Climate Dynamics (2006) 27: 127–147
DOI 10.1007/s00382-006-0121-0

Matthew Collins · Ben B. B. Booth · Glen R. Harris
James M. Murphy · David M. H. Sexton
Mark J. Webb

Towards quantifying uncertainty in transient climate change

Received: 30 May 2005 / Accepted: 4 January 2006 / Published online: 6 April 2006

Abstract Ensembles of coupled atmosphere–ocean global circulation model simulations are required to make probabilistic predictions of future climate change. "Perturbed physics" ensembles provide a new approach in which modelling uncertainties are sampled systematically by perturbing uncertain parameters. The aim is to provide a basis for probabilistic predictions in which the impact of prior assumptions and observational constraints can be clearly distinguished. Here we report on the first perturbed physics coupled atmosphere–ocean model ensemble in which poorly constrained atmosphere, land and sea-ice component parameters are varied in the third version of the Hadley Centre model (the variation of ocean parameters will be the subject of future study). Flux adjustments are employed, both to reduce regional sea surface temperature (SST) and salinity biases and also to admit the use of combinations of model parameter values which give non-zero values for the global radiation balance. This improves the extent to which the ensemble provides a credible basis for the quantification of uncertainties in climate change, especially at a regional level. However, this particular implementation of flux-adjustments leads to a weakening of the Atlantic overturning circulation, resulting in the development of biases in SST and sea ice in the North Atlantic and Arctic Oceans. Nevertheless, model versions are produced which are of similar quality to the unperturbed and un-flux-adjusted version. The ensemble is used to simulate pre-industrial conditions and a simple scenario of a 1% per year compounded increase in CO_2. The range of transient climate response (the 20 year averaged global warming at the time of CO_2 doubling) is 1.5–2.6°C, similar to that found in multi-model studies. Measures of global and large scale climate change from the coupled models show simple relationships with associated measures computed from atmosphere-mixed-layer-ocean climate change experiments, suggesting that recent advances in computing the probability density function of climate change under equilibrium conditions using the perturbed physics approach may be extended to the transient case.

M. Collins (✉) · B. B. B. Booth · G. R. Harris · J. M. Murphy
D. M. H. Sexton · M. J. Webb
Hadley Centre for Climate Prediction and Research,
Met Office, FitzRoy Road, EX1 3PB, Exeter, UK
Tel.: +44-1392-884110
Fax: +44-870-9005050
E-mail: matthew.collins@metoffice.gov.uk
URL: http://www.metoffice.gov.uk

1 Introduction

We cannot predict with certainty future climate. What will global average temperature be in 2050? What will be the frequency of occurrence of land-falling hurricanes in 2100? Predictions are uncertain because of unknown future concentrations of greenhouse gases and other anthropogenic and natural forcing agents (e.g. injections of stratospheric aerosol from explosive volcanic eruptions), because of natural (unforced) climate variations and because our models which we use to make predictions are imperfect[1]. How are we to make predictions, for which there is great demand, in the presence of these uncertainties?

The use of complex modes in ensemble and probabilistic weather forecasting on time-scales of days to weeks is now well established. In weather forecasting, the principal uncertainty that has been dealt with using ensembles is that associated with measuring the initial state of the atmosphere. There exist a number of techniques for perturbing the initial state to take account of those uncertainties (e.g. Molteni et al. 1996; Toth and Kalnay 1997). Other uncertainties, to do with the use of imperfect models, are also being addressed (Palmer 2001). The value to end-users of the probabilistic fore-

[1]Errors in observations of the climate system are a further source of uncertainty, but here we focus on the issue from the climate modelling perspective.

cast in comparison with the purely deterministic (and potentially erroneous) forecast has been clearly demonstrated (Palmer 2002).

The probabilistic decadal-centennial climate change prediction system may, in contrast, be said to be still in development. Nevertheless a number of studies have been performed using models ranging from simple energy balance models (EBMs; Andronova and Schlesinger 2001; Wigley and Raper 2001), through earth system models of intermediate complexity (EMICs; Forest et al. 2002; Knutti et al. 2002) to full global circulation models (GCMs; Murphy et al. 2004; Stainforth et al. 2005; Tebaldi et al. 2005; R. Knutti et al. submitted; Piani et al. 2005). It is the latter we concern ourselves with here, as only GCMs are capable of representing the full nonlinear interactions between physical processes which are responsible for determining future climate change, and therefore of predicting the type of regional multivariate information required by end-users.

We may define two classes of complex-model ensemble generation methods. The multi-model method collects GCM output from different modelling centres into a central repository where it can easily be accessed (e.g. Covey et al. 2003). This method is ad hoc in the sense that it has not been designed to completely span the range of model uncertainty, but it has the advantage that there is a large "gene-pool" of possible model components. The other method has been termed the "perturbed physics" approach (Murphy et al. 2004; Stainforth et al. 2005) in which a single model structure is used and perturbations are introduced to the physical parameterisation schemes in the model. Perturbations are made either to parameters or to the schemes themselves (by switching between different existing options rather than substituting in entirely different routines—we term the latter a "structural" change to the model) and the advantage is that variations in model formulation can be made in a systematic way with greater control over ensemble design. In comparison with the multi-model approach, larger ensembles may be generated in order to explore nonlinearities and extreme behaviour. It is also possible to distinguish clearly the effects of different prior assumptions by showing the sensitivity of the ensemble output to the parameter sampling strategy, and to distinguish the effects of different observational constraints by using different experimental designs (e.g. equilibrium, historical simulations, palaeo-climate simulations) and different observational data sets. Despite the lack of structural perturbations, it has been shown (Webb et al. 2005) that the perturbed physics approach can explore, for example, much of the range of detailed atmospheric feedbacks seen in a multi-model ensemble. We may hope one day to see a combined ensemble made up of a number of perturbed physics ensembles made with different AOGCMs.

Perturbed physics ensemble projects have, to date, employed atmosphere models coupled to thermodynamic mixed layer oceans (Murphy et al. 2004; Stainforth et al. 2005), the advantage being that simulations of these "slab" models can be easily run to equilibrium to investigate atmospheric feedbacks and climate sensitivity (here we use the term climate sensitivity to mean specifically the equilibrium global mean temperature change due to a doubling of CO_2). This set-up cannot be used to simulate changes in ocean circulation or transient climate change. It does not capture the delaying effect of the deep ocean on time dependent climate changes, the impact of potential changes in ocean circulation and associated heat transport (e.g. Wood et al. 1999) nor the potential for changes in the characteristics of El Niño variability (e.g. Collins 2000). All of these are highly uncertain and known to be dependent on model formulation. It is the purpose of this paper to test the validity of the perturbed physics approach in generating ensembles with coupled atmosphere–ocean global circulation models (AOGCMs) and to highlight some of the important issues. We provide a preliminary assessment of the behaviour of a 17-member AOGCM ensemble produced with version three of the Hadley Centre model. The study may be viewed as a small step towards the generation of probabilistic predictions of climate change at global and regional scales.

1.1 An aside on probabilistic prediction

Before getting into the details of ensemble generation using AOGCMs, it is worth setting out the general method for probabilistic climate prediction to provide context. Let us assume we are to make a prediction of the future value of a climate variable s. s may, for example, be the equilibrium climate sensitivity under a doubling of CO_2, the transient climate response (TCR—the 20-year average global mean temperature change at the time of CO_2 doubling under a forcing of 1%/per year compounded CO_2 increase), or an impact-related regional variable such as the risk of wind-speeds in the South West of the UK exceeding some threshold in 30 years time. The basic approach can be characterised by Bayes' Theorem, namely

$$p(s|\text{data}) \propto p(\text{data}|s)p(s), \tag{1}$$

where data is some collection of observed climate variables and p denotes the probability. To statisticians the formula is well known and indeed it is increasingly appearing in the weather and climate literature (e.g. Robertson et al. 2004). Nevertheless, it is worth spending some time over the terms in order to provide an interpretation of Bayes' Theorem for the climate change prediction problem.

The term on the left hand side of Eq. 1 is termed the posterior probability of s (it is read as the probability of s given the data). It is our target prediction e.g. the probability density function of the TCR. The second term on the right hand side is the prior distribution of s. It may simply be taken from the raw output of a multi-model ensemble of AOGCM simulations (recognising that this

may be highly dependent on which models are included), it may be the output from an ensemble of perturbed physics experiments in which the perturbations are sampled from expert-defined distributions (Murphy et al. 2004), or it may be some sub-sample of the ensemble output in which the characteristics of the prior distribution are enforced: an example being a uniform distribution in which all prior values of TCR are considered equally likely. The important point is that it is the uncertainty in s before any observations are used. In the context of this study, it is the output of a small ensemble of simulations augmented by some statistical "emulation" of the TCR at untried combinations of parameter values (see Sect. 4). The first term on the right hand side is termed the likelihood and is perhaps the least familiar. Formally it is the "probability of the data given s" which is not particularly intuitive. Practically it is a measure of data-model fit: how "good" an model ensemble member is in comparison with observations.

Herein lies the major problem with long-term climate prediction. If we wished to predict, for example, the TCR, then the key observation needed to tightly constrain the prediction would be of the TCR itself. Yet we do not have, and will never have, such an observation. In probabilistic forecasting of shorter time scale variations of weather and climate, the relative likelihood of different ensemble members may be determined by examining many verification cycles, although this often neglected. (For example, Doblas-Reyes et al. (2005) find that the reliability of probabilistic seasonal predictions from a multi-model ensemble cannot be significantly improved by assigning different weights to the different models, while the reliability of probabilities derived from an ensemble of simulations of the same model with different initial conditions can be improved simply by inflating the ensemble spread to be consistent with what happened in the past.) In the climate change prediction problem we must, in virtually all cases, determine the likelihood from what we may term indirect observations of future climate change as there are obviously no observations of the future nor are there any direct historical or palaeoclimatic analogues. By indirect observations we mean common (and not so common) tests applied to climate models, e.g. their ability to simulate mean climate and variability, 20th century temperature trends, the response to volcanic eruptions etc. The key issue in probabilistic climate prediction is to use the available observations to attach formal estimates of likelihood to members of an ensemble and to constrain, as tightly as possible, the posterior distribution of future change. There should be a good correlation between the model variable that is being tested and the climate change variable of interest. Such information may be extract from the ensemble of simulations using, for example, a "perfect model" approach in which individual members are used as surrogates for the real world (Senior et al. 2004).

Schematic examples of how Bayes' theorem might be applied in practice are shown in Fig. 1. In Fig. 1a and b, the likelihood function is relatively narrow, that is the data in our possession contains much information about the forecast variable of interest. In these cases it matters little what the prior distribution is: the posterior distribution or prediction is virtually unaffected. The prior may be broad and flat (uniform) or even biased in comparison with the likelihood, although we note that if the prior is biased and relatively narrow then it may "pull" the posterior distribution away from what the data is telling us (thus we see the importance of sampling a wide range of the uncertainties when generating the ensemble).

For most climate prediction variables we are not currently in possession of sharply peaked likelihood functions. In Fig. 1c the likelihood function is broad and flat and has little impact on the posterior. In this case we rely almost entirely on the prior and hence our strategy for generating the ensemble is of critical importance to the prediction. In the case of the multi-model ensemble, the prediction may be heavily dependent on the model simulations available. The most likely scenario is that information in both the prior and the likelihood determine the forecast probability density function (Fig. 1d) and hence care must be taken in both designing and quantitatively assessing the ensemble.

2 Ensembles of "Perturbed Physics" AOGCMs

As stated above, previous ensemble work with version three of the Hadley Centre GCM has utilised a model set-up with a 50 m mixed-layer ("slab") ocean coupled to the atmosphere module of the GCM (Murphy et al. 2004; Stainforth et al. 2005). Parameters were perturbed in the atmosphere and sea-ice components in order to generate the ensemble and sample the uncertainties in climate feedback processes. Experiments were performed with CO_2 levels set at pre-industrial and double pre-industrial levels.

In Murphy et al. (2004), each of the 53 ensemble members was generated by perturbing atmosphere, sea-ice or land-surface parameters away from their standard value or by switching on and off particular options within a parametrisation scheme. Twenty-nine parameters/switches were considered in total and more details are given in the on-line supplementary of Murphy et al. (2004) and in Barnett et al. (2006). Since that work, we have performed a number of ensemble experiments in which multiple parameters are perturbed simultaneously. Here we utilize 129 model versions (the standard model and 128 variants) in which 29 of the model parameters and switches are perturbed simultaneously (Webb et al. 2006) according to the following algorithm.

Using the 53 experiments with single parameter perturbations, we assumed that the effects of multiple parameter perturbations to the model could be inferred from a linear combination of the output from those runs. We predicted the climate sensitivity and the Murphy et al. (2004) index of model skill [the Climate

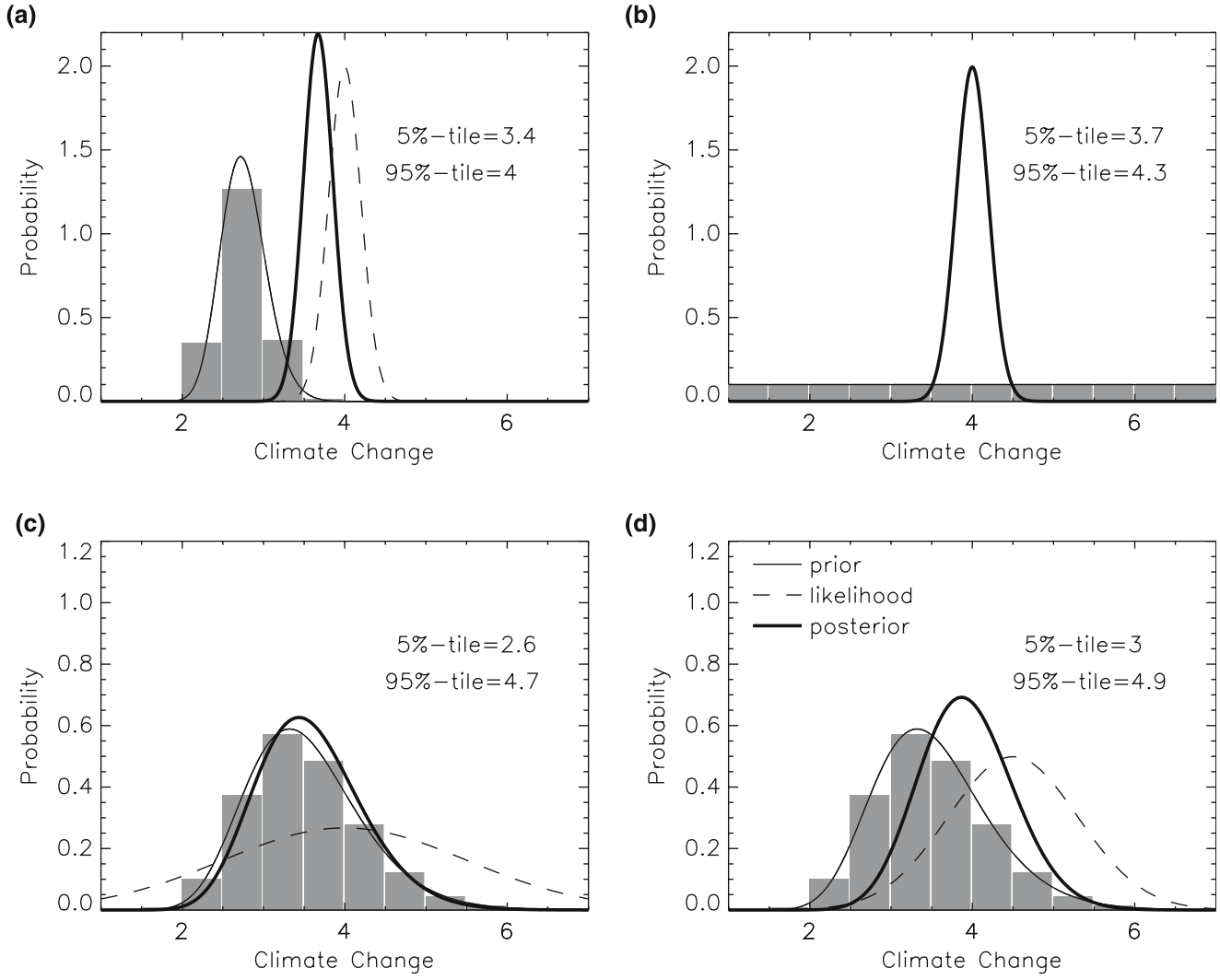

Fig. 1 Four schematic examples of how an ensemble of climate model experiments could by used to make probabilistic predictions of climate change. The *grey histograms* represent the raw output of a climate change variable from an ensemble of models. The *thin line* represents some smoothed version of this histogram generated by statistical means, potentially emulating the response at untried parameter values and is the prior distribution. The *dashed line* represents the likelihood and the thick line represents the posterior. See Sect. 1.1 for more details

Prediction Index, CPI, being the combined normalised root mean squared error (RMSE) of a number of time-mean surface and atmosphere fields for which good observations or re-analysis fields exist] for four million possible combinations of parameters with the parameters drawn from uniform distributions between the minimum and maximum of the expert specified ranges. From those four million hypothetical model versions, we picked 128 versions by splitting the resultant distribution of climate sensitivity into 64 equally probable bins and picking 20 model versions from each bin with the best CPI score. Taking this 1,280 subset of the four million, and starting from the model version with the best CPI score, we then picked the model that was furthest away from that version as measured by a non-dimensional measure of distance in model parameter space. The next model version was picked to be furthest away from these two and the process was iterated subject to picking only two versions from each climate sensitivity bin. The aim was to span the range of climate sensitivities consistent with a uniform prior on parameters but in the process maximise the chance of getting plausible model versions and span a wide range of parameter settings. The 128 model versions were run in slab configuration as in Murphy et al. (2004). We also picked 16 versions to run contemporaneously in fully coupled configuration by taking the parameter settings with the best predicted CPI score in the subset of eight model versions from four adjacent climate sensitivity bins. Again, the aim was to sampling a wide range of (predicted) sensitivities and parameter values from the 128 member slab model ensemble.

Table 1 gives values used in the 17 HadCM3 model versions used here: the standard version (Gordon et al.

Table 1 Parameters values used in the standard version of HadCM3 (labelled 3.0) and in the perturbed versions (labelled 3.1–3.16)

	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10	3.11	3.12.	3.13	3.14	3.15	3.16
V_{fl} (ms^{-1})	1	0.98441	1.12739	1.32784	0.68948	0.93976	1.37049	1.9116	0.6439	0.80071	0.50546	1.42339	0.52716	1.52212	1.72538	1.71394	0.98441
C_t (s^{-1})	1e-4	6.3e-5	0.000254	0.000337	0.000376	0.000163	0.000392	0.000275	0.000384	0.000114	0.000329	9.7e-5	0.000197	0.000273	0.000374	7.4e-5	6.3e-5
C_w (kgm^{-3})																	
Land	0.0002	0.0014	0.00148	0.000169	0.000185	0.000175	0.000272	0.000193	0.0002	0.00136	0.00172	0.000149	0.000111	0.000985	0.000518	0.00017	0.0014
Sea	5e-5	0.00035	0.00037	4.07e-5	4.55e-5	4.25e-5	6.8e-5	4.79e-5	5e-5	0.00034	0.00043	3.47e-5	2.33e-5	0.000246	0.000129	4.1e-5	0.00035
Rhcrit	0.7		0.79362	0.81494		0.84077		0.83237	0.66922	0.70242	0.82018	0.72092		0.78591	0.72857	0.7843	
Flow dependent Rh_{crit}	Off	On	Off	Off	On	Off	On	Off	Off	Off	Off	Off	On	off	off	off	on
Cloud fraction at saturation)																	
Boundary layer value	0.5	0.79552	0.6083	0.65616	0.62732	0.79598	0.73237	0.55859	0.5126	0.76827	0.73756	0.54611	0.50077	0.74373	0.73164	0.50049	0.79552
Free troposphere value	0.5	0.64776	0.55415	0.57808	0.56366	0.64799	0.61619	0.5293	0.5063	0.63414	0.61878	0.52306	0.50038	0.62187	0.61582	0.50025	0.64776
Vertical gradient of cloud water in grid box	Off	Off	Off	Off	Off	On	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off
Entrainment rate coefficient	3	2.91705	2.92534	3.48943	3.39838	2.89934	4.29985	4.46905	3.14446	2.66921	2.98215	2.98785	3.75359	2.95648	3.4243	3.07091	2.91705
Time scale for destruction of CAPE (h)	Off	1.02	Off	2.52	3.51	Off	1.54	1.18	Off	1.01	1.4	1.35	Off	Off	Off	Off	1.02
Convective anvils: excluding convective precipitation from cloud water path	Off	On	On	Off	Off	On	On	Off	Off	On	Off	On	Off	On	Off	On	On
Convective anvils: updraught factor	Off	0.28379	0.66215	Off	Off	0.51016	0.81295	Off	Off	0.99001	Off	0.97566	Off	0.2654	Off	0.69757	0.28379
Convective anvils: shape factor	Off	2.43144	1.66421	Off	Off	2.59138	1.06076	Off	Off	1.11988	Off	1.15898	Off	2.25243	Off	2.15719	2.43144
Sea-Ice Albedo																	
Albedo at 0°C	0.5	0.59963	0.58543	0.50884	0.54965	0.50666	0.5095	0.50918	0.63007	0.55931	0.63406	0.52458	0.53872	0.54743	0.50894	0.62347	0.59963
Albedo at Tcold	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Tcold (°C)	−10	−3.8889	−4.4214	−9.3686	−6.4536	−9.5243	−9.3214	−9.3443	−2.7474	−5.7636	−2.5977	−8.2443	−7.2343	−6.6121	−9.3614	−2.9949	−3.8889
Ocean-ice diffusion coefficient (m^2s^{-1})	2.5e-5	2.493e-5	2.393e-5	2.387e-5	2.447e-5	2.447e-5	2.437e-5	2.437e-5	2.013e-5	2.433e-5	2.393e-5	2.460e-5	2.413e-5	2.433e-5	2.480e-5	2.347e-5	2.213e-5
Ice particle size (μ m)	30	35.628	27.713	28.249	33.248	30.761	29.606	31.257	33.075	39.952	29.069	27.611	31.774	33.348	31.544	30.078	35.628
Non-spherical ice particles	Off	Off	Off	On	Off	Off	On	On	On	Off	Off	Off	Off	On	Off	Off	Off
Shortwave water vapour continuum absorption	Off	Off	On	Off	Off	On	Off	On	Off	Off	Off	Off	On	Off	Off	On	Off
Sulphur cycle	On	On	On	On	On	On	On	On	On	On	On	On	On	On	On	On	On
Order of diffusion operator	6	6	4	6	6	4	4	4	4	6	6	6	6	4	4	6	6
Diffusion e-folding time (hours)	12	17.978	21.127	6.926	10.074	8.513	8.569	10.76	9.772	7.067	13.096	14.179	8.795	19.691	19.24	14.887	17.978
Starting level for gravity wave drag	3	3	3	3	5	3	3	5	5	3	3	5	5	3	5	3	3

Table 1 (Contd.)

	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10	3.11	3.12.	3.13	3.14	3.15	3.16:
Surface gravity wave parameters																	
Typical wavelength (m)	20,000	19,800	16,400	11,700	17,100	19,800	11,600	19,500	10,400	16,300	11,800	19,900	19,800	11,100	12,100	10,600	19,800
Trapped lee wave constant ($m^{-3/2}$)	300,000	297,000	246,000	176,000	256,000	297,000	174,000	292,000	156,000	244,000	177,000	298,000	297,000	166,000	182,000	159,000	297,000
Surface-canopy energy exchange	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1
Forest roughness lengths (m)																	
Dense evergreen needleleaf forest	0.78	0.78	0.78	2.0	0.5	0.78	0.78	0.78	2.0	0.78	0.78	2.0	2.0	0.78	2.0	2.0	0.78
Dense deciduous needleleaf forest	0.78	0.78	0.78	2.0	0.5	0.78	0.78	0.78	2.0	0.78	0.78	2.0	2.0	0.78	2.0	2.0	0.78
Dense deciduous broadleaf forest	0.78	0.78	0.78	2.0	0.5	0.78	0.78	0.78	2.0	0.78	0.78	2.0	2.0	0.78	2.0	2.0	0.78
Equatorial rainforest	1.05	1.05	1.05	2.9	1.05	1.05	1.05	1.05	2.9	1.05	1.05	2.9	2.9	1.05	2.9	2.9	1.05
Dependence of stomatal conductance on CO_2	On	Off	Off	On	On	Off	Off	Off	On	Off	Off	On	On	Off	On	Off	Off
No. soil levels accessed for evaoptranspiration																	
Forest	4	2	2	4	4	2	2	2	4	2	2	4	4	2	4	4	2
Grass	3	1	1	3	3	1	1	1	3	1	1	3	3	1	3	3	1
Charnock constant	0.012	0.016	0.0172	0.0121	0.0125	0.0128	0.0128	0.0128	0.0184	0.0182	0.0157	0.0154	0.0132	0.0123	0.0134	0.0162	0.016
Free convective roughness length over sea (m)	0.0013	0.00488	0.00187	0.00285	0.00111	0.00414	0.00381	0.00468	0.00356	0.000579	0.000995	0.00242	0.000936	0.000571	0.00411	0.00334	0.00488
Boundary layer flux profile parameter	10	5.8703	9.8886	12.4727	9.5363	16.9231	8.5368	9.3776	17.7768	6.2373	8.6838	15.7703	6.5362	7.2765	16.8074	12.8178	5.8703
Asymptotic neutral mixing length parameter	0.15	0.30688	0.14272	0.08755	0.22946	0.3356	0.30901	0.30522	0.1424	0.47881	0.48694	0.4892	0.10945	0.40805	0.09932	0.10963	0.30688

More details of the parameters and there individual effect on climate sensitivity can be found in the supporting material of Murphy et al. (2004) and in Barnett et al. (2006)

Table 2 Key global mean quantities relating to the atmosphere-slab (HadSM) and atmosphere-ocean (HadCM) GCM experiments

Version	HadSM flux conv. Wm^{-2}	HadSM TOA Wm^{-2}	HadCM FA Wm^{-2}	HadCM TOA Wm^{-2}	HadSM s K ($\sigma = 0.07$ K)	HadCM TCR K ($\sigma = 0.08$ K)
3	−2.8	2.5	0.0	−0.2	3.4	2.0
3.0	−4.2	4.0	−4.5	4.1	3.5	2.1
3.1	5.9	−6.2	5.5	−6.0	2.6	1.6
3.2	−5.2	4.8	−5.3	5.0	3.1	2.0
3.3	−5.8	5.5	−6.2	5.8	3.8	2.3
3.4	−8.4	8.1	−8.5	8.2	4.6	2.4
3.5	5.6	−6.0	5.2	−5.7	2.2	1.5
3.6	−2.3	2.1	−3.2	2.9	3.8	2.1
3.7	−5.8	5.5	−6.3	5.6	3.3	1.9
3.8	−7.0	6.7	−7.3	6.8	4.9	2.5
3.9	2.1	−2.4	1.6	−2.0	2.2	1.6
3.10	−6.5	6.1	−6.1	6.2	3.9	2.0
3.11	−2.3	2.0	−2.7	2.1	3.2	2.1
3.12	-10.0	9.7	−10.2	9.8	4.9	2.6
3.13	−1.5	1.2	−1.6	1.5	3.2	2.1
3.14	−1.8	1.5	−2.1	1.7	3.1	2.1
3.15	−2.1	1.8	−2.3	1.8	2.9	1.7
3.16	−2.7	2.3	−3.8	2.9	3.0	2.1

Version 3 refers to the standard version of the model, 3.0 to the standard version with interactive sulphur cycle and versions 3.1 to 3.16 the perturbed physics versions (see Table 1 and text for more details). Column 2 is the global mean heat flux divergence in the slab component of the ensemble member. Columns 3 and 5 the top of the atmosphere flux imbalance in the slab and coupled members. Column 4 is the global mean of the flux adjustment term. Column 6 is the equilibrium climate sensitivity for $2xCO_2$ and column 7 is the transient climate response from the 1% per year CO_2 coupled model experiments

2000; Pope et al. 2000; Collins et al. 2001) and the 16 variants. In a further development on Murphy et al. (2004) we have also activated the interactive sulphur cycle component of the model in all our simulations. No attempt was made to perturb parameters in the ocean component of the AOGCM, although a small coding error inadvertently turned off the Richardson number dependence of the diffusivity calculation in the ocean vertical mixing scheme in all ensemble members. This has very little impact on large-scale climate and climate change. Ocean parameter perturbations will be the subject of future research. We adopt the following nomenclature for the ensemble experiments.

Coupled AOGCM	Atmos-Slab model	Description
HadCM3	HadSM3	Standard parametrisation settings
HadCM3.0	HadSM3.0	Standard parametrisation settings with interactive sulphur cycle
HadCM3.1	HadSM3.1	Perturbed parametrisation setting 1—see Table 1
HadCM3.2	HadSM3.2	Perturbed parametrisation setting 2—see Table 1
...	...	...
HadCM3.16	HadSM3.16	Perturbed parametrisation setting 16—see Table 1

The rate of time-dependent global-mean temperature change depends jointly on atmospheric feedbacks associated with climate change (here simply measured by the inverse of the climate sensitivity) and the efficiency of processes which remove heat from the surface to the deep ocean (e.g. Raper et al. 2002). The use of identical ocean components, coupled to atmospheric components with perturbed parameters, means we can isolate the component associated with atmospheric feedbacks. However, it also means that this study cannot attempt to capture the full range of uncertainty in transient climate change possible with HadCM3; a point which we shall return to in Sect. 3.1.

2.1 Use of flux adjustments

The atmosphere-slab set-up hides one of the major difficulties in generating AOGCM ensembles. In the experiments, a calibration phase is first performed in which the model sea surface temperature (SST) field is reset to the annually varying climatology at the end of each day. The heat convergence (or "q-flux") field that would have been required to achieve this reset is stored and a time average is produced at the end of the calibration phase. In subsequent control and double CO_2 experiments, this ocean heat convergence field is applied as a term which varies with position and season but not from year to year. It ensures that time averaged SSTs remain close to observed climatological values in the control simulation, while allowing SSTs to vary through changes to the surface heat flux balance resulting from natural variability or a change in CO_2. The heat convergence field represents the effects of ocean currents which are not explicitly simulated but also corrects biases in the surface heat flux balance that would cause the model to drift if the atmosphere were to be coupled to a dynamical ocean model. In particular, we follow standard practice in allowing it to take a non-zero global mean value.

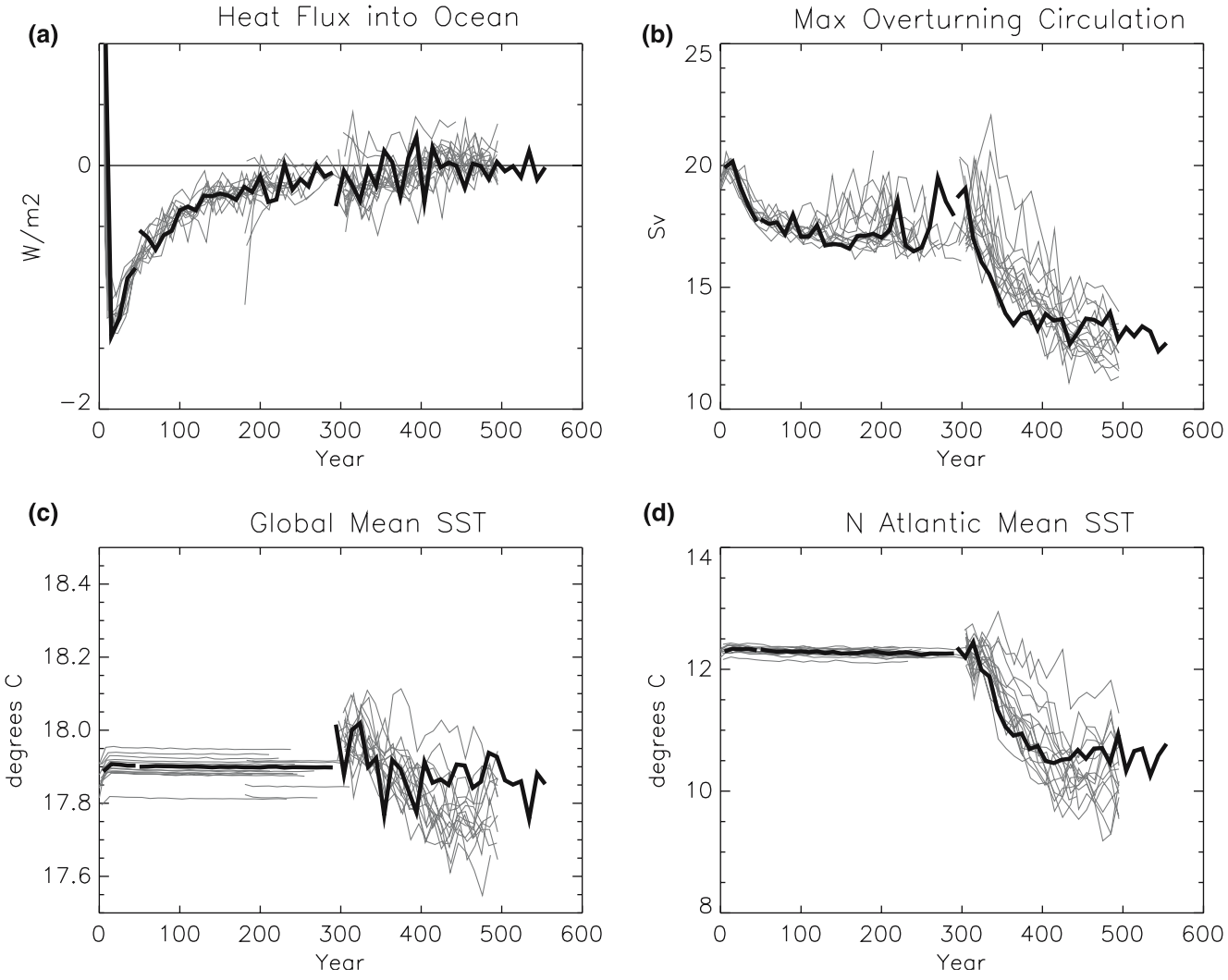

Fig. 2 Time series of key variables during the spin-up of the 17 member perturbed physics AOGCM ensemble. In the first ~300 years of integration a Haney forcing term is included with SSTs and SSSs relaxed to climatology with a relaxation timescale of approximately 15 days. The seasonally-varying relaxation term is averaged over the last 50 years of the phase and used in the flux-adjustment phase for a further 200–300 years. In this phase there is a rapid adjustment of the MOC in each member (see text for more details). **a** Total heat flux into the ocean from the atmosphere, including the relaxation and flux-adjustment term. **b** The maximum of the meridional streamfuntion in the North Atlantic. **c** The global mean SST and **d** the SST averaged in the region 50°E–10°W, 40°N–60°N in the North Atlantic. Units are indicated. The member with the standard parameter setting is highlighted by the *bold black line* and is representative of the behaviour in all ensemble members

In coupled GCMs with a full dynamic ocean component, the equivalent technique is flux adjustment (e.g. Manabe and Stouffer 1988), in which additive adjustments to the ocean surface heat and water fluxes are first calibrated from a preliminary integration with relaxation to observed climatological fields, and then applied during subsequent control and climate change simulations. The flux adjustments arise from errors in ocean transport and atmosphere–ocean exchanges. The IPCC TAR reported, for the first time, results from coupled model simulations performed without reliance on flux adjustments (Cubasch et al. 2001), identifying this as an important step forward in demonstrating the plausibility of large-scale climate change scenarios simulated by such models. Here, however, we revert to the use of flux adjustments, for the three reasons outlined below.

The earth receives an annual average of 341 Wm^{-2} of shortwave energy from the sun. In an equilibrium climate, this is balanced over the long term by the sum of the outgoing longwave and shortwave energy flux. When building an AOGCM, representations of different components of the physics and dynamics of the climate system are often developed independently, and once joined it is hoped that the complex nonlinear relationships between components will conspire to produce precisely 341 Wm^{-2} of outgoing radiation. In other words, we expect the top of the atmosphere (TOA) balance to be an emergent property of the model. In reality this is not the case and when the model is assembled, modification of parameters away from their best-guess values is required in order to achieve the balance. When we perturb parameters and parametri-

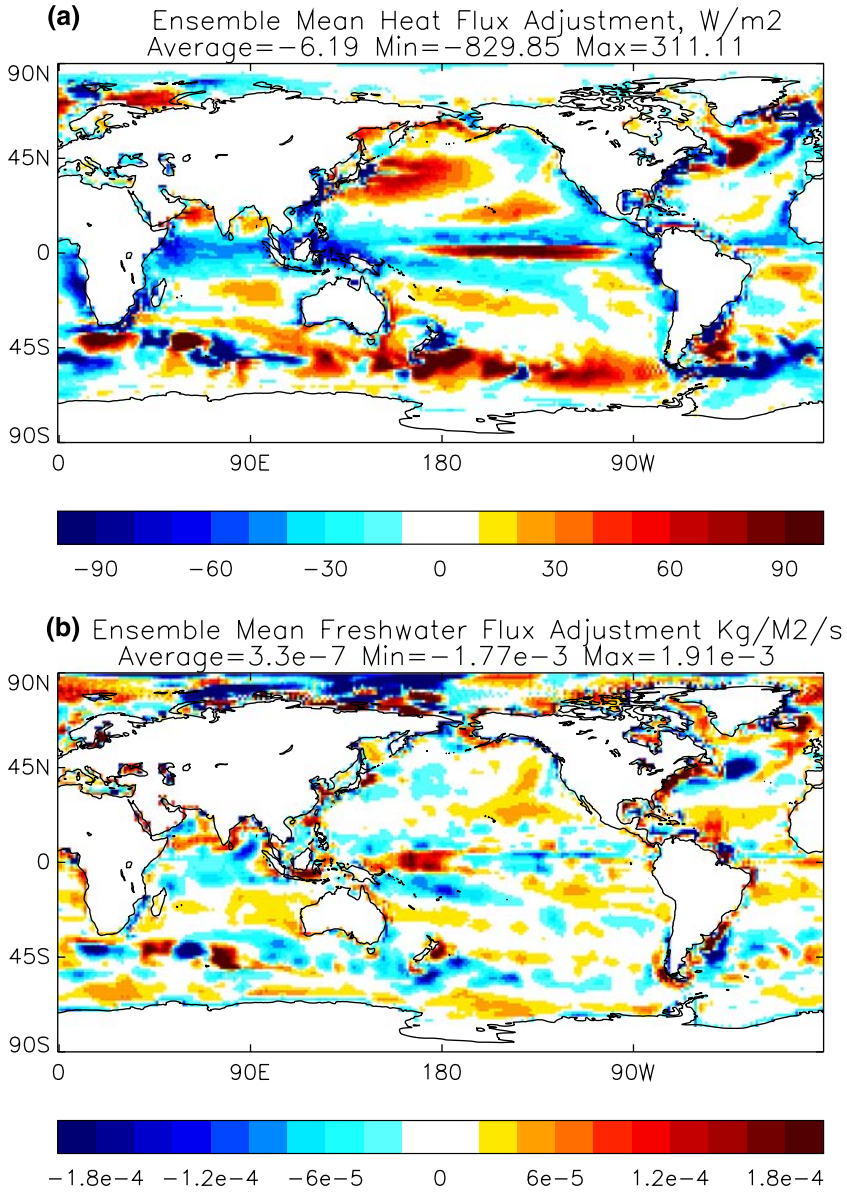

Fig. 3 Ensemble mean heat (**a**) and salinity (**b**) flux-adjustment terms. The ensemble mean is highly representative of the spatial pattern of flux-adjustment terms from the individual ensemble members with relevant differences being in the global mean values shown in Table 2. Similarities between the pattern of heat flux adjustment and the SST biases in the un-flux-adjusted HadCM3 (Fig. 4) are noted

sation schemes to generate our ensemble, the finely balanced TOA radiation budget is upset, typically by a few Wm^{-2} (see Table 2), even though the perturbed values of the parameters are just as plausible as the values which gave rise to the "standard" version of the model with near-perfect TOA balance. (Although note that the standard atmosphere-slab configuration of HadCM3 is still out of balance at the TOA by 2.5 Wm^{-2}; Table 2). It may be argued that we should only make perturbations which result in zero (or small) changes in the model TOA balance. However, even in a complex GCM it is clear that a component of model error arises from missing, poorly-resolved or structurally-deficient representations of physical processes, in which case the process of tuning model parameters may lead to TOA balance being achieved for the wrong reasons (e.g. switching on the more realistic sulphur cycle results in additional 1.5 Wm^{-2} change to the TOA radiation in the standard model version; Table 2). The imposition of TOA balance as a fundamental constraint could artificially restrict the space of allowable parameter combinations consistent with expert knowledge of the individual processes which the parameters control. Using an explicit flux adjustment (calculated separately for each ensemble member), rather than relying on an implicit correction applied by tuning the model parameters, allows us to sample, in principle, the full range of combinations of model parameter values.

Whilst the ability to produce reasonable simulations of continental-scale features of present day climate without flux adjustment is a welcome achievement, this is achieved at the expense of substantial regional biases in SST and salinity (e.g. Gordon et al. 2000). This potentially reduces the plausibility of the simulated

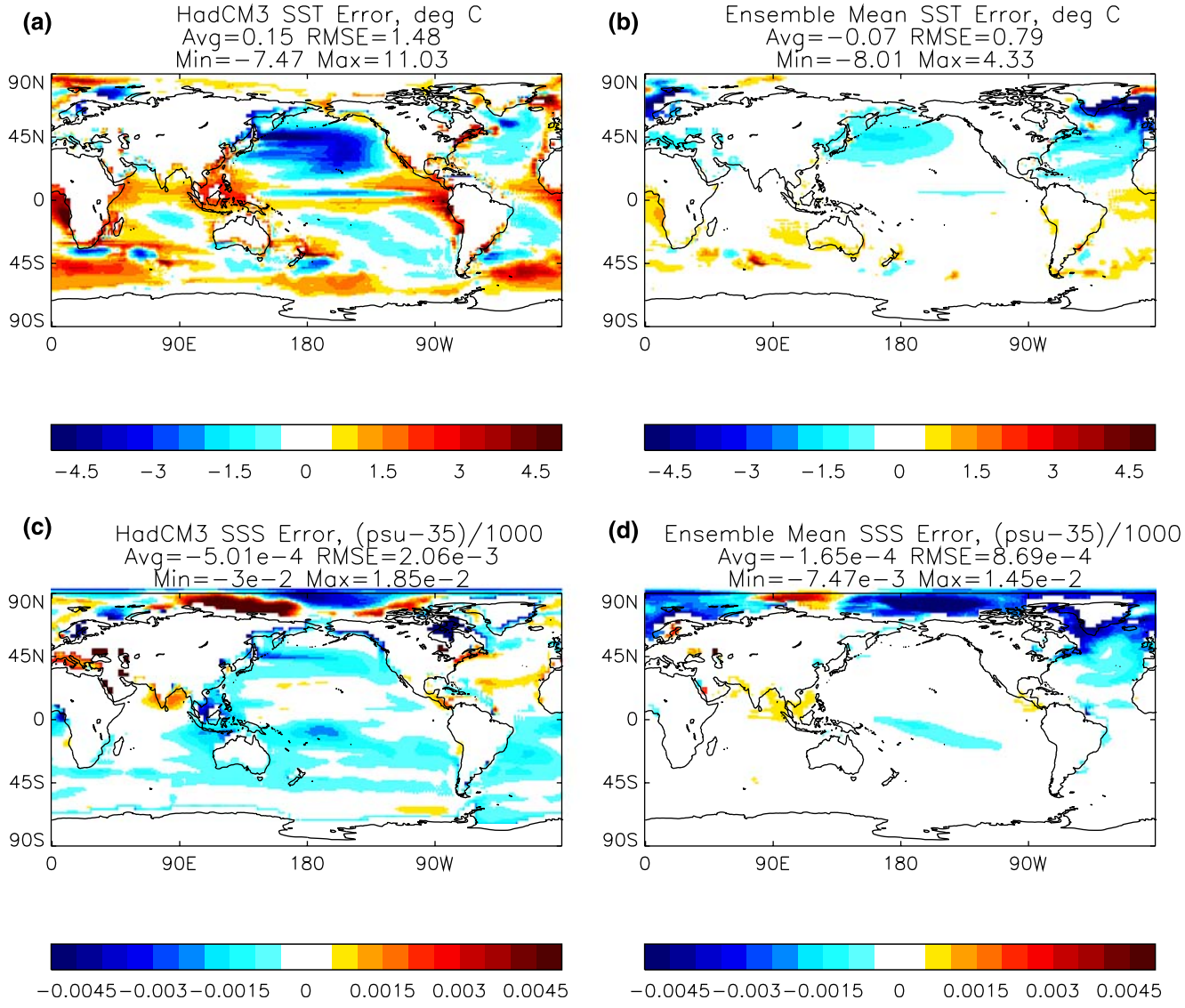

Fig. 4 SST and SSS biases in the un-flux-adjusted HadCM3 (**a**, **c**) and the ensemble mean of the flux-adjusted perturbed physics ensemble (**b**, **d**). Units are as indicated on the figure. The ensemble mean is highly indicative of SST and SSS biases in each ensemble member. One notable feature is the excessively cool and fresh ensemble-mean North Atlantic which is a result of the reduced strength MOC (see text)

changes and natural variability of climate at a regional scale, to the point where the regional modelling community typically generate their detailed scenarios of sub-continental climate change from intermediate "time-slice" atmosphere-only simulations in which the impact of biases in SST are removed (e.g. M. Deque et al. submitted; D.P. Rowell submitted). It is impractical to produce time-slice simulations for each of a large ensemble of model versions, so we take the view that the use of flux adjustments to control regional SST biases is essential to provide a basis for plausible regional simulations. In addition, SST biases may influence the distributions of clouds, sea-ice and other variables which are key in determining the feedbacks which exert a leading-order control on climate sensitivity and TCR, so it could be argued that models with significant SST biases are less plausible than those models which avoid such biases through the use of flux adjustments. However we also recognise that the use of flux adjustments artificially limits the development of simulation biases. It would be therefore be very important to include the magnitude of the flux adjustments as an element in any metric of model skill used to calculate relative likelihoods for alternative model versions, although we reserve this step for future work.

In order to provide (in future) credible regional probability distributions for transient climate change, we will need ensembles much larger than the 17 members considered here. It remains computationally unfeasible to generate ensembles of the required size (perhaps 100

or more members) using in-house computer resources, so our strategy relies on an ability to augment the 17 member ensemble by scaling the results of larger slab model ensembles simulating the equilibrium response to doubled CO_2. This approach, described in detail by G.R. Harris et al. (submitted), relies on an ability to establish robust, traceable relationships between the equilibrium and transient responses of a given model version, hence it is essential that consistent strategies are used to generate the slab and transient simulations. If we were to (effectively) use flux adjustments in the slab ensemble but not the coupled ensemble, the control simulations would differ substantially and our ability to map the equilibrium response to the transient response could be substantially impaired.

The following steps were performed to compute the flux adjustment fields.

1. Each model version was started from an initial state taken from near the end of the long multi-millennial HadCM3 control experiment (Gordon et al. 2000). The greenhouse gases, solar constant, background volcanic aerosol, ozone and sulphur emissions were all set at levels appropriate to pre-industrial conditions.
2. A Haney forcing, i.e. a relaxation to an observed seasonally and spatially varying climatology of SST and sea surface salinity (SSS), was applied. The SST climatology was taken as an average of years 1871–1900 of the HadISST1 data set (Rayner et al. 2003) in which observed SSTs are used to construct fields with full spatial coverage. The SSS climatology is that described in Levitus (1994) and for both a relaxation coefficient of 164 $Wm^{-2}\,K^{-1}$ was employed. This corresponds to an e-folding time of 15 days for a 50 m mixed-layer ocean. This Haney forcing phase was run until the 50 year global average heat flux into the ocean (all terms including the Haney forcing terms) was less than 0.2 Wm^{-2} (Fig. 2a). This required an average of approximately 300 years of model simulation for each member.
3. The SST and SSS Haney terms were then averaged over the last 50 years of the Haney forcing phase to form the seasonally and spatially varying flux adjustment fields. The ensemble annual average of the SST and SSS flux adjustment terms are shown in Fig. 3. The fields for each member are very similar and highlight common biases found in coupled models e.g. in the region of the Gulf Stream and in the equatorial Pacific. The global mean values are given in Table 2 and are consistent with the global means of the equivalent slab model heat flux divergence fields. The global-mean heat flux adjustment terms are of the order of a few Wm^{-2} and hence the errors in the outgoing radiation that result from the cancellation of the spatially-varying positive and negative biases are only between 0.5 and 3%, yet this would be enough to cause significant drift in global-mean quantities.

We would expect each ensemble member run with flux adjustments to have a stable climate, provided the flux-adjustment term is calculated over a time period of relative stability (as is the case here), and assuming that nonlinear effects of climate variability on the time averaged state are small.. However there is some drift which happens in all the members when the flux adjustment phase is started. Figure 2b shows the strength of the Atlantic Meridional Overturning Circulation (MOC) during the Haney and flux adjustment phases. During the Haney phase the circulation in each ensemble member is slightly weaker than that seen in the standard HadCM3 non-flux adjusted control experiment, but once the transition is made to flux adjustments the circulation weakens by several Svedrups before stabilising at a new level.

The standard version of HadCM3 has a tendency to form ocean waters which are too salty at high latitudes in the North Atlantic, and this is apparently true of all the ensemble members here despite the physics perturbations (see the ensemble mean salinity flux adjustment term in Fig. 3b). The strength of the model MOC is closely related to the high-latitude ocean surface density via its connection with the meridional steric height gradient and interior meridional density gradient (Thorpe et al. 2001; Vellinga and Wu 2004). During the Haney phase, the SSS is strongly constrained and the forcing term provides a correction to the weaker than observed (but still positive) fresh water input in the region of oceanic convection. This initial correction results in a slight reduction in MOC strength at the beginning of the Haney phase (Fig. 2b) as the fresh water is mixed down and affects the meridional density gradient.

When the Haney relaxation term is switched off and the atmosphere and ocean are fully coupled at the beginning of the flux adjustment phase, natural variations in the high-latitude fresh water budget are possible. Any variability-induced change in the fresh water input causes a change in surface water density and subsequent vertical mixing, interior density gradient and MOC strength. If the system were linear, we would expect a positive ''pulse'' of freshwater to result in a similar magnitude reduction in MOC strength as the increase in strength that would come from and equal but opposite negative pulse. However, the system is not linear, and indeed appears close to criticality in the sense that a reduction in MOC strength associated with a positive freshwater pulse is very much greater than the increase associated with a negative pulse. A reduction in fresh water input results in a slight increase in mixing and an associated small increase in the overturning whereas as small increase in fresh water input seems to cap the mixing and results in a rapid reduction in each ensemble member. Fortunately the system is not absolutely unstable as the (albeit) reduced strength MOC does transport dense salty water polewards from the tropical regions resulting in a stabilisation of high-latitude surface water density and corresponding stabilisation of the

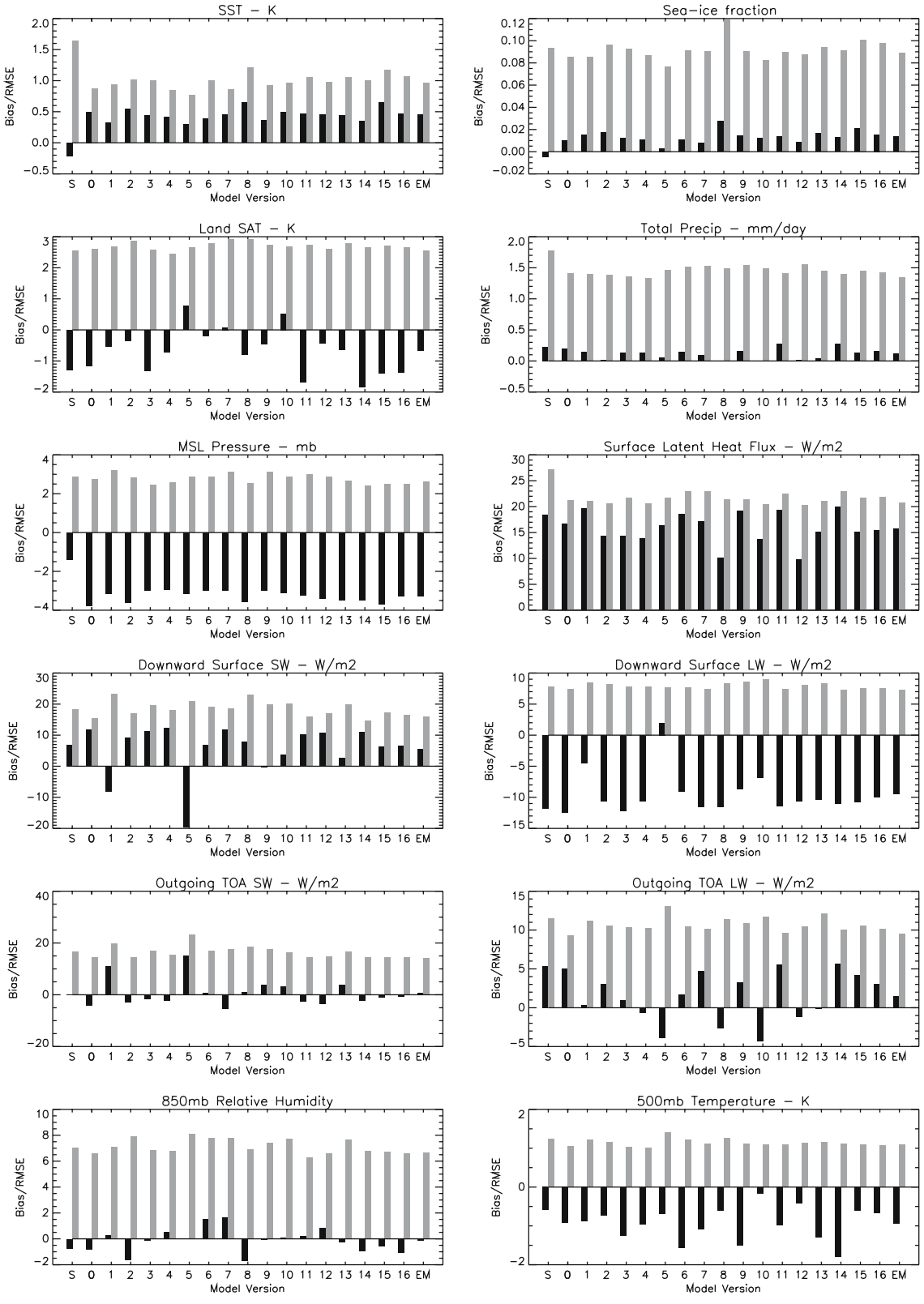

Fig. 5 Global mean bias (*black bars*) and root mean squared error (RMSE, *grey bars*—computed with the global bias removed) for a number of key modelled climate variables computed with respect to appropriate observations or re-analysis fields. Seasonal mean values from the control experiments are first calculated and then averaged to produce an annual mean. *S* indicates the standard (and un-flux-adjusted) version of HadCM3, 0 is the flux-adjusted version of this model with standard parameters and 1–16 are the versions with perturbed physics. EM indicates the bias and RMSE computed for the ensemble mean of the 17 model versions

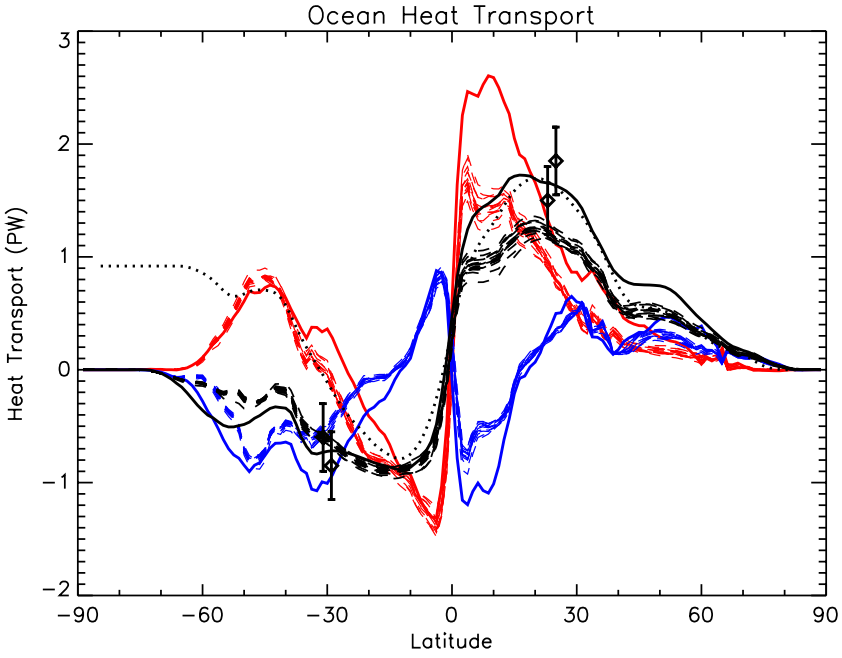

Fig. 6 Zonal mean total ocean heat transport (computed from ocean variables) in the standard un-flux-adjusted control run (*black solid line*) and from the flux-adjusted ensemble control simulations (*black dashed lines*). The total heat transport is further broken down into its main components associated with the overturning (*red*) and gyre (*blue*) circulations. Again the un-flux-adjusted version is indicated by the *solid line* and the ensemble members by the *dashed lines*. The *dotted line* is an estimate computed from an observed surface flux data set (Grist and Josey 2003) adjusted to have a near zero global annual mean and the diamonds with error bars are estimates from observed ocean sections (MacDonald and Wunsch 1996; Ganachaud and Wunsch 2000)

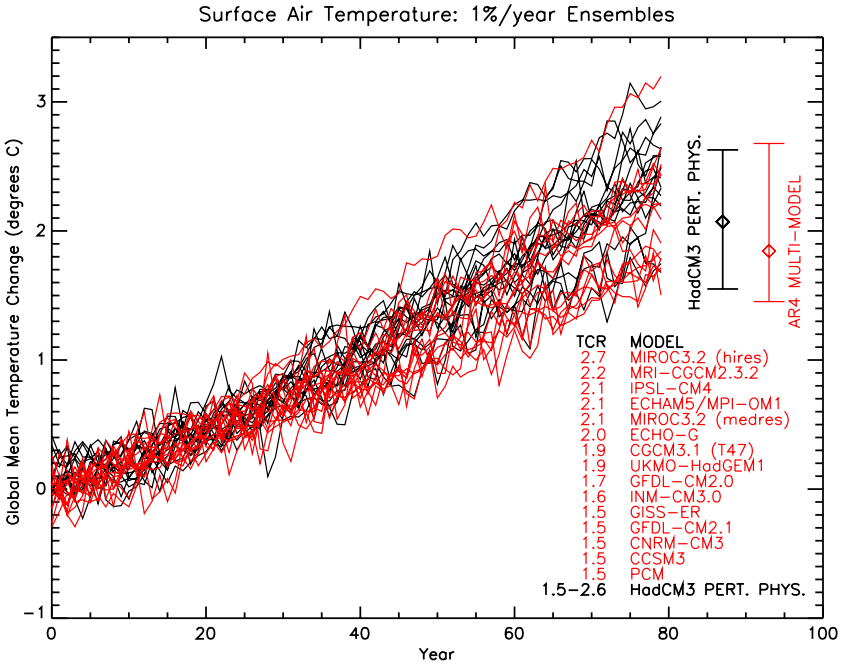

Fig. 7 Global mean surface air temperature change from the perturbed physics ensemble (*black lines*) and from the multi-model ensemble of models submitted to the IPCC AR4 assessment (*red lines*). The forcing is a 1% per year compounded rise in CO_2 and temperatures are expressed in terms of their anomalies with respect to the average of the relevant 80 years of control simulation. The *bars* on the right are the range of Transient Climate Response (TCR, see text for definition) from the perturbed physics and AR4 ensembles respectively with values for individual models indicated in tabular form. For the UKMO-HadGEM1 and MIROC3.2 hires models, only the scenario in which CO_2 concentrations are held fixed after year 70 were available resulting in a likely but small underestimation of the model TCR

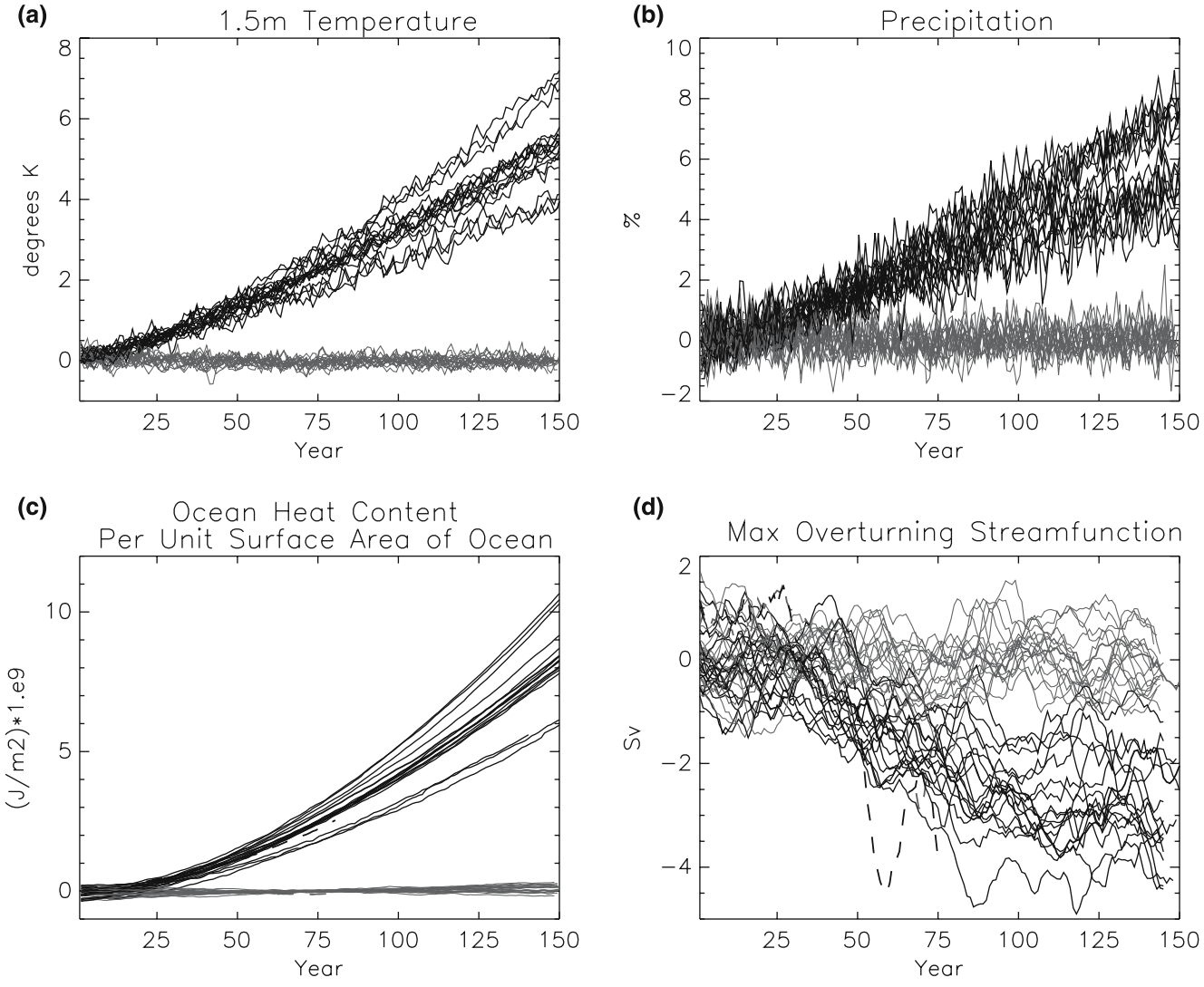

Fig. 8 Time series of key variables from control (*grey lines*) and 1% scenario experiments (*black lines*). *Solid lines* are for the 17 perturbed physics ensemble members and (where it is possible to make it out) the dotted line is from the standard HadCM3 experiment (run only to year 80 of the 1% scenario). Anomalies for each ensemble member computed with respect to its corresponding control experiment. **a** Annual, global mean surface air temperature in K. **b** Annual, global mean precipitation change (%). **c** Annual, global mean ocean heat content per unit surface area of ocean (Jm^{-2}). **d** Annual mean maximum of the Atlantic meridional streamfunction (Svedrups) smoothed with a simple 10-year running mean filter

MOC (the feedback described in Thorpe et al. 2001 and Vellinga and Wu 2004).

In each member, the slowing of the MOC results in a reduced northward heat transport and a cooling of North Atlantic SSTs (Fig. 2d), although global SSTs are relatively stable (Fig. 2c). On average, 200–300 years of AOGCM simulation were required to spin up each individual ensemble member in the flux adjustment phase. As a result, the SST and SSS fields for each member contain unexpected regional biases in the North Atlantic. These biases are shown in Fig. 4 (the ensemble mean is very representative of individual members) and are compared with the biases in the standard un-flux adjusted version of HadCM3. With the exception of the North Atlantic, biases in the flux adjusted versions of the model are less than those seen in the un-flux adjusted model (note the substantial reduction in annual RMSE), demonstrating that in most parts of the world ocean the flux adjustment successfully achieves our aim of limiting regional biases compared to simulations without flux adjustment. However, note that an additional effect of the North Atlantic cooling is an excessive build-up of sea ice in the polar-regions in each member (see Fig. 5 described later).

2.2 Ensemble experiments

We perform the following set of experiments with the spun-up standard and perturbed physics ensemble.

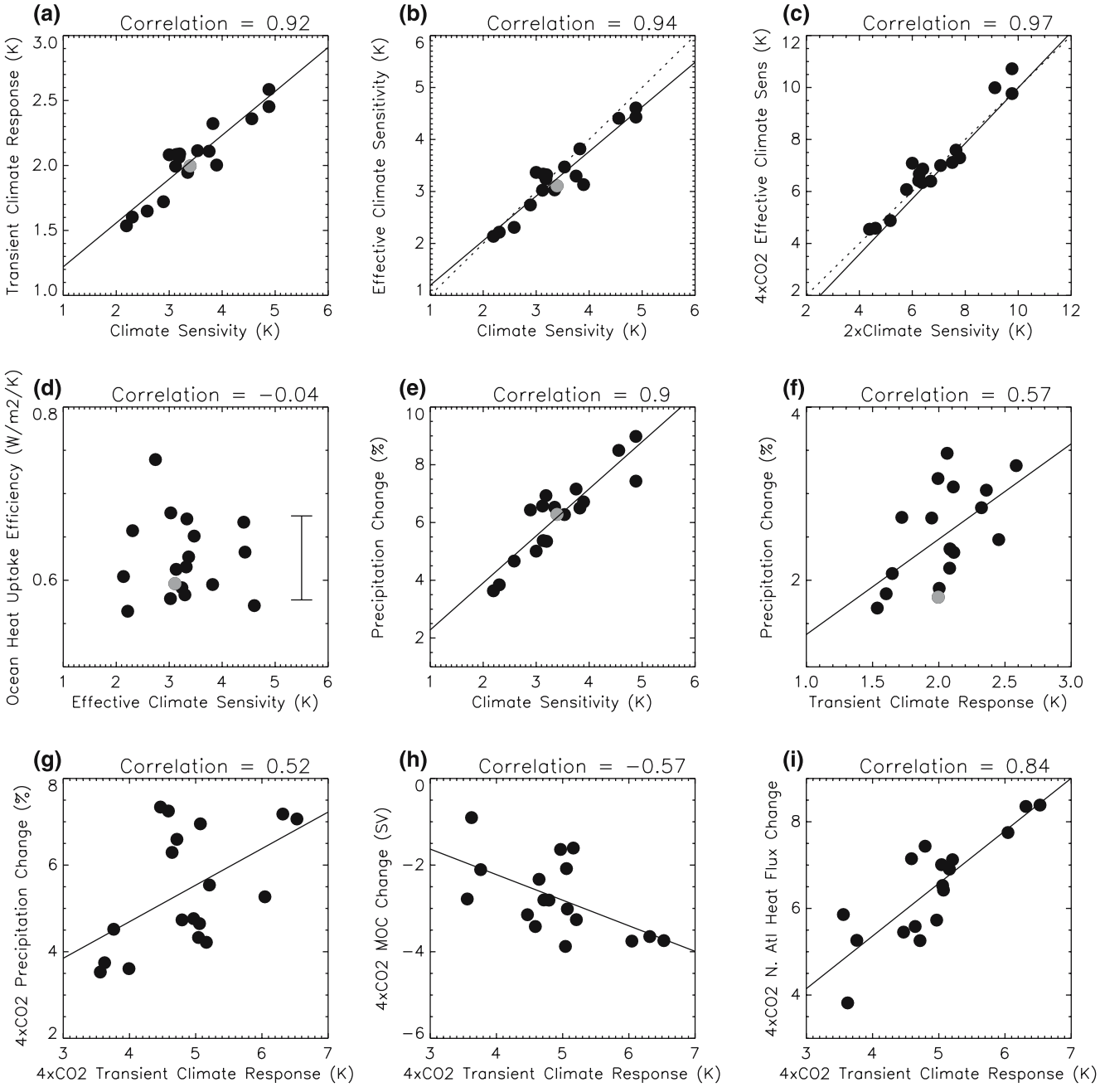

Fig. 9 Relationships between various key global mean diagnostics from the perturbed physics ensemble (*black dots*) and from the standard un-flux-adjusted HadCM3 experiment (*grey dot*). **a** Climate sensitivity versus TCR with fitted regression line. **b** Climate sensitivity versus effective climate sensitivity, with regression (*solid*) and unit gradient line (*dotted*). **c** Twice the climate sensitivity versus the effective climate sensitivity at the time of quadrupled CO_2 with lines as in **b**. **d** Effective climate sensitivity versus the ocean heat uptake efficiency with ± 2 standard deviations in the calculation indicated by the error bar. **e** Climate sensitivity versus percentage precipitation change from the corresponding slab model experiments. **f** TCR versus percentage precipitation change at the time of doubling of CO_2. **g** TCR at the time of CO_2 quadrupling versus percentage precipitation change at that time. **h** TCR at the time of CO_2 quadrupling versus MOC change at that time. **i** TCR at the time of CO_2 quadrupling versus change in total heat flux into the ocean in the North Atlantic region 70°E–10°W, 0–80°N at that time. More details are given in the text

1. AOGCM control experiments with CO_2 and other forcing agents fixed at pre-industrial values. Each experiment is 240 years in length.
2. AOGCM experiments with CO_2 increased at a rate of 1% per year compounded. Each experiment is 150 years in length (i.e. to four times pre-industrial concentrations).
3. Atmosphere-slab model experiments with pre-industrial CO_2. Each experiment was run to equilibrium and model output was averaged over 20 years. The

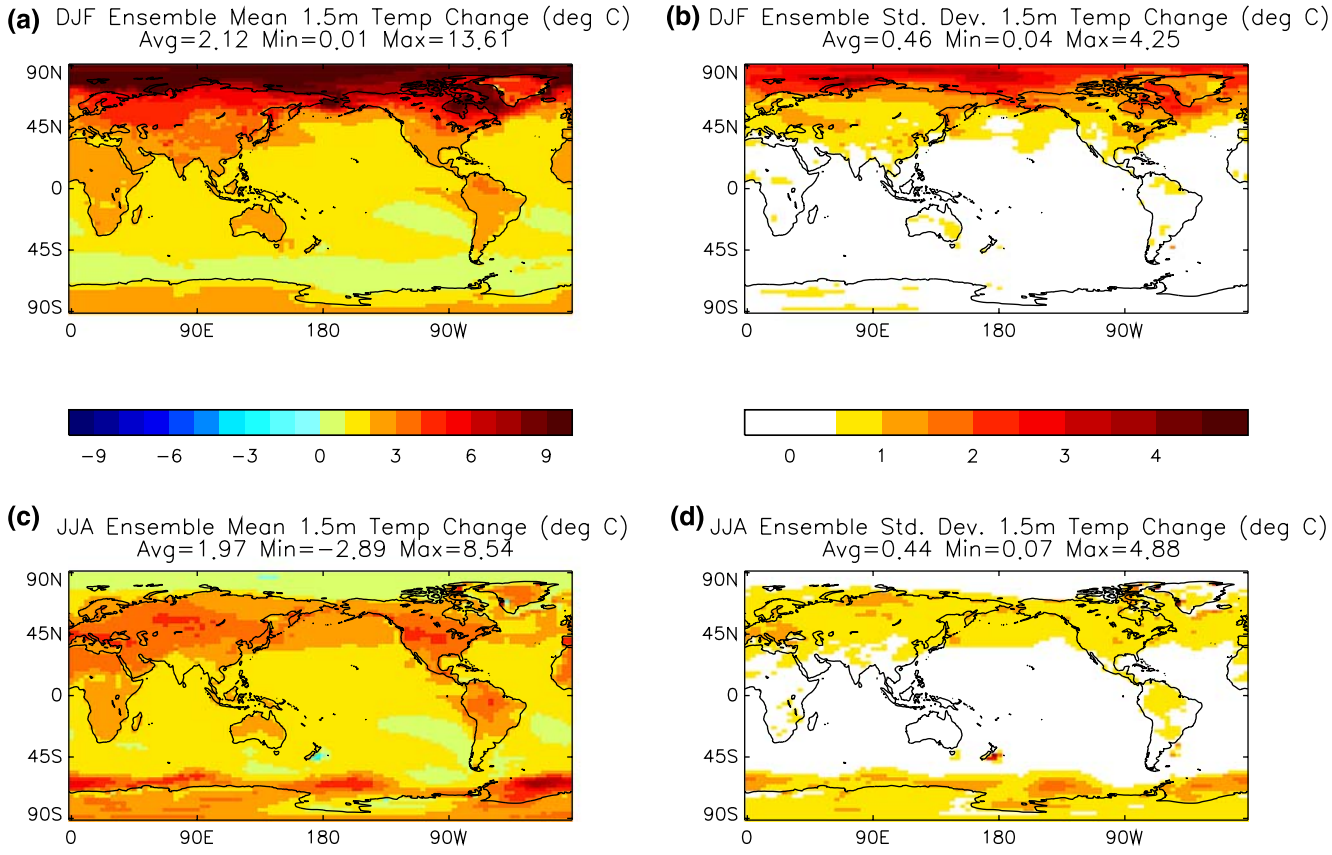

Fig. 10 The ensemble mean and standard deviation of temperature change from years 61 to 80 of the 1% per year ensemble. **a** and **b** for December to February and **c** and **d** for June to August. For each member, anomalies are computed with respect to the corresponding control simulation

corresponding 17 members are a subset of the bigger 128 member ensemble.

4. Atmosphere-slab model experiments with twice pre-industrial CO_2. Each experiment was run to equilibrium and model output was averaged over 20 years.

2.2.1 *Evaluation of control climates*

In this section, we evaluate the simulated climates of the AOGCM ensemble and compare them with the standard un-flux-adjusted version of HadCM3. There are a large number of possible observational data sets and evaluation methods we could employ to validate both mean climate and variability, but space considerations limit us to a small number of summary statistics. In Sect. 1 it was emphasised that model evaluation should be used to produce a formal estimate of likelihood in the Bayesian prediction. This is the subject of on-going work so here we present a more "traditional" evaluation with the simpler aim of highlighting the relative skill of the perturbed and flux-adjusted ensemble in comparison with the standard and well-used model version.

Figure 5 shows summary statistics for a number of different climate variables for which observed (or re-analysis) data exists with relatively complete global coverage. The variables are chosen as a combination of user-related quantities and other variables which are less directly relevant to users but are potentially important determinants of the credibility of simulated climate feedbacks, such as those which affect the radiation balance of the members. We show both the globally averaged RMSE and the global mean bias. The former is a useful measure of the average regional fit between the model and the observations but we note that for the latter it is often difficult to compute the absolute value of the observed global mean for many variables. Hence the bias should be interpreted mainly as a measure of how different the model versions are from each other rather than from the real world.

In broad terms the flux-adjusted ensemble members are not significantly worse, and in some cases can be better than the un-flux-adjusted standard version of HadCM3. For SST, each member shows a similar RMS error although this is dominated by the North Atlantic region in the perturbed physics ensemble due to the MOC adjustment highlighted above (see Fig. 4). For all the ensemble members, the sea-ice bias is positive (i.e. too much sea-ice) which is again related to the MOC slow-down and associated cool North Atlantic. RMS errors in land-surface air temperature, total precipitation rate and mean sea level pressure, all relevant for

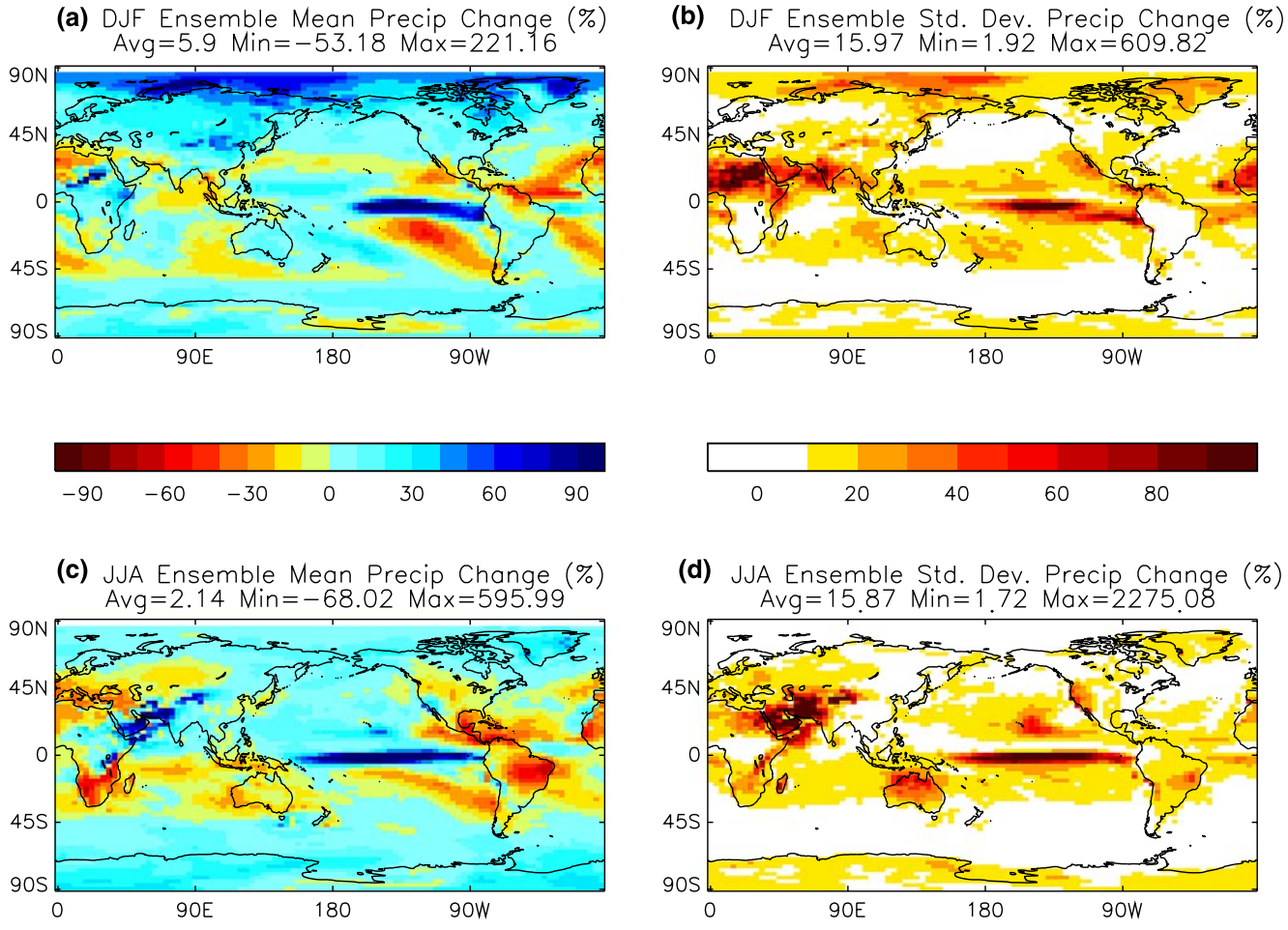

Fig. 11 As in Fig. 10 but for total precipitation expressed in terms of the percentage change with respect to the corresponding control simulation

user applications, are similar between the ensemble and standard version. [Global mean biases in mean sea level pressure are due in part to a drift in atmospheric mass in the spin-up phase. This drift is eliminated from the control and 1% runs (see later) by making a small modification to the model code.] While RMS errors in the energy-flux diagnostics are similar, there are significant differences in the global mean biases which are ultimately responsible for the small but significant variations in the TOA radiation balance which lead to our need to use flux adjustments.

Also shown in Fig. 5 are summary statistics for the ensemble mean of the model simulations. It is commonly found that the ensemble mean outperforms individual ensemble members, either in the sense that it produces smaller (local) biases in long-term climate fields (e.g. Lambert and Boer 2001), or in the sense that it verifies better in short term forecasts (e.g. Hagedorn et al. 2004). This would indicate that the process of ensemble averaging removes the random component of the model error. For all the variables assessed in Fig. 5, the RMS error of the ensemble mean is less than the RMS error averaged over all the ensemble members, with the range of improvement being between approximately 3 and 15%. Only for a few of the variables shown is there more than a single member that is better than the ensemble mean in this RMS sense. Thus the ensemble-mean appears to out perform individual members, although the improvement is perhaps not as dramatic as that seen in multi-model studies.

Figure 6 shows the horizontal transport of heat in the un-flux-adjusted HadCM3 and in the flux-adjusted standard and perturbed physics ensemble members compared with two estimates from observations. Each of the flux-adjusted models shows very similar but reduced northern hemisphere northward heat transport in comparison with the standard un-flux-adjusted HadCM3. This is largely due to a reduction in the overturning component associated with the reduced MOC strength in each member rather than any significant heat transport by the flux-adjustment term itself. While there are considerable uncertainties in the observations of this quantity, the northward heat transport in the flux-adjusted ensemble simulations is at the lower end of the estimates. Nevertheless, the ocean heat transport is within the range suggested by these and other observations and is consistent with that seen in other AOGCMs.

3 Climate change experiments

The 1%/year CO_2 increase scenario (hear after referred to as "the 1% scenario") was chosen as it is relatively simple to implement and understand and because it facilitates the comparison with other AOGCMs. We highlight the gross features of global mean climate change in comparison with other models in the following section before going on to look at regional changes.

3.1 Global and large scale climate change

Global mean surface air temperature (Figs. 7, 8a) shows an approximately linear trend due to the dependence of the radiative forcing on the logarithm of the CO_2 concentration. We use the TCR, the global mean anomaly in surface air temperature change relative to the equivalent control experiment averaged over years 61–80 at the time of CO_2 doubling, as a measure of the magnitude of the change. The perturbed physics ensemble has a TCR range of 1.5–2.6°C (Table 2, Fig. 9). Natural variability diagnosed from 20 year segments of the control experiments implies a one standard deviation error in a single TCR estimate of 0.08°C so the range is not simply due to statistical fluctuations, as is clearly evident by the increase of ensemble spread with time seen in the figure. The range may be compared to that from the models submitted as part of the multi-model assessment for the IPCC 4th assessment report (AR4) which is 1.5–2.7°C (Fig. 7). The equivalent range quoted in the IPCC 3rd assessment report (TAR) was 1.1–3.1°C and that from the current database of the Coupled Model Inter-comparison Project is 1.0–3.6°C, although a single model is responsible for the high upper figure and the range would be 1.0–2.1°C with this model excluded. Thus the spread of TCR values in the perturbed physics ensemble is comparable with that from other multi-model studies and is very close to that seen in the latest AR4 ensemble.

There is a strong relationship between the TCR of the given perturbed physics coupled ensemble member and the climate sensitivity from the slab-model experiment with the corresponding set of parameter perturbations (Fig. 9a). In general, we would expect a nonlinear relationship between transient and equilibrium change as models with higher climate sensitivities and the same ocean heat uptake efficiency (see below) have a slower relative warming rate (i.e. realising less of their total equilibrium warming at any particular time e.g. Raper et al. 2002). The relationship is approximately linear here because the ensemble does not explore extreme values of sensitivity. Taking figures from Table 9.1 of the TAR, the multi-model correlation is slightly weaker (0.73 in comparison with 0.92 in Fig. 9a) with three of the models exhibiting less warming at the time of CO_2 doubling than would be expected from their respective equilibrium sensitivity. In those three models the time-dependent effective climate sensitivity (Murphy 1995) at the time of CO_2 doubling is smaller than that computed from the corresponding atmosphere-mixed-layer experiment [see e.g. Senior and Mitchell (2000) for a specific example].

The time-dependent effective climate sensitivity from the perturbed physics ensemble is shown in Fig. 9b and c at two points of the time evolution. This was calculated as the ratio of the difference between the radiative forcing and net TOA flux and the temperature change, with the assumption that the radiative forcing due to the 1% per year increase in CO_2 increases linearly in time at a rate corresponding to a forcing of 3.8 Wm^{-2} at the time of doubling. Raper et al. (2002) show the calculation of effective climate sensitivity to be only very weakly dependent on this assumption, and we would expect very little variation across the ensemble as the same radiation scheme is employed in each member. Again there is a high correlation between the slab-model equilibrium sensitivity and the effective climate sensitivity, and the fitted regression line at the time of doubling and quadrupling of CO_2 both have slopes which are statistically indistinguishable from unity. This is reassuring, as the reduced strength MOC and excessive sea-ice in the coupled model experiments could have modified the strength of the feedbacks in the coupled-model when compared to those from the equivalent slab-model members. We may conclude that at the levels of CO_2 forcing and time scales considered, there are no changes in climate within the ensemble which lead to the sort of large variations in atmospheric feedbacks which might affect the climate sensitivity to the extent that other authors have seen (albeit on the long time scales associated with stabilisation: Senior and Mitchell 2000; Raper et al. 2001; Gregory et al. 2004).

Anomalies in ocean heat content from the control and 1% scenario experiments (Fig. 8c) demonstrate the stability of the integrations. The rate of ocean heat content increase in the 1% CO_2 increase experiments can be quantified by examining the relationship between the ocean heat uptake efficiency (see Raper et al. 2002) and climate sensitivity (Fig. 9d). We calculate the heat uptake efficiency by taking the 20 year average of the change in heat flux at the surface of the ocean at the time of CO_2 doubling and the TCR. The uncertainty bar in Fig. 9d is the ± 2 standard deviation error in the calculation expected from natural variability. Thus we interpret this figure as implying the ocean heat uptake efficiency is the same in each member (i.e. statistically significantly differences cannot be found). This is unsurprising as no ocean parameters were perturbed in the ensemble. Previous multi-model studies (Cubasch et al. 2001; Raper et al. 2002; Meehl et al. 2004) have found a correlation between the effective climate sensitivity and ocean heat uptake efficiency which tended to reduce the uncertainty range of the TCR. This is not the case here. They speculate that this is because the high-latitude warming is greater relative to the global mean in higher sensitivity models leading to greater heat uptake efficiency. In this ensemble there is only a weak corre-

lation between (for example) the ratio of high latitude warming to TCR and the TCR itself, which tends not to support this mechanism.

Figure 9e, f and g show the relative change in global mean total precipitation rate from the slab and coupled ensemble experiments versus the temperature change (see also Fig. 8b). Global precipitation change is closely related to global mean temperature change through the Clausius–Clapeyron relationship. The year 61–80 range in total precipitation change is 1.7–3.5% in the perturbed physics ensemble. Interestingly, the correlation is much weaker in the coupled model case than in the slab model case. This is not a simple signal to noise issue as the correlation across the coupled model simulations is also weak at the time of quadrupling of CO_2 (Fig.9g). Nor is it the case that coupled models have significantly more variability in 20 year mean precipitation than the slab models. This remains an issue for future research.

Also shown in Fig. 8d are anomalies in the strength of the MOC in the control and 1% ensemble. All members show a relatively stable MOC in the control phase (post the MOC adjustment highlighted above) and all members show a reduction in MOC strength as the CO_2 forcing increases. There is a moderate correlation between strength of the MOC weakening and the strength of the global mean surface temperature change (Fig. 9h). This is a manifestation of a stronger correlation ($r = 0.84$) between the global mean response and the change in the total heat flux into the North Atlantic ocean which is tempered by the relatively large decadal variability in the MOC (Fig. 7d and Vellinga and Wu 2004). There also a weaker ($r = 0.48$) correlation between the MOC slow down and the change in fresh water flux in the North Atlantic region indicating that both processes are important (Thorpe et al. 2001); although the heat-flux provides the large contribution. Despite starting from a state of relatively weak overturning, none of the members exhibit a complete collapse of the circulation under the 1% per year increase in CO_2 forcing.

3.2 Spatial variation of change and uncertainty

Examples of regional patterns of change and their uncertainty can be assessed by examining the perturbed physics ensemble average and standard deviation of seasonal mean surface air temperature and precipitation anomalies (Figs. 10, 11). These figures come with two caveats. Firstly, they are not formal indicators of the mean and width of some probability density function (under the 1% scenario) as no attempt has been made to apply the Bayesian algorithm (Sect. 1.1). Secondly, the figures present information at the model grid-scale which we might be wary of as GCMs only resolve the large scale dynamics. Nevertheless, they are good examples of the type of information which we hope to present in the future with formal estimates of probabilities.

Mean temperature changes show the familiar pattern of more warming at high latitudes in comparison to the tropics and more warming over land areas in comparison to ocean areas. In December to February (DJF) the uncertainty, as measured by the ensemble standard deviation, is largest in the northern high-latitude regions (the magnitude and pattern remain similar if the contribution form different global mean warming in different ensemble members is removed). In June to August (JJA) there is a similar uncertainty in the winter high latitudes, but also considerable ensemble spread over northern hemisphere land regions.

Ensemble mean precipitation changes are largest in the tropics in absolute terms but expressed as a percentage of the corresponding control simulation (Fig. 11a, c) significant changes at all latitudes are evident. There is, for example, a strengthening of the Asian Summer Monsoon and a weakening of the rainy season in DJF in the vicinity of the Amazon basin. The latter has been implicated in the potential for Amazon die-back as simulated in a version of the Hadley Centre model with an interactive carbon cycle included (Cox et al. 2000). While uncertainties in the predicted monsoon strength are large (Fig. 11d) they are relatively smaller in north-eastern South America indicating, at least for this model, some robust signal. The impact of uncertainties in Amazonian rainfall change on the probability for die-back will be a key application of this ensemble as coupled atmosphere–ocean processes are known to be implicated (Cox et al. 2004). For example, there is some indication of an El Niño-like signal in precipitation which has also been seen in other models (Collins et al. 2005).

4 Discussion and future work

There have been several recent attempts to produce probabilistic predictions of both equilibrium and transient global mean temperature climate change (Andronova and Schlesinger 2001; Wigley and Raper 2001; Forest et al. 2002; Knutti et al. 2002, submitted; Murphy et al. 2004; Piani et al. 2005). The response of society to the issue of climate change requires quantitative predictions at the regional level and for variables other than temperature. There are many potentially complex combinations of variables such as the frequency of extreme events and ultimately we require quantitative predictions of, for example, the impact of climate change on biological systems and economies. We believe that such detailed information about changes in the physical Earth System can only come from complex climate models. Moreover, in the absence of a perfect model of the Earth System and of a perfect knowledge of future forcing agents, the ensemble probabilistic approach is the best way to provide those predictions.

We have made some headway recently in generating ensemble simulations of equilibrium change using the perturbed physics approach (Murphy et al. 2004;

Stainforth et al. 2005). We require a systematic method for generating ensembles in order to produce reasonably large ensemble sizes to adequately explore possible nonlinearities, to specify a recognisable prior (see Sect. 1.1) and to test the sensitivity of the predictions to that prior. None of these are currently possible with the multi-model ensemble. However, there is clearly a requirement to move forward from the equilibrium change case to quantifying uncertainty in time-dependent future climate change. Thus the main motivation for this study was to test the validity of the perturbed physics approach for generating ensembles of transient climate change with fully coupled atmosphere–ocean models and to highlight some of the main issues and pitfalls. We have shown that it is possible to generate an ensemble in this way and moreover that the ensemble behaves in a way which is consistent with simple physical relationships between several global mean and large scale measures of equilibrium and transient climate change (e.g. Fig. 9) for the range of responses explored. This point is important as although we have generated some relatively large (and one very large) ensemble of perturbed physics atmosphere-mixed-layer-ocean models (Murphy et al. 2004; Stainforth et al. 2005; Webb et al. 2005), the spin-up time and length of integration of coupled model climate change experiments will limit our ability to generate similar sized ensembles. Hence it is our hope to infer or "emulate" the behaviour of any version of the coupled model by scaling its pattern of equilibrium climate change by the global mean transient change produced using an energy balance model, and allowing for differences between the transient and equilibrium patterns of change (G.R. Harris et al. submitted). This emulation step will be further extended to infer responses at untried parameter values.

One caveat is that we have only perturbed atmosphere, land and sea-ice parameters. We are currently working on assessing and running experiments with perturbed ocean parameters and these will be the subject of future papers. The assessment of uncertainties in ocean heat uptake is a crucial component of the time-dependent problem.

While the use of flux adjustments will seem to some a retrograde step, we feel that we must revisit its use. Improvements in models (e.g. in meridional ocean heat transport—Gordon et al. 2000) have opened up the possibility of running without flux adjustments, but we do not believe that we have resolved all the issues. Two specific examples are uncovered in this paper. Firstly, HadCM3 can apparently run without flux adjustments, yet in the atmosphere-mixed-layer configuration there is still a 2.5 Wm^{-2} TOA flux imbalance. Secondly, by adding a more complex representation of a physical process (an interactive sulphur cycle) this TOA imbalance is made worse. We could have rejected model versions which do not have a finely tuned TOA balance, but we have no evidence that those particular parameter combinations are not credible as there may be missing or poorly represented processes which could correct the balance if included. In addition, we intend to use these ensembles to make policy-relevant predictions, to provide boundary conditions for embedded regional models and as baseline models for the assimilation of data to make initial value predictions. For these applications, the existence of temperature and salinity biases significantly reduces utility.

Despite these arguments, the use of Haney forcing and flux adjustments as we have applied them (i.e. in the usual way) was not entirely successful. Each ensemble member exhibits a rapid adjustment in the strength of the Atlantic overturning in the initial flux adjustment phase which leads to a reduction in ocean northward heat transport, cooling of North Atlantic SSTs and build-up of excessive northern hemisphere sea-ice. While these biases do not appear to excessively influence global and large-scale climate change, they do perhaps limit the usefulness of the ensemble in other applications. We are currently working on techniques to limit the impact of ocean overturning instabilities in flux-adjusted ensembles.

We have made a small but significant step on the road to making estimates of the probability density function of future transient climate change. There are still many challenges ahead; the generation of sufficiently large ensembles to adequately span the range of uncertainty, constraining the ensemble with observations, augmenting the ensemble using statistical methods to produce PDFs, producing information at the regional scales demanded by stakeholders and communicating the probabilistic predictions in a way which stakeholders can interpret. These are all areas in which we are actively working and hope to make headway in the near future.

Acknowledgements This work could not have been possible without the input from numerous and dedicated Hadley Centre staff, in particular Peter Good who downloaded the AR4 model data and Ian Culverwell who supplied the ocean heat transport data. Chris Brierley also helped in the analysis of ocean heat uptake efficiency. The work was supported by the UK Department of the Environment, Food and Rural Affairs under Contract PECD/7/12/37 and by the European Community ENSEMBLES (GOCE-CT-2003–505539) and DYNAMITE (GOCE-003903) projects under the Sixth Framework Programme. We acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy.

References

Andronova NG, Schlesinger ME (2001) Objective estimation of the probability density function for climate sensitivity. J Geophys Res 106:22605–22611

Barnett DN, Brown SJ, Murphy JM, Sexton DMH, Webb MJ (2006) Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations. Clim Dyn (in press). DOI 10.1007/s00382-005-0097-1

Collins M (2000) Understanding Uncertainties in the response of ENSO to Greenhouse Warming. Geophys Res Lett 27:3509–3513

Collins M, Tett SFB, Cooper C (2001) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 17:61–81

Collins M., the CMIP2 Modelling Groups (2005) El Niño- or La Niña-like climate change? Clim Dyn 24:89–104

Covey C, AchutaRao KM, Cubasch U, Jones P, Lambert SJ, Mann ME, Phillips TJ, Taylor KE (2003) An overview of results from the Coupled Model Intercomparison Project. Glob Planet Change 37:103–133

Cox PM, Betts RA, Jones CD, Spall SA, Totterdell IJ (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408:184–187

Cox PM, Betts RA, Collins M, Harris PP, Huntingford C, Jones CD (2004) Amazonian forest dieback under climate-carbon cycle projections for the 21st century. Theoret Appl Climatol 78:137–156

Cubasch U, Meehl GA, Boer GJ, Stouffer RJ, Dix M, Noda A, Senior CA, Raper S, Yap KS (2001) Projections of future climate change. In: Climate Change 2001: The Scientific Basis. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden P, Dai X, Maskell K, Johnson CI (eds.) Contribution of Working Group I to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, 525–582

Doblas-Reyes FJ, Hagedorn R, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and combination. Tellus (in press)

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

Ganachaud A, Wunsch C (2000) Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. Nature 408:453–457

Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 16:147–168

Gregory JM, Ingram WJ, Palmer MA, Jones GS, Stott PA, Thorpe RB, Lowe JA, Johns TC, Williams KD (2004) A new method for diagnosing radiative forcing and climate sensitivity. Geophs Res Lett 31:L03205

Grist JP, Josey SA (2003) Inverse analysis adjustment of the SOC air-sea flux climatology using ocean heat transport constraints. J Clim 20:3274–3295

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2004) The rationale behind the success of multimodel ensembles in seasonal forecasting. Part I: Basic concept. Tellus (in press)

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

Lambert SJ, Boer GJ (2001) CMIP1 evaluation and intercomparison of coupled climate models. Clim Dyn 17:83–106

Levitus S, Boyer T (1994) World Ocean Atlas 1994. NOAA Atlas NESDIS, U.S. Department of Commerce, Washington

MacDonald AM, Wunsch C (1996) An estimate of global ocean circulation and heat fluxes. Nature 382:436–439

Manabe S, Stouffer RJ (1988) Two stable equilibria of a coupled ocean-atmosphere model. J Clim 1:841–866

Meehl GA, Washington WM, Arblaster JM, Hu A (2004) Factors affecting climate sensitivity in global coupled models. J Clim 17:1584–1596

Molteni F, Buzzia R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: methodology and validation. Quart J Roy Met Soc 122:73–119

Murphy JM (1995) Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: Analysis of global mean response using simple models. J Clim 8:496–514

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

Palmer TN (2001) A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. Quart J Roy Met Soc 127:279–304

Palmer TN (2002) The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. Quart J Roy Met Soc 128:747–774

Piani C, Frame DJ, Stainforth DA, Allen MR (2005) Constraints on climate change from a multi-thousand member ensemble of simulations. Geophys Res Lett 32:L23825. DOI 10.1029/2005GL024452

Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact pf new physical parametrizations in the Hadley Centre climate model—HadAM3. Clim Dyn 16:123–146

Raper SCB, Gregory JM, Osborn TJ (2001) Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. Clim Dyn 17:601–613

Raper SCB, Gregory JM, Stouffer RJ (2002) The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. J Clim, 15:124–130

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. J Geophys Res 108:10.1029/2002JD002670

Robertson AW, Lall U, Zebiak SE, Goddard L (2004) Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. Month Weather Rev 132:2732–2744

Senior CA, Mitchell JFB (2000) The time-dependence of climate sensitivity. Geophys Res Letts 27:2685–2688

Senior C, Wielicki B, McAvaney B, Boer G (2004) Report on the joint WCRP CFMIP/IPCC expert meeting on climate sensitivity and feedbacks. Annex 5 of IPCC Working Group 1 report of the Workshop on Climate Sensivity

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433:403–406

Tebaldi C, Smith RL, Nychka D, Mearns LO (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. J Clim 18:1524–1540

Thorpe RB, Gregory JM, Johns TC, Wood RA, Mitchell JFB (2001) Mechanisms determining the Atlantic thermohaline circulation response to greenhouse gas forcing in a non-flux-adjusted coupled climate model. J Clim 14:3102–3116

Toth Z, Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. Month Weather Rev 125:3297–3319

Vellinga M, Wu PL (2004) Low-latitude freshwater influence on centennial variability of the Atlantic thermohaline circulation. J Clim 17:4498–4511

Webb MJ, Senior CA, Williams KD, Sexton MDH, Ringer MA, McAvaney BJ, Colman R, Soden BJ, Andronova NG, Emori S, Tsushima Y, Ogura T, Musat I, Bony S, Taylor K (2006) On uncertainty in feedback mechanisms controlling climate sensitivity in two GCM ensembles. Clim Dyn (in press). DOI 10.1007/s00382-006-0111-2

Wigley TML, Raper SCB (2001). Interpretation of high projections for global-mean warming. Science, 293:451–454

Wood RA, Keen AB, Mitchell JFB, Gregory JM (1999) Changing spatial structure of the thermohaline circulation in response to atmospheric CO_2 forcing in a climate model. Nature 399:572–575