



An efficient method to generate a perturbed parameter ensemble of a fully coupled AOGCM without flux-adjustment

P. J. Irvine¹, L. J. Gregoire², D. J. Lunt², and P. J. Valdes²

¹Institute for Advanced Sustainability Studies IASS, Potsdam, Germany

²School of Geographical Sciences, University of Bristol, UK

Correspondence to: P. J. Irvine (peter.irvine@iass-potsdam.de)

Received: 9 January 2013 – Published in Geosci. Model Dev. Discuss.: 6 February 2013

Revised: 10 July 2013 – Accepted: 29 July 2013 – Published: 10 September 2013

Abstract. We present a simple method to generate a perturbed parameter ensemble (PPE) of a fully-coupled atmosphere-ocean general circulation model (AOGCM), HadCM3, without requiring flux-adjustment. The aim was to produce an ensemble that samples parametric uncertainty in some key variables and gives a plausible representation of the climate. Six atmospheric parameters, a sea-ice parameter and an ocean parameter were jointly perturbed within a reasonable range to generate an initial group of 200 members. To screen out implausible ensemble members, 20 yr pre-industrial control simulations were run and members whose temperature responses to the parameter perturbations were projected to be outside the range of $13.6 \pm 2^\circ\text{C}$, i.e. near to the observed pre-industrial global mean, were discarded. Twenty-one members, including the standard unperturbed model, were accepted, covering almost the entire span of the eight parameters, challenging the argument that without flux-adjustment parameter ranges would be unduly restricted. This ensemble was used in 2 experiments; an 800 yr pre-industrial and a 150 yr quadrupled CO_2 simulation. The behaviour of the PPE for the pre-industrial control compared well to ERA-40 reanalysis data and the CMIP3 ensemble for a number of surface and atmospheric column variables with the exception of a few members in the Tropics. However, we find that members of the PPE with low values of the entrainment rate coefficient show very large increases in upper tropospheric and stratospheric water vapour concentrations in response to elevated CO_2 and one member showed an implausible nonlinear climate response, and as such will be excluded from future experiments with this ensemble. The outcome of this study is a PPE of a fully-coupled AOGCM

which samples parametric uncertainty and a simple methodology which would be applicable to other GCMs.

1 Introduction

1.1 Background on perturbed parameter ensembles

PPEs of general climate models (GCMs) are becoming more common as a means to assess the range of uncertainty in climate model projections (Murphy et al., 2004; Stainforth et al., 2005; Collins et al., 2006; Sanderson, 2011; Yokohata et al., 2010; Shiogama et al., 2012; Klocke et al., 2011). This PPE approach is a complement to the Multi-Model Ensemble (MME) approach notably applied in the Intergovernmental Panel on Climate Change (IPCC) assessments (Solomon et al., 2007; Meehl et al., 2007b; Taylor et al., 2012). These two approaches address two aspects of model uncertainty; in MMEs, the structural uncertainty associated with the understanding, discretization and parameterization of the climate system as a GCM and in PPEs, the parametric uncertainty associated with the uncertain values of the parameters within a GCM. The MME approach has the advantage of having independent modelling schemes (although the fact there is a somewhat common heritage amongst models and they are developed by a group of experts sharing similar knowledge, limits their independence, Masson and Knutti, 2011), but as the number of possible models is indefinable, any MME will represent an unquantifiable and incomplete sampling of the structural uncertainty in climate model predictions (Meehl et al., 2007b). The PPE approach has the advantage that members of the ensemble differ in a well-defined way and

the “parameter space” of all possible parameter combinations can be precisely defined. It is not possible to generate a large number of models with different structures without a long programme of model development, however it is possible to generate a very large number of different versions of one model by perturbing parameters, with the availability of computing resources being the only effective limit. For these reasons, PPE experiments are a useful tool for assessing uncertainty in climate model projections.

As greater computing resources have become available, larger and more complex perturbed parameter ensembles of GCMs have become possible (Frame et al., 2009). There are many hundreds of uncertain parameters in a GCM and so expert elicitation is needed to select which parameters are important and to indicate a reasonable range for these parameters (Murphy et al., 2004). The early perturbed parameter ensembles consisted of single-parameter perturbations, in effect a sensitivity test of parametric uncertainty (Murphy et al., 2004). However, many parameters in a GCM will interact in complex, nonlinear ways, and so parameters must be perturbed simultaneously to explore the full range of response implied by the prior parametric uncertainty (Stainforth et al., 2005; Sanderson, 2011; Shiogama et al., 2012). The space of all uncertain parameters can be very large indeed for GCMs and so many studies have taken subsets of the most important parameters to achieve a more thorough coverage of the parameter space (Stainforth et al., 2005; Knight et al., 2007; Shiogama et al., 2012).

Most PPEs to date have used atmosphere-only or slab-ocean versions of GCMs as these take a few years or a few decades of model simulation to reach equilibrium respectively, as opposed to the millennia required to fully spin up a fully dynamic coupled atmosphere-ocean GCM, although some parametric sensitivity studies have used coupled oceans (Collins et al., 2007; Brierley et al., 2010; Shiogama et al., 2012). Most PPE studies with fully coupled models have used flux-adjustment to keep the ensemble members from drifting too far from observed climatology. This flux-adjustment is applied as either a heat, water or momentum flux into the ocean surface designed to correct for model biases (Collins et al., 2006). Top-of-atmosphere (TOA) radiative balance is an emergent property in GCMs and the fact that the models of the IPCC Assessment Report 4 did not need flux-adjustment was seen as an improvement over earlier models (Solomon et al., 2007).

Numerous methods to test the “realism” of members of a perturbed parameter ensemble of a GCM have been developed and these are often used to exclude or weight the members of a PPE for the purposes of making projections (Edwards et al., 2011; Murphy et al., 2004; Rodwell and Palmer, 2007). Murphy et al. (2004) analysed a perturbed parameter ensemble of the UK Met Office Hadley Centre Model version 3 (HadCM3) using the climate prediction index, a method which applies a set of comparisons to observational data that gives each member a weight, and which has also been applied

to other PPE studies (Collins et al., 2010). An alternative approach is to run the GCM in forecast mode, i.e. starting from observed initial atmospheric conditions, and measure the deviation of the simulated atmospheric column from observations over the course of a few days of simulation (Rodwell and Palmer, 2007). If the PPE member changes the structure of the variables throughout the atmospheric column substantially from observations the member can be ruled unrealistic and excluded or down-weighted. Another, more simple, approach is that of Edwards et al. (2011), who outlined a “pre-calibration” approach for testing the “plausibility” of model output; a set of lenient physical criteria are defined such that the member should be deemed implausible if it fails to satisfy any of these loose criteria and those members which remain should be considered plausible representations of the system. In this study we do not attempt to rank the ensemble members but we follow the spirit of the approach of Edwards et al. (2011) and test whether or not the ensemble members are “plausible” representations of the climate system.

1.2 Objectives of this study

In this study, we develop a perturbed parameter ensemble (PPE) using the fully-coupled AOGCM HadCM3 without applying flux adjustments. Our study follows on from the work of Gregoire et al. (2010) who used a Latin Hypercube sampling scheme to tune a low resolution GCM, the Fast Met Office/UK universities Simulator (FAMOUS). Here we adapted this approach to a more computationally expensive GCM, by estimating the equilibrium temperature response to the parameter perturbations using the method of Gregory et al. (2004). We test an efficient approach to initially select members, which excludes ensemble members that are expected to deviate too far from the observed global mean temperature of the pre-industrial in response to their parameter perturbations. The objective is to produce an ensemble of tens of members which have “plausible” behaviour when compared against the European Centre for medium-range weather forecasting atmospheric reanalysis dataset (ERA-40) for the period 1961–1990, and additional observational data. For comparison, we include results from members of the World Climate Research Program’s (WCRP’s) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset. An application of the ensemble is made to a quadrupled CO₂ experiment. The methodology, selection approach and results are then discussed. The rest of the paper is laid out as: methodology in Sect. 2, results and evaluation in Sect. 3, and discussion in Sect. 4. The Supplement is included which consists of 2 tables that detail the parameter values and some measures of performance for all members of the ensemble.

2 Methodology

2.1 HadCM3 model description

The fully coupled atmosphere-ocean general circulation model (AOGCM) used in this paper is HadCM3 (Gordon et al., 2000). HadCM3 has been used in the IPCC third and fourth assessment reports (Houghton et al., 2001; Solomon et al., 2007) and performs well in a number of tests relative to other global GCMs (Solomon et al., 2007; Covey et al., 2003). The speed of HadCM3 compared to the newer state-of-the-art Met Office Hadley Centre Global Environmental Model version 2 (HadGEM2) (Collins et al., 2011), makes it a powerful tool for multi-millennial scale climate studies. It is also ideal for uncertainty analysis studies using perturbed physics ensembles such as the one presented here. The horizontal resolution of the atmospheric model is 2.5° in latitude by 3.75° in longitude, with 19 vertical layers. The atmospheric model has a time step of 30 minutes and includes many parameterizations representing sub grid-scale effects, such as convection (Gregory and Rowntree, 1990) and boundary-layer mixing (Smith, 1993). The spatial resolution in the ocean is 1.25° by 1.25° , with 20 vertical layers. The ocean model component uses the Gent and McWilliams (1990) mixing scheme, and there is no explicit horizontal tracer diffusion. The sea-ice model uses a simple thermodynamic scheme and contains parameterizations of sea-ice drift and leads (Cattle and Crossley, 1995). We employ the Met Office Surface Exchange Scheme (MOSES) 1 land surface scheme (Cox et al., 1999), which accounts for terrestrial surface fluxes of temperature, moisture and radiation. MOSES includes 4 soil layers recording temperature, moisture and phase changes, a canopy layer and a representation of lying snow. The representation of evaporation includes the dependence of stomatal resistance on temperature, vapour pressure and CO_2 concentration (Cox et al., 1999). Each grid cell has surface properties; roughness length, snow-free albedo, etc., which reflect the vegetation cover present, as derived from the Wilson and Henderson-sellers (1985) dataset.

2.2 Ensemble design

A relatively small number of simulations will be possible as we are using a fully-coupled AOGCM which will require a considerable spinup. Therefore, to allow for a reasonable coverage of parameter space only a small number of parameters are chosen. The greater the number of parameters included in an ensemble the more aspects of the parametric uncertainty in the model can be assessed; however, with a greater number of parameters there is a larger parameter space. One way to quantify the coverage of parameter space that a given ensemble represents is to imagine dividing each parameter range into two halves, “low” and “high”, thus there are 2^p combinations of “low” and “high” for p parameters. If

we start with an ensemble of 200 members, a number judged to be computationally feasible for short runs of this model, 78 % of the “halves” of an 8 parameter space can be covered but only 20 % of the “halves” of a 10 parameter space and only 5 % of a 12 parameter space. We chose to start with an initial ensemble of 200 members and chose to modify only 8 parameters to strike a balance between coverage of parameter space and the number of important parameters.

We chose to vary the atmospheric, oceanic and sea ice parameters listed in Table 1. These include the 6 atmospheric parameters modified in Stainforth et al. (2005), the lateral entrainment rate coefficient from the environment into convective clouds (ENTCOEF), the ice-fall speed (VF1), the critical relative humidity (RHCRIT), the droplet to rain conversion rate (CT), the droplet to rain conversion threshold over land and sea (CW_LAND/SEA, two parameters that are perturbed as one), the empirically adjusted cloud fraction (EACF); the sea-ice minimum albedo at melting point (ALPHAM); and the background vertical ocean diffusivity parameter (VDIFF, consisting of two parameters perturbed as one) used in Collins et al. (2007). The 6 parameters modified in Stainforth et al. (2005) were chosen for the large impact that these parameters have on climate sensitivity (Rougier et al., 2009). The sea-ice minimum albedo (ALPHAM) parameter was added as it is expected that this ensemble will be used for paleo-climate simulations of glacial times where sea-ice parameters may play a more important role than in the modern day or future (Gregoire et al., 2011). The vertical ocean diffusivity parameter was added, as this was the ocean parameter found to have the most significant effect on the transient climate response of HadCM3 (Collins et al., 2007; Brierley et al., 2010).

The range for all parameters except for VDIFF were taken from the expert elicitation in Murphy et al. (2004); however, the lower ranges of EACF and ALPHAM were extended by 20 % as the standard version of HadCM3 sits at the lower limit for these parameters. It was reasoned that if the parameter values of the standard version of HadCM3 are reasonable, small deviations from these values should be reasonable too. The VDIFF parameter consists of the initial surface background diffusivity and a rate of increase of diffusivity with depth which were varied together as in Collins et al. (2007) and Brierley et al. (2010). All parameters except one are sampled using a uniform prior on parameter value. For VDIFF the initial diffusivity and the rate of increase of diffusivity vary as 2^x and 4^x respectively, where x varies uniformly from -1 to 1 . This choice for the VDIFF parameter was made after discussions with the author of a study which presented an expert elicited range for this parameter (C. Brierley, personal communication, 2011).

To select parameter combinations a maximin latin hypercube sampling technique was used and 200 combinations of the 8 parameters drawn (Gregoire et al., 2010; Tang, 1994). To generate a latin hypercube each parameter range is divided into 200 sections with one point drawn from each of

Table 1. Shows a list of the eight parameters perturbed in the experiment. Shown are the value of the parameter in the standard configuration, the minimum and maximum for the parameter range, and a short description of that parameter.

Parameter Name	Standard Value	Minimum Value	Maximum Value	Description
ENTCOEF	3.0	0.6	9.0	Lateral entrainment rate coefficient
VFI	1.0	0.5	2.0	Ice-fall speed
CT	1.0×10^{-4}	5.0×10^{-5}	4.0×10^{-4}	Cloud droplet to rain conversion rate
CW (Land/Sea)	2.0×10^{-4} 5.0×10^{-5}	1.0×10^{-4} 2.0×10^{-5}	2.0×10^{-3} 5.0×10^{-4}	Cloud droplet to rain conversion threshold over land and sea
EACF	0.5	0.47	0.65	Empirically adjusted cloud fraction at saturation
RHCRIT	0.7	0.6	0.9	Threshold of relative humidity for cloud formation
ALPHAM	0.5	0.47	0.65	Sea-ice albedo at 0 °C
VDIFF (Min/Max)	1.0–15.0	0.5–4.0	2.0–58.0	Background vertical diffusivity, which varies as a function of depth

the sections of each parameter, ensuring that there is no repetition, and giving good univariate separation between members. There are many possible latin hypercubes which satisfy these conditions and a better sampling is possible with the maximin latin hypercube approach. Maximin latin hypercube sampling adds the requirement that each point drawn must be as far from previous points as possible, thus ensuring a greater multivariate separation of the ensemble members. At this stage each point is defined as a small region of parameter space between the minima and maxima of its respective parameter sections. To get a definitive value for each of the point's parameter co-ordinates a random value between the minimum and maximum of each section of each parameter is found in turn. Thus, we have 200 well-spaced parameter value drawn from across the 8 dimensional parameter space.

2.3 Experimental setup

To select members for our final ensemble, we applied a low-cost selection criterion to these initial 200 members. Instead of running each one of the 200 ensemble members for several hundred years to equilibrium, we only ran them for 20 yr. We then projected the equilibrium temperature of the model runs using the approach of Gregory et al. (2004) and discarded all ensemble members that had projected temperature outside a plausible temperature range. All simulations were started from the end of a many thousand years long pre-industrial spinup of the standard version of HadCM3 (standard model), i.e. with standard parameter values. Around half of the simulations failed to complete these first 20 yr and these failed

members could not be used for further simulations. HadCM3 is known to be not entirely stable across its parameter space (Rougier et al., 2009), and without flux-adjustment some otherwise stable simulations have been found to give a simulation so unrealistic that they eventually became numerically unstable (Murphy et al., 2004).

To make the equilibrium temperature response projections, we assume that the change in parameters caused an instantaneous change in radiative forcing, an approach which has previously been applied to perturbed parameter ensembles (Joshi et al., 2010; Shiogama et al., 2012). The projection of temperature and the initial radiative forcing is made from a linear regression of the change in temperature and the change in TOA radiative imbalance from the standard model's control mean, in the manner of Gregory et al. (2004). Note that not all models exhibit radiative balance in equilibrium, some members of the CMIP3 ensemble show persistent radiative imbalances of up to 4.0 W m^{-2} , if so it is necessary to project the equilibrium temperature response using the TOA radiative imbalance anomaly from the standard model. HadCM3 has only a small persistent TOA radiative imbalance of -0.13 W m^{-2} , and so we adopted the absolute TOA radiative imbalance. We kept only members which were projected to have equilibrium pre-industrial global-mean temperature within 2°C of the estimated pre-industrial temperature of 13.6°C (Jones et al., 1999; Brohan et al., 2006), which form the PPE. The range of $\pm 2.0^\circ\text{C}$ was decided upon as being approximately equal to the largest difference between the pre-industrial absolute temperature of a member of the

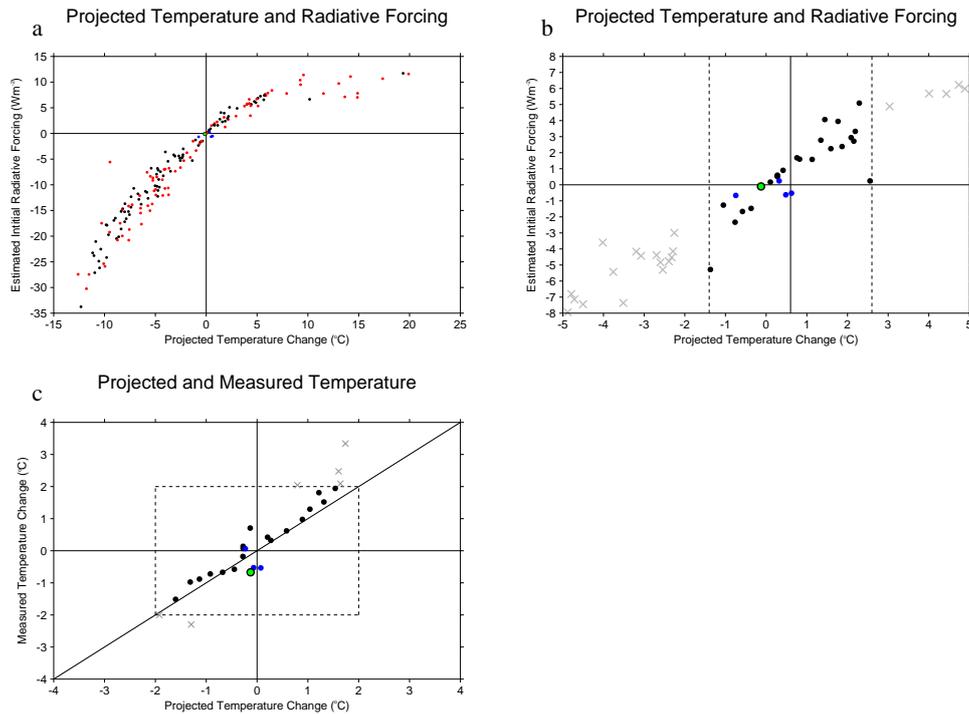


Fig. 1. Projections of equilibrium temperature and initial radiative forcing for the initial ensemble generated by applying the Gregory et al. (2004) approach to 20 yr of pre-industrial simulation (**a**, **b**). Panel (**b**) shows the acceptable range of temperatures with dashed lines, i.e. within $\pm 2^\circ\text{C}$ of the observed pre-industrial temperature of 13.6°C (Brohan et al., 2006; Jones et al., 1999). Panel (**c**) shows a comparison between the projected temperature and the simulated temperature at the end of the 800 yr control run. Simulations which completed the first 20 yr are shown in black and those which failed to complete are shown in red, the large green and black point is for the standard model, the crosses in (**b**) and (**c**) show runs which were too warm or cold. The projection method failed for some runs which are shown in blue where the blue dot shows the temperature and radiative imbalance of the last simulated year rather than the projection.

CMIP3 ensemble (i.e. -1.8°C) and similar to the spread of 3.3°C (Meehl et al., 2007b).

The members which passed this selection criterion formed the final PPE ensemble and were used for further simulations. As we are modifying the ocean and atmosphere of the model it will take thousands of years for the model to equilibrate fully. Due to computational constraints we cannot run the model this long, and so we follow Collins et al. (2007), and run a 500 yr spinup to allow some degree of adjustment to the altered conditions. After the spinup 2 further simulations were started, a 300 yr pre-industrial control run and a 150 yr simulation with an instantaneous quadrupling of CO_2 ($4*\text{CO}_2$).

3 Results

3.1 Initial selection on projected temperature response

The initial selection of the PPE was based on the projected temperature, Fig. 1a and b shows the projected temperature and estimated initial TOA radiative imbalance of the initial 200 members. A large number of the simulations failed

to complete, but there was no clear relation between failure to complete this first 20 yr and any individual parameter. Around three quarters of the simulations which completed the 20 yr pre-industrial control simulations had very large changes in TOA radiative balance and were projected to warm or cool rapidly, deviating greatly from the observed global-mean pre-industrial temperature. Figure 1c shows the projected temperature from the first 20 yr and the temperature after 800 yr of pre-industrial control run for each of the 27 members which were projected to be within $\pm 2.0^\circ\text{C}$ of the observed pre-industrial temperature of 13.6°C (Brohan et al., 2006; Jones et al., 1999). Most of the members of the PPE are close to their respective projected temperatures, but two warmer members, and a single cold member, are clearly outside of the range, with three further runs within 0.2°C of the target range. Of the 27 members selected by the Gregory method $\approx 80\%$ remained within the target window and most are within a few tenths of a degree of their projected values. The application of this approach avoided the need to run the tens of initially rejected members to equilibrium, saving substantial amounts of computing time.

The final ensemble (hence, PPE) consists of 21 accepted members, including the standard configuration, with an

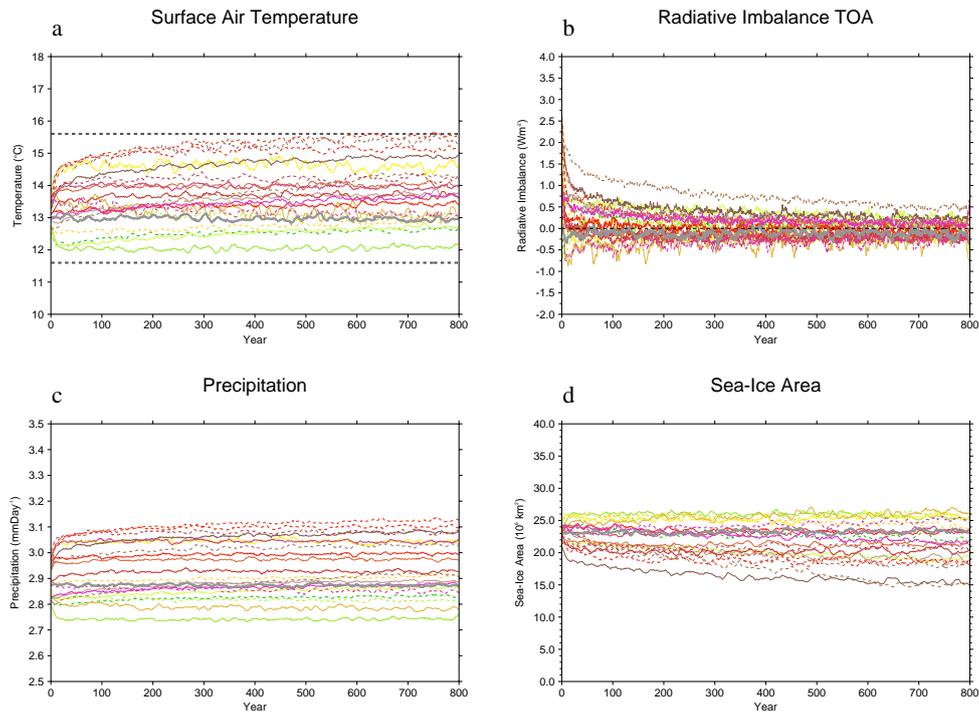


Fig. 2. Evolution over the course of the 800 yr pre-industrial control simulation of the global annual means of surface air temperature (a), top-of-atmosphere radiative balance (b), precipitation (c) and annual-mean sea-ice area (d). The standard version of HadCM3 is plotted with a thick gray line.

additional 6 failed members. The failed members will be retained, but shown only in plots that illustrate the role of the parameters. Supplement Table 1 lists the parameter values and the pre-industrial temperature anomaly from observations of each of the members of the PPE with the members which failed the selection criterion marked.

3.2 Pre-industrial spinup

Overall, we ran 800 yr of pre-industrial conditions with the final ensemble of 21 successful and 6 failed members. Figure 2 shows the evolution of a number of variables over the course of the 800 yr pre-industrial control runs, note that the same colour scheme is used throughout this study to aid the identification of ensemble members across plots. Figure 2a and b show that most members of the PPE behave as if an instantaneous radiative forcing had been applied, in other words, they follow an asymptotic approach to a new equilibrium temperature and the radiative imbalance is decaying to zero. One member has a markedly higher radiative imbalance which is at 0.5 W m^{-2} at the end of the control run but remains within the target temperature range after 800 yr. The change in precipitation, Fig. 2c, shows a rapid adjustment to the altered atmospheric conditions followed by a temperature driven change in precipitation (Bala et al., 2010). The sea-ice area, Fig. 2d, changes quite significantly, with the warmer

members losing up to a third of their sea-ice, and some members gaining sea-ice area.

Figure 3a shows that the deep ocean has not adjusted fully to the parameter perturbations by the end of the pre-industrial control; all the members of the ensemble show deep-ocean temperature trends which change little over the 800 yr control run. In fact, even the standard HadCM3 simulation, still shows a slight cooling. Figure 3b shows the evolution of the maximum meridional overturning circulation in the Atlantic; most members remain close to the standard model's condition with an overturning strength of $\approx 18 \text{ Sv}$, but 3 of the members show increased overturning of around $\approx 25 \text{ Sv}$ and some also show a large increase in variability. Although the ocean is not in equilibrium, significant changes have occurred by the end of the 800 yr pre-industrial control. Figure 3c and d show the depth profile of the ocean temperature and salinity for the end of the simulations, showing that the condition of the ocean has changed markedly across the ensemble. We find that changes in ocean temperature at depth are determined more by the atmospheric variables than by the ocean vertical diffusivity parameter, VDIFF (not shown). Previous results with a PPE of HadCM3 with only atmospheric parameters perturbed found that the members with the warmest control climate had substantially reduced Atlantic overturning (Jackson et al., 2012). However, we find that the maximum Atlantic overturning strength is

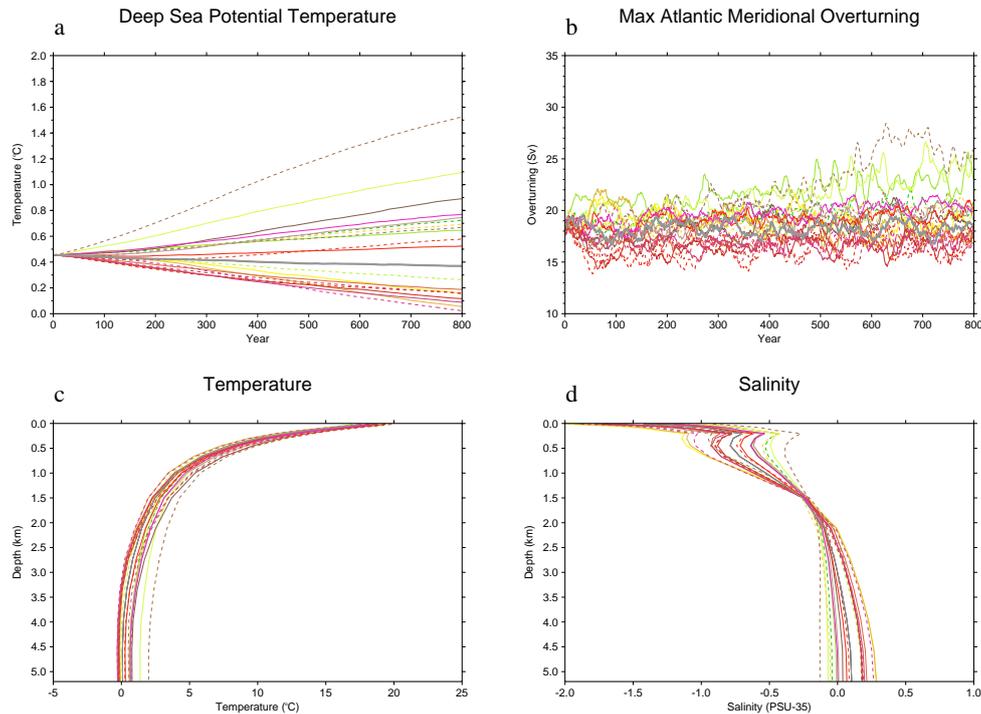


Fig. 3. Evolution of the global annual means of potential temperature at a depth of 2700 m (a), maximum Atlantic overturning (b) over the course of the 800 yr pre-industrial control simulation. The figure also shows ocean depth profile of annual and area average temperature (c) and salinity (d), averaged over years 501 to 800 of the pre-industrial control. The standard version of HadCM3 is plotted with a thick gray line.

most strongly associated with the value of the VDIFF parameter rather than the control temperature, with the members with the highest values of VDIFF showing a large increase in overturning whereas the members with a standard or low value of VDIFF showing little change. This strong response of Atlantic overturning to changes in VDIFF in HadCM3 was also found by Brierley et al. (2010).

The effect of the perturbed atmospheric and sea-ice parameters on the HadCM3 model have been explored in detail by a number of other studies (Collins et al., 2006; Murphy et al., 2004; Sanderson et al., 2008a; Knight et al., 2007), as have the effects of the perturbed ocean parameter (Collins et al., 2007; Brierley et al., 2010), and so the interested reader should refer to these for further information. However, we note here a number of relevant correlations we find between some parameters and the resulting pre-industrial climate. We find that the ENTCOE and VF1 parameters that have previously been found to have the largest role in controlling climate sensitivity in the HadCM3 model are also found to exert significant control over the equilibrium pre-industrial temperature (Rougier et al., 2009; Sanderson et al., 2008a), with low values of both parameters tending to give warmer conditions. We also find that higher values of ocean vertical diffusivity (VDIFF) are associated with more positive radiative imbalance in the pre-industrial control, despite not directly

affecting the global energy budget. For high values of the VDIFF parameter more energy is absorbed by the ocean (up to $\approx 1.4 \text{ W m}^{-2}$ compared with $\approx 0.6 \text{ W m}^{-2}$ in the standard model), absorbing energy that would otherwise have warmed the model surface, and vice versa for low values of VDIFF. It seems likely that this association between high values of VDIFF and higher pre-industrial temperatures is due to VDIFF mitigating the initial rate of temperature change (C. Brierley, personal communication, 2011). It has been found that high values of VDIFF cause an increase in the flux of energy into the oceans in the initial years and so may act to keep members, that would have otherwise warmed too fast, close to the observed pre-industrial temperature (Brierley et al., 2010). We also note that the sea-ice minimum albedo parameter (ALPHAM) has a much smaller effect than surface air temperature on the pre-industrial sea-ice fraction.

3.3 Comparison of the PPE with ERA-40 and the CMIP3 ensemble

Figure 4 shows the annual and zonal mean state of the pre-industrial climate for the PPE and compares this with the ERA-40 1961–1990 average and the CMIP3 ensemble (Meehl et al., 2007b). Figure 4a shows that the zonal mean temperatures of the PPE and the ERA-40 dataset show a similar distribution. The PPE zonal precipitation, Fig. 4b, is

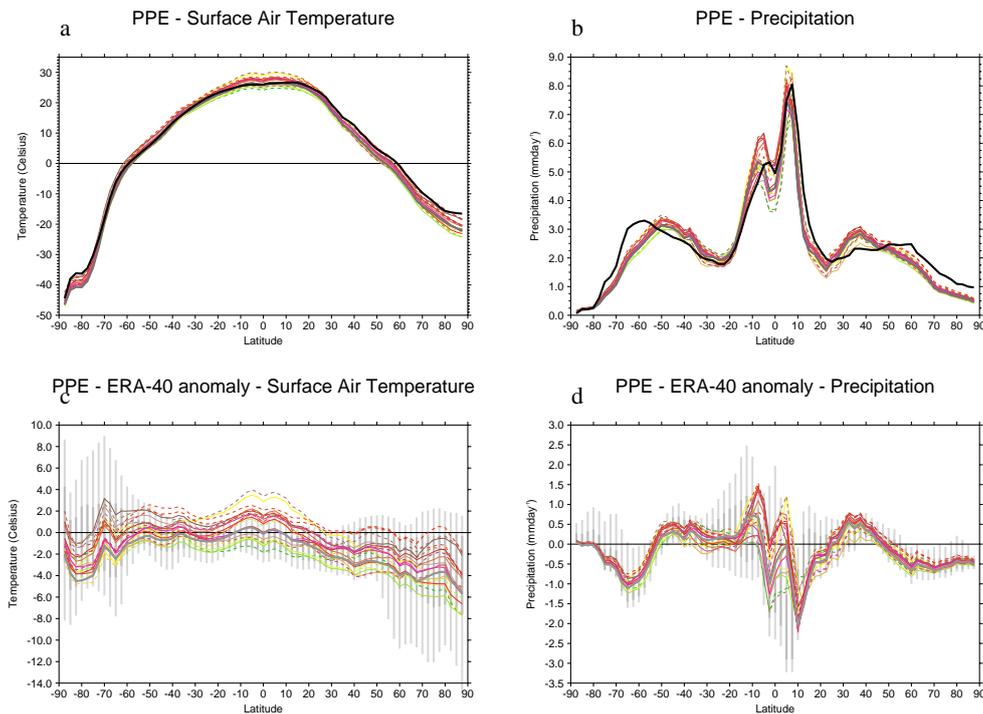


Fig. 4. The zonal-mean surface air temperature (a) and precipitation (b) for the PPE simulations with the ERA-40 1961–1990 mean plotted for comparison. The zonal-mean anomaly between each member of the PPE and the ERA-40 mean for surface air temperature (c) and precipitation (d). The standard version of HadCM3 is plotted in dark gray and the ERA-40 mean is plotted in black. In (c) and (d) the range for the CMIP3 ensemble is plotted with light gray bars and the standard deviation of the CMIP3 ensemble around the ensemble mean is plotted with darker gray bars.

broadly similar to the ERA-40 dataset, however there are noticeable differences south of the equator and polewards of the sub-tropical dry regions. To put the differences between the PPE and the ERA-40 dataset in context, Fig. 4c and d show the anomaly from the ERA-40 dataset for both the PPE ensemble and the CMIP3 ensemble for SAT and precipitation, respectively. Most members of both ensembles are colder than the ERA-40 dataset, particularly at high northern latitudes, as one would expect from the warming that occurred between the pre-industrial and the late 20th century. Some members of the PPE are up to 4.0 °C warmer than the ERA-40 dataset in the Tropics, substantially warmer than any of the CMIP3 members. The low-latitude ocean heat transport in HadCM3 has been found to be relatively ineffective and on long timescales this can effectively act as a positive feedback on radiative imbalances at low-latitudes (Vellinga and Wu, 2008), which could explain why the anomalous behaviour is limited to this region. Both the PPE and CMIP3 ensembles show reduced precipitation at high latitudes compared with the ERA-40 dataset, which again fits with the changes expected between the pre-industrial and the late 20th century.

Most of the parameters that were perturbed in the PPE were related to uncertain atmospheric processes, particularly convective and cloud processes, and so one would expect differences throughout the atmospheric column. Figure 5 shows

a comparison between the ERA-40 1961–1990 mean vertical temperature and specific humidity profiles and both the PPE and the CMIP3 ensemble (Meehl et al., 2007b). The PPE follows the vertical temperature profile of the ERA-40 dataset, with all members remaining within ≈ 5 °C throughout whereas the CMIP3 ensemble shows a wider spread particularly at higher altitudes where models differ by up to 10 °C, see Fig. 5a and b. Both the PPE and the CMIP3 ensemble follow the ERA-40 humidity profile with humidity declining with altitude until it reaches around a few ppm at the tropopause. The ERA-40 dataset and many CMIP3 members show almost constant humidity throughout the stratosphere whereas some CMIP3 members and all PPE members show a continuing decline in humidity with altitude. All members of the PPE thus have a stratospheric water vapour content that is of the order of one tenth of the ERA-40 value.

We now evaluate the behaviour of the HadCM3 PPE with the “plausibility” approach of Edwards et al. (2011) in mind, using a small number of global-scale metrics of the PPE performance. We use the 1961–1990 ERA-40 average and other relevant datasets as the basis for judgments of plausibility, with the ranges from the CMIP3 ensemble shown for comparison. Supplement Table 2 shows the response of every ensemble member for the following global-scale metrics: global mean temperature, pole to equator temperature

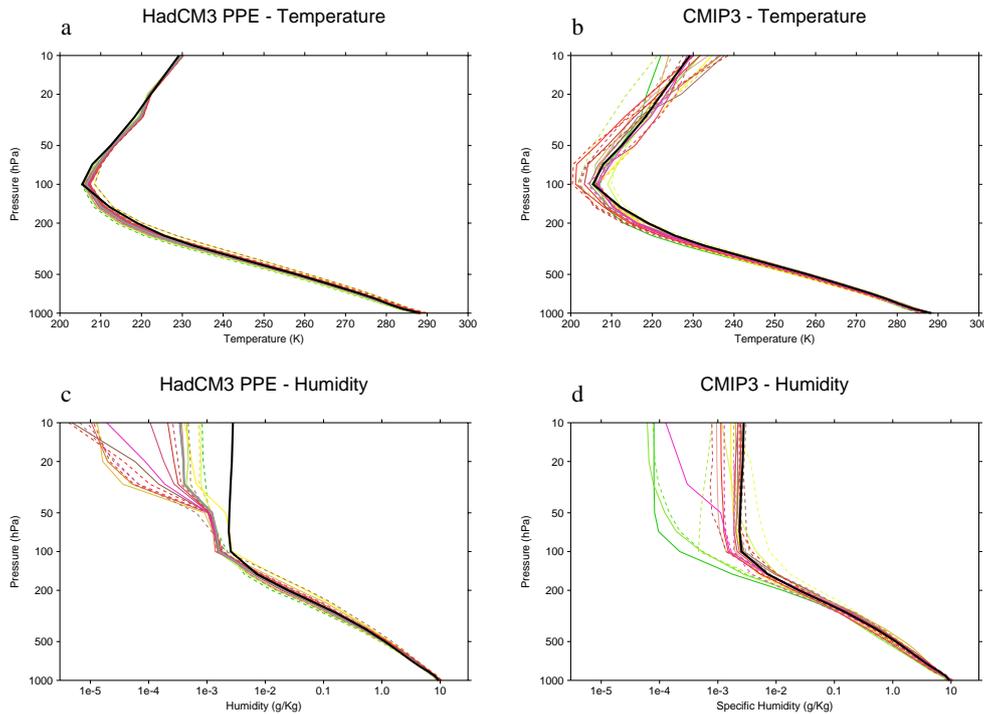


Fig. 5. The temperature (a, b) and specific humidity (c, d) throughout the atmospheric column for the PPE simulations and the CMIP3 ensemble. The standard version of HadCM3 is plotted in dark gray for the PPE simulations and the ERA-40 1961–1990 mean is plotted in black. Note that for the PPE cells below ground level, the values are extrapolated using an average lapse rate and included in the level mean.

difference (i.e. the average from 60° N to 90° N and between 30° S and 30° N), global mean precipitation, maximum overturning strength in the Atlantic, and global-mean pre-industrial humidity at 100 hPa and at 10 hPa. After 800 yr of pre-industrial control 6 of the PPE members were found to fall outside the target temperature of $13.6 \pm 2^\circ\text{C}$. The average pole to equator temperature difference for the PPE is 42.0°C which is greater than the ERA-40 average of 39.6°C for the period 1961–1990, however this may be partly due to the warming of the 20th century which will be greatest at high latitudes reducing the pole to equator temperature range. For the Atlantic overturning circulation we find a number of members show a substantially stronger overturning than the observed value of 18.7 Sv reported by Rayner et al. (2011), but only 2 accepted members of the PPE exceed the range of 12 to 24 Sv, given as the largest observational range in the IPCC Assessment Report 4 (Meehl et al., 2007a).

As was shown in Fig. 5 the specific humidity of the PPE is fairly close to the ERA-40 reanalysis at 100 hPa with an average value that is 70 % of the ERA-40 value but at 10 hPa the PPE has between 30 % and less than 1 % of the ERA-40 value. As absorption of longwave radiation scales approximately with the logarithm of the concentration of water vapour (Forster and Shine, 2002; IPCC, 2007), one could expect large changes in the water vapour greenhouse effect in the PPE (Held and Soden, 2000; Forster and Shine,

2002; Joshi et al., 2010). However, the absolute water vapour content is extremely low and major changes in the stratospheric radiation budget would be reflected in the upper atmospheric temperatures (Forster and Shine, 2002), but these have changed little from the standard model values, see Fig. 5a. Furthermore, the CMIP3 ensemble also includes many models with very dry stratospheres and this does not seem to be a critical shortcoming in these models.

Finally, as was shown in Fig. 4, the zonal mean climatology of the PPE generally compares well to the ERA-40 reanalysis, although some exceptions to this are found in the zonal temperature where some accepted members stand out clearly from the ERA-40 reanalysis and are well beyond the CMIP3 ensemble range. It is clear that the PPE shows a narrower range of behaviour than the CMIP3 ensemble and that the PPE shares many biases with the standard HadCM3 model configuration. Overall, we judge that the 21 accepted members of the PPE are plausible representations of the pre-industrial climate, and we retain all members which passed the initial temperature selection criterion.

3.4 Elevated CO₂ experiments

Figure 6 shows the change in the vertical profile of some atmospheric variables at the end of the 150 yr 4*CO₂ simulations. All members of the PPE show a temperature response to CO₂ that is broadly in line with the CMIP3 models

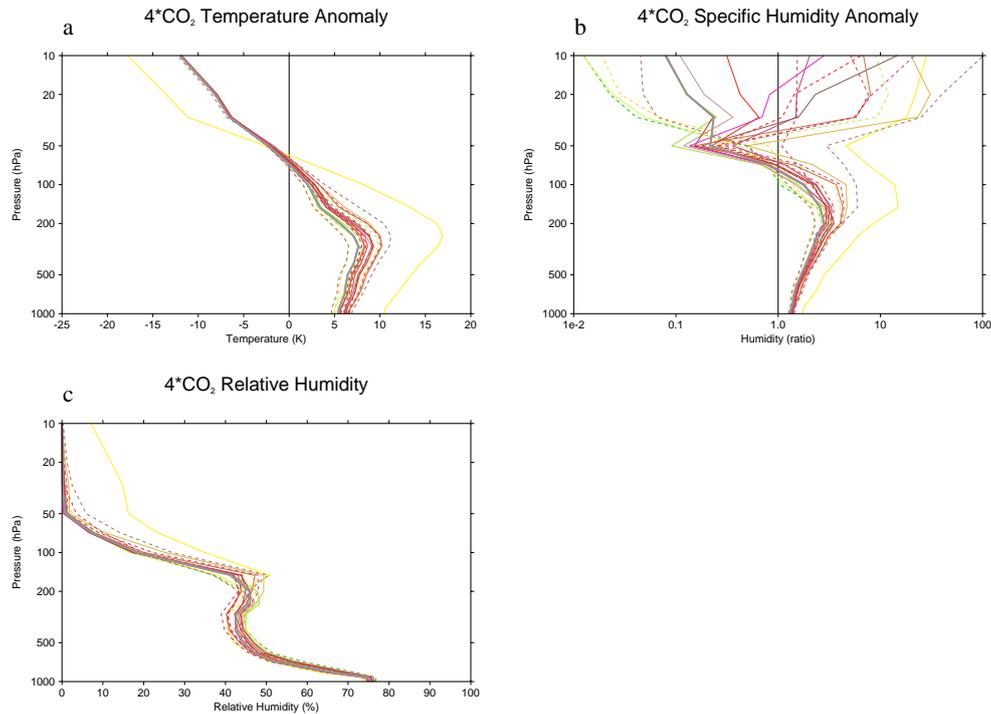


Fig. 6. The response of the PPE at 4*CO₂ throughout the atmospheric column; the anomaly between 4*CO₂ and the pre-industrial control for temperature (a); the ratio of specific humidity between 4*CO₂ and the pre-industrial control (b); and the absolute relative humidity at 4*CO₂ (c). The standard version of HadCM3 is shown in dark gray. Note that FOR cells below ground level the values are extrapolated using an average lapse rate and included in the level mean.

(IPCC, 2007), i.e. a warming in the troposphere, a rise of the tropopause, and a cooling of the stratosphere. There is a wide spread of temperature response, but all members show a peak warming in the mid-troposphere which is roughly 50 % greater than the surface warming. At higher altitudes most members of the PPE show the same cooling of ≈ 12 °C despite a broader range of surface temperature responses, $\approx 6 \pm 1.5$ °C; however, one accepted member shows a surface warming of around 11 °C and a high altitude cooling of around 18 °C, much greater than any other model. Figure 6b shows the change in specific humidity; up to 100 hPa the humidity increases for all members in a similar way, with the warmer runs showing a greater increase in humidity. However, at higher altitudes there is a very broad range of response with many members, including the standard model, showing humidity decreasing to a tenth, or even a hundredth, of the pre-industrial value and others showing a ten to a hundred fold increase in humidity. However, it is the absolute humidity that determines the magnitude of the radiative contribution of high altitude humidity and all but the two warmest PPE members remain below the ERA-40 1961–1990 stratospheric value of ≈ 3 ppm. The warmest PPE member shows a specific humidity of ≈ 10 ppm, i.e. more than three times greater than the 1961–1990 ERA-40 stratospheric humidity. For most members over most of the atmospheric column the

absolute change in relative humidity is less than 5 %, excluding around 150 hPa, where changes in tropopause height are evident. The warmest and second warmest accepted members (the solid yellow and dashed dark brown lines) stand out in the specific and relative humidity plots at altitudes above 100 hPa, showing specific humidity levels of order 100 and 10 times greater than the mean response and relative humidities of order 10 % and 1 % where other models show effectively 0 % relative humidity, see Fig. 6c.

The entrainment rate coefficient (ENTCOEF) plays the greatest role of any of the parameters in controlling high altitude humidity, as it controls the mixing of deep convective plumes with their surroundings (Sanderson et al., 2008a; Rougier et al., 2009; Murphy et al., 2004), and thus the mechanism by which water vapour can reach the upper atmosphere. High values of ENTCOEF are associated with high climate sensitivities in HadCM3 (Sanderson et al., 2008a; Forster and Shine, 2002; Joshi et al., 2010; Sanderson, 2011), and it has been suggested that changes in high altitude humidity contribute to this (Joshi et al., 2010). Figure 7a shows how the specific humidity at 30 hPa in the pre-industrial control simulations varies as a function of ENTCOEF; both low and high values of ENTCOEF are associated with higher specific humidities in the upper atmosphere. Figure 7b shows that the effect of ENTCOEF is much more marked in the

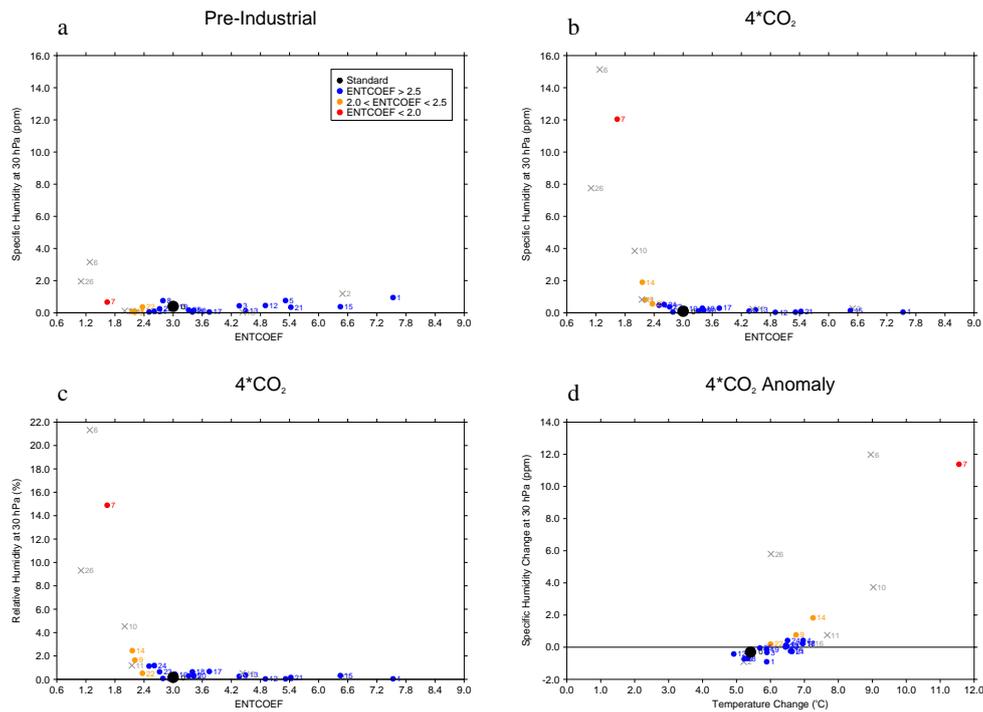


Fig. 7. The specific humidity at 30 hPa for the pre-industrial (a) and 4*CO₂ (b) as a function of ENTCOEF, the relative humidity at 30 hPa for 4*CO₂ as a function of ENTCOEF (c), and the change in specific humidity at 30 hPa and temperature between the 4*CO₂ and pre-industrial simulations (d). The standard model is shown as a larger black dot and failed runs were included in this plot as gray crosses to make clearer the role of the parameter. Values of ENTCOEF are also indicated with colours as indicated in the legend.

4*CO₂ simulation; members with a value of ENTCOEF below about 2.5 show specific humidities of between 0.5 and 15 ppm whereas others show very low specific humidities of less than 0.5 ppm (Joshi et al., 2010; Sanderson et al., 2008a; Sanderson, 2011). ENTCOEF also affects the relative humidity of the stratosphere which increases sharply for values of ENTCOEF below about 2.5 in the 4*CO₂ experiment, see Fig. 7c. Figure 7d shows that generally the members with the largest changes in high altitude humidity also show the largest increases in temperature at 4*CO₂.

Figure 8a shows how TOA radiative imbalance and temperature change evolve over the course of the 4*CO₂ experiment for the PPE ensemble. The method of Gregory et al. (2004) involves carrying out a regression on the joint evolution of temperature and radiative imbalance and is expected to provide an estimate of the initial radiative forcing perturbation and a final equilibrium temperature after only a few years or decades of such an instantaneous forcing experiment. Most ensemble members roughly follow the expected linear trend, although there is a common drift to higher temperatures towards the end of the runs as was seen by Gregory et al. (2004) for coupled models. The warmest ensemble member does not follow a linear evolution of TOA radiative imbalance and temperature at all, instead after a number of years the radiative imbalance ceases to reduce whilst temperature continues to rise at ≈ 0.4 K per decade

over the last 50 yr. This nonlinear climate response is also seen in a number of the failed ensemble members, i.e. those which fell outside of the temperature limits after 800 yr of pre-industrial control. The projected equilibrium temperatures of the 4*CO₂ simulations (4*CS) are shown in Fig. 8b, these are found by applying the Gregory method to the entire 150 yr timeseries. This long fitting period was applied to capture some of the deviation that the members with the greatest warming show. Most accepted ensemble members have a 4*CS in the range of 6.5–10.5 °C, with the warmest accepted member having an estimated 4*CS of 35 °C, although this is likely an underestimate due to the breakdown of the linear relation between increasing temperatures and decreasing TOA radiative imbalance. We also note a weak correlation between high pre-industrial temperature and climate sensitivity.

Figure 9a shows how precipitation and temperature evolve over the course of the 4*CO₂ experiment for the PPE ensemble. All members seem to follow the expected evolution of a rapid reduction in precipitation followed by a recovery as temperatures rise, which is approximately linear for all models including the warmest member (Bala et al., 2010). The response of precipitation to changes in radiative forcing has been conceptualized as consisting of a fast component or “precipitation adjustment”, corresponding to a change in the patterns of latent and specific heating particular to the

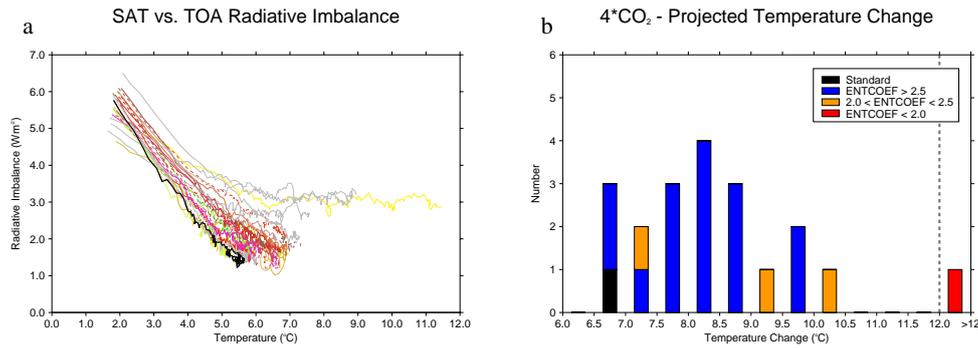


Fig. 8. The evolution of the TOA radiative imbalance and temperature over the full 150 yr of the 4*CO₂ simulation (a) and a histogram of projections of the equilibrium temperature for the 4*CO₂ simulation (b). The projections of equilibrium temperature are made using the Gregory method for all 150 yr of the simulations. The members which failed the pre-industrial temperature selection are included as gray lines in panel a. All changes are relative to the pre-industrial control simulations.

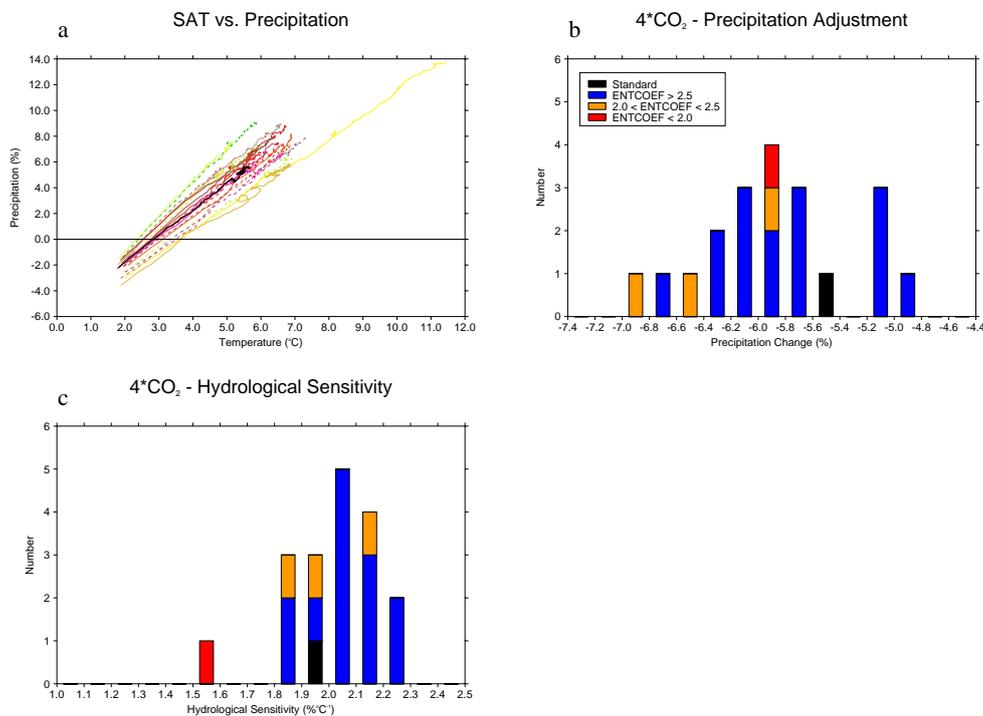


Fig. 9. The evolution of percentage precipitation change and temperature change over the full 150 yr of the 4*CO₂ simulation relative to the pre-industrial averages (a), and histograms of the estimated precipitation adjustment (b) and hydrological sensitivity (c) for the 4*CO₂ simulation. Linear fits to the first 50 yr of each simulation are used to calculate the precipitation adjustments, which are found from the intercept where $\Delta SAT = 0$, and the hydrological sensitivities, which are found from the gradients of the linear fits.

type of forcing, and a more or less independent slow component, that depends on the global mean temperature (Andrews et al., 2010; Bala et al., 2010). This slow, temperature-driven, component has been called the “hydrological sensitivity” and is measured in percentage change per degree of warming ($\% \text{ } ^\circ\text{C}^{-1}$) (Bala et al., 2010; Andrews et al., 2010). Calculating these values by applying a linear fit to the first 50 yr of the 4*CO₂ experiment we find that the PPE shows a range

of both fast and slow behaviour to the 4*CO₂ forcing with a precipitation adjustment of between -4.8 to -7.0 %, and a hydrological sensitivity of between 1.8 to 2.3 $\% \text{ } ^\circ\text{C}^{-1}$ (excepting the warmest accepted member which has a value of less than 1.6 $\% \text{ } ^\circ\text{C}^{-1}$), see Fig. 9b and c. At 2*CO₂ Andrews et al. (2009) showed an ensemble mean hydrological sensitivity of 2.8 $\% \text{ } ^\circ\text{C}^{-1}$ and a mean precipitation adjustment of 2.5 % for the CMIP3 models they considered, but in line

with our HadCM3 PPE results they find a hydrological sensitivity of $2.2\% \text{ } ^\circ\text{C}^{-1}$ and a precipitation adjustment of 3.0% (roughly half the $4*\text{CO}_2$ value shown here) for the HadSM3 model.

4 Discussion

To produce this non-flux adjusted perturbed parameter ensemble of HadCM3, an initial ensemble was produced using a maximin latin hypercube sampling approach and then a simple selection criteria was applied, based on the projected temperature response of the members from a 20 yr pre-industrial simulation. This approach allowed a large number of initial parameter combinations to be screened to exclude ensemble members that would produce an unrealistically warm or cold pre-industrial climate. This projection approach, based on the Gregory method (Gregory et al., 2004), was largely successful. Only 6 of the 27 members of the ensemble members failed to remain within the target temperature range of $13.6 \pm 2.0^\circ\text{C}$ after 800 yr of pre-industrial simulation (see Fig. 1c), corresponding to a success rate of 80 %, and those that failed were mostly within a few tenths of a degree of the target range. To aid future applications of this approach we highlight a number of issues with this approach. Firstly, many GCMs do not have a perfect radiative balance in equilibrium as they contain energy sources or sinks of up to a few W m^{-2} and so we suggest using the anomaly in TOA radiative imbalance rather than absolute radiative imbalance for making these projections (Mauritsen et al., 2012). Secondly, internal variability affects projections based on short timeseries, but longer runs or additional short simulations obviously increase computational cost and we judge that 20 yr was a good compromise. Thirdly, we assumed that a change in parameter values would be realised as an instantaneous change in radiative forcing and a change in the feedback processes of the model, but this is not necessarily the case. Analysis by Joshi et al. (2010), indicate that perturbations of the ENTCOEF parameter induce changes in the climate that do not follow the linear relation between temperature and radiative forcing that is commonly assumed (Gregory et al., 2004). Finally, if ocean parameters are perturbed the energy balance between atmosphere and ocean can change independently of the TOA radiative imbalance (Brierley et al., 2010). Our results suggest that the perturbation of the vertical ocean diffusivity parameter lead to an adjustment of the ocean-atmosphere energy balance, which affected our short-term temperature projections. Despite these difficulties, we believe our use of short-term projections using the Gregory et al. (2004) approach has been very successful, as running all 200 initial members for 800 yr would have required $\approx 160\,000$ model years as opposed to the $\approx 25\,000$ model years required with our approach.

The pre-industrial climates of the PPE were evaluated on the basis of a comparison to the ERA-40 dataset and by

consideration of the spread of the CMIP3 ensemble, see Sect. 3.3. Most PPE members show behaviour that is reasonably close to the ERA-40 data, but a number of members showed tropical temperatures up to 4°C warmer than the ERA-40 data and some showed high values for the Atlantic overturning circulation. All members of the PPE seemed to share some biases with the standard version of the HadCM3 model. This under-dispersion in the range of behaviour in the PPE is a common problem for PPEs, particularly those which perturb only a limited number of parameters (Yokohata et al., 2012). Overall, the initial selection criteria appears to have been very successful in that it removed all the members of the PPE which exhibited pre-industrial climatic conditions that are clearly implausible, but the small number of perturbed parameters has limited the range of behaviour within the PPE.

We find very high climate sensitivities for PPE members with low values of the ENTCOEF parameter, including one accepted member, and a number of the members which failed the temperature selection criterion, which show a clearly nonlinear response with unchecked warming in the later years of the $4*\text{CO}_2$ experiment, see Fig. 8. The mechanism behind this unchecked warming has not been definitively identified, but one plausible hypothesis presented in Sanderson (2011) is that the large increases in upper atmospheric humidity in response to warming in the warmest member constitutes a very large, positive, clear-sky longwave feedback which comes to dominate at higher temperatures. The climate sensitivity of HadCM3 has been found to rise rapidly for ENTCOEF values below the standard value of 3.0 (see Fig. 6 of Sanderson et al., 2008a and Fig. 6 of Rougier et al., 2009), and here we find that the stratospheric humidity response to elevated CO_2 also rises rapidly below ENTCOEF values of about 2.5. Most GCMs simulate a weak stratospheric humidity response to warming and small changes in relative humidity throughout the atmospheric column (Colman, 2001; Stuber et al., 2005), which is backed up by observations of recent warming (IPCC, 2007).

The presence of a nonlinear climate response in this PPE of HadCM3, suggests that other PPEs of HadCM3 may also include members which exhibit nonlinear climate responses which raises a number of issues for the climate sensitivity estimates made using these (Sanderson et al., 2008b; Yamazaki et al., 2013; Piani et al., 2005). Firstly, it is typical to assess the realism of PPE members with a control simulation and to discard unrealistic members of the PPE or down-weight them, however our nonlinear member did not perform too poorly in the pre-industrial and so would likely be included in estimates of climate sensitivity made in these previous studies. Secondly, estimates of the equilibrium temperature response are typically made by either applying the Gregory method or by fitting an exponential to the temperature timeseries; if a PPE member exhibits a nonlinear climate response these methods will not produce reliable estimates. Finally and perhaps most importantly there is the

question of how should a nonlinear climate response be interpreted? Are such runaway climate responses plausible? Or should PPE methods include a test of the climate response to elevated CO₂ concentrations which screens out members with nonlinear climate responses? We would suggest that the PPE members which produced climate sensitivity estimates towards the extreme upper end may be suspect; either under-estimating climate sensitivity if such nonlinear climate responses in HadCM3 are plausible or over-estimating climate sensitivity if they are implausible and should be excluded. We thus agree with the conclusion of Joshi et al. (2010) that the very high climate sensitivities found for low values of ENTCOEF are very unlikely in light of the observed response to warming. We suggest that future PPEs of HadCM3 should test whether a linear climate response best describes the response of the ensemble members and should consider restricting the range of ENTCOEF from 0.6–9.0 to 2.0–9.0 to mitigate these issues.

5 Conclusions

This study presents the methodology and some initial results from a perturbed parameter ensemble (PPE) of a non-flux adjusted, fully-coupled CMIP3-era GCM. The purpose has been to create a modestly-sized PPE to explore the effects of parametric uncertainty on climate and paleo-climate experiments. 200 different versions of the HadCM3 model were generated with 8 continuous parameters varied. 21 ensemble members of the HadCM3 model (Gordon et al., 2000), including the standard configuration, were selected from these 200 using an estimation of the equilibrium pre-industrial temperature to constrain the ensemble, i.e. models with projected temperatures within $13.6 \pm 2^\circ\text{C}$ were kept (Brohan et al., 2006; Jones et al., 1999). However, an additional 6 members which were projected to be within the target temperature range were either warmer or colder after the 800 yr control than the target temperature range and thus were excluded from the ensemble. Despite the ocean not reaching equilibrium after 800 yr the pre-industrial control surface climatology of the ensemble compares well on the whole to the ERA-40 dataset and the CMIP3 ensemble, except in the Tropics for some members (Meehl et al., 2007b). We find that not using flux-adjustment and instead constraining our ensemble on the pre-industrial equilibrium temperature has not led to a serious curtailment of parameter space as has been suggested previously (Collins et al., 2006). Applying the ensemble to a quadrupled CO₂ experiment reinforced earlier findings of links between low values of the entrainment rate coefficient, large increases in high altitude humidity and high climate sensitivities in HadCM3 (Joshi et al., 2010). In fact, one member with a low value of the entrainment rate coefficient exhibited a clearly nonlinear climate response at 4*CO₂ after a few decades, showing a rapid warming without a reduction in TOA radiative imbalance. This raises the question

of whether the plausibility of ensemble members' response to elevated CO₂ concentrations should be evaluated alongside historical performance in perturbed parameter ensemble studies.

Supplementary material related to this article is available online at <http://www.geosci-model-dev.net/6/1447/2013/gmd-6-1447-2013-supplement.zip>.

Acknowledgements. P. J. Irvine acknowledges support in the form of a NERC PhD studentship. We acknowledge the CMIP3 modelling groups for making their model output available for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving this data, and the WCRP's Working Group on Coupled Modelling (WGCM) for organising the model data analysis activity. The WCRP CMIP3 multi-model dataset is supported by the Office of Science, US Department of Energy. ECMWF ERA-40 data used in this study have been obtained from the ECMWF Data Server.

Edited by: R. Redler

References

- Andrews, T., Forster, P. M., and Gregory, J. M.: A Surface Energy Perspective on Climate Change, *J. Climate*, 22, 2557–2570, doi:10.1175/2008jcli2759.1, 2009.
- Andrews, T., Forster, P. M., Boucher, O., Bellouin, N., and Jones, A.: Precipitation, radiative forcing and global temperature change, *Geophys. Res. Lett.*, 37, L14701, doi:10.1029/2010gl043991, 2010.
- Bala, G., Caldeira, K., and Nemani, R.: Fast versus slow response in climate change: implications for the global hydrological cycle, *Clim. Dynam.*, 35, 423–434, doi:10.1007/s00382-009-0583-y, 2010.
- Brierley, C. M., Collins, M., and Thorpe, A. J.: The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model, *Clim. Dynam.*, 34, 325–343, doi:10.1007/s00382-008-0486-3, 2010.
- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.-Atmos.*, 111, D12106, doi:10.1029/2005jd006548, 2006.
- Cattle, H. and Crossley, J.: MODELING ARCTIC CLIMATE-CHANGE, *Philos. T. Roy. Soc. London Series a*, 352, 201–213, 1995.
- Collins, M., Booth, B. B. B., Harris, G. R., Murphy, J. M., Sexton, D. M. H., and Webb, M. J.: Towards quantifying uncertainty in transient climate change, *Clim. Dynam.*, 27, 127–147, doi:10.1007/s00382-006-0121-0, 2006.
- Collins, M., Brierley, C. M., MacVean, M., Booth, B. B. B., and Harris, G. R.: The sensitivity of the rate of transient climate change to ocean physics perturbations, *J. Climate*, 20, 2315–2320, doi:10.1175/jcli4116.1, 2007.

- Collins, M., Booth, B., Bhaskaran, B., Harris, G., Murphy, J., Sexton, D., and Webb, M.: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles, *Clim. Dynam.*, 36, 1737–1766, doi:10.1007/s00382-010-0808-0, 2010.
- Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, *Geosci. Model Dev.*, 4, 1051–1075, doi:10.5194/gmd-4-1051-2011, 2011.
- Colman, R. A.: On the vertical extent of atmospheric feedbacks, *Clim. Dynam.*, 17, 391–405, doi:10.1007/s003820000111, 2001.
- Covey, C., AchutaRao, K. M., Cubasch, U., Jones, P., Lambert, S. J., Mann, M. E., Phillips, T. J., and Taylor, K. E.: An overview of results from the Coupled Model Intercomparison Project, *Global Planet. Change*, 37, 103–133, doi:10.1016/s0921-8181(02)00193-5, 2003.
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, *Clim. Dynam.*, 15, 183–203, 1999.
- Edwards, N. R., Cameron, D., and Rougier, J.: Pre-calibrating an intermediate complexity climate model, *Clim. Dynam.*, 37, 1469–1482, 2011.
- Forster, P. M. D. and Shine, K. P.: Assessing the climate impact of trends in stratospheric water vapour, *Geophys. Res. Lett.*, 29, 1086, doi:10.1029/2001gl013909, 2002.
- Frame, D. J., Aina, T., Christensen, C. M., Faull, N. E., Knight, S. H. E., Piani, C., Rosier, S. M., Yamazaki, K., Yamazaki, Y., and Allen, M. R.: The climateprediction.net BBC climate change experiment: design of the coupled model ensemble, *Philos. T. Roy. Soc. a*, 367, 855–870, doi:10.1098/rsta.2008.0240, 2009.
- Gent, P. R. and McWilliams, J. C.: Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr.*, 20, 150–155, 1990.
- Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dynam.*, 16, 147–168, 2000.
- Gregoire, L., Valdes, P., Payne, A., and Kahana, R.: Optimal tuning of a GCM using modern and glacial constraints, *Clim. Dynam.*, 37, 705–719, doi:10.1007/s00382-010-0934-8, 2010.
- Gregory, D. and Rowntree, P. R.: A Mass Flux Convection Scheme with Representation of Cloud Ensemble Characteristics and Stability-Dependent Closure, *Mon. Weather Rev.*, 118, 1483–1506, 1990.
- Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, *Geophys. Res. Lett.*, 31, L03205, doi:10.1029/2003gl018747, 2004.
- Held, I. M. and Soden, B. J.: Water vapour feedback and global warming, *Annu. Rev. Energ. Environ.*, 25, 441–475, doi:10.1146/annurev.energy.25.1.441, 2000.
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C.: Climate change 2001: the scientific basis, Cambridge University Press Cambridge, 2001.
- IPCC: Climate Change 2007: The Physical Science Basis, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, 996 pp., 2007.
- Jackson, L., Vellinga, M., and Harris, G.: The sensitivity of the meridional overturning circulation to modelling uncertainty in a perturbed physics ensemble without flux adjustment, *Clim. Dynam.*, 39, 277–285, doi:10.1007/s00382-011-1110-5, 2012.
- Jones, P. D., New, M., Parker, D. E., Martin, S., and Rigor, I. G.: Surface air temperature and its changes over the past 150 years, *Rev. Geophys.*, 37, 173–199, 1999.
- Joshi, M. M., Webb, M. J., Maycock, A. C., and Collins, M.: Stratospheric water vapour and high climate sensitivity in a version of the HadSM3 climate model, *Atmos. Chem. Phys.*, 10, 7161–7167, doi:10.5194/acp-10-7161-2010, 2010.
- Klocke, D., Pincus, R., and Quaas, J.: On Constraining Estimates of Climate Sensitivity with Present-Day Observations through Model Weighting, *J. Climate*, 24, 6092–6099, doi:10.1175/2011JCLI4193.1, 2011.
- Knight, C. G., Knight, S. H. E., Massey, N., Aina, T., Christensen, C., Frame, D. J., Kettleborough, J. A., Martin, A., Pascoe, S., Sanderson, B., Stainforth, D. A., and Allen, M. R.: Association of parameter, software, and hardware variation with large-scale behaviour across 57,000 climate models, *P. Natl. Acad. Sci. USA*, 104, 12259–12264, doi:10.1073/pnas.0608144104, 2007.
- Masson, D. and Knutti, R.: Climate model genealogy, *Geophys. Res. Lett.*, 38, L08703, doi:10.1029/2011gl046864, 2011.
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, 4, M00A01, doi:10.1029/2012MS000154, 2012.
- Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., Kitoh, A., Knutti, R., Murphy, J., and Noda, A.: Global climate projections. Chapter 10 in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S. and Qin, D. M., Cambridge University Press, Cambridge, United Kingdom, and New York, NY, USA, 996 pp., 2007a.
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J., and Taylor, K. E.: The WCRP CMIP3 multimodel dataset – A new era in climate change research, *B. Am. Meteorol. Soc.*, 88, 1383, doi:10.1175/bams-88-9-1383, 2007b.
- Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., and Collins, M.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768–772, doi:10.1038/nature02771, 2004.
- Piani, C., Frame, D. J., Stainforth, D. A., and Allen, M. R.: Constraints on climate change from a multi-thousand member ensemble of simulations, *Geophys. Res. Lett.*, 32, L23825, doi:10.1029/2005gl024452, 2005.

- Rayner, D., Hirschi, J. J. M., Kanzow, T., Johns, W. E., Wright, P. G., Frajka-Williams, E., Bryden, H. L., Meinen, C. S., Baringer, M. O., Marotzke, J., Beal, L. M., and Cunningham, S. A.: Monitoring the Atlantic meridional overturning circulation, *Deep-Sea Res. II*, 58, 1744–1753, doi:10.1016/j.dsr2.2010.10.056, 2011.
- Rodwell, M. J. and Palmer, T. N.: Using numerical weather prediction to assess climate models, *Q. J. Roy. Meteorol. Soc.*, 133, 129–146, doi:10.1002/qj.23, 2007.
- Rougier, J., Sexton, D. M. H., Murphy, J. M., and Stainforth, D.: Analysing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments, *J. Climate*, 22, 3540–3557, doi:10.1175/2008jcli2533.1, 2009.
- Sanderson, B. M.: A Multimodel Study of Parametric Uncertainty in Predictions of Climate Response to Rising Greenhouse Gas Concentrations, *J. Climate*, 24, 1362–1377, doi:10.1175/2010jcli3498.1, 2011.
- Sanderson, B. M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D. J., Ingram, W. J., Piani, C., Stainforth, D. A., Stone, D. A., and Allen, M. R.: Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes, *J. Climate*, 21, 2384–2400, doi:10.1175/2008jcli1869.1, 2008a.
- Sanderson, B. M., Piani, C., Ingram, W. J., Stone, D. A., and Allen, M. R.: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations, *Clim. Dynam.*, 30, 175–190, doi:10.1007/s00382-007-0280-7, 2008b.
- Shiogama, H., Watanabe, M., Yoshimori, M., Yokohata, T., Ogura, T., Annan, J., Hargreaves, J., Abe, M., Kamae, Y., O’ishi, R., Nobui, R., Emori, S., Nozawa, T., Abe-Ouchi, A., and Kimoto, M.: Perturbed physics ensemble using the MIROC5 coupled atmosphere-ocean GCM without flux corrections: experimental design and results, *Clim. Dynam.*, 39, 3041–3056, doi:10.1007/s00382-012-1441-x, 2012.
- Smith, R. N. B.: Experience and developments with the layer cloud and boundary layer mixing schemes in the UK Meteorological Office Unified Model, *Proc. ECMWF/GCSS workshop on parameterisation of the cloud-topped boundary layer*, ECMWF, Reading, England, 1993.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L.: *Climate Change 2007: The Physical Science Basis*, in: *Climate Change 2007: The Physical Science Basis. Contribution of working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, 996 pp., 2007.
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R.: Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, 433, 403–406, doi:10.1038/nature03301, 2005.
- Stuber, N., Ponater, M., and Sausen, R.: Why radiative forcing might fail as a predictor of climate change, *Clim. Dynam.*, 24, 497–510, doi:10.1007/s00382-004-0497-7, 2005.
- Tang, B. X.: A theorem for selecting oa-based Latin Hypercubes using a distance criterion, *Commun. Stat.-Theory Methods*, 23, 2047–2058, doi:10.1080/03610929408831370, 1994.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *B. Am. Meteorol. Soc.*, 93, 485–498, 2012.
- Vellinga, M. and Wu, P.: Relations between northward ocean and atmosphere energy transports in a coupled climate model, *J. Climate*, 21, 561–575, 2008.
- Wilson, M. F. and Henderson-Sellers, A.: A global archive of land cover and soils data for use in general-circulation climate models, *J. Climatol.*, 5, 119–143, 1985.
- Yamazaki, K., Rowlands, D. J., Aina, T., Blaker, A. T., Bowery, A., Massey, N., Miller, J., Rye, C., Tett, S. F. B., Williamson, D., Yamazaki, Y. H., and Allen, M. R.: Obtaining diverse behaviours in a climate model without the use of flux adjustments, *J. Geophys. Res.-Atmos.*, doi:10.1002/jgrd.50304, 2013.
- Yokohata, T., Webb, M. J., Collins, M., Williams, K. D., Yoshimori, M., Hargreaves, J. C., and Annan, J. D.: Structural Similarities and Differences in Climate Responses to CO₂ Increase between Two Perturbed Physics Ensembles, *J. Climate*, 23, 1392–1410, doi:10.1175/2009jcli2917.1, 2010.
- Yokohata, T., Annan, J., Collins, M., Jackson, C., Tobis, M., Webb, M., and Hargreaves, J.: Reliability of multi-model and structurally different single-model ensembles, *Clim. Dynam.*, 39, 599–616, doi:10.1007/s00382-011-1203-1, 2012.