Geosci. Model Dev. Discuss., 6, 841–892, 2013
www.geosci-model-dev-discuss.net/6/841/2013/
doi:10.5194/gmdd-6-841-2013

Geoscientific Model Development Discussions
Open Access

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

P. J. Irvine[1], L Gregoire[2], D. J. Lunt[2], and P. J. Valdes[2]

[1]Institute for Advanced Sustainability Studies, Potsdam, Germany
[2]School of Geographical Sciences, University of Bristol, UK

Received: 9 January 2013 – Accepted: 15 January 2013 – Published: 6 February 2013

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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 displays a similar range of behavior as seen in multi-model ensembles (MMEs). 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 response to the parameter perturbations was projected to be outside the range of $13.6 \pm 2\,^{\circ}C$, i.e. near to the observed pre-industrial global mean, were discarded. 21 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 3 experiments; a 800 yr pre-industrial, a 150 yr quadrupled CO_2, and a 150 yr 1 % CO_2 rise per annum simulation. The behavior of the PPE for the pre-industrial control compared well to 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 vapor concentrations in response to elevated CO_2 and some show implausibly high climate sensitivities, and as such some of these members 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 with a range of behavior similar to the CMIP3 ensemble and a simple methodology which would be applicable to other GCMs.

GMDD
6, 841–892, 2013

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 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 modeling 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, unless a long program of model development is begun, 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

GMDD
6, 841–892, 2013

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, non-linear 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 try and achieve a 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 between 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 AR4 did not need flux-adjustment was seen as an improvement over earlier models (Solomon et al., 2007).

1.2 Background on assessment of ensembles

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) analyzed a perturbed parameter ensemble of HadCM3 using the climate prediction index, a method which

GMDD
6, 841–892, 2013

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). The climate prediction index aggregates a large number of different tests of model "realism", or similarity to observations, and so highly unrealistic behaviour in one aspect can be counteracted by reasonably realistic behaviour in others. For example, a member of the ensemble with a very low value of the entrainment rate coefficient was kept in the ensemble of Murphy et al. (2004) despite being numerically unstable without flux-adjustment and which showed a stratospheric water vapour response to warming that was substantially larger than the observed response, and hence arguably unrealistic (Joshi 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, although defining a multi-variate measure that excludes unrealistic models but retains an appropriate fraction of the ensemble may be challenging. 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. For example, in their study on the Atlantic overturning circulation response to warming, a model was judged plausible based on whether its ocean temperature, salinity, and maximum Atlantic streamfunction fell within a broad range of "physical" values. In this study we do not attempt to rank the ensemble members but we apply the simple and transparent pre-calibration approach of Edwards et al. (2011) to test whether or not the ensemble members display "plausible" representations of the climate system.

To illustrate the importance of such ensemble selection criteria on climate projections we will take the example of climate sensitivity, i.e. the equilibrium temperature response

GMDD
6, 841–892, 2013

to doubling CO_2, which is one of the key uncertain aspects of the climate system. The GCMs investigated by the IPCC AR4 surveyed climate models that showed a reasonable similarity to observed climate and these models had a range of climate sensitivities of between 2.1 °C and 4.4 °C, with a mean value of 3.2 °C (IPCC, 2007). These GCM results, along with other evidence, led to the IPCC stating that climate sensitivity was "likely to be in the range 2.0 to 4.5 °C with a best estimate of about 3 °C", i.e. greater than 66 % probability of being in this range (IPCC, 2007). Perturbed parameter ensembles have reported much broader ranges of climate sensitivity than those in the IPCC AR4 multi-model ensemble, e.g. Stainforth et al. (2005) found acceptable model versions with climate sensitivities from 2.0–11.0 °C, a range which Piani et al. (2005) constrained using observational data finding a mean climate sensitivity of 3.3 °C, with 5 % and 95 % confidence bounds of 2.2 and 6.8 °C, respectively. The choice of priors for the parameter ranges plays an important role for PPEs, as does the choice of using a uniform prior on the climate sensitivity or feedback parameter (Frame et al., 2005). Flux adjustments effectively removes constraints on the model behavior, by for example cancelling an imbalanced top-of-atmosphere radiative budget by adding a heat flux to the surface ocean, and allow models with arguably implausible behavior to be retained.

1.3 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 (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" behavior when compared against

GMDD
6, 841–892, 2013

a number of key metrics drawn from a comparison with observations and the members of the World Climate Research Program's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset, e.g. global mean temperature, pole to equator gradient and Atlantic Overturning strength. The behavior of the remaining members of the ensemble is then evaluated at elevated CO_2 levels. The methodology and selection approach 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. Supplement is included which consists of 3 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 Model HadGEM2 (Collins et al., 2011), makes it a powerful tool for multi-millenial 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 min 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

GMDD
6, 841–892, 2013

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 atmospheric, oceanic and sea ice parameters (Table 1). These include the 6 atmospheric parameters modified in Stainforth et al. (2005), the entrainment rate coefficient (ENTCOEF), the ice-fall speed (VF1), the critical relative humidity (RHCRIT), the droplet to rain conversion rate (CT), the droplet to rain conversion

GMDD
6, 841–892, 2013

threshold over land and sea (CW_LAND/SEA, two parameters that are perturbed as one), the empirically adjusted cloud fraction (EACF); the minimum sea-ice 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. (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 low 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. (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 are sampled using a uniform prior on parameter value except for VDIFF which uses a uniform prior on the power of the parameter value, i.e. 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 communications, 2012).

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 the sections of each parameter, ensuring

GMDD
6, 841–892, 2013

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 spin-up of the standard version of hadCM3 (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

GMDD
6, 841–892, 2013

et al., 2010; Shiogama et al., 2012). The projection of temperature and the initial radiative forcing is made from a linear regression of the temperature and the TOA radiative imbalance, in the manner of Gregory et al. (2004). 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 ±2.0 °C was decided upon as being approximately equal to the largest difference between the pre-industrial absolute temperature of a member of the 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 ran some further simulations. As we are modifying the ocean and atmosphere of the model we follow the procedure outlined in Collins et al. (2007), of running a 500 yr spin-up to allow the model's temperature to approach close to the equilibrium value but due to computational constraints not long enough to equilibrate the deep ocean (this is even true for the standard published HadCM3 which also exhibits some small drift in deep ocean temperatures). After the spin-up 3 further simulations were started, a 300 yr pre-industrial control run, a 150 yr simulation with an instantaneous quadrupling of CO_2 ($4 \times CO_2$) and a 150 yr simulation with CO_2 rising by 1 % per annum (1 %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, b shows the projected temperature and estimated initial top-of-atmosphere 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

GMDD
6, 841–892, 2013

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 ±2.0 °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. This discrepancy between the projected and realized temperature response arises for three reasons: internal variability that affects the projections, a change in the atmosphere-ocean heat exchange (discussed in the next section) and the assumption that the parameter perturbation acts like an instantaneous radiative forcing perturbation does not hold completely and breaks down for some parameter perturbations (Joshi et al., 2010). Of the 27 members selected by the Gregory method ~80 % remained within the target window, 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 additional 6 failed members. The failed members will be retained and shown only in plots that illustrate the role of the parameters. Supplementary 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. Figure 2a, b shows that most members of the PPE behave as if an instantaneous radiative forcing had been applied,

GMDD
6, 841–892, 2013

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 both fast and slow changes in precipitation. We see 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 2e shows the ocean temperature at 2.7 km; all the members of the ensemble show a trend in temperature which is not diminished at the end of the 800 yr. In fact, even the standard HadCM3 simulation, run from the end of a very long spin-up without parameters being perturbed still shows a slight cooling. Figure 2f shows the evolution of the maximum meridional circulation in the Atlantic; most members remain close to the standard model's condition with an overturning strength of ~ 18 Sv but a number of members show increased overturning of around ~ 25 Sv and some also show a large increase in variability. Although the surface variables appear to be near equilibrium after 800 yr of pre-industrial control simulation, the deep ocean temperature and the overturning are not.

The 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., 2008). Figure 3 shows that the entrainment rate coefficient (ENTCOEF) and the ice fall speed (VF1) have the greatest influence on pre-industrial temperature, with low values of both parameters tending to give warmer conditions. However the coldest failed member of the PPE has the lowest value of ENTCOEF and one of the highest value of VF1, which finds some agreement with the results of Sanderson et al. (2008) who found that this combination led to a member with very low climate sensitivity. Figure 4 shows that VF1 also exerts a strong control over the cloud fraction in the control runs with lower values of VF1 having a cloud fraction greater than 55 % and the highest values of VF1

GMDD
6, 841–892, 2013

having a cloud fraction lower than 50 % as seen in other studies (Wu, 2002; Sanderson et al., 2008). A regression of the sea-ice low albedo parameter (ALPHAM) and surface air temperature against the pre-industrial sea-ice fraction shows that surface air temperature plays a much more significant role in determining the sea-ice fraction than ALPHAM. The effect of these and other atmospheric parameters on the HadCM3 model have been explored in detail by a number of other studies and so the interested reader should refer to these for further information (Collins et al., 2006; Murphy et al., 2004; Sanderson et al., 2008; Knight et al., 2007).

Figure 5 shows 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 $\sim 1.4\,\mathrm{W\,m^{-2}}$ compared with $\sim 0.6\,\mathrm{W\,m^{-2}}$ in the standard model), absorbing energy that would otherwise have warmed the model surface (the opposite is also true for low values of VDIFF with up to $\sim 0.15\,\mathrm{W\,m^{-2}}$ less energy absorbed by the ocean). However, high values of VDIFF on its own have been found to have an opposite effect, raising global mean temperatures even if by only a few tenths of a degree (Collins et al., 2007; Brierley et al., 2010). It is 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 communications, 2012). 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), and vice versa.

The deep ocean has not adjusted fully to the parameter perturbations by the end of the pre-industrial control; Fig. 2 shows that the ocean is not in equilibrium but significant changes in the ocean have occurred by the end of the 800 yr pre-industrial control. Figure 6 shows the depth profile of the ocean potential temperature and salinity at the end of the simulations, showing that the condition of the ocean has changed markedly across the ensemble. Changes in ocean temperature at depth are determined more by

GMDD
6, 841–892, 2013

the atmospheric variables than by the vertical diffusivity parameter (not shown). However, Fig. 7a shows that Atlantic overturning strength is associated with VDIFF, 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 show little change; this matches the results of Brierley et al. (2010). This is despite the fact that members of the PPE with the highest values of VDIFF also have the warmest control run simulations and a warmer climate is expected to weaken overturning due to the increased freshwater flux at high latitudes and reduced sea-ice formation (see Fig. 7) (Solomon et al., 2007).

3.3 Comparison of PPE with CMIP3 ensemble

Figure 8 shows the annual and zonal mean state of the pre-industrial climate in the PPE and compares this with the CMIP3 ensemble (Meehl et al., 2007b). Figure 8a, b shows that the zonal mean temperatures of the PPE and the CMIP3 ensemble show a similar distribution. However, there are PPE members that are a few degrees warmer in the tropics than any of the CMIP3 ensemble. The zonal precipitation, Fig. 8c, d, of the PPE shows a similar overall structure but much less variance than the CMIP3 ensemble. All members of the PPE show a sharp peak in precipitation north of the equator with many of the ensemble members exceeding the CMIP3 range in this region. The TOA radiative imbalance, Fig. 8e, f, of the PPE is similar to the CMIP3 ensemble except near the equator where there is a greater spread than in the CMIP3 ensemble. Figure 9 shows the anomaly between the PPE members and the CMIP3 ensemble and how this compares to the spread seen in the CMIP3 ensemble for the same variables shown in Fig. 8. This figure makes clear that some members show behavior in the Tropics that is clearly outside the range of behavior seen in the CMIP3 ensemble. Overall the temperature constraint has been sufficient to produce a PPE that shares similar features with the CMIP3 ensemble and only a few of the members which passed the selection criteria would stand out amongst the CMIP3 models. We believe that the members of the

GMDD
6, 841–892, 2013

PPE are thus plausible representations of the climate system and useful as a means of investigating parametric uncertainty in model studies of the climate.

Most of the parameters that were perturbed in the PPE were related with uncertain atmospheric properties, particularly convection and clouds, and so differences within the ensemble are expected to be greatest in the atmospheric column. Figure 10 shows a comparison between the vertical temperature and specific humidity profile of the PPE and the CMIP3 ensemble (Meehl et al., 2007b). The PPE shows little spread in the vertical temperature profile except at the surface in contrast to the CMIP3 ensemble which shows the opposite, with general agreement at the surface and a large spread of temperatures at higher altitudes. This must in part be due to the wider spread in control surface air temperatures in the PPE and indicates that the PPE has not substantially perturbed the dynamics that control upper atmosphere temperature in HadCM3. In Fig. 10c the PPE shows a broad spread of behavior in the humidity profile, with the greatest spread at the highest altitudes (note that humidity is plotted on a logarithmic scale). Both the PPE and the CMIP3 ensemble show humidity declining with altitude until around 100 mB where it reaches a value of order one ppm. Beyond this point most members of the CMIP3 ensemble show near constant humidity or variations in humidity smaller than one order of magnitude whereas the PPE members' response varies from a more or less constant humidity to a continued decline to parts per billion of water vapour. The parameters perturbed in the PPE have had little effect on the vertical temperature profile, clearly not covering the range of behavior shown in the CMIP3 ensemble. For humidity the parameter perturbation has had a greater effect with the PPE covering a range of specific humidity concentrations broader than that seen in the CMIP3 ensemble. These changes indicate that although six key atmospheric parameters have been perturbed, the PPE does not cover the range of different upper atmospheric behavior seen in the CMIP3 ensemble, as has been noted for other HadCM3 ensembles (Collins et al., 2010).

Water vapor accounts for about 60 % of the natural greenhouse effect for clear skies (Kiehl and Trenberth, 1997), and water vapor at high altitudes is disproportionately

GMDD
6, 841–892, 2013

responsible for this natural greenhouse effect despite the much lower concentrations of water found at high altitudes (Solomon et al., 2007). Absorption of longwave radiation scales approximately with the logarithm of concentration of water vapour, an absolute change in water vapor concentration at the drier high altitudes, where the water vapor absorption bands are not saturated, is more significant than at the moist lower altitudes (Forster and Shine, 2002; IPCC, 2007). As the planet warms in response to elevated CO_2 concentrations the water vapor concentration in the atmosphere is expected to rise, increasing the water vapor greenhouse effect, and warming the planet further; this water vapor feedback is believed to be one of the most important radiative forcing feedbacks in the climate system (IPCC, 2007). For the PPE there is a large range of specific humidity concentrations at high altitudes which constitutes a large range of changes in humidity from the standard model configuration and so a large change in the water vapor greenhouse effect (Held and Soden, 2000; Forster and Shine, 2002; Joshi et al., 2010). Despite these large changes in the water vapor greenhouse effect across the PPE most members have remained within 2.0 °C of the pre-industrial observed surface air temperature.

3.4 Evaluation of pre-industrial performance

We now evaluate the behavior of the HadCM3 PPE with the "plausibility" approach of Edwards et al. (2011) in mind, using the CMIP3 ensemble as the basis for judgments of plausibility. The behavior of the PPE will be evaluated with a small number of global-scale metrics and by comparing the vertical profiles and zonal-mean averages of the PPE to the CMIP3 ensemble. Supplementary Table 2 shows the response of every ensemble member for the following global-scale metrics: global mean temperature, pole to equator temperature 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, pre-industrial humidity at 100 mB and at 10 mB. After 800 yr of pre-industrial control 6 of the PPE members were found to fall outside the target temperature of 13.6 ± 2 °C. If we take the Atlantic overturning circulation as another constraint we find

GMDD
6, 841–892, 2013

2 accepted members of the PPE with a relatively high value of 25 Sv but all others are within the range of 12–24 Sv, given as the largest observational range in the IPCC AR4 (Meehl et al., 2007a). At high altitude a wide range of values for specific humidity are found in both the PPE and the CMIP3 ensemble, with the PPE results mostly within the range of the CMIP3 ensemble, the exception being that almost half the members of the PPE have a specific humidity at 10 mB lower than any of the members of the CMIP3 ensemble. This does not seem problematic as the CMIP3 ensemble spans a broad range of very low humidity values from ∼ 10 ppb to ∼ 1000 ppb at 10 mB. A crude comparison of the zonal mean climatology of the PPE and the CMIP3 ensemble shows that the behavior of most of the members of the PPE is within the range seen in the CMIP3 ensemble and so could be judged to be consistent with previous work, see Fig. 9. Exceptions to this are found in the zonal temperature and radiative balance plots where some accepted members stand out clearly from the CMIP3 ensemble range, and for precipitation where many members show a peak in precipitation north of the equator that is beyond the range seen in the CMIP3 ensemble. Even though some members fall outside the CMIP3 range, we judge that they are still plausible, and we retain all members which passed the initial selection criterion.

3.5 Elevated CO_2 experiments

Figure 11 shows the change in the vertical profile of some atmospheric variables with height at the end of the 150 yr quadrupled CO_2 level simulations. All members of the PPE show a temperature response to CO_2 that is broadly in line with the CMIP3 models (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 proportional to the surface warming. At higher altitudes most members of the PPE show the same cooling of ∼ 12 °C despite the spread of ∼ 4 °C in the surface temperature signals of these members; however, one accepted member shows a much greater surface warming and a greater high altitude cooling than all other accepted members. Figure 11c, d

GMDD
6, 841–892, 2013

shows the changes in specific humidity; up to 100 mb the humidity increases for all members in a similar way, with the warmer runs showing a greater increase in humidity. 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. The specific humidity in Fig. 11 is plotted as the logarithm of the change from the pre-industrial, as the change in radiative effect scales with the logarithm of the moisture content (IPCC, 2007). However, the absolute humidity level determines the absolute magnitude of the radiative contribution of high altitude humidity and only the two warmest accepted models show specific humidities greater than 1 ppmv. Most models of the CMIP3 ensemble show a doubling or tripling of upper level humidity on doubling CO_2 concentrations (although this excludes the HadCM3 model which shows a drying) (Meehl et al., 2007b), much smaller than the 10 to 100 fold increase of some members of the PPE. Figure 11e, f shows that for most members over most of the atmospheric column the absolute change in relative humidity is less than 5 % (excluding around 150 mb, where changes in tropopause height are evident) but above 50 mB the warmest ensemble member shows relative humidity values greater than 10 % where most members have a relative humidity approaching zero and very low concentrations of water vapour (~ 100 ppbv). 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 mB, 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. This suggests that the mechanism behind the high climate sensitivity of these models is related to the high altitude humidity response.

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 warm, moist, convecting air packets with their surroundings (Sanderson et al., 2008; Rougier et al., 2009; Murphy et al., 2004), and thus the mechanism by which water vapour can reach

GMDD
6, 841–892, 2013

the upper atmosphere. These changes in upper atmospheric humidity constitute a radiative feedback which could amplify the warming from elevated CO_2 levels (Sanderson et al., 2008; Forster and Shine, 2002; Joshi et al., 2010; Sanderson, 2011). Figure 12a–c shows the specific and relative humidity in the pre-industrial control simulations vary as a function of ENTCOEF; low values of ENTCEOF are associated with high specific humidities in the upper atmosphere as has been seen in other studies, for example at 100 mB members with values of ENTCOEF below 2.0 show specific humidities around double the values of the rest of the ensemble (Joshi et al., 2010; Sanderson et al., 2008; Sanderson, 2011). Figure 12d–f shows how ENTCOEF is associated with high altitude humidity in the $4 \times CO_2$ simulation; members with a value of ENTCOEF below 2.0 show specific humidities of order 10 ppmv at 30 mB whereas for moderate and high values of ENTCOEF specific humidity is less than 1 ppmv at this altitude. Figure 12g, h shows that only the members with the largest increases in temperature at $4 \times CO_2$ show very large changes in high altitude humidity. Such large changes in humidity at high altitude will increase the member's greenhouse effect and will therefore act as a positive feedback warming the members further (Held and Soden, 2000; Forster and Shine, 2002; IPCC, 2007; Joshi et al., 2010). These amplified changes in high altitude humidity seem to occur at values of the entrainment coefficient of 2.5 or less with the largest changes occurring in members with an entrainment coefficient of 2.0 of less. The role of the ice fall speed parameter (VF1) must be mentioned as this interacts strongly with ENTCOEF by controlling the sink for high altitude humidity through ice particle sedimentation. The climate sensitivity of HadCM3 has been found to be much greater for these low values of the entrainment coefficient, and the climate sensitivity rises rapidly for levels below the standard value of 3.0 (see Fig. 6 of Sanderson et al. (2008) and Fig. 6 of Rougier et al. (2009)), seemingly agreeing with the suggestion of Joshi et al. (2010) that the high altitude humidity of a low entrainment rate coefficient HadCM3 simulation was responsible for its extremely high climate sensitivity.

GMDD
6, 841–892, 2013

In the quadrupled CO_2 simulations all members of the PPE are still warming somewhat after 140 yr with the warmest accepted run showing a rate of warming of ~ 0.4 K per decade in the last 50 yr, see Fig. 13a. The warmest accepted run has a residual radiative imbalance more than 1 W m^{-2} greater than the ensemble mean of ~ 1.6 W m^{-2} after 140 yr. A similar trend can be seen in the 1 % CO_2 per annum experiment, Fig. 13b, d, with the warmest run building up a greater radiative imbalance by the end of the run than the rest of the PPE, and showing a greater than linear trend in temperature rather than the linear response seen in the other members. These radiative forcing and temperature results suggest that the warmest run is not following the linear relation between initial radiative forcing and equilibrium temperature response suggested by Gregory et al. (2004), as seen by Joshi et al. (2010) for a low ENTCOEF run. For both the $4 \times CO_2$ and 1 % CO_2 experiments the precipitation response roughly follows the temperature trend (Fig. 13e, f), with the $4 \times CO_2$ members showing an initial sharp reduction in ensemble mean precipitation to around -0.12 mm day^{-1} which recovers as the temperature rises (Bala et al., 2010). We also find that the members with the highest pre-industrial temperature are the members which warm the most at $4 \times CO_2$, see Fig. 14 for details.

Estimates of the equilibrium temperature response and initial radiative forcing of the $4 \times CO_2$ simulations are possible by following the method of Gregory et al. (2004) and applying a linear fit over the first 50 yr, as was applied for the 20 yr pre-industrial pre-calibration. Carrying out a regression on the joint evolution of temperature and radiative imbalance is expected to provide an estimate of the initial radiative forcing perturbation and a final equilibrium temperature. This approach works for most members of the PPE, which show some deviation from the 50 yr linear fit, but for a number of the low ENTCOEF members (most of which failed the initial selection criterion) this relationship breaks down. The warmest run deviates substantially from the initial 50 yr linear fit and in the later years of the simulation shows increases in temperature without the expected reduction in TOA radiative imbalance (not shown). The mechanism causing this deviation is not certain but may be due to the very large increases in humidity in the

GMDD
6, 841–892, 2013

upper atmosphere of members with low ENTCOEF members in response to warming. Most of the low ENTCOEF runs, which show this deviant warming, are outside of the pre-industrial target temperature window after the 800 yr pre-industrial control run but the warmest accepted run shows the greatest increase in temperature and is well within the temperature window in the pre-industrial.

The projected equilibrium temperatures of the $4 \times CO_2$ simulations ($4 \times CS$) are shown in Fig. 15a. These are found from a 150 yr linear fit applied in the same way as in the temperature and radiative forcing projections discussed above. This longer fitting period was applied to capture some of the deviation that the members with the greatest warming show. Most accepted ensemble members have a $4 \times CS$ in the range of 6.5–10.5 °C, with the warmest accepted member having an estimated $4 \times CS$ of 35 °C due to the breakdown of the linear relation between increasing temperatures and decreasing TOA radiative imbalance. The temperature of the 1 % CO_2 simulations at the time at which CO_2 levels have quadrupled (year 140), $4 \times TCR$, is shown in Fig. 15b. We find that most PPE members have a ratio of roughly 2 : 3 between their $4 \times TCR$ and $4 \times CS$ values, see Fig. 14. This comparison does not work for the warmest accepted run due to the breakdown of the "Gregory plot", linear relationship between temperature and TOA radiative imbalance (Gregory et al., 2004). If the low ENTCOEF runs are excluded, we find a clear correlation between low values of $4 \times TCR$ and high values of VDIFF despite there being no correlation between VDIFF and $4 \times CS$ (again excluding the low ENTCOEF members). This follows from the fact that higher values of VDIFF should lead to a greater transport of heat to depth and matches the results of an earlier study into the effects of VDIFF and other ocean parameters (Collins et al., 2007).

It is possible to estimate the initial precipitation response and an equilibrium "hydrological sensitivity" for the $4 \times CO_2$ simulations by following the method outlined in Bala et al. (2010), for these calculations a 50 yr linear fit is made to the joint evolution of precipitation and temperature. The response of precipitation to changes in radiative forcing has been considered to consist of a fast component or "precipitation adjustment", corresponding to a change in the patterns of latent and specific heating particular to the type

GMDD
6, 841–892, 2013

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 (Bala et al., 2010; Andrews et al., 2010). The PPE shows a range of both fast and slow behavior to the $4 \times CO_2$ forcing with a "fast" reduction in precipitation of between −4.8 to −7.0 %, and a hydrological sensitivity of between 1.8 to 2.3 % $^{\circ}C^{-1}$ (excepting the warmest accepted member which has a value of less than 1.6 % $^{\circ}C^{-1}$), see Fig. 15c, d. At $2 \times CO_2$ Andrews et al. (2009) showed an ensemble mean hydrological sensitivity of 2.8 % $^{\circ}C^{-1}$ 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 % $^{\circ}C^{-1}$ and a precipitation adjustment of 3.0 % (roughly half the $4 \times CO_2$ value shown here as expected) for the HadSM3 model.

3.6 Evaluation of response at $4 \times CO_2$

The PPE response to the $4 \times CO_2$ and 1 %CO_2 simulations is now evaluated for plausibility and compared to the CMIP3 ensemble response to elevated CO_2 concentrations. Supplementary Table 3 lists the climate sensitivity and transient climate response at $4 \times CO_2$ ($4 \times$ CS and $4 \times$ TCR) and the global mean humidity response at 100 mB and 10 mB. Excluding one member, the accepted members of the PPE show a range of $4 \times$ CS between 6.5 and 10.5 $^{\circ}C$ which is on the high side when compared to the likely range reported in the IPCC: 4.0 to 9.0 $^{\circ}C$ (after doubling to account for the higher CO_2 level used here). One accepted member shows a very high projected value of $4 \times$ CS value of $> 35\,^{\circ}C$, in fact the member does not appear to be converging on an equilibrium temperature in the way anticipated by the Gregory method (Gregory et al., 2004), i.e. with TOA radiative imbalance not reducing despite rapidly rising temperature. This warmest member also shows by far the lowest value for the hydrological sensitivity, with a value of less than 1.6 % $^{\circ}C^{-1}$ far below the CMIP3 average of 2.8 % $^{\circ}C^{-1}$ and the standard HadCM3 value of 2.2 % $^{\circ}C^{-1}$ (Andrews et al., 2009).

GMDD
6, 841–892, 2013

The high altitude humidity response at $4 \times CO_2$ differs substantially between the members of the PPE with a number of the accepted members (and most of the failed members) showing an upper atmosphere humidity response that is well outside the range of responses seen in the CMIP3 ensemble at $2 \times CO_2$ (after doubling the values to make them comparable). At 100 mB, some members with high values of $4 \times CS$ show very large increases in specific humidity that are beyond the range seen in the CMIP3 ensemble but at 10 mB, only the warmest member shows an increase in specific humidity beyond the range seen in the CMIP3 ensemble. This increase in high altitude humidity is likely one of the drivers of the high values of climate sensitivity in these members as it has a large effect on the TOA radiative balance (Forster and Shine, 2002; IPCC, 2007). The ENTCOEF parameter has been shown to be linked to these large changes in high altitude humidity and climate sensitivity in this paper and in others (Joshi et al., 2010; Sanderson et al., 2008), therefore it is unsurprising that the warmest member has the lowest value of ENTCOEF of any of the accepted members. The warmest member and the member with the second highest climate sensitivity – the two members with the lowest values of ENTCOEF – are also the members with the highest values of high altitude specific and relative humidity at $4 \times CO_2$, as can be seen in Fig. 11 (these runs are shown as the solid yellow and dashed dark brown lines). 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). Thus these implausibly large increases in upper atmospheric humidity in response to elevated CO_2 levels, and the associated high climate sensitivities, seen in the ensemble members with low values of ENTCOEF may be unrealistic (Joshi et al., 2010).

GMDD
6, 841–892, 2013

4 Discussion

This study has presented the methodology used to generate a perturbed parameter ensemble using the fully-coupled AOGCM HadCM3 without flux adjustment and presented some analysis of these results. The goal has been to develop a research tool that can be used to explore the role of parametric uncertainty in climate studies where flux-adjustment is not appropriate. Heat, water and momentum flux adjustments correct for biases developing which prevents the model, or a member of an ensemble, deviating too far from observations (Collins et al., 2006). However by not allowing models to drift, arguably unrealistic members of an ensemble will be maintained; for the ENTCOEF parameter in HadCM3 this has arguably led to exaggerated climate sensitivity ranges as ensemble members with unrealistic climates are retained (Murphy et al., 2004; Sanderson et al., 2008; Joshi et al., 2010). By removing flux adjustment from the model, an additional physical constraint is applied to the model and by only preserving model versions that are close to the observed pre-industrial temperature, unrealistic model versions can be excluded. This approach has been criticized on the grounds that it is an overly strong constraint, discarding many informative members of an ensemble and would limit the range of parameters too much (Collins et al., 2006). Our results challenge this criticism as we have produced an ensemble of 21 members, without using flux-adjustment that spans most of the range of the 8 atmospheric, ocean and sea-ice parameters we perturbed (supplementary Table 1 lists the parameter values and pre-industrial temperatures of the PPE after an 800 yr spinup).

To produce this perturbed parameter ensemble, an initial simple selection criteria, based on the projected temperature response of the members from a 20 yr pre-industrial simulation, was applied which allowed a large number of initial parameter combinations to be screened to exclude unreasonable members. 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 $13.6 \pm 2.0\,^{\circ}\mathrm{C}$ after 800 yr of pre-industrial simulation (see Fig. 1c),

GMDD
6, 841–892, 2013

corresponding to a success rate of 80 %. These deviations from the projection are likely due to three different mechanisms: internal variability, the parameter perturbation not acting simply as a change in forcing or feedback processes, and short-term absorption of energy by the ocean. Firstly, Internal variability affects projections based on short timeseries and can be addressed with statistical tools or with repeated simulations. Secondly, in this study we assumed that a change in parameter values would be realized as an instantaneous change in radiative forcing and a change in the feedback processes of the model (and thus amenable to the Gregory method of projecting equilibrium temperature). However 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). Thirdly, we find evidence of an indirect interaction between the vertical ocean diffusivity parameter and the projected pre-industrial equilibrium temperature. It is believed that the perturbation of the vertical ocean diffusivity parameter leads to a short-term adjustment of the surface ocean heat content which for high values of the parameter leads to a substantial absorption of energy from the atmosphere, see Fig. 5. We thus advise that caution be exercised when applying the Gregory method for projecting equilibrium temperature response to parameter perturbation experiments due to potential non-linear feedbacks that can arise and the short-term ocean adjustments that ocean parameter perturbations can induce (Gregory et al., 2004). However, overall 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 ~ 160 000 model years as opposed to the ~ 25 000 model years required with our approach.

This study presents one of the only perturbed parameter ensembles to use a fully coupled AOGCM without flux adjustment and as such the long-term nature of the response of the ocean must be considered (Gregoire et al., 2010). We find that even after an 800 yr control the oceans of the members of the PPE are not fully adjusted to these new conditions and so both the pre-industrial simulations and the elevated

GMDD
6, 841–892, 2013

CO_2 simulations were run with an ocean not in equilibrium. The implication is that the results found in this study may not perfectly match those that would be obtained after a full multi-thousands year spin-up. For high values of the ocean vertical diffusivity parameter, VDIFF, increased absorption of energy into the oceans and lowered values of TCR have been found in a single-parameter perturbation experiment (Collins et al., 2007). In this ensemble we find a selection artifact that connects high values of this VDIFF parameter and higher pre-industrial temperatures, for the reasons outlined above, and in addition we find an association between high pre-industrial temperatures and higher values of climate sensitivity (see Fig. 9). As high values of VDIFF should lead to lower values for TCR, due to the ocean absorbing more energy, and as our selection artifact links high climate sensitivity and high values for VDIFF, this leads to the PPE having a narrower range of TCR than would be the case if this selection artifact was not present.

This initial selection criterion based on a temperature projection appeared to produce an ensemble of models that behaved plausibly in the pre-industrial when compared to the CMIP3 ensemble, however at elevated CO_2 levels it was found that one seemingly plausible member produced an unphysical warming response at $4 \times CO_2$. The pre-industrial climates of the PPE were judged to be plausible representations of the climate in the sense that they were roughly in line with the range of responses seen in the CMIP3 ensemble but with a number of common biases shared with the parent model, HadCM3. The exception to this general picture is that some members showed temperature, precipitation and TOA radiative balance responses in the tropics that were clearly beyond the range seen in the CMIP3 ensemble. 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.

The PPE response to quadrupled CO_2 and 1 % CO_2 per annum simulations was also evaluated for plausibility and compared to the CMIP3 ensemble response to elevated CO_2 concentrations and at this stage the warmest member was identified as an outlier. The greatest differences between members of the PPE are found in high altitude

GMDD
6, 841–892, 2013

humidity where a number of the members, with low values of ENTCOEF, have specific humidity values 10–100 times higher than the standard model version between 100 and 10 mb which has a large effect on the TOA radiative balance (Forster and Shine, 2002; IPCC, 2007). The very high climate sensitivities of the low ENTCOEF runs are linked to these temperature driven increases in upper tropospheric and stratospheric specific and relative humidity (Joshi et al., 2010; Sanderson et al., 2008). The mechanism behind this unchecked warming has not been definitively identified but one plausible hypothesis presented in Sanderson et al. (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 . 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). Our results for low entrainment rate members agree with those of Joshi et al. (2010) and we concur with their assessment that the kind of response seen in the low ENTCOEF members is unrealistic and as such the very high climate sensitivities produced by these members should be viewed with extreme caution, see Fig. 15. On these grounds we suggest revising the range for ENTCOEF for future perturbed parameter studies with HadCM3 from 0.6–9.0 to 2.0–9.0, which would exclude our warmest ensemble member. We note however, that plausible ensemble members may be found for lower values of ENTCOEF given certain values for other parameters, particularly VF1, given a sufficiently large sample of parameter perturbations.

The ensemble generation and selection process applied in this study is straightforward and generally successful. However a number of problems were identified. The initial temperature selection criteria, was based on a straightforward projection from a 20 yr simulation and was very successful in that it excluded models that were clearly implausible due to being either too warm or too cold in the pre-industrial. The simple evaluation of the longer pre-industrial simulations with global mean indicators failed to exclude members which showed an implausible response at elevated CO_2 levels.

GMDD
6, 841–892, 2013

Some members deviated markedly beyond the range of behavior seen in the CMIP3 ensemble in the Tropics but this was not considered in our simple selection approach. Figure 9 shows that in the pre-industrial the members with the lowest values of ENTCOEF, the warmest member and the member with the second highest climate sensitivity, both stand well outside the CMIP3 range in the Tropics, showing some of the largest differences for temperature, precipitation and TOA radiative balance (the yellow solid line and the dashed dark brown line). The simple global-scale evaluation of the PPE at elevated CO_2 levels allowed implausible behavior in the ensemble to be identified, specifically implausibly large increases in upper atmospheric humidity and extreme warming in one member. These findings suggest that simple plausibility tests can be used, firstly, to narrow down a broad range of parameter combinations and, secondly, to exclude members with implausible climatic responses. We'd advise that a simple plausibility approach, along the lines outlined in Edwards et al. (2011), is applicable to PPEs of AOGCMs but that tropical conditions and the response to elevated CO_2 levels should be included in such evaluations of plausibility.

5 Conclusions

This study presents the methodology and some initial results from the first 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 $13.6 \pm 2\,^{\circ}C$ 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

GMDD
6, 841–892, 2013

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 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). In fact one member of the ensemble with a low value of the entrainment rate coefficient remain close to pre-industrial top of atmosphere radiative balance despite the fact that at quadrupled CO_2 levels it shows an unrealistic increase in stratospheric and upper tropospheric humidity levels and a non-linear temperature increase in response to CO_2 radiative forcing. This suggests that the plausibility of ensemble members' response to elevated CO_2 levels should be evaluated in perturbed parameter ensemble studies.

5.1 Future work

Future work will involve applying this ensemble to climate and paleo-climate studies which require a model without flux-adjustment. In particular this PPE will be applied to the geoengineering model intercomparison project (geoMIP) experiments to investigate the effects of parametric uncertainty on Solar Radiation Management (SRM) geoengineering results (Kravitz et al., 2011). The ensemble that will be used in these future studies will consist of the accepted ensemble discussed above but with the warmest member excluded. The methodology developed in this study could be applied easily to other coupled AOGCMs; however, a selection criterion based on the projected equilibrium temperature can be affected by changes in atmosphere-ocean heat-exchange when perturbing ocean parameters. A similar methodology could be applied to atmosphere-only GCMs, as in Shiogama et al. (Shiogama et al., 2012). An alternative would be to constrain the models on the magnitude of energy, water and momentum flux-correction required to fit observations. The resulting atmospheric parameter perturbations could then be applied to a fully-coupled model hopefully without producing large drifts in the mean climate. Flux-adjustment is no longer used in

GMDD
6, 841–892, 2013

the standard configurations of coupled AOGCMs, although it does provide some advantages for perturbed parameter studies, we have shown that a tractable ensemble design with a broad range of model behaviour is possible without flux-adjustment.

Supplementary material related to this article is available online at: http://www.geosci-model-dev-discuss.net/6/841/2013/gmdd-6-841-2013-supplement.zip.

Acknowledgements. PI acknowledges support in the form of a NERC PhD studentship. We acknowledge the CMIP3 modeling 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 organizing the model data analysis activity. The WCRP CMIP3 multi-model dataset is supported by the Office of Science, US Department of Energy.

References

Andrews, T., Forster, P. M., and Gregory, J. M.: A surface energy perspective on climate change, J. Climate, 22, 2557–2570, 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, 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, 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.

GMDD
6, 841–892, 2013

Cattle, H. and Crossley, J.: Modeling arctic climate-change, Philos. T. Roy. Soc. A, 352, 201–213, 1995.
Collins, M., Booth, B. B. B., Harris, G. R., Murphy, J. M., Sexton, D. M. H., an Webb, M. J.: Towards quantifying uncertainty in transient climate change, Clim. Dynam., 27, 127–147, 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, 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., doi:10.1007/s00382-010-0808-0, 1–30, 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, 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, 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.: Precalibrating an intermediate complexity climate model, Clim. Dynam., 37, 7, 1469–1482, 2011.
Forster, P. M. D. and Shine, K. P.: Assessing the climate impact of trends in stratospheric water vapor, Geophys. Res. Lett., 29, 3309–3312, doi:10.1029/1999GL010487 , 2002.
Frame, D. J., Booth, B. B. B., Kettleborough, J. A., Stainforth, D. A., Gregory, J. M., Collins, M., and Allen, M. R.: Constraining climate forecasts: the role of prior assumptions, Geophys. Res. Lett., 32, L09702, doi:10.1029/2004gl022241, 2005.
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

GMDD
6, 841–892, 2013

experiment: design of the coupled model ensemble, Philos. T. Roy. Soc. A, 367, 855–870, 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, 1–15, 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 vapor feedback and global warming, Annu. Rev. Energ. Env., 25, 441–475, 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 2007: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assesment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Quin, D., Manning M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., Miller, H. L., Cambridge University Press, Cambridge, 2007.
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.
Kiehl, J. and Trenberth, K. E.: Earth's annual global mean energy budget, B. Am. Meteorol. Soc., 78, 197–208, 1997.

GMDD
6, 841–892, 2013

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, 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 behavior across 57 000 climate models, P. Natl. Acad. Sci. USA, 104, 12259–12264, 2007.

Kravitz, B., Robock, A., Boucher, O., Schmidt, H., Taylor, K. E., Stenchikov, G., and Schulz, M.: The Geoengineering Model Intercomparison Project (GeoMIP), Atmos. Sci. Lett., 12, 162–167, 2011.

Masson, D. and Knutti, R.: Climate model genealogy, Geophys. Res. Lett., 38, L08703, doi:10.1029/2011gl046864, 2011.

Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., Kitoh, A., Knutti, R., Murphy, J., and Noda, A.: Global climate projections, Chapt. 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., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK, 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, 1383–1394, 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, 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, 23, L23825, doi:10.1029/2005gl024452, 2005.

Rodwell, M. J. and Palmer, T. N.: Using numerical weather prediction to assess climate models, Q. J. Roy. Meteor. Soc., 133, 129–146, 2007.

Rougier, J., Sexton, D. M. H., Murphy, J. M., and Stainforth, D.: Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments, J. Climate, 22, 3540–3557, 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, 2011.

GMDD
6, 841–892, 2013

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, 2008.

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, 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, 1993 ECMWF, Reading, England, 8–11 June 1993.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (Eds.): Climate Change 2007: The Physical Science Basis, Contribution of working Group I to the Fourth Assesment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 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, 2005.

Stuber, N., Ponater, M., and Sausen, R.: Why radiative forcing might fail as a predictor of climate change, Clim. Dynam., 24, 497–510, 2005.

Tang, B. X.: A theorem for selecting oa-based latin hypercubes using a distance criterion, Commun. Stat. Theory, 23, 2047–2058, 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, 2012.

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.

Wu, X. Q.: Effects of ice microphysics on tropical radiative-convective-oceanic quasi-equilibrium states, J. Atmos. Sci., 59, 1885–1897, 2002.

GMDD
6, 841–892, 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 CO_2 increase between two perturbed physics ensembles, J. Climate, 23, 1392–1410, 2010.

GMDD
6, 841–892, 2013

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	Entrainment rate coefficient
VF1	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 [1]	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 [1]	0.65	Sea-ice albedo at 0 °C
VDIFF (min/max)	1.0–15.0	0.5–4.0	2.0–58.0 [2]	Background vertical diffusivity, which varies as a function of depth

GMDD
6, 841–892, 2013

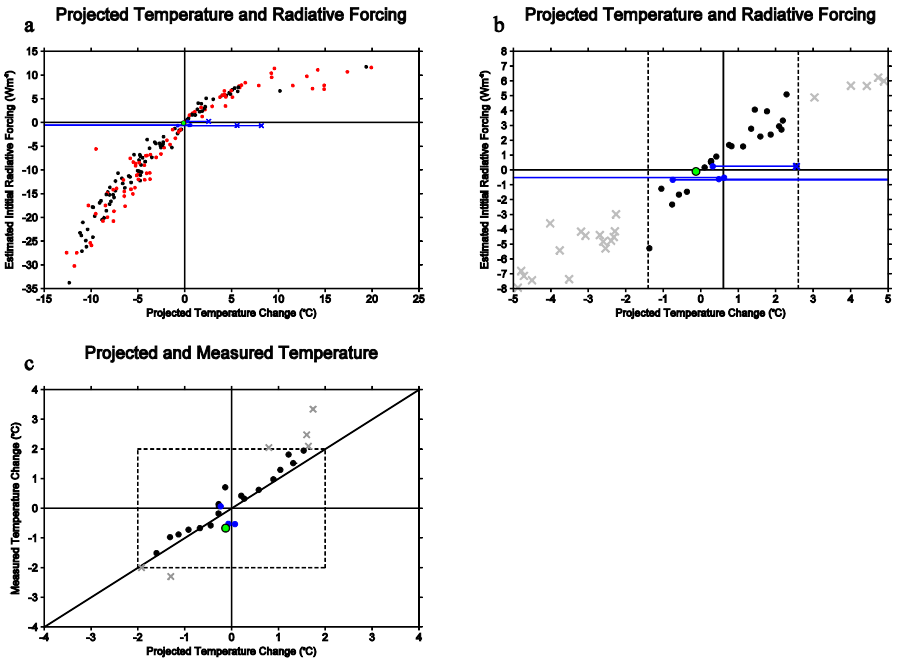

Fig. 1. Projections of equilibrium temperature and initial radiative forcing for the ensemble generated made by applying the Gregory et al. (2004) approach to the initial 20 yr of simulation **(a, b)**. **(b)** shows the acceptable range of temperatures with dashed lines, i.e. within ±2 °C of the observed pre-industrial temperature of 13.6 °C. **(c)** shows a comparison between the projected temperature and the simulated temperature at the end of the 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 shown in blue with a cross at the anomalous projected temperature and a dot for the temperature of the 20th simulated year.

GMDD
6, 841–892, 2013

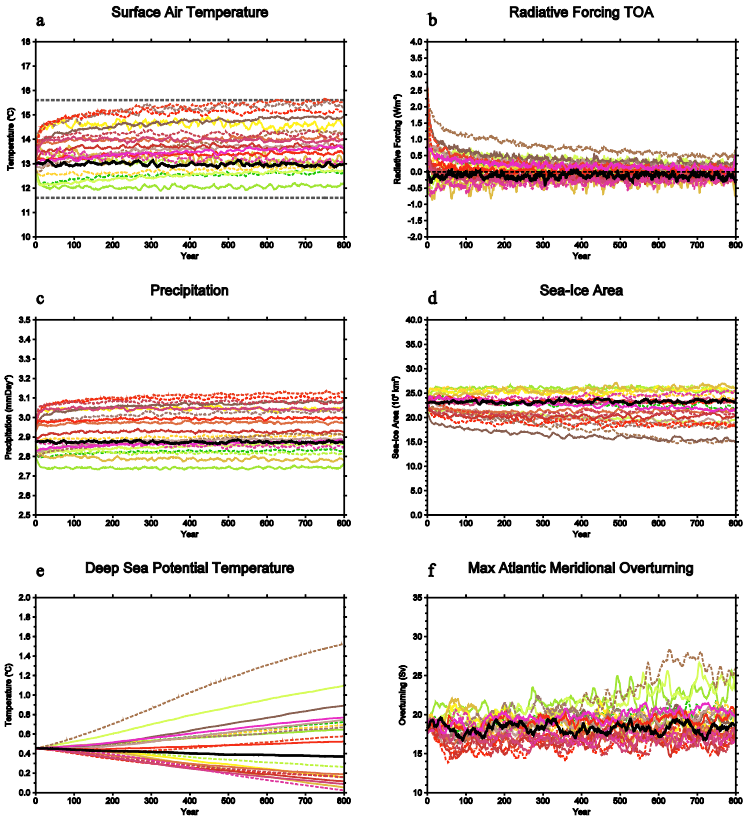

Fig. 2. Evolution of the global annual means over the course of the 800 yr pre-industrial control simulation of surface air temperature **(a)**, top-of-atmosphere radiative balance **(b)**, precipitation **(c)**, annual-mean sea-ice area **(d)**, potential temperature at 2700 m **(e)**, maximum Atlantic overturning **(f)**. The standard version of HadCM3 is plotted with a black line.

GMDD
6, 841–892, 2013

PPE of AOGCM without flux-adjustment

P. J. Irvine et al.

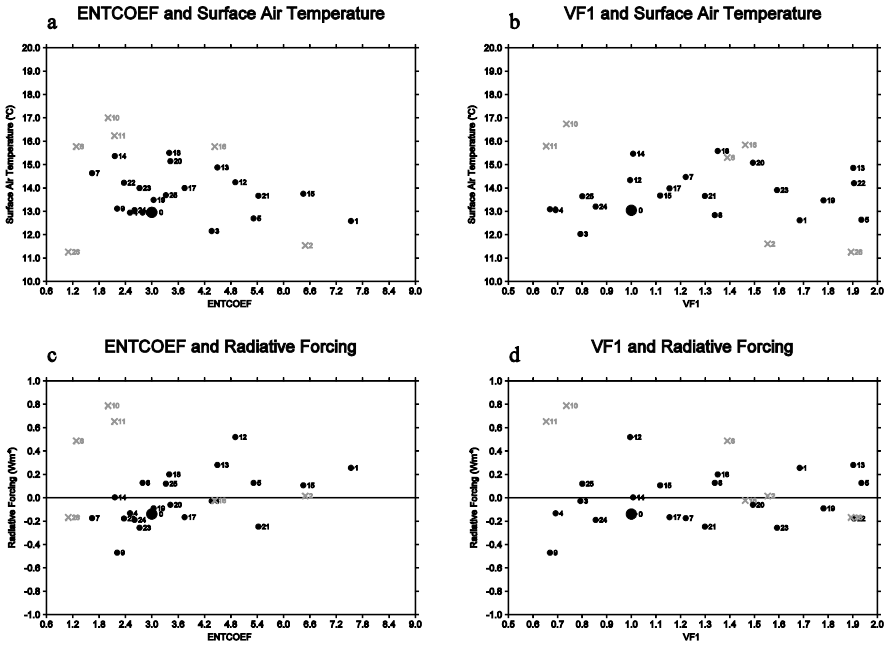

Fig. 3. Scatter plots of the pre-industrial control temperature **(a, b)** and the top-of-atmosphere radiative forcing at the end of the control run against the entrainment coefficient (ENTCOEF) and the ice fall speed (VF1), respectively. The standard model is shown as a larger black dot and the failed runs were included in this plot as gray crosses to make clearer the role of the parameters.

GMDD
6, 841–892, 2013

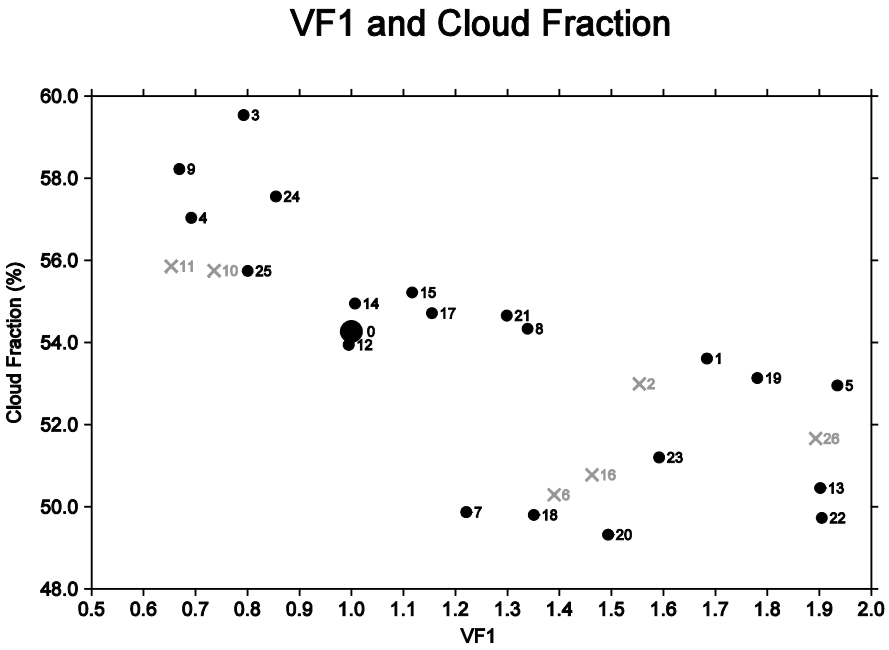

Fig. 4. The control global cloud fraction after 800 yr as a function of the ice fall speed parameter. The standard run is plotted with a larger point and labeled with a zero. The standard model is shown as a larger black dot and the failed runs were included in this plot as gray crosses to make clearer the role of the parameter.

GMDD
6, 841–892, 2013

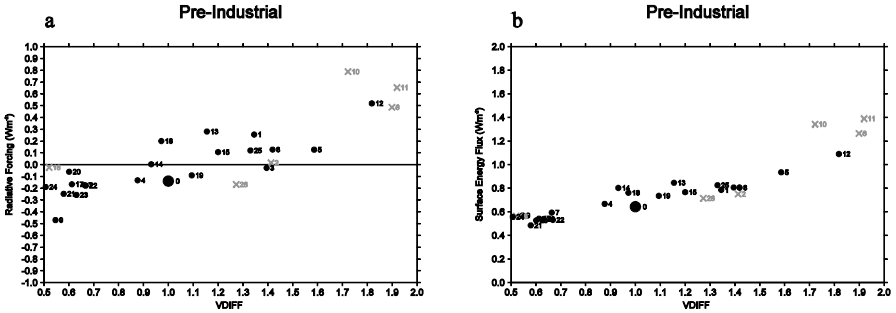

Fig. 5. The control global mean temperature after 800 yr **(a)** and the global mean heat flux into the ocean **(b)**, as a function of the vertical diffusivity parameter. The standard run is plotted with a larger point and labelled with a zero. The standard model is shown as a larger black dot and the failed runs were included in this plot as gray crosses to make clearer the role of the parameter.

GMDD
6, 841–892, 2013

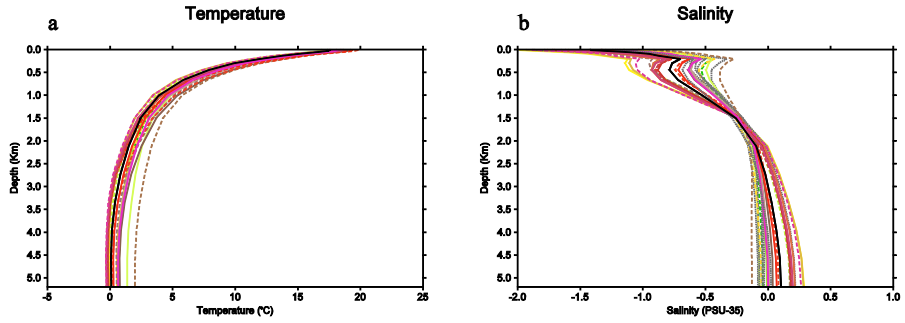

Fig. 6. The annual and ocean area average temperature **(a)** and salinity **(b)** of the ocean with depth. The standard version of HadCM3 is plotted with a black line.

GMDD
6, 841–892, 2013

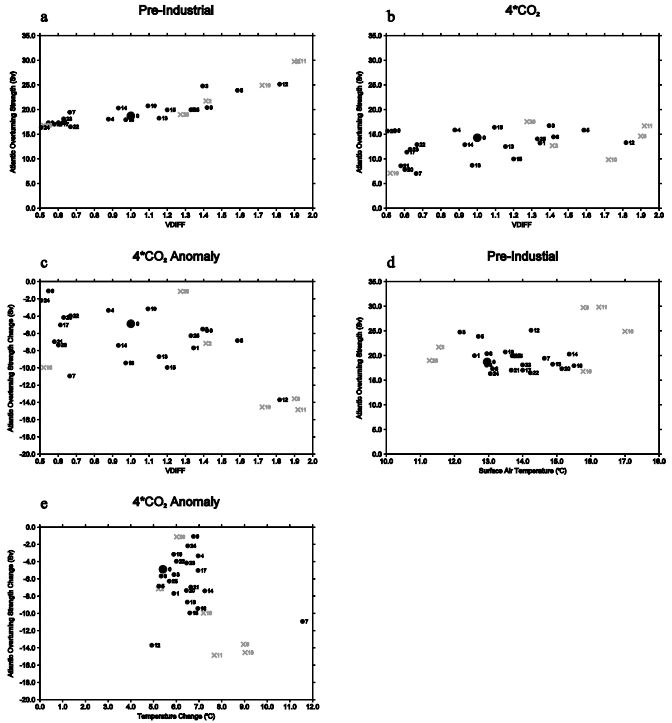

Fig. 7. The relationship between the maximum Atlantic overturning and both the background vertical diffusivity parameter and global-mean temperature for the ensemble. **(a–c)** show how vertical diffusivity affects the control overturning **(a)**, the overturning at $4 \times CO_2$ **(b)** and the change in overturning between $4 \times CO_2$ and the control **(c)**. **(d, e)** show the effect of pre-industrial global-mean temperature **(d)** and the effect of warming to $4 \times CO_2$ **(e)** on Atlantic overturning. The standard model is shown as a larger black dot and the failed runs were included in this plot as gray crosses to make clearer the role of the parameters.

GMDD
6, 841–892, 2013

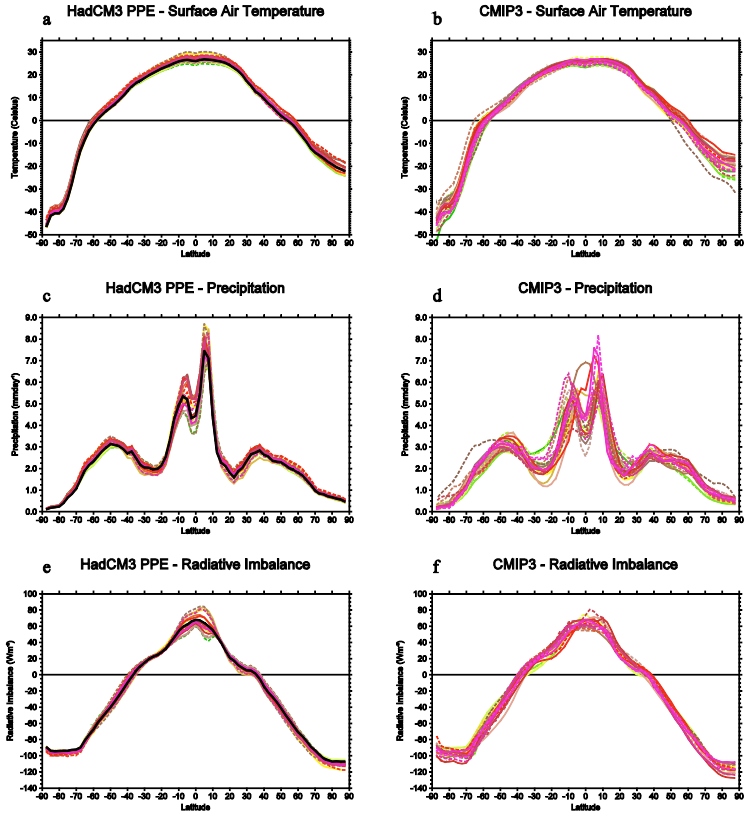

Fig. 8. The surface air temperature **(a, b)**, precipitation **(c, d)**, and top-of-atmosphere radiative balance **(e, f)** for the PPE simulations and for the CMIP3 ensemble. The standard version of HadCM3 is plotted in black for the PPE simulations.

GMDD
6, 841–892, 2013

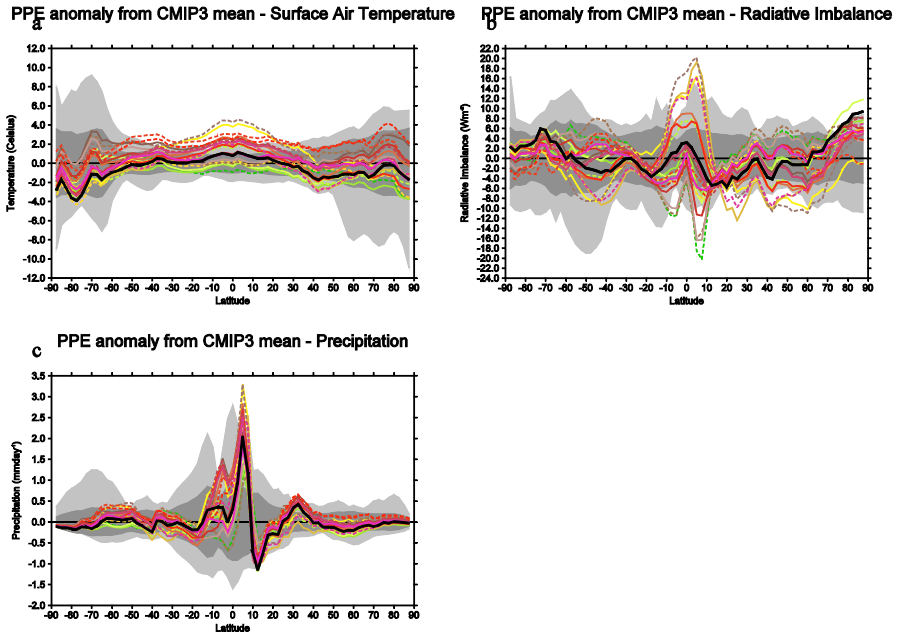

Fig. 9. The anomaly between each member of the PPE and the CMIP3 ensemble mean for the surface air temperature **(a)**, top-of-atmosphere radiative balance **(b)** and precipitation **(c)**. The standard version of HadCM3 is plotted in black. The range for the CMIP3 ensemble is plotted in light grey and the standard deviation of the CMIP3 ensemble is plotted in dark grey.

GMDD
6, 841–892, 2013

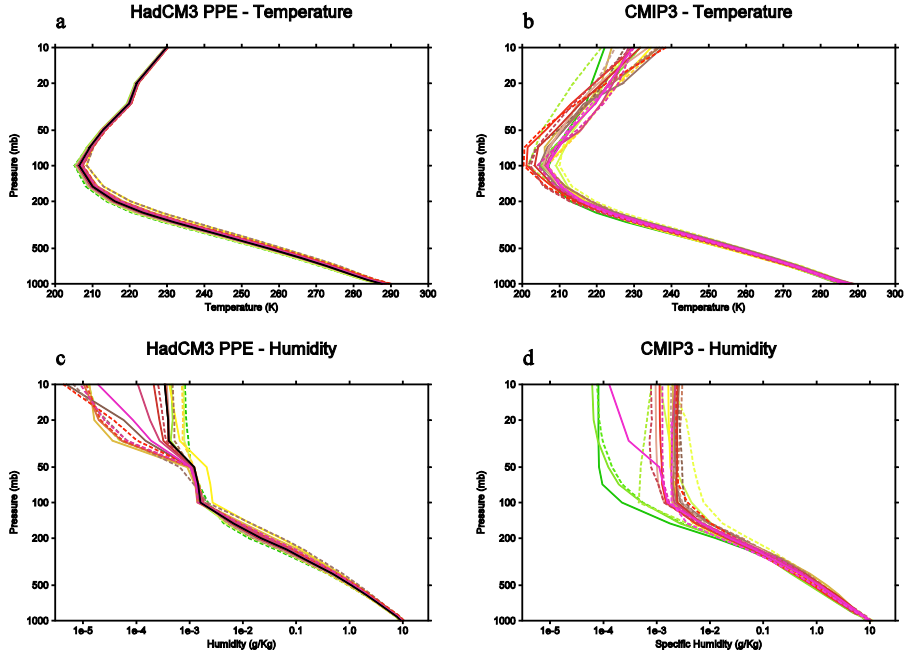

Fig. 10. The temperature **(a, b)** and specific humidity **(c, d)** throughout the atmospheric column for the PPE simulations and the CMIP3 ensemble. The standard version of HadCM3 is plotted in black for the PPE simulations. Note that for the PPE cells below ground level the values are extrapolated using an average lapse rate and included in the level mean.

GMDD
6, 841–892, 2013

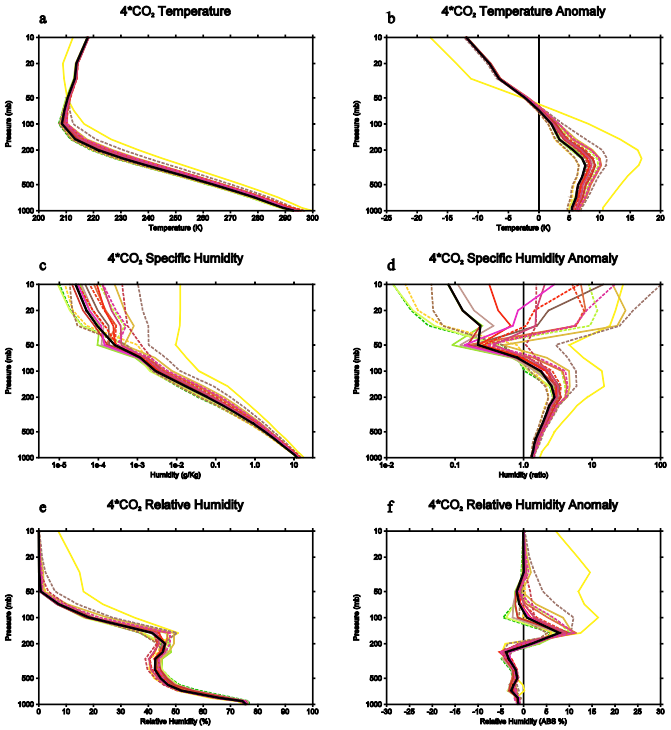

Fig. 11. Shows the mean at 4 × CO_2 and the anomaly between 4 × CO_2 and the pre-industrial control for temperature **(a, b)**, specific humidity **(c, d)** and relative humidity **(e, f)** for the PPE simulations. The standard version of HadCM3 is shown in black for the PPE plot. Note that cells below ground level the values are extrapolated using an average lapse rate and included in the level mean.

GMDD
6, 841–892, 2013

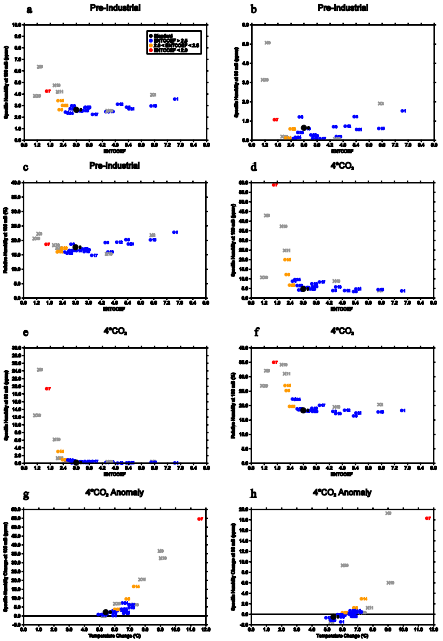

Fig. 12. Specific and relative humidity (in ppmv) at stratospheric levels against the entrainment rate coefficient **(a–f)** and temperature **(g, h)**. **(a–f)** show for the pre-industrial **(a–c)** and 4 × CO_2 **(d–f)** the relationship between the entrainment rate coefficient and specific humidity at 100 mB and 30 mB, and relative humidity at 100 mB, respectively. **(g, h)** show the relationship between the surface air temperature anomaly between 4 × CO_2 and the pre-industrial and the change in specific humidity at 100 mB **(g)** and 30 mB **(h)**. 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.

GMDD
6, 841–892, 2013

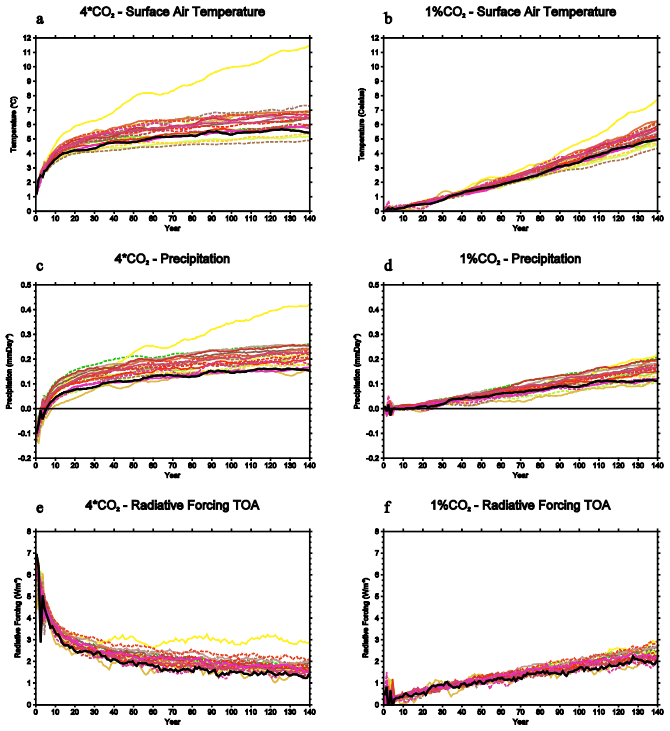

Fig. 13. Shows the evolution of temperature **(a, b)**, top-of-atmosphere radiative forcing **(c, d)**, and precipitation **(e, f)**, for the $4 \times CO_2$ and 1 % CO_2 simulations. The variables are plotted as anomalies from the start of the runs, i.e. the end of each of the pre-industrial spin-ups. The standard run is plotted in black. A ten year running mean is applied to the data without extrapolation at the beginning of the run.

GMDD
6, 841–892, 2013

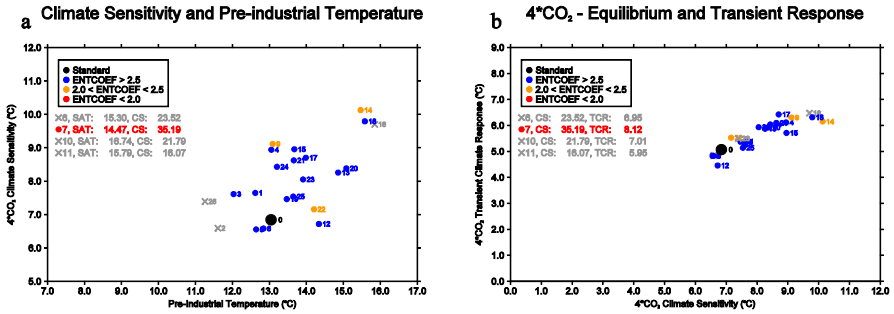

Fig. 14. A scatter plot of the pre-industrial control temperature against the estimated $4 \times CO_2$ climate sensitivity **(a)** and a scatter plot of the $4 \times CO_2$ climate sensitivity against the $4 \times CO_2$ transient climate response. The values are global means averaging over the last 10 yr of the 150 yr $4 \times CO_2$ simulations and the same point in the control simulations. The standard model is shown as a larger black dot and the failed runs were included in this plot as gray crosses to make clearer the role of the parameters. The values of simulations which exceed the bounds of the plot are listed below the legend.

GMDD
6, 841–892, 2013

PPE of AOGCM without flux-adjustment

P. J. Irvine et al.

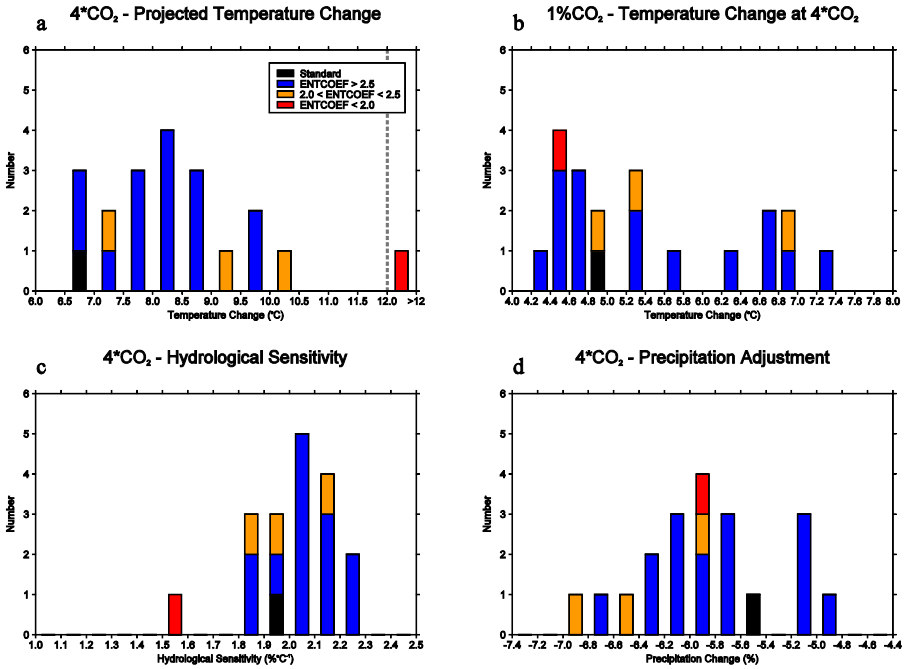

Fig. 15. Shows histograms of the projected temperature change of the 4 × CO_2 simulations **(a)**, the temperature change of the 1 % CO_2 simulations at year 140 **(b)**, and the hydrological sensitivity **(c)** and precipitation adjustment **(d)** of the 4 × CO_2 simulations. The members which failed the pre-industrial temperature selection are excluded from this plot. All changes are relative to the pre-industrial control simulations.

GMDD
6, 841–892, 2013