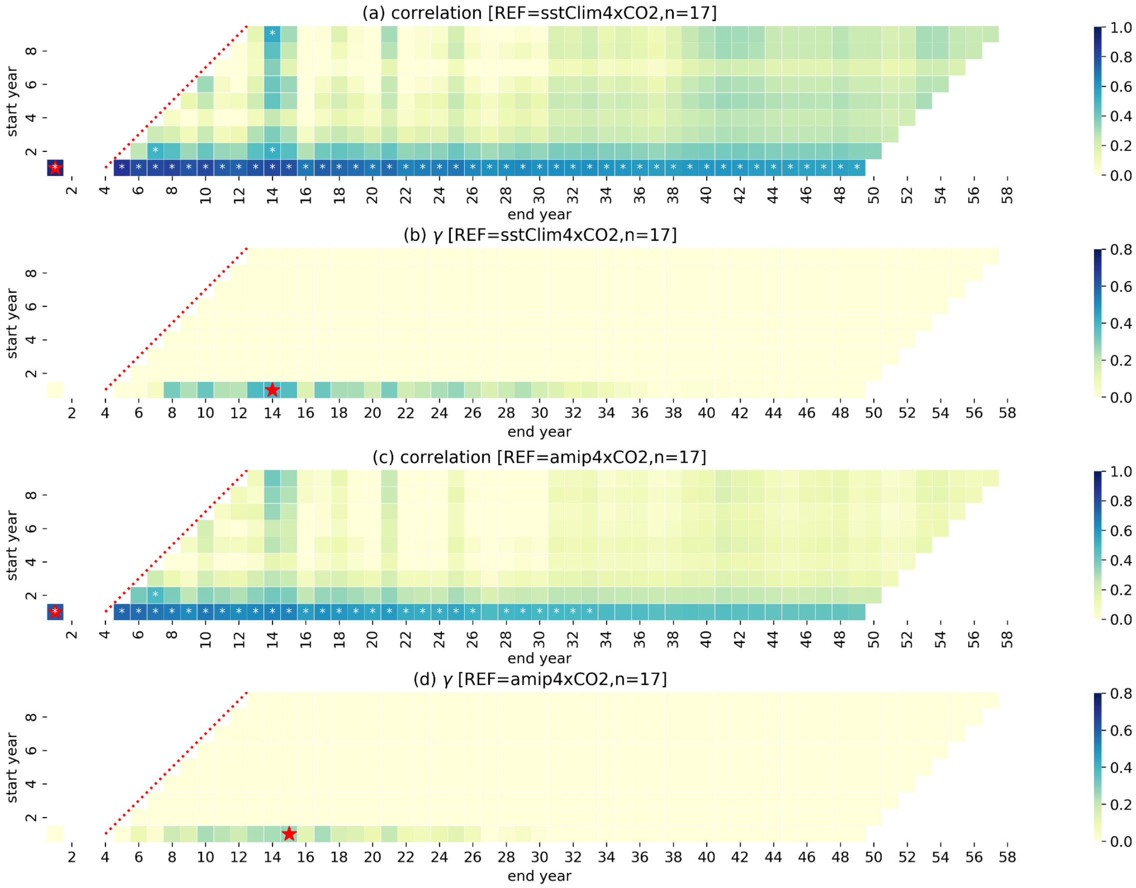

Figure S4. As in Figure S3, but for those models with both sstClim4xCO_2 and amip4xCO_2 experiments.

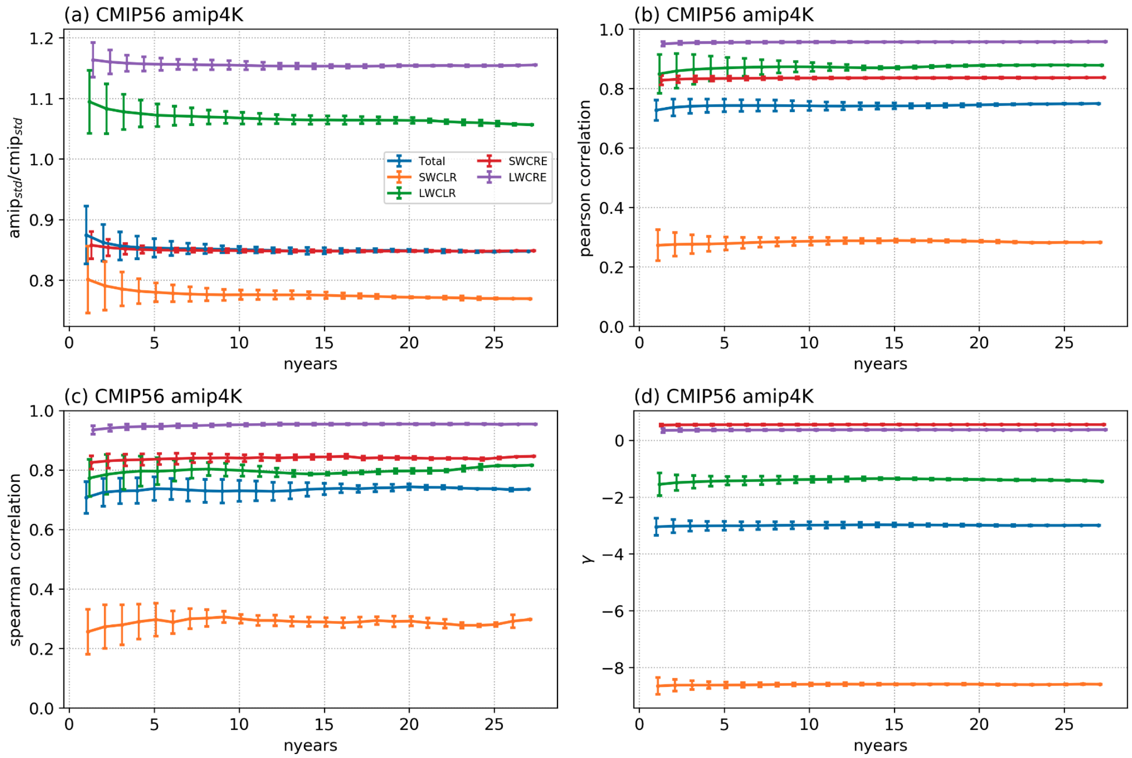

Figure S5. (a) the ratio of amip4K across-model standard deviation (std) to abrupt-4xCO_2 across-model standard deviation, (b) Pearson linear correlation coefficient, (c) Spearman rank correlation coefficient and (d) γ for (blue) total, (orange) SWCLR, (green) LWCLR, unadjusted (red) SWCRE and (purple) LWCRE feedbacks between abrupt4xCO_2 and amip4K experiments as a function of years used to calculate amip4K cloud feedbacks. The error bar denotes the standard deviation of each variable due to the variation of selected time slices.

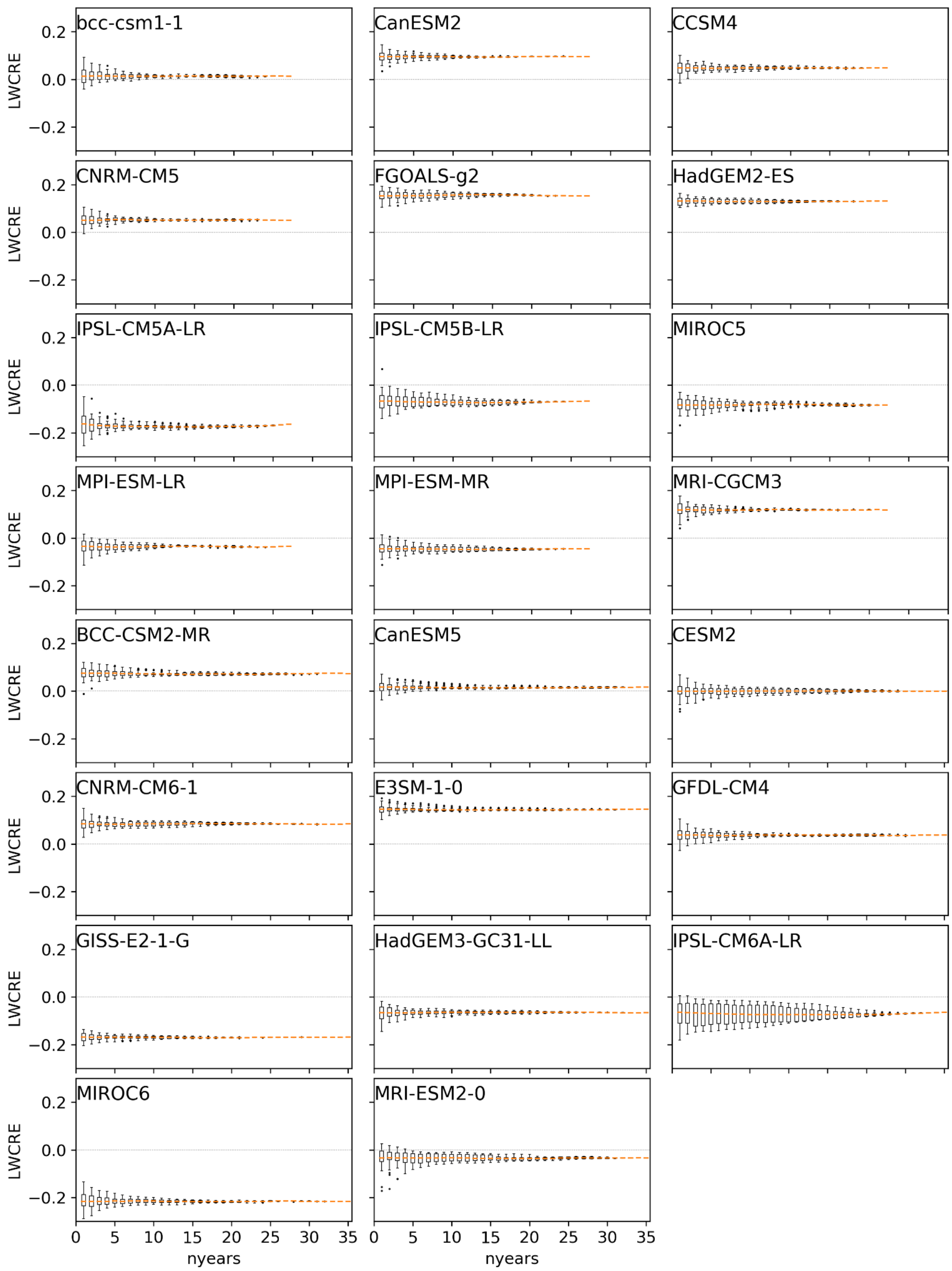

Figure S6. The adjusted LWCRE feedback (W/m^2/K) difference between amip4K and coupled experiments as a function of years used in calculating amip4K feedbacks for available CMIP models. The box extends from the 25th percentile (Q1) and 75th percentile (Q3) with the horizontal line at the mean. The whiskers indicate the range of the nonoutliers [outliers are either > (Q3 + 1.5 * IQR) or < (Q1 - 1.5*IQR); IQR=Q3-Q1]. Outliers are plotted as separate dots.

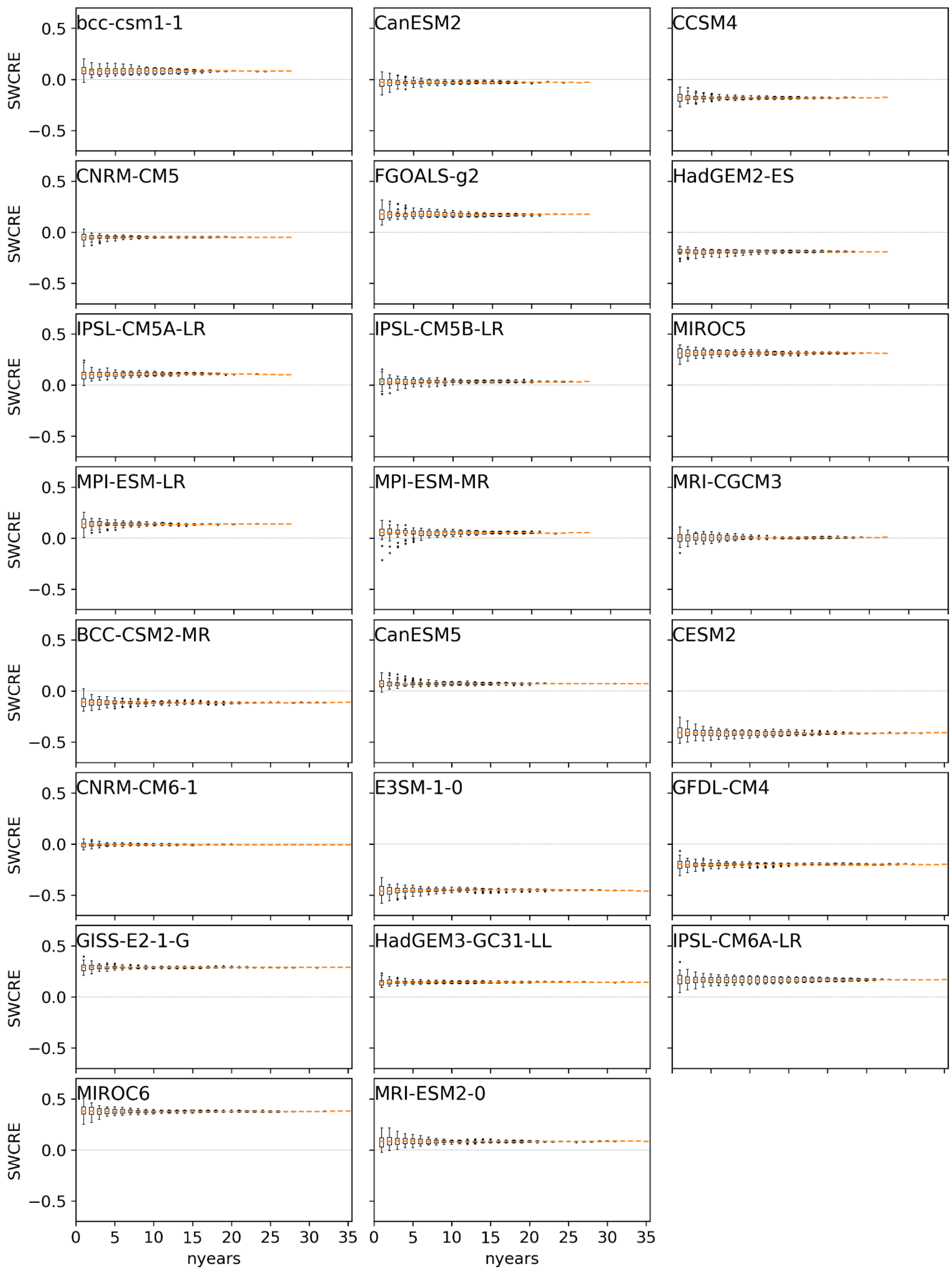

Figure S7. As in Figure S6, but for adjusted SWCRE feedback.

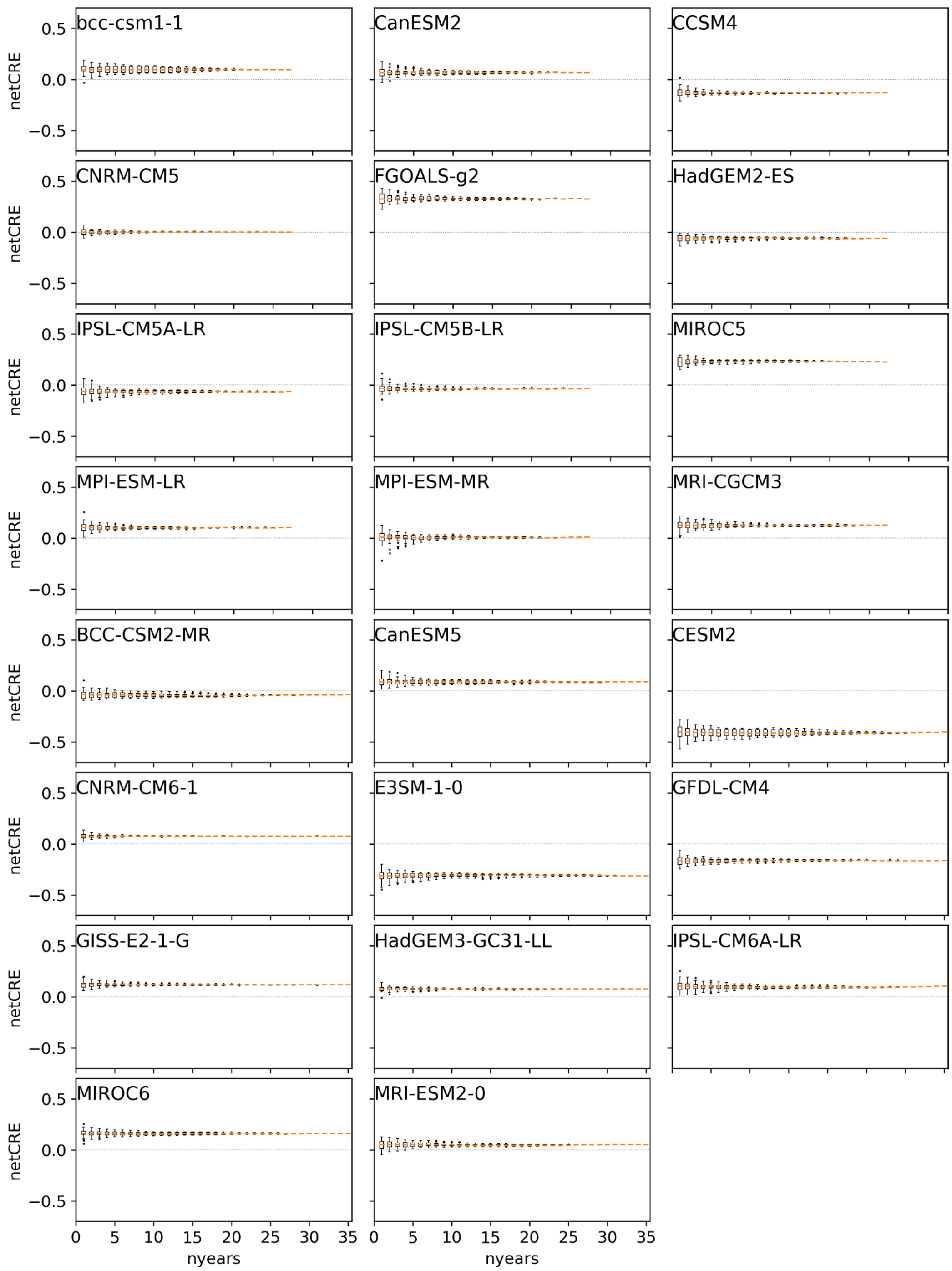

Figure S8. As in Figure S6, but for adjusted net CRE feedback.

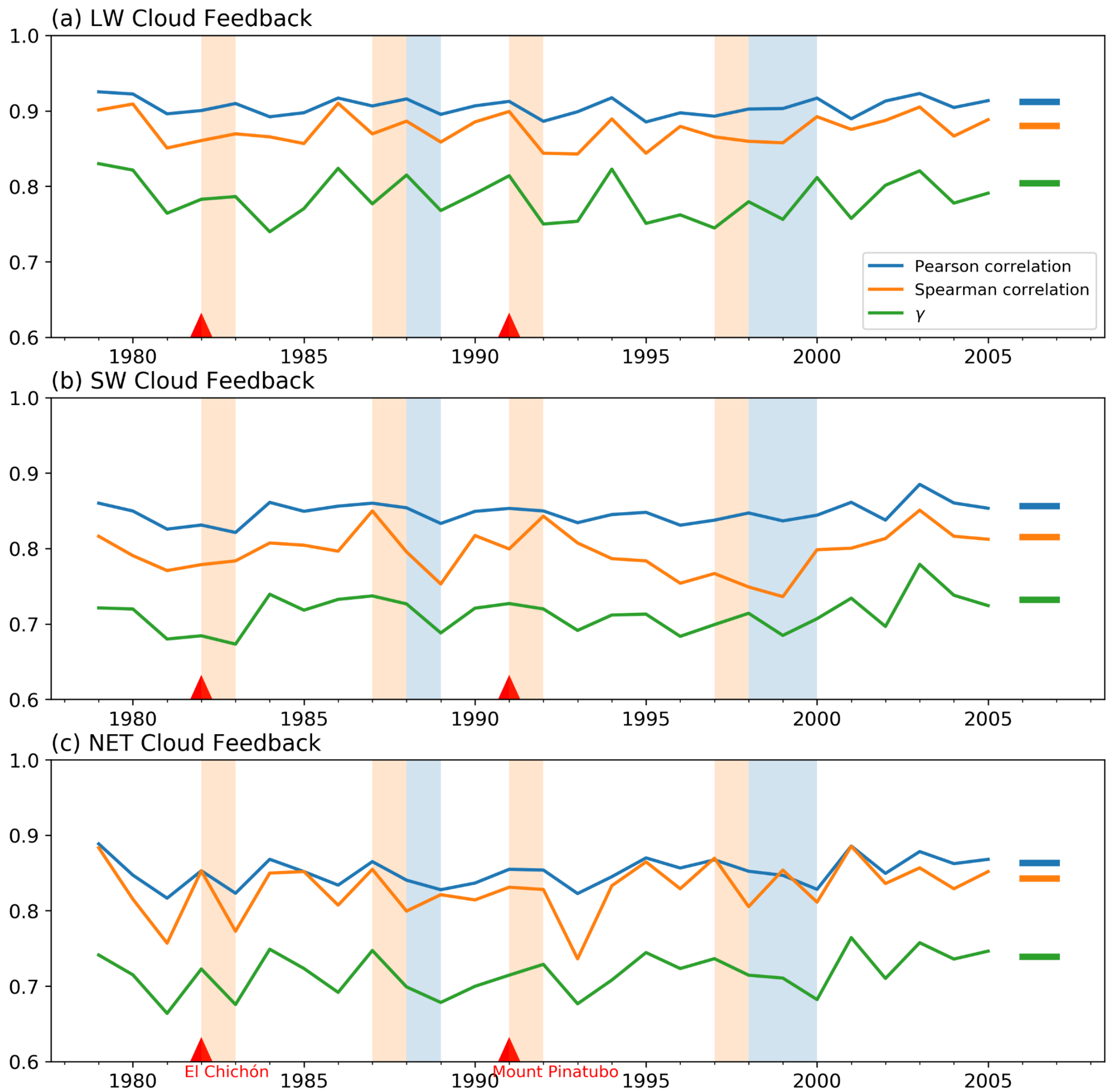

Figure S9. The (blue) Pearson linear correlation, (orange) Spearman rank correlation, and (green) γ for adjusted (a) LW, (b) SW and (c) net CRE feedbacks between 1-year amip and coupled experiments as a function of the simulation year varying from 1979 to 2005. Two volcanoes (El Chichon in 1982 and Mount Pinatubo in 1991) are labelled by red triangles. Red shades denote El Nino events, and blue shades denote La Nina events.

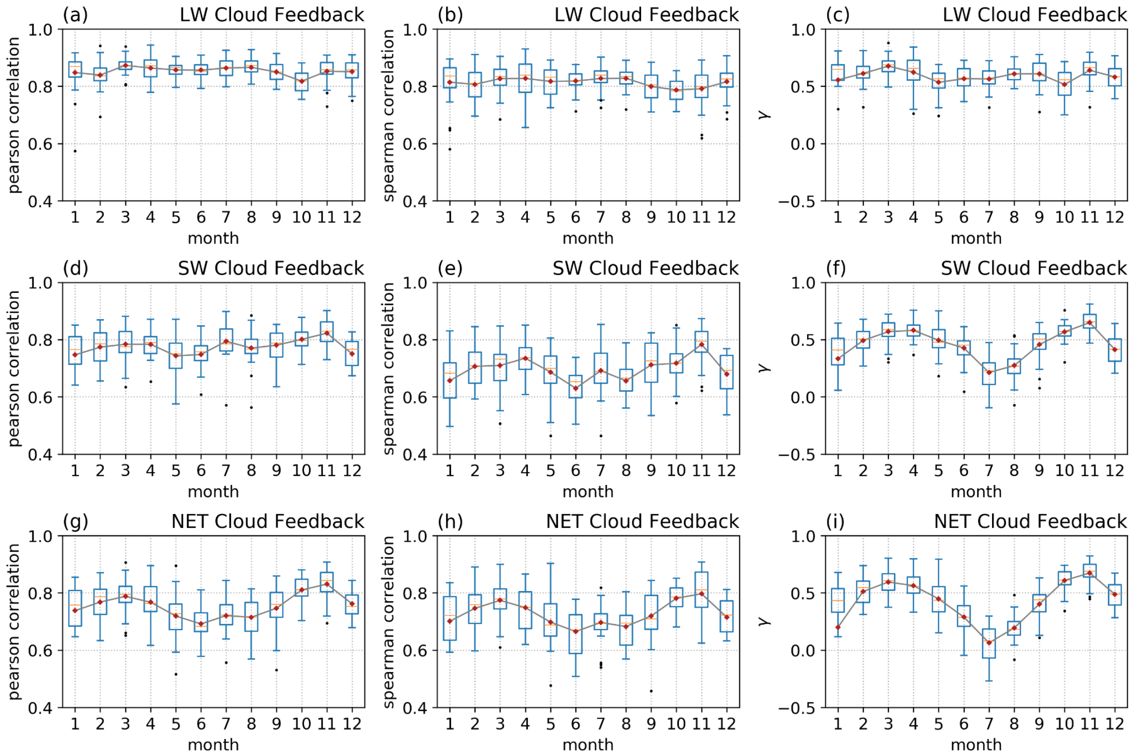

Figure S10. The (a, d, g) Pearson correlation, (b, e, h) Spearman correlation, and (c, f, i) γ for adjusted (a-c) SW, (d-f) LW and (g-i) net CRE feedbacks between 1-month amip and coupled experiments as a function of the used month for amip feedback varying from January to December. The box extends from the 25th percentile (Q1) and 75th percentile (Q3) with the horizontal line at the median and the red dot at the mean. The whiskers indicate the range of the nonoutliers [outliers are either > (Q3 + 1.5 * IQR) or < (Q1 - 1.5*IQR); IQR=Q3-Q1]. Outliers are plotted as separate black dots.