Perturbed physics ensemble using the MIROC5 coupled atmosphere–ocean GCM without flux corrections: experimental design and results

Parametric uncertainty of climate sensitivity

Hideo Shiogama · Masahiro Watanabe · Masakazu Yoshimori · Tokuta Yokohata · Tomoo Ogura · James D. Annan · Julia C. Hargreaves · Manabu Abe · Youichi Kamae · Ryouta O'ishi · Rei Nobui · Seita Emori · Toru Nozawa · Ayako Abe-Ouchi · Masahide Kimoto

Abstract In this study, we constructed a perturbed physics ensemble (PPE) for the MIROC5 coupled atmosphere–ocean general circulation model (CGCM) to investigate the parametric uncertainty of climate sensitivity (CS). Previous studies of PPEs have mainly used the atmosphere-slab ocean models. A few PPE studies using a CGCM applied flux corrections, because perturbations in parameters can lead to large radiation imbalances at the top of the atmosphere and climate drifts. We developed a method to prevent climate drifts in PPE experiments using the MIROC5 CGCM without flux corrections. We simultaneously swept 10 parameters in atmosphere and surface schemes. The range of CS (estimated from our 35 ensemble members) was not wide (2.2–3.2 °C). The shortwave cloud feedback related to changes in middle-level cloud albedo dominated the variations in the total feedback. We found three performance metrics for the present climate simulations of middle-level cloud albedo, precipitation, and ENSO amplitude that systematically relate to the variations in shortwave cloud feedback in this PPE.

Keywords Climate sensitivity · Cloud feedback · General circulation model · Perturbed physics ensemble · Metrics

H. Shiogama (✉) · T. Yokohata · T. Ogura · M. Abe ·
S. Emori · T. Nozawa
National Institute for Environmental Studies, Tsukuba,
Ibaraki 305-8506, Japan
e-mail: shiogama.hideo@nies.go.jp

M. Watanabe · M. Yoshimori · Y. Kamae · R. O'ishi ·
R. Nobui · S. Emori · A. Abe-Ouchi · M. Kimoto
Atmosphere and Ocean Research Institute,
The University of Tokyo, Kashiwa, Japan

J. D. Annan · J. C. Hargreaves
Japan Agency for Marine-Earth Science and Technology,
Yokohama, Japan

Y. Kamae
Graduate School of Life and Environmental Sciences,
University of Tsukuba, Tsukuba, Japan

1 Introduction

Climate sensitivity (CS), which is defined as the global mean surface air temperature response to a doubling of the atmospheric CO_2 concentration, is a crucial piece of information that informs the adaptation and mitigation policies for anthropogenic climate change. Despite the considerable efforts of climate scientists and technical advances, the ranges of the CS have not been narrowed (Knutti and Hegerl 2008). In the multi-model ensemble (MME) of general circulation models (GCMs) used for the Intergovernmental Panel on Climate Change Fourth Assessment Report, the range of CS was 2.1–4.4 °C (Randall et al. 2007). The variation in the CS in the MME is caused by the use of different model structures, i.e., different physical parameterization schemes and resolutions. Therefore, this is known as "structural uncertainty" (Murphy et al. 2004, 2007).

The "parametric uncertainty" is another substantial uncertainty. The present climate biases and future climate changes in a single model may be sensitive to changes in parameter values in the model physical schemes (Murphy et al. 2004). The Met Office Hadley Centre's project "Quantifying Uncertainty of Model Predictions" (QUMP) and http://climateprediction.net are the first and most comprehensive projects to investigate the parametric uncertainties of climate responses to external forcing (Murphy et al. 2004, 2007; Stainforth et al. 2005; Webb et al. 2006;

Collins et al. 2006, 2007, 2011; Brierley et al. 2010; Jackson et al. 2011). These investigators constructed several perturbed physics ensembles (PPEs) in which they swept uncertain parameters within the HadCM3 model (Gordon et al. 2000). The variation in the CS in their PPEs was comparable to or greater than the variation in the CS in the MME (Murphy et al. 2004; Stainforth et al. 2005; Collins et al. 2011).

A PPE that uses a different GCM, ECHAM5, also has a CS variation that is comparable to the MME (Klocke et al. 2011). However, these wide ranges in CS are not universal. The PPEs used versions of the CAM3 GCM (Jackson et al. 2008; Sanderson 2011) and the EGMAM GCM (Niehörster and Collins 2009) consistently yielded a CS of less than 3 °C, whereas the PPEs of MIROC3 produced a CS greater than 4 °C (Annan et al. 2005; Yokohata et al. 2010; Yoshimori et al. 2011). The distributions of CS in PPEs are contingent upon the model structures as well as the experimental design. Inter-comparison studies of multi-PPEs have only recently begun and can facilitate further understanding of the structural and parametric uncertainties of climate responses to external forcing (Yokohata et al. 2010; Sanderson 2011).

Although previous PPE approaches have been useful, they have limitations. Most of the previous PPE studies used atmosphere/slab-ocean (mixed layer ocean) GCMs (ASGCMs) rather than coupled atmosphere/full-ocean GCMs (CGCMs) (Murphy et al. 2004; Stainforth et al. 2005; Annan et al. 2005; Sanderson 2011; Klocke et al. 2011). One of the reasons for this is that the computational costs required to reach equilibrium for the ASGCMs are lower than those for the CGCMs. However, the climate feedback may differ between ASGCMs and CGCMs (Boer and Yu 2003a; Yokohata et al. 2008; Williams et al. 2008).

A few studies have performed CGCM PPEs to move beyond this limitation of PPEs in ASGCMs (Collins et al. 2006, 2007, 2011; Brierley et al. 2010; Jackson et al. 2011; Rowlands et al. 2012). However, another problem remains. Changes in values of atmosphere and surface parameters can lead to larger net radiation imbalance at the top of atmosphere (TOA) and the climate drifts. To prevent large climate drifts, most previous CGCM PPE studies have applied corrections for ocean surface heat and salinity fluxes (note that ASGCM PPEs also require flux corrections). However, flux corrections can affect the CS because of changes in the climatology of the sea surface temperature, cloud distribution, sea ice and other parameters in the control simulation.

Jackson et al. (2011) performed a QUMP CGCM PPE without flux corrections using atmosphere and surface parameter values in the members of the ASGCM PPE with a relatively small TOA imbalance. Because their approach was based on the existence of the ASGCM PPE, other modeling groups cannot easily apply it.

In this study, we developed a CGCM PPE without flux corrections. Our method utilizes a preliminary ensemble of atmospheric GCMs (AGCMs), which have much lower computational costs to reach equilibrium than ASGCMs or CGCMs. Therefore, this method would be useful for other modeling groups for performing CGCM PPEs without flux corrections.

Metrics, which are defined as measurements of the ability of the models to simulate the present climate, have been investigated for their use in constraining the uncertainty of future climate projections (Murphy et al. 2004; Piani et al. 2005; Hall and Qu 2006; Boe et al. 2009, 2010; Yokohata et al. 2010; Shiogama et al. 2011). Klocke et al. (2011) found metrics of specific cloud regions related to the variations in the CS in their PPE. We also sought, within our PPE, some metrics related to the variations in feedback.

This paper is organized as follows: Sect. 2 introduces the preliminary AGCM ensemble and the design of the CGCM PPE; Sect. 3 contains the results of the CGCM PPE; and Sect. 4 presents the discussion and conclusions.

2 Experimental design and AGCM experiments

2.1 Description of the standard model

We use the MIROC5 CGCM (Watanabe et al. 2010). The atmospheric component of MIROC5 used here has T42 horizontal resolution with 40 vertical levels, whereas the original version of MIROC5 has T85 resolution with 40 vertical levels. The ocean component model (Center for Climate System Research Ocean Component Model version 4.5; Hasumi 2006) has a horizontal resolution of approximately 1° and 49 vertical levels with an additional bottom boundary layer. This model requires no flux correction with the standard set of physics parameter values.

First, using the standard CGCM, we performed 20-year preindustrial control runs (C-CTL) after long spin-up runs (Watanabe et al. 2010) and 20-year abrupt $4\times CO_2$ concentration runs (C-CO_2); combinations of these runs are called Gregory-style experiments (Gregory et al. 2004). We computed the ordinal least-squares regression of the global averaged anomalies of longwave (LW) plus shortwave (SW) radiation fluxes at TOA on the global averaged annual mean anomalies of surface air temperature (ΔT). When $\Delta T \to 0$, the LW + SW intercept of the regression line indicates the stratosphere-troposphere adjusted radiative forcing (RF) for $4\times CO_2$. The slope of the regression line indicates the feedback parameter. Effective climate sensitivities for $4\times CO_2$ are estimated by the ΔT-intercept (as SW + LW $\to 0$). The CS for $2\times CO_2$ is half of the

value of the CS for $4\times CO_2$. The CS for $2\times CO_2$ of the standard model estimated using the Gregory method is 2.85 °C.

2.2 Selection of parameters that influence RF and feedback

As a next step, we chose the relatively long list of parameters that were expected to be the important determinants of CS. Previous studies of QUMP have shown that perturbations in the physics parameters of the atmosphere and surface components induced larger ranges in CS than the ocean component (e.g., Collins et al. 2011). Therefore, we selected physics parameters for the atmosphere and surface components of MIROC5 (Table 1) for a total of 20 continuous variables and 1 logical switch. The min–max range of each physics parameter value was also determined. These procedures were used in earlier PPE studies. However, it was also suggested that only a few parameters from the long list can dominate variations in the CS (Sanderson et al. 2008a) (note that these results may depend on the chosen parameters and/or ranges). Therefore, it is more efficient to select the most important parameters before performing the CGCM PPE.

To investigate how each physics parameter influences the feedback and RF of the MIROC5 model, we performed perturbed single-parameter ensembles of the AGCM. The following three types of AGCM runs were computed with the minimum and maximum values of each parameter:

A-CTL: AGCM forced by the monthly long-term averages of sea surface temperature (SST) and sea ice from the C-CTL run of the standard model and by pre-industrial CO_2 concentrations.

A-SST: AGCM forced by SST and sea ice (the average of last 10 years) from the C-CO_2 run (i.e., SST and sea ice in the warmer climate) of the standard model and by pre-industrial CO_2 concentrations.

A-CO2: AGCM forced by SST and sea ice from the C-CTL run of the standard model and by $4\times CO_2$ concentrations.

For each run, 6-year integrations were performed, whereas 20-year integrations were performed for the standard model. We considered the first year as the spin-up, and analyzed the climatology after the second year. Figure 1a, b show the changes in the net radiative fluxes at the TOA in the A-CO_2 and A-SST runs for each parameter (max minus min), which reveal how each parameter affected the RF and feedback (inflated by temperature changes). Generally, we selected parameters that have large effects on the RF and/or feedback (Fig. 1a, b; see the "Appendix"). We did not use the radiation imbalance at TOA (Fig. 1c) as the criterion for the parameter selection

because our new method of building the CGCM runs, as described below, did not require small radiation imbalances for each parameter. These selections reduced the list of parameters from 21 to 10 (Table 1; Fig. 1).

It should be noted that these analyses of the AGCM runs have the following limitations:

(i)   non-linear responses between different parameters cannot be investigated;

(ii) our analysis period was limited to 5 years because of limited computational resources; therefore, the influence of the natural variability may be significant;

(iii) because sea ice patterns are fixed in the AGCM runs, we could not correctly estimate the effects of changes in the sea ice parameters on the feedback; and

(iv) because SST is fixed in the AGCM runs, feedback loops between the SST patterns and radiative fluxes cannot occur.

Despite these limitations, the careful selection of a relatively short list of important parameters from the original long list that includes similar and insensitive parameters is worthwhile. When analyzing CGCM experiments to estimate the relative contributions of each parameter to the variance of CS, the necessary ensemble sizes of CGCM runs will be smaller if only 10 parameters. Furthermore, the exclusion of similar parameters can make it easier to interpret the effect of each parameter on the uncertainty of CS. We should note that the parameter sub-selection depends on "expert judgment" in choosing original parameters and/or limits. At present, there is no objective way to overcome the necessity of these subjective expert judgments.

2.3 Efficient sampling of the influential parameter combinations

We developed a methodology for choosing sets of parameter values for the CGCM runs. Because we chose to sweep multiple parameters simultaneously, it is necessary to sample the parameter space as efficiently as possible, given a finite number of runs. To allow a statistical investigation of the effects of each parameter, parameters with minimal covariance should be chosen. Latin hypercube sampling (LHS) is useful for fulfilling these requirements (McKay et al. 1979). When we perform the M member ensembles using the LHS method, we divide the range of each parameter into M strata of equal intervals $1/M$ and sample once from each stratum. The selected values of all 10 parameters are randomly paired to form the M 10-dimensional input vectors. Klocke et al. (2011) and Sanderson (2011) applied LHS to select parameter values for their ASGCM PPE with flux corrections. However,

Table 1 List of physics parameters that were varied in the AGCM runs

Name	Category	Description	Standard	Min	Max
wcbmax[a]	Cumulus	Maximum cumulus updraft velocity at cloud base (m/s)	1.7	0.7	2.8
precz0[a]	Cumulus	Base height for cumulus precipitation (m)	500	200	1,000
preczh[a]	Cumulus	Reference height for cumulus precipitation (m)	4,500	3,000	6,000
clmd[a]	Cumulus	Entrainment efficiency (ND)	0.51	0.4	0.6
meltau[a]	Cumulus	Timescale of ice melting (s)	10	1	15
evatau[a]	Cumulus	Timescale of liquid evaporation (s)	2	0.1	4
rcfact[b]	Cloud	Random overlapping factor in ice cloud falling (ND)	0.2	0	1
vicec[b]	Cloud	Factor for ice falling speed ($m^{0.474}$/s)	38	25	40
b1[c]	Cloud	Berry parameter (m^3/kg)	0.09	0.07	0.11
b2[c]	Cloud	Berry parameter (s)	0.095	0.07	0.12
faz1[d]	Turbulence	Factor for PBL overshooting (ND)	1.5	1	3
alp1[d]	Turbulence	Factor for length scale L_T (ND)	0.23	0.16	0.3
alp3[d]	Turbulence	Factor for length scale L_B (ND)	5	2	8
octei[d]	Turbulence	Switch for cloud top entrainment instability	OFF	ON	
tnuw[c]	Aerosol	Timescale for nucleation (s)	18,000	14,400	21,600
rcmax[c]	Aerosol	Maximum radius of cloud droplet (liquid, ice) (m)	30×10^{-6}, 185×10^{-6}	25×10^{-6}, 150×10^{-6}	35×10^{-6}, 200×10^{-6}
ucmin[c]	Aerosol	Minimum cloud droplet number (liquid) (m^{-3})	2.5×10^7	2.2×10^7	3.0×10^7
alb[e]	Surface	Albedo of ice and snow[f]	Medium	Low	High
talsnw[e]	Surface	Temperature thresholds for albedo function[g] (K)	268.15, 273.15	253.15, 271.15	258.15, 273.15
wsscl[e]	Surface	Lifetime of puddle over land ice (s)	216,000	108,000	432,000
tauage[e]	Surface	Snow aging time scale (s)	2×10^6	2×10^5	2×10^7

Variables in italics indicate that they were also swept in the CGCM runs

[a] Chikira and Sugiyama (2010)

[b] Wilson and Ballard (1999)

[c] Takemura et al. (2005, 2009)

[d] Nakanishi and Niino (2004)

[e] Takata et al. (2003) and Watanabe et al. (2010)

[f] "alb" indicates a collection of 8 parameters corresponding to albedo for ice and snow over sea and land

[g] Because the standard values are the maximum, we performed AGCM runs with two lower value sets

LHS does not constrain the net radiative imbalance at the TOA. Therefore, LHS leads to large climate drifts in the C-CTL runs of CGCMs without flux corrections. To reduce the radiative imbalance at the TOA and prevent large climate drifts, we developed a new method called suppressed imbalance sampling (SIS):

(a) As described above, we performed A-CTL runs with the minimum or maximum values for each physics parameter, and we estimated the changes in the net radiation imbalance at the TOA (Fig. 1c).

(b) We used LHS to generate large potential sets of parameter values (5,000 samples), in which the first sample is set equal to the parameter values of the standard model.

(c) We emulated the radiative imbalance at the TOA for each sample by applying piecewise linear interpolations of the changes in the imbalance of A-CTL runs (Fig. 1c).

(d) We selected the sample with the lowest amplitude of anomalies in imbalance relative to the standard model as the combination of the parameter values to be used in the CGCM ensemble (thus, the first-selected sample must be the standard model).

(e) From the initial 5,000 sets of potential parameters, we removed the selected sample described in (d), and we also deleted all samples with parameter values that were "very close" (defined below) to the selected sample.

(f) We repeated steps (d) and (e) to choose N subsets.

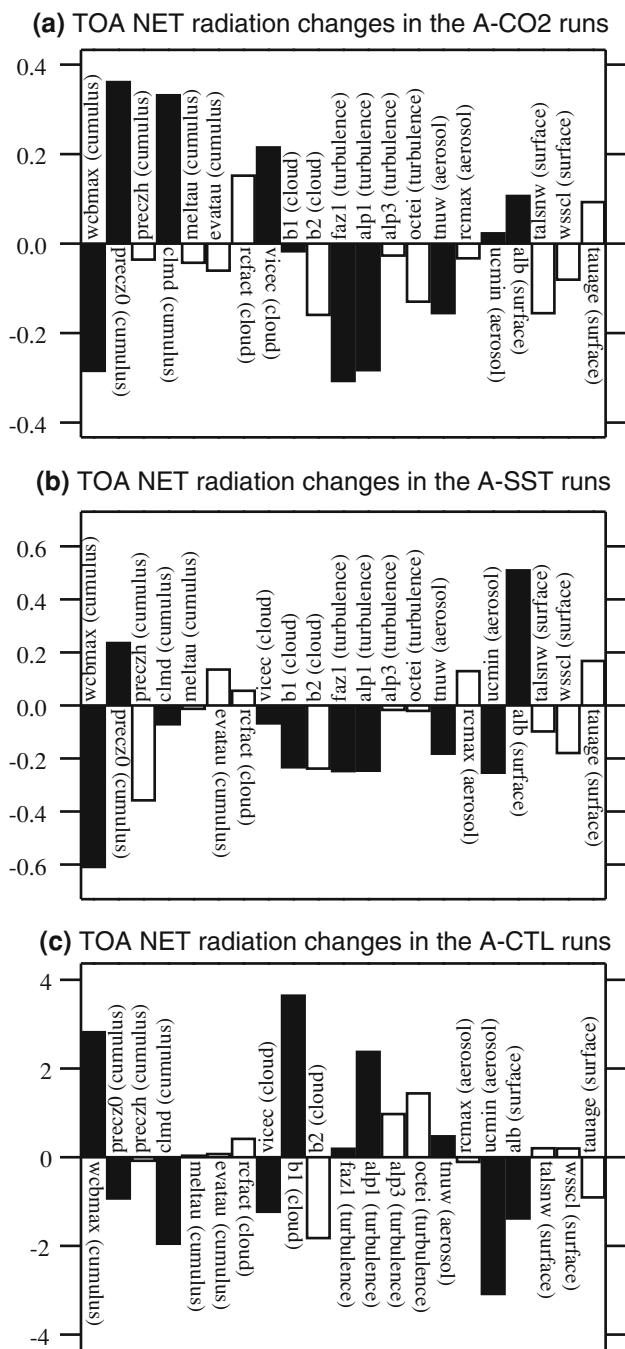Here, N = 100, but this algorithm has the flexibility to change N depending on the available computational resources.

(a) TOA NET radiation changes in the A-CO2 runs

(b) TOA NET radiation changes in the A-SST runs

(c) TOA NET radiation changes in the A-CTL runs

Fig. 1 Differences in net radiative flux at TOA (W/m^2) between the AGCM runs with maximum and minimum values (max. minus min.; "on" minus "off" in the *octei* case) of each physics parameter in **a** A-SST runs, **b** A-CO$_2$ runs and **c** A-CTL runs. *Black bars* indicate that these parameters were selected for perturbation in the CGCM runs

In QUMP, more elaborate emulation techniques were applied to *predict* various metrics to select parameter values in their experiments (Webb et al. 2006; Rougier and Sexton 2007; Rougier et al. 2009). However, recent studies found that it is difficult to define metrics well correlating with the variations of climate projections in advance

(Shiogama et al. 2011; Abe et al. 2009, 2011; Knutti 2010). Therefore, we included only the emulated TOA imbalance that is necessary to conduct simulations without flux corrections as the *prior* metric in our experimental design. We investigate metrics correlating with the variations of feedback in Sect. 3.4. The possible effects of the TOA imbalance constraint on the distribution of the CS are discussed in Sect. 4.

Figure 2 shows the selected parameter values in LHS and SIS. It is important that the sampled values in SIS are not concentrated in small subspaces. Figure 3a shows the Euclidean distance of two different samples in the normalized 10-dimensional parameter space. We defined "very close" samples in the above step (e) as those that fell in the lowest 4 % of the probability distribution of parameter differences. It is clear that, compared with LHS, the SIS method sampled the parameter space relatively evenly.

Figure 3b shows the emulated changes in the radiative imbalance at the TOA. These changes are large in LHS but sufficiently small (less than 1 W/m^2) in SIS. Although the suppression of changes in the TOA imbalance resulted in correlations between different parameters, the effect was not large (the maximum of the absolute correlation values is approximately 0.3) (Fig. 3c). In this algorithm, the definition of a "very close" distance is critical for determining the amplitude of the emulated TOA imbalance and the correlation between parameters. There is a trade-off between the parameter correlation and the amplitude of the TOA imbalance. Here, we chose the definition of a "very close" distance so that the amplitude of the emulated TOA imbalance and the correlation are both sufficiently small.

It should be noted that the SIS method also has the limitations (i)–(iv) mentioned in Sect. 2.2. Therefore, to confirm that this method works well, it was necessary to actually perform the CGCM runs with the selected parameter values.

3 CGCM experiments

3.1 CGCM control experiments

Here, we present the results from ongoing C-CTL and C-CO$_2$ runs of the CGCM without flux corrections. We have completed 35 members of the C-CTL and C-CO$_2$ ensembles so far that cover wide ranges of parameter values (Fig. 2).

In the C-CTL runs, all of the members showed only small changes in radiative imbalance and little drift in surface air temperature (Fig. 4), which demonstrates that our SIS method works well to prevent large drift and avert long spin-up runs. This method allows modeling groups to overcome the limitations of previous PPE studies, i.e., the requirements of the ASGCM and flux corrections. We defined the
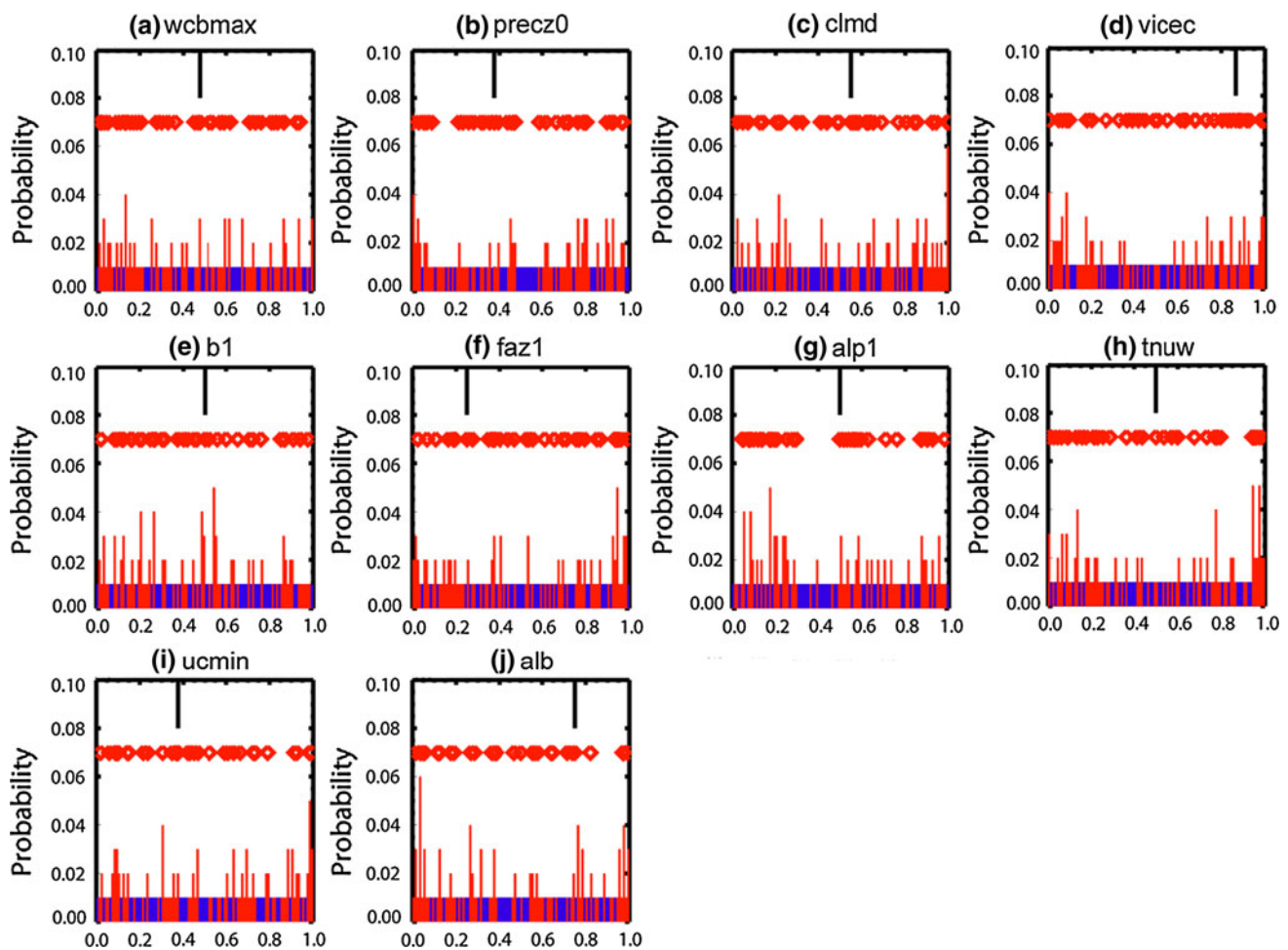
Fig. 2 Histograms of selected parameter values for each parameter in the Latin hypercube sampling (*blue*) and suppressed imbalance sampling (*red*). *Horizontal axes* indicate normalized parameter values (0 is minimum and 1 is maximum). *Vertical axes* show probabilities, which are sampled in 0.01-width bins. *Black vertical lines* indicate the parameter values of the standard model. *Red diamonds* indicate the parameter values of 35 CGCM runs
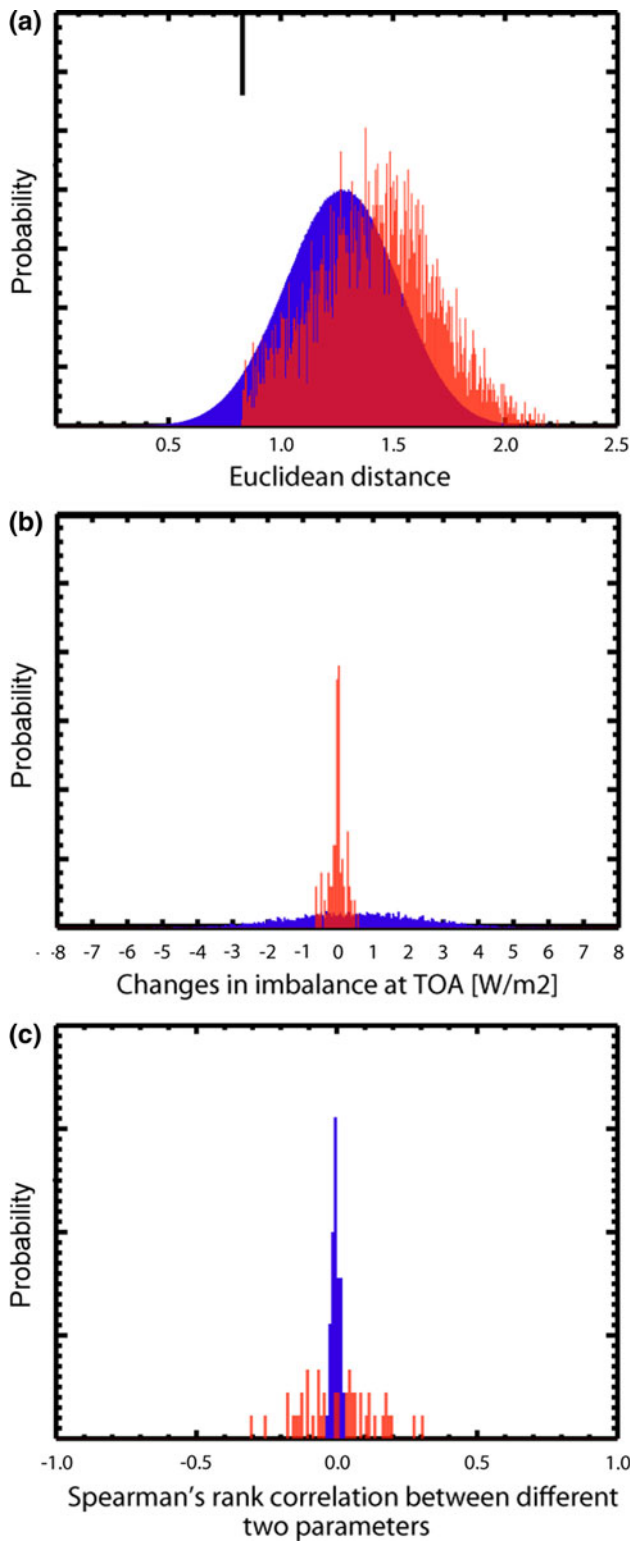
first 10-year period as spin-up and years 11–30 as the C-CTL runs. The 20-year C-CO$_2$ runs were performed beginning with year 11 of the corresponding spin-up/C-CTL runs. The 10-year length of the spin-up was chosen because of the limited computational resources. However, the ocean time scales of equilibration with the altered parameters could be centuries. Not surprisingly, some models have slight drifts. Hereafter, we removed the annual mean values of the C-CTL runs from the values of the C-CO$_2$ runs to reduce the effects of the slight drifts. In other words, we assumed that the drift involved in the C-CO$_2$ run is the same as that in the CTL run for each model.

3.2 Climate sensitivity, feedback and radiative forcing

Figure 5 shows the histogram of effective climate sensitivity estimated by using the Gregory method. The range of CS in our ensemble is 2.2–3.2 °C, which occupies the lower range of the MME but not the upper range. Here, we accounted for uncertainty arising from the natural variability in the estimation of CS of the standard model by applying the bootstrap method (Efron 1979; Wilks 1995). We implemented the bootstrap method as follows:

– We computed a linear regression between changes in the annual mean global average surface air temperature and changes in the annual mean global average radiative flux at the TOA (C-CO$_2$ minus C-CTL).
– We randomly resampled 10 anomalies (each 2 years in length with considering the persistence in a first-order auto-regression; Wilks 1995) from the original regression line with replacement.
– We added the randomly resampled anomalies to the original regression line and re-computed the new regression (called the bootstrap sample).
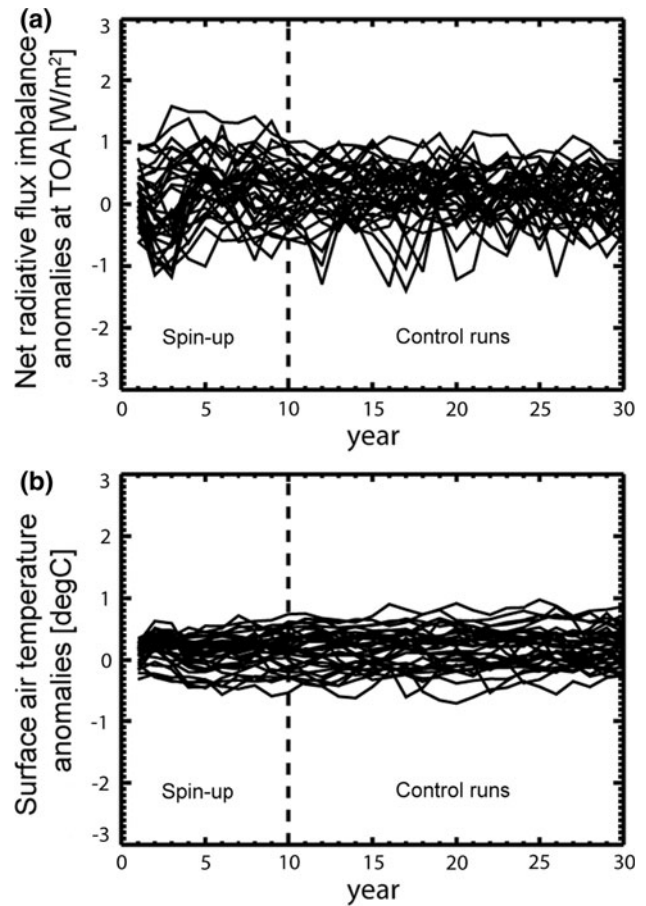– We repeated the above two steps to generate 1,000 bootstrap samples.

Fig. 4 **a** Annual mean time series of global averaged net radiative flux imbalance anomalies at the TOA (W/m^2) from the standard model. The first 10 years are spin-up, and the following 20 years are the control runs. **b** Annual averaged time series of global mean surface air temperature anomalies from the standard model (°C)

– We measured uncertainties using the probability of the bootstrap samples.

The range of the CS in this PPE was greater than that generated by the natural variability alone.

Figure 6 shows a scatter plot of the RF and feedback parameters. There is a clear anti-correlation (−0.71) between the RF and feedback in this ensemble. This anti-correlation is one of the reasons why the range of the CS is relatively narrow. The range of the CS would be larger with the same range of feedback parameters but with a positive RF-feedback correlation (see the blue contours of Fig. 6). For each parameter set, we performed only one run with perturbed initial conditions; therefore, the influence of the natural variability may be important. By defining $\bar{x}$, $\bar{y}$, λ and F as the 20-year mean of ΔT, the 20-year mean of

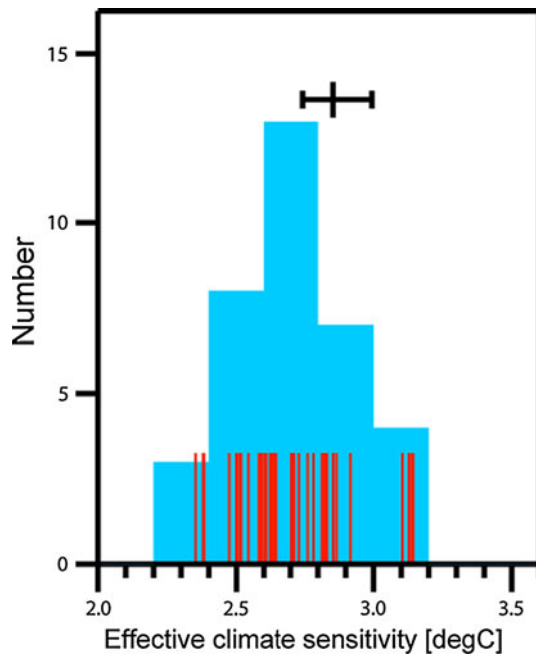**Fig. 5** Histogram of effective climate sensitivity for a doubling of CO_2 in the CGCM ensemble (°C). *Vertical red lines* indicate each CGCM run. The *black error bar* indicates the 10–90 % range and the best estimate of the standard model (estimated using the bootstrap method)
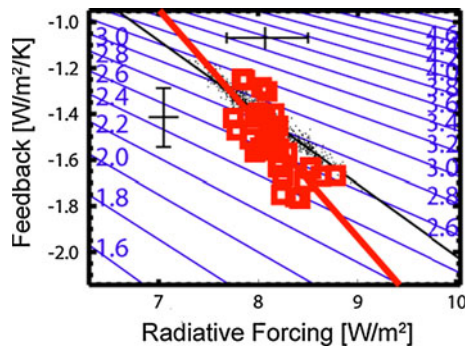

Fig. 6 The *red squares* represent a scatter plot of radiative forcing (W/m^2; *horizontal axis*) and feedback parameters (W/m^2/K; *vertical axis*); the *red line* represents their total least-squares regression. *Blue contours* indicate the effective climate sensitivity for a doubling of CO_2 (°C). *Black dots* are the bootstrap samples of the standard model; the *black line* is their total least squares regression. The *black error bars* are the 10–90 % ranges and the best estimates of RF and feedback in the standard model

SW + LW, the feedback parameter and RF, respectively, we can state that $F = \bar{y} - \lambda\bar{x}$ in the Gregory method. The natural variability is expected to have only a small effect on $\bar{x}$ and $\bar{y}$ because they are computed as 20-year averages and because the global average climate responses to the $4\times$ CO_2 concentration are greater than the natural variability. In contrast, the natural variability can significantly affect λ and F. The fluctuations of λ and F due to the natural variability (λ_u and F_u) tend to be anti-correlated,
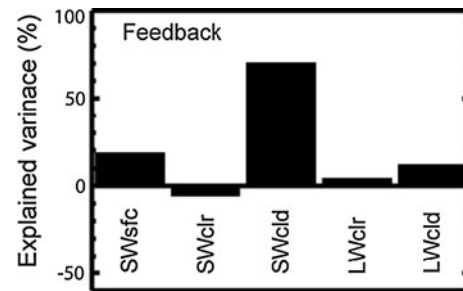

Fig. 7 The contributions of SWsfc, SWclr, SWcld, LWclr and LWcld to the total feedback variance between the CGCM runs (%)

i.e., $F_u = -\lambda_u\bar{x}$ ($\bar{x}$ is positive). Therefore, the natural variability also induces some anti-correlations between RF and feedback (the black dots and the regression line of Fig. 6), as has been reported previously (Webb et al. 2012). However, we found that the slope of the ensemble members (the red regression line) differed from that of the bootstrap samples (the black regression line), which suggests that this anti-correlation of ensemble members is not caused only by the natural variability. It is difficult to isolate the effect of natural variability and the relationship between RF and feedback. We found that the models with higher amplitudes of the El Niño-like warming pattern in the feedback process tend to have greater magnitudes of rapid La Niña-like warming in the adjustment process (not shown). These opposing responses of the tropical SST may result in the anti-correlation between RF and feedback because they induce dynamical changes in cloud cover in directions opposite to each other. However, it is not clear yet whether these tropical SST opposing responses are artifacts that result from the natural variability.

3.3 SW cloud feedback

The total feedback was decomposed into components of surface SW (SWsfc), clear-sky SW (SWclr), cloud-sky SW (SWcld), clear-sky LW (LWclr) and cloud-sky LW (LWcld). The SW components were estimated by applying the approximate partial radiative perturbation (APRP) method (Taylor et al. 2007; Yokohata et al. 2008). The APRP method provides an estimate of the SW radiative perturbation by using monthly mean model outputs. Biases in the estimates of the SWcld feedback arising from the cloud masking that occur in the conventional cloud radiative forcing method (Soden et al. 2004) do not appear in the APRP method. For LWclr and LWcld, the conventional cloud radiative forcing method (Cess et al. 1990) was used. We computed the contributions of each component to the total feedback variance by applying the following method (Boer and Yu 2003b; Webb et al. 2006; Williams and Webb 2009; Yokohata et al. 2010; Yoshimori et al. 2011):
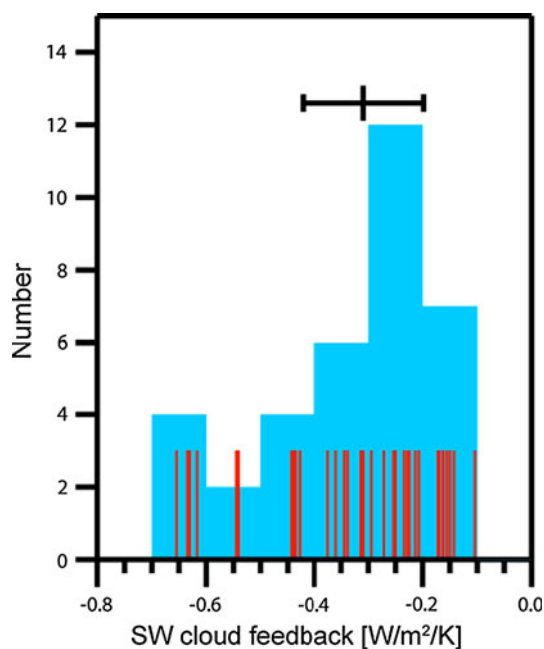
Fig. 8 Histogram of SWcld feedbacks (W/m²/K). *Vertical red lines* indicate each CGCM run. The *black error bar* shows the 10–90 % range and the best estimate of the standard model, estimated using the bootstrap method

– We defined $\vec{Y}$ as the vector of total feedback for all ensemble members and $\vec{X}_i$ as each component (i = SWsfc, SWclr, SWcld, LWclr, LWcld), where $\vec{Y} = \Sigma \vec{X}_i$.

– We removed the ensemble averages from $\vec{Y}$ and $\vec{X}_i$.

– We computed a scalar $\vec{Y}^t \vec{X}_i / \vec{Y}^t \vec{Y}$, which indicates the fraction of the total variance explained by each component.

The largest fraction of the total variance in feedback is explained by large differences between the ensemble members in SWcld (Fig. 7).

Figure 8 shows the histogram of the SWcld feedback across all ensemble members. It is interesting that all the ensemble members had negative SWcld feedback, resulting in the low climate sensitivities of this ensemble. The middle-level cloud albedo increased and resulted in negative SWcld feedback values (not shown). To investigate the mechanism of different SWcld feedback among the models, we analyzed the differences between the models with the 10 lowest (more negative) and the 10 highest (less negative) values for the global mean SWcld feedback (i.e., we subtracted the 10 highest feedback values from the 10 lowest values). The differences in the local SWcld feedback appear mainly in the tropical oceans (Fig. 9a). Here we broke down the global mean feedback to the local feedback by regressing the local radiation (and other
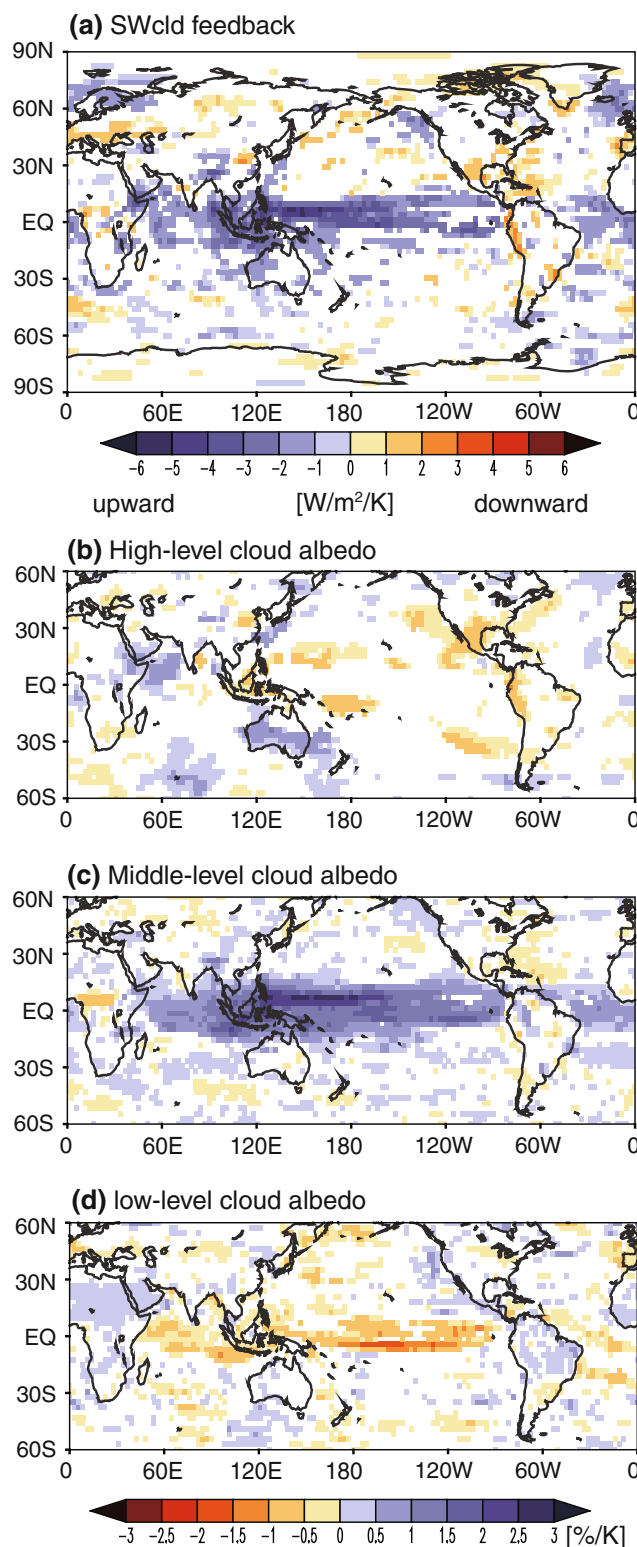

Fig. 9 a Local differences in SWcld feedbacks (W/m²/K) between models with the 10 lowest (more negative) and 10 largest (less negative) global mean SWcld feedback parameters. Differences in cloud albedo feedbacks with cloud top height of **b** high, **c** middle and **d** low (%/K). We show significant differences based on ±10% levels of t tests
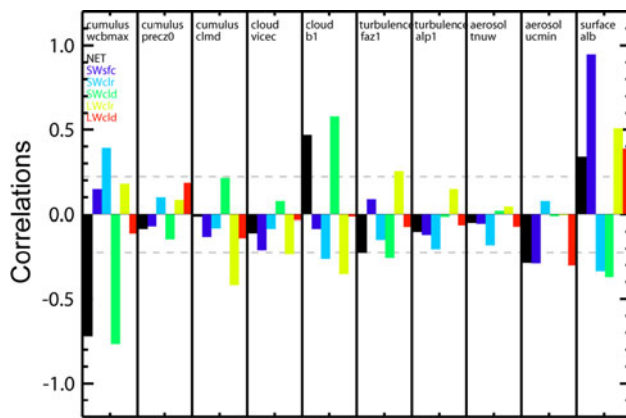
Fig. 10 Correlations between feedback parameters and the values of physics parameters. *Dashed lines* indicate correlations that are significant at the 10 % level in a *t* test

variables) on the global averaged ΔT (Boer and Yu 2003b; Webb et al. 2006).

We compared the spatial pattern of the SWcld feedback with the patterns of the cloud albedo changes (Fig. 9b–d). Here, the cloud albedo was computed with outputs from the online observation simulator of the International Satellite Cloud Climatology Project (ISCCP; Rossow and Schiffer 1999), which mimics the satellite view from space (along with certain ISCCP retrieval assumptions: Klein and Jakob 1999; Webb et al. 2001). The total cloud albedo was decomposed into contributions from clouds with three different levels of cloud-top pressure: low level (1,000–680 hPa), middle level (680–440 hPa) and high level (440–50 hPa) (Rossow and Schiffer 1999). The more negative SWcld feedback values in our simulations are mainly attributable to greater increases in the middle-level cloud albedo, which are partly offset by decreases in the low-level cloud albedo.

Previous studies have shown that a few parameters dominate variations in feedback (Sanderson et al. 2008a, b, 2010; Rougier et al. 2009). To investigate the parameters that result in changes in feedback, we applied single linear correlation analyses between each parameter and each component of feedback (Fig. 10). Models with higher values of the cumulus parameter *wcbmax* have more negative SWcld feedback. The second most important contributor to variations in SWcld feedback is the cloud parameter *b1*. The surface parameter *alb* (albedo values of snow and sea ice) has a strong effect on the SWsfc feedback (surface albedo feedback). It should be noted that multiple regression analyses or non-linear analyses, which require larger ensemble sizes, are necessary to accurately compare the contributions of each parameter to the variations of feedback. Instead we confirmed that the differences in cloud albedo feedback between the AGCM runs with the maximum and minimum values of *wcbmax* were similar to
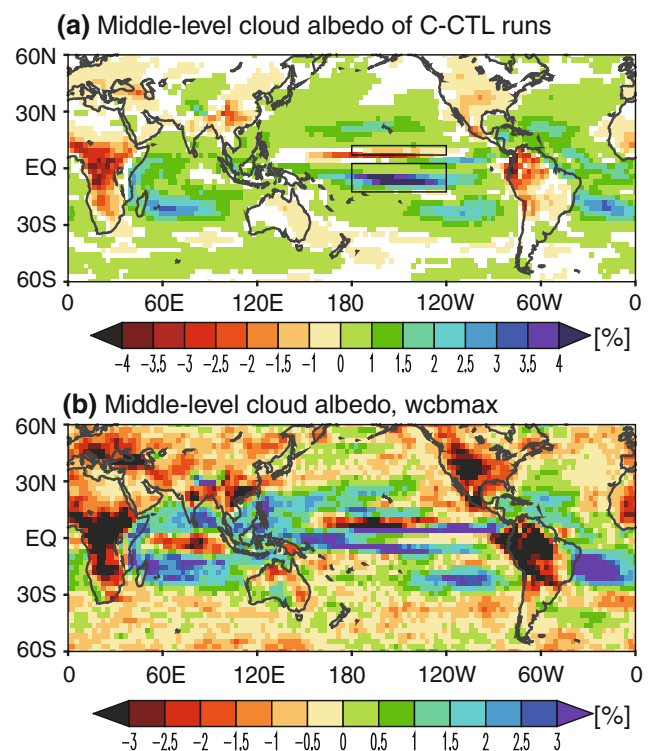

Fig. 11 **a** Differences in middle-level cloud albedo (%) between the C-CTL runs with the 10 lowest (more negative) and 10 largest (less negative) global mean SWcld feedback parameters. These colored differences are significant at the 10 % level in a *t* test. *Black boxes* indicate regions where the Normalized Cloud Index is defined (see text). **b** Differences in middle-level cloud albedo (%) between the A-CTL runs with maximum and minimum *wcbmax* parameter values

the patterns presented in Fig. 9b–d (not shown). This result justifies our conclusion, based on the simple composite and linear correlation analyses, that the single parameter *wcbmax* was the main driver of the variations of the SWcld feedback.

In this section, we describe our findings that changes in the middle-level clouds result in the greatest spread in the SWcld and total feedback. This finding is very different from those of previous studies, where differences in the low- and/or high-level cloud feedback mainly resulted in the variations of the CS in MMEs and PPEs (Bony and Dufresne 2005; Webb et al. 2006; Medeiros et al. 2008; Williams and Webb 2009; Yokohata et al. 2010; Watanabe et al. 2011c, d). We found that the perturbations in the single parameter of the newly developed cumulus scheme (Chikira and Sugiyama 2010) drove the significant sensitivity of the middle-level cloud albedo feedback. This parameterization is characterized by a vertically variable entrainment rate that depends on the surrounding environment and for which the population of the middle-level cumulus congestus is greater than that in the Arakawa-Schuber scheme (Chikira 2010). The implementation of
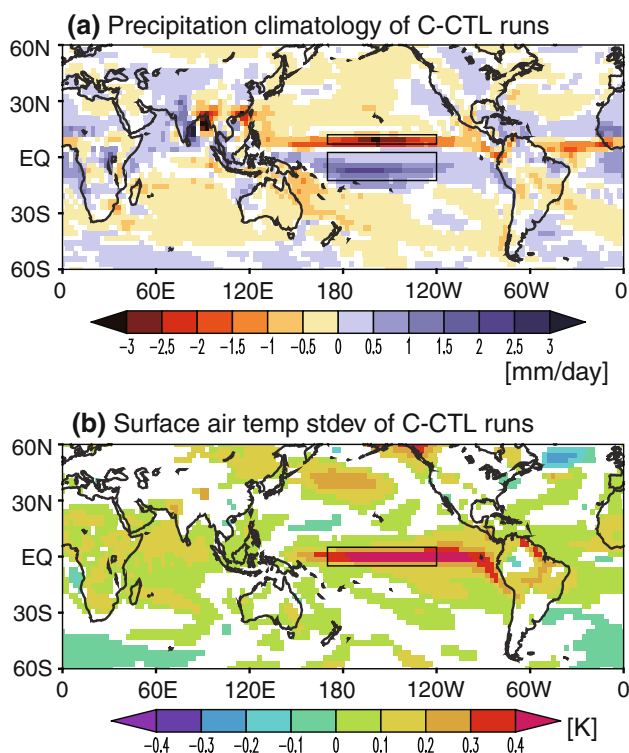
(a) Precipitation climatology of C-CTL runs

[mm/day]

(b) Surface air temp stdev of C-CTL runs

[K]

Fig. 12 **a** Differences in precipitation (mm/day) between the C-CTL runs with the 10 lowest (more negative) and 10 largest (less negative) global mean SWcld feedback parameters. *Black boxes* represent regions where the Normalized Precipitation Index is defined (see text). **b** Differences in the standard deviation of the annual mean surface air temperature (K) between the C-CTL runs with the 10 lowest (more negative) and 10 largest (less negative) global mean SWcld feedback parameters. The *black box* indicates the Nino 3.4 region. Both *panels* show differences that are significant at the 10 % level in a *t* test

this new parameterization scheme results in the middle-level cloud having a greater impact on the variations in CS in the MIROC5 model than in the other GCMs.

3.4 Metrics related to the SW cloud feedback

Here, we sought metrics that are related to the variations in SWcld feedback to provide observational constraints within our ensemble. Figure 11a shows the differences in middle-level cloud albedo between the C-CTL runs of models with the 10 lowest (more negative) and the 10 highest (less negative) SWcld feedback values. Models with greater middle-level cloud albedo in the tropical and subtropical oceans tend to have more negative SWcld feedback. However, this trend does not hold in the Pacific Intertropical Convergence Zone (ITCZ), where the cloud albedo is low. The lower albedo of the middle-level clouds was also observed over land surfaces. The differences in the middle-level cloud albedo between the A-CTL runs with the maximum and minimum *wcbmax* values (Fig. 11b) are

similar to those shown in Fig. 11a, which provides collateral evidence that changes in the cumulus parameter *wcb-max* dominate the variations in SWcld feedback.

In the C-CTL runs, middle-level cloud albedo differences (the black boxes in Fig. 11a) accompany significantly less precipitation over the Pacific ITCZ and significantly more precipitation over the southern part of the Pacific ITCZ (Fig. 12a). The cumulus scheme of MI-ROC5 is able to realistically represent the population of middle-level cumulus congestus (Chikira and Sugiyama 2010; Chikira 2010). Larger increases in the middle-level cloud albedo in the C-CO_2 runs are related to more cumulus congestus over the southern area of the Pacific ITCZ and less cumulus congestus over the ITCZ in the C-CTL runs. A wetter atmosphere over the southern part of the Pacific ITCZ increases precipitation anomalies related to the El Niño-Southern Oscillation (ENSO) (i.e., stronger Bjerknes feedback; Watanabe et al. 2011a), which results in higher amplitudes of the ENSO (Fig. 12b).

Here, we define the metric that measures the meridional contrast of middle-level cloud albedo in the C-CTL runs according to Watanabe et al. (2011a):

- A indicates the middle-level cloud albedo averaged over the area of the north black box in Fig. 11a (120°W, 180°W, 7°N, 12°N).
- B is the average over the area of the south black box (120°W, 180°W, 12.5°S, 2.5°S).
- The Normalized Cloud Index (NCI) is defined as $(B - A)/(B + A)$.

The Normalized Precipitation Index (NPI), which measures the meridional contrast of precipitation, is defined similarly to NCI. The standard deviation of the annual mean SST averaged over the Nino 3.4 region (170°E, 120°W, 5°S, 5°N) is considered to be the metric of the ENSO amplitude. These metrics, in the C-CTL runs, are negatively correlated with the SWcld feedback in the C-CO_2 runs, which is expected given their definitions (Fig. 13). For example, the models with higher ENSO amplitudes in the C-CTL runs had more negative SWcld feedback values. We confirmed the robustness of these results by expanding the length of three C-CTL runs of the standard model as well as that of models with minimum and maximum ENSO amplitudes from 20 years to 100 years (the triangles of Fig. 13). We compared the metrics of each model to the metrics from the observational datasets of Rossow and Schiffer (1999) (ISCCP cloud albedo), Xie and Arkin (1997) (precipitation) and the Nino 3.4 index of the Climate Prediction Center of the National Oceanic and Atmospheric Administration. The values of these metrics obtained from observational data (red lines in Fig. 13) are close to the values from the standard model (red squares), which implies that the SWcld feedback of
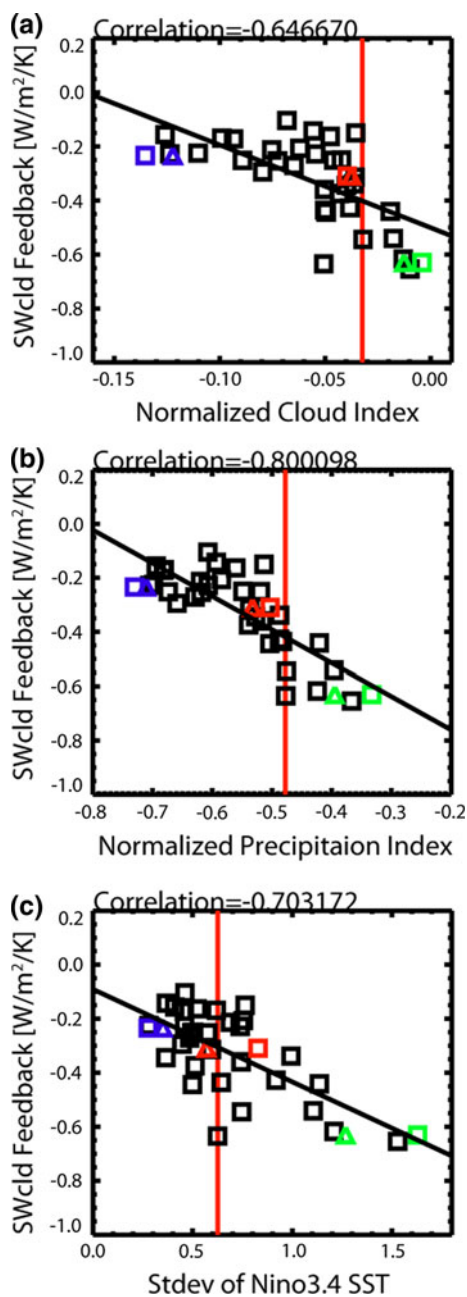
Fig. 13 Scatter plots (*squares*) of SWcld feedbacks (W/m²/K) in the C-CO₂ runs and **a** Normalized Cloud Index (no dimension), **b** Normalized Precipitation Index (no dimension) and **c** the standard deviation of annual mean Nino3.4 SST (°C) in the C-CTL runs. *Black lines* are the ordinal least square regressions. *Red lines* represent observations. *Red squares* are output from the standard model; *blue* and *green squares* are outputs from models with the minimum and maximum amplitude of ENSO, respectively. *Red*, *blue* and *green triangles* represent 100-year C-CTL runs of the standard model and the models with the minimum and maximum amplitude of ENSO, respectively

the standard model is more reliable than that of the other models in this ensemble. However, non-negligible uncertainty remains, which is denoted by the spreads of the

anomalies from the regression lines, i.e., other models perform comparably to the standard model with respect to at least these three metrics. To determine the likelihood of model's feedback value, an analysis of many more climate variables is required.

It should be noted that the use of ISCCP data for the evaluation of the middle-level cloud may be problematic. It has been reported that the ISCCP observational data have greater middle-level cloud fractions than the MODIS measurements, and it is likely that the ISCCP data overestimate the fraction of middle-level clouds (Pincus et al. 2012, and references therein). Furthermore, satellite measurement simulators do not completely mimic the satellite instruments (Pincus et al. 2012). To estimate the effects of the observational uncertainty on the model performance metrics, it is better to compare the metric based on ISCCP simulator data and that from other satellite measurements simulators.

It should be noted that we evaluated only the performance of each PPE member of our single model, and we do *not* claim that the three metrics defined here can necessarily be adapted to other GCMs. Within this PPE, models with a higher ENSO amplitude in the C-CTL runs had more negative SWcld feedback values. Toniazzo et al. (2008) also found a negative correlation between the ENSO amplitudes and CS within the flux-corrected QUMP ensemble. By contrast, we did not identify any significant correlations between the ENSO amplitudes and the CS within the MME contributing to Coupled Model Intercomparison Project Phase 5 (not shown). ENSO amplitudes are determined by several processes (e.g., Guilyardi et al. 2009; Collins et al. 2010; Watanabe et al. 2011a), and processes that are related to the uncertainty of the CS may differ among GCMs. Klocke et al. (2011) identified a performance metric of specific cloud regions relating to climate sensitivity within the ECHAM5 PPE, but that relationship did not carry into the MME. These analyses of the similarities and differences of MMEs and PPEs will promote a further understanding of the structural and parametric uncertainties of climate responses to external forcing.

4 Summary and discussion

The PPE approach facilitates a greater understanding of the abilities and limitations of a particular climate model structure as well as the systematic exploration of the uncertainties in processes and feedbacks, apart from the limitation that MMEs are not designed to sample modeling uncertainty in a systematic fashion (Allen and Ingram 2002; Murphy et al. 2004, 2007; Knutti 2010). Recent comparisons of two PPEs revealed significant differences between them (Yokohata et al. 2010; Sanderson 2011). It would
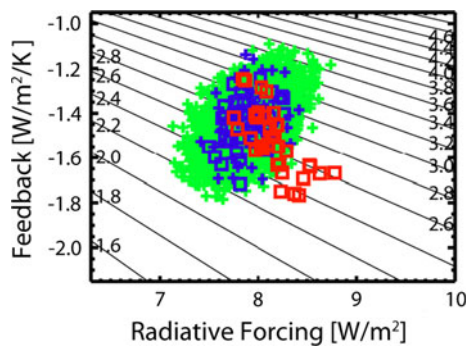
Fig. 14 *Red squares* represent a scatter plot of radiative forcing (W/m^2; *horizontal axis*) and feedback parameters (W/m^2/K; *vertical axis*), which are the same as in Fig. 6. *Green crosses* indicate emulations of LHS members. *Blue squares* indicate emulations of SIS members that have been performed, and *blue crosses* indicate the remaining SIS members. *Black contours* show the effective climate sensitivity for a doubling of CO_2 (°C)

therefore be desirable for modeling groups to perform PPEs of their own GCMs and compare them to identify and understand the processes driving the structural and parametric uncertainty of climate change projections.

Previous studies of PPEs mainly used ASGCMs and flux corrections, which can significantly affect the climate biases and projections. Previously, no methodology has been useful across modeling groups to perform PPE with CGCM and without flux corrections. The greatest challenge is a considerable TOA imbalance, which leads to large climate drifts. In this study, we developed a method to constrain the TOA imbalance in the CGCM PPE without flux corrections. Although this method requires AGCM control runs, such runs did not greatly increase the difficulty of the procedure or the required computational resources. To sweep M parameters, it is necessary to compute only $2M + 1$ short control runs of AGCM, i.e., one run of the standard model plus maximum and minimum value runs for each parameter. According to the results from this AGCM ensemble, our SIS method provides parameter sets with a low TOA imbalance. One critical point for PPEs is that hardly any of those models would be ever selected for standard climate integration because they are far from balanced. We succeeded in building a PPE without this limitation (other metrics might also not be passed). We hope that this method can help other modeling groups to perform CGCM PPEs and thus enables comparisons of multi-PPEs referred to as "super-ensembles" by Murphy et al. (2004).

The range of CS in the MIROC5 PPE was 2.2–3.2 °C, and one may wonder whether constraining the TOA imbalance narrowed this CS range. Figure 14 shows the scatter plots of RF and feedback in the CGCM PPE, as well as the emulations of SIS and LHS estimated by applying piecewise linear interpolations of the A-CO_2 and A-SST runs. Constraining the TOA imbalance does not greatly reduce the range of CS in SIS compared with LHS, at least in these linear emulations.

It is notable that the anti-correlation between RF and feedback was not found in the emulations of LHS and SIS (Fig. 14). In fact, weak positive correlations were present, as could be expected from the RF and feedback responses of the AGCM runs (i.e., they responded in the same direction for most parameters) (Fig. 1a, b). Therefore, the anti-correlation in the CGCM PPE is likely caused by an air–sea interaction or the effect of the natural variability.

Variations in the SWcld feedback dominate the uncertainty of the total feedback. Increases in the middle-level cloud albedo result in more negative SWcld feedback. This finding differs from those of previous studies of cloud feedback uncertainty that highlighted the importance of variations of low- and high-level cloud feedback (Hartmann and Larson 2002; Bony and Dufresne 2005; Webb et al. 2006; Medeiros et al. 2008; Williams and Webb 2009; Yokohata et al. 2010; Zelinka and Hartmann 2010; Watanabe et al. 2011c, d). Because the cumulus parameter correlates well with the SWcld feedback, it seems that the newly developed cumulus scheme implemented in MIROC5 (Chikira and Sugiyama 2010; Chikira 2010) causes the significant responses of middle-level cloud albedo. Within the MIROC5 PPE, the performance metrics of the middle-level cloud albedo pattern, precipitation pattern and ENSO amplitude are related to the SWcld feedback values. Without the aforementioned analyses, one may expect these metrics to be independent even though they are physically related to each other. When different metrics are combined with observational constraints (Knutti et al. 2002; Forest et al. 2002; Annan and Hargreaves 2006), a careful consideration of independence is necessary.

Although the CS of MIROC5 PPE is low (2.2–3.2 °C), the CS within the PPE of MIROC3 ASGCM (an old version of MIROC) is high (>4 °C) (Annan et al. 2005). Because many of the physical schemes in MIROC3 and MIROC5 differ (Watanabe et al. 2010), it is not clear which schemes are largely responsible for determining the structural dependency. The differences in the spatial resolution (and the time step) may also affect the CS of the standard models (Seiffert and von Storch 2008) and the ranges of the CS in the PPEs. Sensitivity experiments in which single or multiple schemes are transferred between the old and new models (with the same spatial resolution and the same time step) would provide insight into the effects of each scheme and the interactions between different schemes (Gettelman et al. 2012). This multi-physics ensemble approach comparing MIROC3 and MIROC5 is reported by Watanabe et al. (2011b).

Appendix: The selection of the parameters involved in the CGCM ensemble

In the cumulus scheme, *wcbmax*, *precz0* and *clmd* induce large changes in RF, and *wcbmax* also affects the feedback (Fig. 1a, b). Therefore, these three parameters were selected to be swept in the CGCM runs. Although *preczh* is important for the feedback, we removed it because a similar parameter, *precz0* had already been selected (both *precz0* and *preczh* were included in the cumulus precipitation process). *meltau* and *evatau* were also omitted from the cumulus parameters because they are not important determinants of either RF or feedback (even though the experts expected significant influence). In the cloud scheme, *vicec* and *b1* were selected. Although *b2* is also an important parameter, we did not select it. This is because *b1* and *b2* are included in the numerator and denominator, respectively, of the equation describing the conversion rate of cloud droplets to raindrops in the water cloud scheme (Takemura et al. 2005; Shiogama et al. 2010). Because it is easier to understand model behaviors when we swept the *b1* values in the numerator rather than the *b2* values in the denominator, we usually selected *b1* rather than *b2* to vary in the CGCM runs. Among the turbulence parameters, *faz1* and *alp1* dominated the changes in RF and feedback. Among the three aerosol parameters, *rcmax* was removed because it had a smaller effect than the other two parameters (*tnuw* and *ucmin*). Among the surface parameters, *alb* was selected because of its large effect on feedback.

References

Abe M, Shiogama H, Hargreaves JC, Annan JD, Nozawa T, Emori S (2009) Correlation between inter-model similarities in spatial pattern for present and projected future mean climate. Sola 5:133–136

Abe M, Shiogama H, Nozawa T, Emori S (2011) Estimation of future surface temperature changes constrained using the future-present correlated modes in inter-model variability of CMIP3 multi-model simulations. J Geophys Res. doi:10.1029/2010JD015111

Allen MR, Ingram WJ (2002) Constraints on future changes in climate and the hydrologic cycle. Nature 419:224–232

Annan JD, Hargreaves JC (2006) Using multiple observationally-based constraints to estimate climate sensitivity. Geophys Res Lett 33:L06704

Annan JD, Hargreaves JC, Ohgaito R, Abe-Ouch A, Emori S (2005) Efficiently constraining climate sensitivity with paleoclimate simulations. Sola 1:181–184

Boe JL, Hall A, Qu X (2009) Current GCMs' unrealistic negative feedback in the Arctic. J Clim 22:4682–4695

Boe JL, Hall A, Qu X (2010) Sources of spread in simulations of Arctic sea ice loss over the twenty-first century. Clim Chang 99:637–645

Boer GJ, Yu B (2003a) Dynamical aspects of climate sensitivity. Geophys Res Lett 30:1135. doi:10.1029/2002GL016549

Boer GJ, Yu B (2003b) Climate sensitivity and response. Clim Dyn 20:415–429

Bony S, Dufresne JL (2005) Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. Geophys Res Lett 32:L20806

Brierley CM, Collins M, Thorpe AJ (2010) The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. Clim Dyn 34:325–343

Cess RD et al (1990) Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. J Geophys Res 95:16601–16615

Chikira M (2010) Acumulus parameterization with state-dependent entrainment rate. Part II: impact on climatology in a general circulation model. J Atmos Sci 67:2194–2211

Chikira M, Sugiyama M (2010) A cumulus parameterization with state-dependent entrainment rate. Part I: description and sensitivity to temperature and humidity profiles. J Atmos Sci 67:2171–2193

Collins M, Booth BBB, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2006) Towards quantifying uncertainty in transient climate change. Clim Dyn 27:127–147

Collins M, Brierley CM, MacVean M, Booth BBB, Harris GR (2007) The sensitivity of the rate of transient climate change to ocean physics perturbations. J Clim 20:2315–2320

Collins M, An SI, Cai W, Ganachaud A, Guilyardi E, Jin FF, Jochum M, Lengaigne M, Power S, Timmermann A, Vecchi G, Wittenberg A (2010) The impact of global warming on the tropical Pacific and El Niño. Nat Geosci 3:391–397

Collins M, Booth BBB, Bhaskaran B, Harris G, Murphy JM, Sexton DMH, Webb MJ (2011) Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. Clim Dyn 36:1737–1766

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26

Forest CE, Stone PH, Sokolov AP, Allen MR, Webster MD (2002) Quantifying uncertainties in climate system properties with the use of recent climate observations. Science 295:113–117

Gettelman A, Kay JE, Shell KM (2012) The evolution of climate sensitivity and climate feedbacks in the community atmosphere model. J Clim 25:1453–1469. doi:10.1175/JCLI-D-11-00197.1

Gordon C et al (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 16:147–168

Gregory JM et al (2004) A new method for diagnosing radiative forcing and climate sensitivity. Geophys Res Lett 31:L03205

Guilyardi E, Wittenberg A, Fedorov A, Collins M, Wang C, Capotondi A, van Oldenborgh GJ (2009) Understanding El

Niño in ocean–atmosphere general circulation models: progress and challenges. Bull Am Meteorol Soc 90:325–340

Hall A, Qu X (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophys Res Lett 33:L03502

Hartmann DL, Larson K (2002) An important constraint on tropical cloud—climate feedback. Geophys Res Lett 29:1951. doi:10.1029/2002GL015835

Hasumi H (2006) CCSR Ocean Component Model (COCO), version 4.0. Center for Climate System Research Report 25, 103 pp. Available online at http://www.ccsr.u-tokyo.ac.jp/~hasumi/COCO/coco4.pdf

Jackson CS, Sen MK, Huerta G, Deng Y, Bowman KP (2008) Error reduction and convergence in climate prediction. J Clim 21:6698–6709

Jackson LC, Vellinga M, Harris GR (2011) The sensitivity of the meridional overturning circulation to modelling uncertainty in a perturbed physics ensemble without flux adjustment. Clim Dyn. doi:10.1007/s00382-011-1110-5

Klein SA, Jakob C (1999) Validation and sensitivities of frontal clouds simulated by the ECMWF model. Mon Weather Rev 127:2514–2531

Klocke D, Pincus R, Quaas J (2011) On constraining estimates of climate sensitivity with present-day observations through model weighting. J Clim 24:6092–6099. doi:10.1175/2011JCLI4193.1

Knutti R (2010) The end of model democracy? Clim Chang 102:395–404

Knutti R, Hegerl GC (2008) The equilibrium sensitivity of the earth's temperature to radiation changes. Nat Geosci 1:735–743

Knutti R, Stocker TF, Joos F, Plattner GK (2002) Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature 416:719–723

McKay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21:239–245

Medeiros B, Stevens B, Held IM, Zhao M, Williamson DL, Olson JG, Bretherton CS (2008) Aquaplanets, climate sensitivity, and low clouds. J Clim 21:4974–4991

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430:768–772

Murphy JM, Booth BBB, Collins M, Harris GR, Sexton DMH, Webb MJ (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philos Trans R Soc A Math Phys Eng Sci 365:1993–2028

Nakanishi M, Niino H (2004) An improved Mellor–Yamada level-3 model with condensation physics: its design and verification. Bound Layer Meteorol 112:1–31

Niehörster F, Collins M (2009) A comparison of perturbed physics ensembles constructed with different models, ENSEMBLES deliverable report D1.14 of working group I, p 8

Piani C, Frame DJ, Stainforth DA, Allen MR (2005) Constraints on climate change from a multi-thousand member ensemble of simulations. Geophys Res Lett 32:L23825

Pincus R, Platnick S, Ackerman SA, Hemler RS, Hofmann RJP (2012) Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. J Clim. doi:10.1175/JCLI-D-11-00267.1

Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer RJ, Sumi A, Taylor KE (2007) Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007—the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Rossow WB, Schiffer RA (1999) Advances in understanding clouds from ISCCP. Bull Am Meteorol Soc 80:2261–2287

Rougier J, Sexton DMH (2007) Inference in ensemble experiments. Philos Trans R Soc A Math Phys Eng Sci 365:2133–2143

Rougier J, Sexton DMH, Murphy JM, Stainforth D (2009) Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. J Clim 22:3540–3557

Rowlands DJ, Frame DJ, Ackerley D, Aina T, Booth BBB, Christensen C, Collins M, Faull N, Forest CE, Grandey BS, Gryspeerdt E, Highwood EJ, Ingram WJ, Knight S, Lopez A, Massey N, McNamara F, Meinshausen N, Piani C, Rosier SM, Sanderson BM, Smith LA, Stone DA, Thurston M, Yamazaki K, Hiro Yamazaki Y, Allen MR (2012) Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nat Geosci 5:256–260. doi:10.1038/ngeo1430

Sanderson BM (2011) A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. J Clim 24:1362–1377

Sanderson BM, Piani C, Ingram WJ, Stone DA, Allen MR (2008a) Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. Clim Dyn 30:175–190

Sanderson BM et al (2008b) Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. J Clim 21:2384–2400

Sanderson BM, Shell KM, Ingram W (2010) Climate feedbacks determined using radiative kernels in a multi-thousand member ensemble of AOGCMs. Clim Dyn 35:1219–1236

Seiffert R, von Storch JS (2008) Impact of atmospheric small-scale fluctuations on climate sensitivity. Geophys Res Lett 35:L10704

Shiogama H, Emori S, Takahashi K, Nagashima T, Ogura T, Nozawa T, Takemura T (2010) Emission scenario dependency of precipitation on global warming in the MIROC3.2 model. J Clim 23:2404–2417

Shiogama H, Emori S, Hanasaki N, Abe M, Masutomi Y, Takahashi K, Nozawa T (2011) Observational constraints indicate risk of drying in the Amazon basin. Nat Commun 2, Article No. 253

Soden BJ, Broccoli AJ, Hemler RS (2004) On the use of cloud forcing to estimate cloud feedback. J Clim 17:3661–3665

Stainforth DA et al (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433:403–406

Takata K, Emori S, Watanabe T (2003) Development of the minimal advanced treatments of surface interaction and runoff. Glob Planet Chang 38:209–222

Takemura T, Nozawa T, Emori S, Nakajima TY, Nakajima T (2005) Simulation of climate response to aerosol direct and indirect effects with aerosol transport-radiation model. J Geophys Res 110:D02202

Takemura T, Egashira M, Matsuzawa K, Ichijo H, O'ishi R, Abe-Ouchi A (2009) A simulation of the global distribution and radiative forcing of soil dust aerosols at the Last Glacial Maximum. Atmos Chem Phys 9:3061–3073

Taylor KE, Doutriaux C, Crucifix M, Hewitt CD, Mitchell JFB, Webb MJ, Braconnot P, Broccoli AJ (2007) Estimating shortwave radiative forcing and response in climate models. J Clim 20:2530–2543

Toniazzo T, Collins M, Brown J (2008) The variation of ENSO characteristics associated with atmospheric parameter perturbations in a coupled model. Clim Dyn 30:643–656

Watanabe M et al (2010) Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. J Clim 23:6312–6335

Watanabe M, Chikira M, Imada Y, Kimoto M (2011a) Convective control of ENSO simulated in MIROC. J Clim 24:543–562

Watanabe M, Shiogama H, Yokohata T, Ogura T, Yoshimori M, Emori S, Kimoto M (2011b) Constraints to the tropical low-cloud trends in historical climate simulations by MIROC. Atmos Sci Lett 12:288–293

Watanabe M, Shiogama H, Yokohata T, Kamae Y, Yoshimori M, Ogura T, Annan JD, Hargreaves JC, Emori S, Kimoto M (2011c) Using a multi-physics ensemble for exploring diversity in cloud shortwave feedback in GCMs. J Clim. doi:10.1175/JCLI-D-11-00564.1

Watanabe M, Shiogama H, Yoshimori M, Yokohata T, Ogura T, Okamoto H, Emori S, Kimoto M (2011d) Fast and slow timescales in the tropical low-cloud response to increasing CO_2 in two climate models. Clim Dyn (in press)

Webb M, Senior C, Bony S, Morcrette JJ (2001) Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. Clim Dyn 17:905–922

Webb MJ, Senior CA, Sexton DMH, Ingram WJ, Williams KD, Ringer MA, McAvaney BJ, Colman R, Soden BJ, Gudgel R, Knutson T, Emori S, Ogura T, Tsushima Y, Andronova NG, Li B, Musat I, Bony S, Taylor KE (2006) On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. Clim Dyn 27:17–38

Webb M, Lambert FH, Gregory JM (2012) Origins of differences in climate sensitivity, forcing and feedback in climate models. Clim Dyn. doi:10.1007/s00382-012-1336-x

Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, New York

Williams KD, Webb MJ (2009) A quantitative performance assessment of cloud regimes in climate models. Clim Dyn 33:141–157

Williams KD, Ingram WJ, Gregory JM (2008) Time variation of effective climate sensitivity in GCMs. J Clim 21:5076–5090

Wilson DR, Ballard SP (1999) A microphysically based precipitation scheme for the UK Meteorological Office unified model. Quart J R Meteorol Soc 125:1607–1636

Xie P, Arkin PA (1997) Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull Am Meteorol Soc 78:2539–2558

Yokohata T et al (2008) Comparison of equilibrium and transient responses to CO_2 increase in eight state-of-the-art climate models. Tellus 60:946–961

Yokohata T, Webb MJ, Collins M, Williams KD, Yoshimori M, Hargreaves JC, Annan JD (2010) Structural similarities and differences in climate responses to CO_2 increase between two perturbed physics ensembles. J Clim 23:1392–1410

Yoshimori M, Hargreaves JC, Annan JD, Yokohata T, Abe-Ouchi A (2011) Dependency of feedbacks on forcing and climate state in perturbed parameter ensembles. J Clim (in press). doi:10.1175/2011JCLI3954.1

Zelinka MD, Hartmann DL (2010) Why is longwave cloud feedback positive? J Geophys Res Atmos 115:D16117