

G. R. Harris · D. M. H. Sexton · B. B. B. Booth
M. Collins · J. M. Murphy · M. J. Webb

Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations

Received: 26 August 2005 / Accepted: 20 February 2006 / Published online: 25 April 2006
© British Crown Copyright 2006

Abstract Large ensembles of coupled atmosphere–ocean general circulation model (AOGCM) simulations are required to explore modelling uncertainty and make probabilistic predictions of future transient climate change at regional scales. These are not yet computationally feasible so we have developed a technique to emulate the response of such an ensemble by scaling equilibrium patterns of climate change derived from much cheaper “slab” model ensembles in which the atmospheric component of an AOGCM is coupled to a mixed-layer ocean. Climate feedback parameters are diagnosed for each member of a slab model ensemble and used to drive an energy balance model (EBM) to predict the time-dependent response of global surface temperature expected for different combinations of uncertain AOGCM parameters affecting atmospheric, land and sea-ice processes. The EBM projections are then used to scale normalised patterns of change derived for each slab member, and hence emulate the response of the relevant atmospheric model version when coupled to a dynamic ocean, in response to a 1% per annum increase in CO₂. The emulated responses are validated by comparison with predictions from a 17 member ensemble of AOGCM simulations, constructed from variants of HadCM3 using the same parameter combinations as 17 members of the slab model ensemble. Cross-validation permits estimation of the spatial and temporal dependence of emulation error, and also allows estimation of a correction field to correct discrepancies between the scaled equilibrium patterns and the transient response, reducing the emulation error. Emulated transient responses and their associated errors are obtained from the slab ensemble for 129 pseudo-HadCM3

versions containing multiple atmospheric parameter perturbations. These are combined to produce regional frequency distributions for the transient response of annual surface temperature change and boreal winter precipitation change. The technique can be extended to any surface climate variable demonstrating a scaleable, approximately linear response to forcing.

1 Introduction

In order to provide an objective basis for estimates of uncertainty in projections of future climate change, it is necessary to use results from ensembles of simulations of comprehensive global climate models. Historically a small ensemble of models developed at different centres has been available, with uncertainties in the response quantified either using simple diagnostics such as the standard deviation (Cubasch et al. 2001) or, more recently, from techniques for the estimation of probabilities (Furrer et al. 2006; Tebaldi et al. 2005). Another recent development has been the advent of larger ensembles designed to sample modelling uncertainties in a systematic manner. Ideally, such ensembles would sample both alternative options for structural elements of the model (for example grid resolution and the basic physical assumptions used in its parameterisations of sub grid-scale processes), and also alternative values for poorly-constrained parameters contained within the parameterisation schemes. Early examples have consisted of perturbed physics ensembles with a more limited aim of varying uncertain parameters within a single model framework, given a fixed set of structural options (Murphy et al. 2004; Stainforth et al. 2005; Annan et al. 2005).

In Murphy et al. (2004), 29 key parameters of the atmospheric model HadAM3 (Pope et al. 2000) were varied, one at a time, to produce an ensemble of 53 members. The atmospheric model was coupled to a

G. R. Harris (✉) · D. M. H. Sexton · B. B. B. Booth
M. Collins · J. M. Murphy · M. J. Webb
Hadley Centre for Climate Prediction and Research,
Met Office, FitzRoy Road, Exeter EX1 3PB, UK
E-mail: glen.harris@metoffice.gov.uk

mixed-layer “slab” ocean (Williams et al. 2001) (hereafter referred to as a “slab model”), and used to study the equilibrium response to a doubling of CO₂. Murphy et al. (2004) diagnosed climate sensitivity s (the equilibrium change in global mean surface temperature) for their ensemble members, and then used linear statistical methods to infer s for four million combinations of multiple parameter values sampled from uniform prior distributions to arrive at a prior predictive distribution (Collins et al. 2006) for s . Murphy et al. (2004) also developed a likelihood measure, based on the ability of each model version to simulate present-day climate observations, and this too was inferred for all parameter combinations using linear statistical methods. The prior predictive distribution was then weighted by likelihood to derive a posterior probability distribution function (pdf) for climate sensitivity, conditional upon the assumed distributions and ranges for uncertain parameters, assumptions in the selection of observational constraints and their conversion to estimated values of likelihood, linearity assumptions in the emulation of unsampled regions of parameter space, and the neglect of structural uncertainties and dynamical ocean feedbacks.

One of the limitations of the approach in Murphy et al. (2004) is that the physics perturbations were applied singly, neglecting non-linear interactions between different parameter combinations (although a simple estimate of the related uncertainty was included in the resulting pdfs). Using the same climate model, Stainforth et al. (2005) created an ensemble sampling multiple perturbations of a reduced parameter set, finding evidence of significant non-linear effects. Here we present results based on a new ensemble in which all 29 of the parameters of Murphy et al. (2004) are simultaneously perturbed. Parameter combinations were chosen by using the results of Murphy et al. (2004) to predict a set of parameter combinations which would span a wide range of climate sensitivities, possess reasonable skill in simulating present-day climate, and maximise the coverage of parameter space. Using this design strategy (described more fully in Webb et al. 2006), a 129 member ensemble was created, consisting of one version using the “standard” parameter settings of Pope et al. (2000) and 128 versions containing multiple parameter perturbations applied to the parameterisations of atmospheric, surface and sea-ice processes. Although members of this ensemble were selected by using a linear statistical method to predict values of sensitivity and model skill at a large sample of points in parameter space (Murphy et al. 2004), the range of climate changes actually simulated by the ensemble members will incorporate the effects of non-linear interactions between model parameters. Note also that this set of model versions does not represent an unbiased sample relative to some expert choice of the prior distribution of either model parameter values or climate sensitivity, so distributions of climate change derived from it should be regarded as sample-dependent frequency distributions,

rather than probability distributions consistent with a particular prior.

Use of a slab model permits this ensemble to examine only the equilibrium climate response to a given change in forcing. However, a key goal for climate research is to predict pdfs of transient climate change for plausible scenarios of future climate forcing. In particular, pdfs of transient changes at regional scales are required by the impacts community for risk assessment (Pittock et al. 2001). Only by coupling atmospheric general circulation models to full dynamical ocean models (hereafter AOGCMs) can this be achieved. We have therefore selected a subset of 16 parameter combinations, plus the unperturbed standard model version, from the 129 member slab ensemble to create a first perturbed physics ensemble of AOGCMs, consisting of variants of the HadCM3 model (Gordon et al. 2000; Johns et al. 2003).

The setup and production of this ensemble is described in detail by Collins et al. (2006). Description and choices for all parameter combinations are given in Table 1 in Collins et al. (2006). Flux adjustments (calibrated separately for each ensemble member) are applied to limit systematic biases in sea surface temperature (SST) and salinity, resulting in stable control simulations despite significant perturbation of the atmospheric model, and successfully preventing the development of regional SST biases in most parts of the world. However, all ensemble members suffer from a weakening of the thermohaline circulation, leading to cold SST biases in the North Atlantic in the control simulations. The applied forcing scenario in the climate change simulations is a 1% per annum increase in CO₂ concentration for 150 years (i.e. to four times pre-industrial concentrations).

Atmosphere–ocean general circulation model experiments are much more expensive than slab model experiments, thereby limiting us to 17 members. In order to provide robust regional pdfs for transient climate change, we need larger ensembles. This paper describes a technique to augment the 17 member AOGCM ensemble by scaling patterns of climate change obtained from the 129 member slab ensemble. In this way we can “emulate” the transient regional response for a larger fraction of parameter space in a cost-effective way. Emulation here refers to the use of a statistical model to predict the outputs of a climate model, which one could in principle run (Currin et al. 1991). Emulators are built from a sample of simulations and provide an efficient means of predicting the response of a complex model, and the error in emulated response, for any point in input parameter space.

The 17 sets of parameter combinations for our AOGCM ensemble have been chosen to be identical to 17 members of the slab ensemble. We can then validate predictions of the transient response obtained by scaling patterns of climate change from slab ensemble members with predictions simulated by corresponding AOGCM versions with identical physical parameterisations in the atmospheric component. Cross validation allows

quantification of uncertainty associated with the scaling method, and therefore forms an essential component of the emulation technique.

Below we describe in detail how scaling equilibrium patterns of climate change from an ensemble of slab model simulations can be used to emulate the transient response of an equivalent ensemble of AOGCM simulations. We emulate patterns of change for annual surface temperature for a 1% per annum increase in CO₂, validating the technique using the AOGCM ensemble described by Collins et al. (2006). The technique can be applied however to any forcing scenario, and to any surface climate variable that has a strong response to radiative forcing, and whose dependence on global surface temperature anomaly is approximately linear. A good test of the technique is to predict future transient regional precipitation changes, and some results of doing so are also presented.

2 Pattern scaling

Pattern scaling was proposed by Santer et al. (1990) as a way of inferring transient regional responses of surface temperature to increases in greenhouse gas forcing from equilibrium simulations with slab models. Such “classical” pattern scaling assumes the transient response for a climate variable of interest equals the normalized spatial pattern $s(x)$ obtained from a slab model, scaled linearly by the mean global surface temperature change $\Delta T(t)$:

$$\Delta F(x, t) = \Delta T(t)s(x) \quad (1)$$

$\Delta F(x, t)$ is the emulated temporal and spatial anomaly in the specified climate field compared to control climatology, and x refers to the vector of model surface grid-points. The spatial pattern of change is defined by

$$s^{\text{slab}}(x) = \frac{F_{2 \times \text{CO}_2}^{\text{slab}}(x) - F_{1 \times \text{CO}_2}^{\text{slab}}(x)}{\langle T_{2 \times \text{CO}_2}^{\text{slab}}(x) - T_{1 \times \text{CO}_2}^{\text{slab}}(x) \rangle}, \quad (2)$$

where angle brackets $\langle \rangle$ denote area-weighted global mean quantities, and all slab model quantities in Eq. 2 are temporally averaged (20 years in this paper).

This approach was necessary since AOGCM’s had yet to be developed. Following their advent, pattern scaling was then used as an efficient technique for inferring the transient response of an AOGCM to different forcing scenarios (Mitchell et al. 1999; Huntingford and Cox 2000; Mitchell 2003). This was achieved by identifying a single spatial pattern which, when scaled by the global temperature response, maximises the explained variance of temporally and spatially varying changes found during the full period of a transient simulation. The optimum spatial pattern, obtained by minimising the mean square error between the AOGCM and the scaled response, is given by

$$s^{\text{gcm}}(x) = \frac{\sum_{i=1}^N \Delta F^{\text{gcm}}(x, t_i) \Delta T^{\text{gcm}}(t_i)}{\sum_{i=1}^N [\Delta T^{\text{gcm}}(t_i)]^2} \quad (3)$$

where $\Delta T^{\text{gcm}}(t_i) = \langle \Delta T^{\text{gcm}}(x, t_i) \rangle$, and anomalies in the surface field F are defined with respect to some temporal mean over a control climate simulation, i.e.: $\Delta F^{\text{gcm}}(x, t_i) = F(x, t_i) - \bar{F}_{\text{ctl}}(x)$. The optimum spatial pattern can then be used to emulate the expected response to alternative forcing scenarios. Mitchell et al. (1999) found that decadal averaging reduced contamination of the response pattern by internal variability. In this paper we use 20 year averages centred on each decade ($t_i = t_0 + 10i$) to further increase the signal-to-noise ratio. These optimal patterns perform best when emulating the regional temperature response, especially if one uses patterns derived from a strong radiative forcing scenario (with high signal-to-noise ratio) to predict the response for a scenario with weaker forcing (Mitchell 2003). Pattern scaling for precipitation changes is less successful, but is useful for seasons and regions where the signal in the forced precipitation anomaly is larger than internal variability (Mitchell et al. 1999).

Our goal is to emulate the transient response of a large number of perturbed AOGCM versions from the equilibrium response of the corresponding slab model versions. In this case use of the slab pattern scaling of Eq. 1 would seem necessary, rather than the more accurate approach of Eq. 3. However, for our 17-member AOGCM ensemble we can obtain both optimal and slab patterns. The difference between these provides a correction field $c_j(x)$ that reduces discrepancies between the transient response, and emulated estimates of it obtained directly from the scaled slab pattern. These arise due to the effects of ocean dynamics not accounted for with the slab model. For the $j = 1, \dots, 17$ members of the AOGCM ensemble, Eq. 1 is rewritten

$$\Delta F_j(x, t_i) = \Delta T_j(t_i) [s_j^{\text{slab}}(x) + c_j(x)], \quad (4)$$

where $c_j(x) = s_j^{\text{gcm}}(x) - s_j^{\text{slab}}(x)$.

Figure 1 compares $s_j^{\text{slab}}(x)$, $c_j(x)$, and their absolute ratio for one member of the ensemble. The slab pattern demonstrates the well-known enhancement in warming over land relative to the oceans (Huntingford and Cox 2000), and the amplification of warming at high latitudes, especially in the Arctic. Over much of the globe $c_j(x)$ is considerably smaller than $s_j^{\text{slab}}(x)$. For example, the correction field is less than 0.2 times the slab pattern over 74% of the globe, and the mean value of the absolute ratio is 0.15. This confirms the assumption in Santer et al. (1990) that the slab pattern explains much of the variation in regional response. Locally, the largest values of $c_j(x)$ relative to $s_j^{\text{slab}}(x)$ are located over the oceans, in regions where slab model patterns fail to capture a dynamical oceanic response to greenhouse gas forcing. For example, in the North Atlantic, and in parts of the southern oceans, the slab model temperatures are too warm. These are regions where significant mixing of

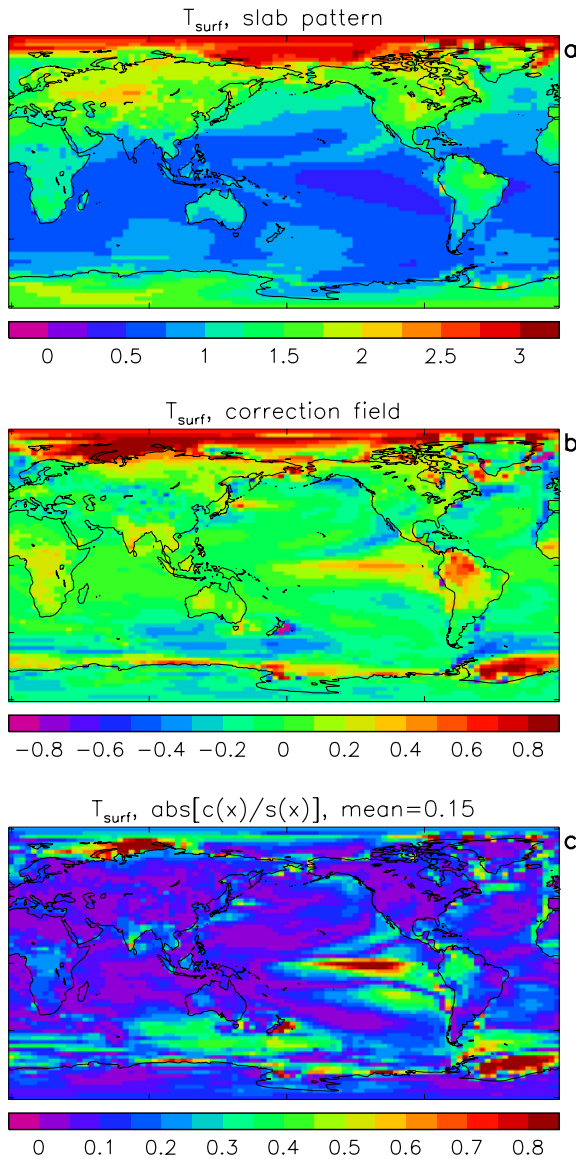


Fig. 1 **a** Pattern of the equilibrium annual surface temperature response to doubled CO_2 change for one member of the perturbed physics slab model ensemble. **b** Correction field that minimises the MSE between the scaled slab pattern in **a** and the same perturbed atmosphere coupled to a dynamic ocean, with a 1% increase in CO_2 for 150 years. **c** Absolute ratio of the correction field (**b**) to the slab pattern (**a**)

shallow and deep water occurs (e.g., compression of the Antarctic Circumpolar Current in the Drake Passage, and thermohaline mixing in the North Atlantic). The AOGCM possesses a more effective heat sink in these regions than the mixed-layer ocean, due to dynamical coupling between the surface layers and the deep ocean.

In the tropical East Pacific, the mixed-layer ocean predicts insufficient warming. This likely due to the slab model not representing sub-surface ocean thermocline dynamics and changes in cold water upwelling which lead to a local enhanced warming of this part of the Pacific in the AOGCM simulations. Associated with this, the correction field shows a positive signal,

implying insufficient warming over the Amazon in the slab model.

As well as assessing the validity of scaling slab patterns as a technique for emulating the transient response, information from the 17 correction fields can be used to improve the emulated transient response patterns inferred from the other 112 members of the slab model ensemble. Both the slab patterns and corresponding correction fields vary as physics parameters are perturbed, and we would ideally search for statistical relationships between the parameter perturbations and the correction fields to interpolate between the $c_j(x)$ for all members. This is not feasible with just 17 simulations over a 29-parameter phase space. We therefore average the correction fields obtained from each of the 17 members of the AOGCM ensemble and assume that the correction is constant across parameter space. The ensemble-mean correction field is given by

$$c_{\text{mean}}(x) = \frac{1}{M} \sum_{j=1}^M c_j(x). \quad (5)$$

This is plotted in Fig. 2 for annual surface temperature change, and is broadly similar in structure to the correction field in Fig. 1 derived for just one member.

For one AOGCM ensemble member, Fig. 3 plots the RMS error between scaled pattern predictions and simulated changes of annual global surface temperature. Three pattern scalings are tested here: (1) slab pattern, (2) optimal pattern, (3) slab plus ensemble mean correction field. Also plotted here is the RMS value of the simulated surface temperature anomaly, a measure of the expected error for an emulation technique with zero skill. The RMS error for the slab pattern scaling rises from an initial value of 0.23°C (an estimate of the model 20-year internal variability for global surface temperature) to 1.1°C by 140 years ($4 \times \text{CO}_2$), which can be compared with the RMS anomaly of 5.7°C for the same period. The variance explained by the slab pattern scaling is 96% by the end of the emulation period in Fig. 3, while for all members of the ensemble, at least

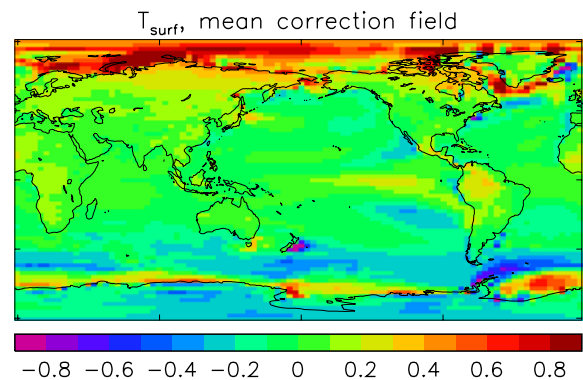


Fig. 2 Ensemble mean correction field for annual surface temperature change for the 17 members of the atmosphere–ocean general circulation model (AOGCM) ensemble

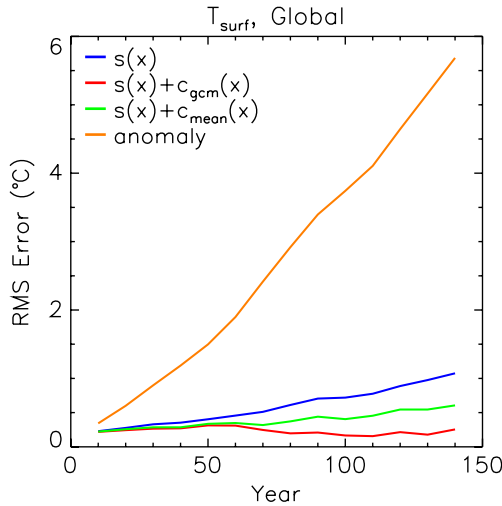


Fig. 3 Comparison of RMS errors between mean annual global surface temperature anomalies predicted by the AOGCM and three possible scaling patterns

92% of the variance is explained. Scaling the optimum pattern for surface temperature change gives RMS errors close to the level of model internal variability throughout the simulation period (Huntingford and Cox 2000). Inclusion of the ensemble mean correction leads to errors intermediate between the slab pattern and the optimum pattern. In this case the variance explained in Fig. 3 is 99% by the end of the emulation period, and for all members at least 95% of the variance is explained. These results confirm that for surface temperature, use of the mean correction field reduces the uncertainty of the scaling predictions.

3 Emulation of the global temperature response

In Fig. 3, spatial patterns were scaled by global temperature anomalies simulated by the AOGCM. For the full ensemble of slab models, however, corresponding AOGCM simulations are not available, so the transient global temperature response needs to be inferred from the equilibrium response. For this we shall use a two-box energy balance model (EBM), similar to Huntingford and Cox (2000), to predict globally averaged land and ocean surface temperature changes in response to imposed radiative forcing anomalies. Depth-dependent ocean temperatures are assumed to satisfy a heat conduction equation, though we modify Huntingford and Cox to include upwelling and downwelling processes, following Schlesinger et al. (1997). Thermal advection between land and ocean is assumed to depend linearly on the land–ocean temperature difference (although later we shall assume a constant ratio of land to ocean warming).

The most important EBM parameters for determining the temperature response are the land and ocean climate feedback parameters λ_l and λ_o (Hoffert et al.

1980). These are related to climate sensitivity (see below) and will vary as atmospheric parameters are perturbed. The EBM climate feedbacks are therefore parameterised separately for each member of the slab model ensemble. The ocean parameters in the EBM also need to be specified. These include effective thermal diffusivity κ , upwelling velocity w , thermal advection coefficient α and others (Raper et al. 2001). These parameters could in principle be calibrated separately for each of the 17 slab model ensemble members for which we possess corresponding AOGCM simulations, but not for the other 112 members. We therefore decided to determine a single set of EBM ocean parameters from the 17 member AOGCM ensemble and assume that they do not vary with perturbations to the atmospheric model physics when emulating the transient responses of the full 129 member ensemble. Any error resulting from this assumption is accounted for in the cross validation of the technique.

Huntingford and Cox (2000) diagnosed the ratio of land to ocean warming (defined here as ν) in transient simulations with HadCM3, and demonstrated it to be relatively constant in time, unlike the diagnosed thermal advection coefficient α . Adopting $\nu = \text{constant}$ as an alternative constraint defines $\Delta T_{\text{land}}(t) = \nu \Delta T_{\text{ocean}}(t)$, and effectively reduces the EBM to a one-box model. Both the one-box and two-box EBM formulations shall be tested here. Figure 4a verifies that ν is indeed relatively constant in time for all 17 members of the AOGCM ensemble, except for the first few decades. At early times the initial transient response over the oceans is delayed compared to the land, due to the smaller thermal inertia of the land surface. The values for the corresponding slab model versions (ν_{slab}) are also shown in Fig. 4a. The ν_{slab} values are always smaller than those for the corresponding AOGCM version (ν_{gcm}), due to enhanced thermal inertia arising from dynamical coupling between the surface and the deep ocean in the AOGCM. We find, therefore, that it is possible to emulate the transient response from the equilibrium response more accurately if we scale ν_{slab} by the mean ratio of $\nu_{\text{gcm}}/\nu_{\text{slab}}$ (equal to 1.09 for our ensemble). The EBM surface boundary condition for ocean temperature anomaly (Eq. 10 in Huntingford and Cox 2000) then becomes

$$-\kappa \frac{\partial \Delta T_o}{\partial z} = \frac{\Delta Q(t)}{f} - \left(1 + 1.09 \frac{1-f}{f}\right) \lambda_o^{\text{slab}} \Delta T_o, \quad (6)$$

where $\Delta Q(t)$ is the radiative forcing, λ_o^{slab} is the ocean climate feedback parameter and f the fraction of the Earth's surface covered by oceans.

We obtain λ_o^{slab} from $\lambda_o^{\text{slab}} = \Delta Q_{2 \times \text{CO}_2} / \Delta T_{o,2 \times \text{CO}_2}^{\text{slab}}$, where the forcing $\Delta Q_{2 \times \text{CO}_2}$ arising from doubling CO_2 in the standard model version is assumed independent of model parameter settings and equal to 3.74 W m^{-2} (Shine et al. 1990), and $\Delta T_{o,2 \times \text{CO}_2}^{\text{slab}}$ is the equilibrium response of ocean surface temperature in the relevant slab model ensemble member. Figure 4b compares λ_o^{slab} to

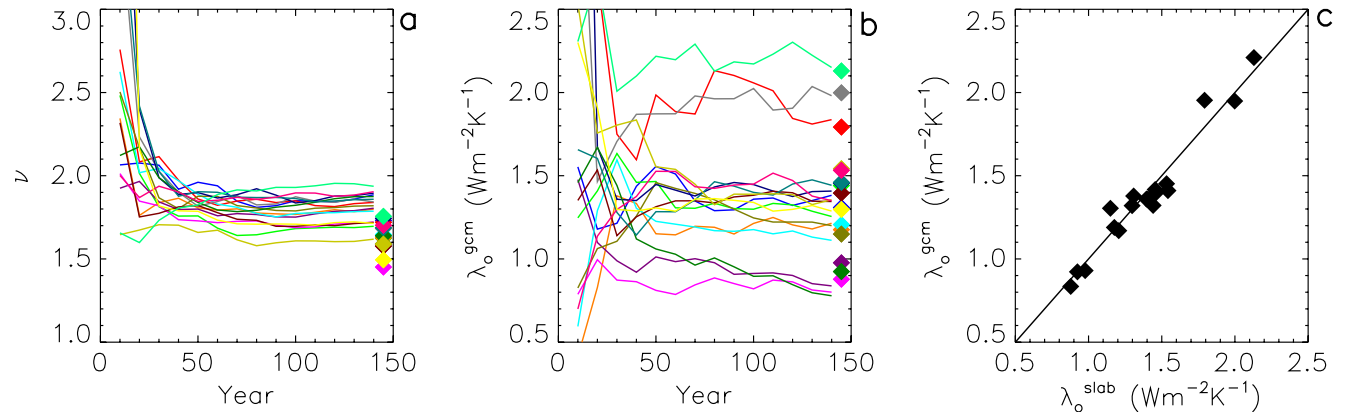


Fig. 4 **a** Evolution of the ratio of land to ocean surface warming $\nu_{\text{gcm}}(t)$, for the AOGCM ensemble. *Plotted symbols* show the equivalent equilibrium ratio for the slab model ensemble member with identical physics parameterisations. **b** Evolution of the ocean climate feedback parameter for the AOGCM ensemble driven by a

1% per annum increase in CO_2 . **c** Mean of the ocean climate feedback parameter over the last 100 years for each of the AOGCM ensemble members, plotted against the equivalent feedback parameter for corresponding members of the slab model ensemble

the effective ocean climate feedback parameter $\lambda_{\text{o}}^{\text{gcm}}(t)$ (Murphy 1995) diagnosed from the transient response, for the 17 model versions for which both slab and AOGCM experiments were available. After the first few decades, when simulated variability is large relative to the warming signal, $\lambda_{\text{o}}^{\text{gcm}}(t)$ settles down to values that evolve only slowly in time, and agree well with the corresponding values of $\lambda_{\text{o}}^{\text{slab}}$. The mean value of $\lambda_{\text{o}}^{\text{gcm}}$ (excluding the first 30 years) is plotted against $\lambda_{\text{o}}^{\text{slab}}$ in Fig. 4c, and compared to the line $y=x$. The correlation is strong ($R=0.97$), with the coefficient of regression equal to 1.03 and its standard error equal to 0.06, confirming the assumption inherent in our approach that feedbacks diagnosed from the slab ensemble accurately represent the effective climate feedbacks found in the corresponding AOGCM versions, at least for this ensemble with 1% per annum CO_2 forcing.

We specify the remaining EBM ocean parameters by using it to emulate the responses of our 17 member AOGCM ensemble. Downhill simplex techniques (Press et al. 1992) are used to select optimum values that minimise the mean square error \overline{E}^2 across the ensemble between the EBM and AOGCM projections for 20 year average global surface temperature changes:

$$\overline{E}^2 = \frac{1}{N} \sum_{i=1}^N E^2(t_i), \quad (7)$$

where

$$E^2(t_i) = \frac{1}{M} \sum_{j=1}^M [e_j(t_i)]^2,$$

$$e_j(t_i) = \Delta T_j^{\text{ebm}}(t_i) - \Delta T_j^{\text{gcm}}(t_i).$$

In Eq. 7 M ($=17$) is the number of ensemble members and N ($=14$) is the number of 20 year periods available, where these were constructed by averaging years 1–20, 11–30, ..., 131–150 of the transient simulations. Figure 5a

plots the error $e_j(t_i)$ for all 17 AOGCM members, following minimisation of \overline{E}^2 in Eq. 7 with respect to two parameters: ocean thermal diffusivity κ and upwelling velocity w . Values of $386 \text{ W m}^{-1} \text{ K}^{-1}$ and 3.1 m s^{-1} are selected respectively for these parameters by the optimisation. Error remains below 0.25°C for the majority of EBM projections until the last few decades when it rises to a maximum of 0.5°C , albeit for one member only. There is a small cold bias in the EBM projections for the first two decades, due to the high initial land-to-ocean warming ratio in the AOGCM that is not captured by the constant ν ratio assumption. The RMS error $E(t_i)$ for the one-box EBM with these values for κ and w is shown in Fig. 5b (blue curve). For comparison the RMS error obtained for the two-box version of the EBM, optimising with respect to κ , w and advection coefficient α , is also shown (red curve). This confirms that due to the relative constancy of ν in HadCM3, the simpler EBM is better at reproducing global surface temperature anomalies than the two-box version, and the $\nu = \text{constant}$ formulation is therefore used for the remainder of this work.

4 Emulation of the regional temperature response

Here we consider projections of regional surface temperature averaged over the well-known Giorgi and Francisco (GF) sub-continental-sized regions (Giorgi and Francisco 2000a, b). The names of these regions are listed in Fig. 13. Following Giorgi and Francisco (2000b), the Australian region is split into two at the 30°S parallel, while Antarctica (acronym ANT) is defined here as all land south of 60°S . Furthermore, all land points are lumped together into a region called ‘All Land’, with acronym ‘LND’, to give a total of 24 GF regions for this study. We consider 20 year mean changes in annual temperature.

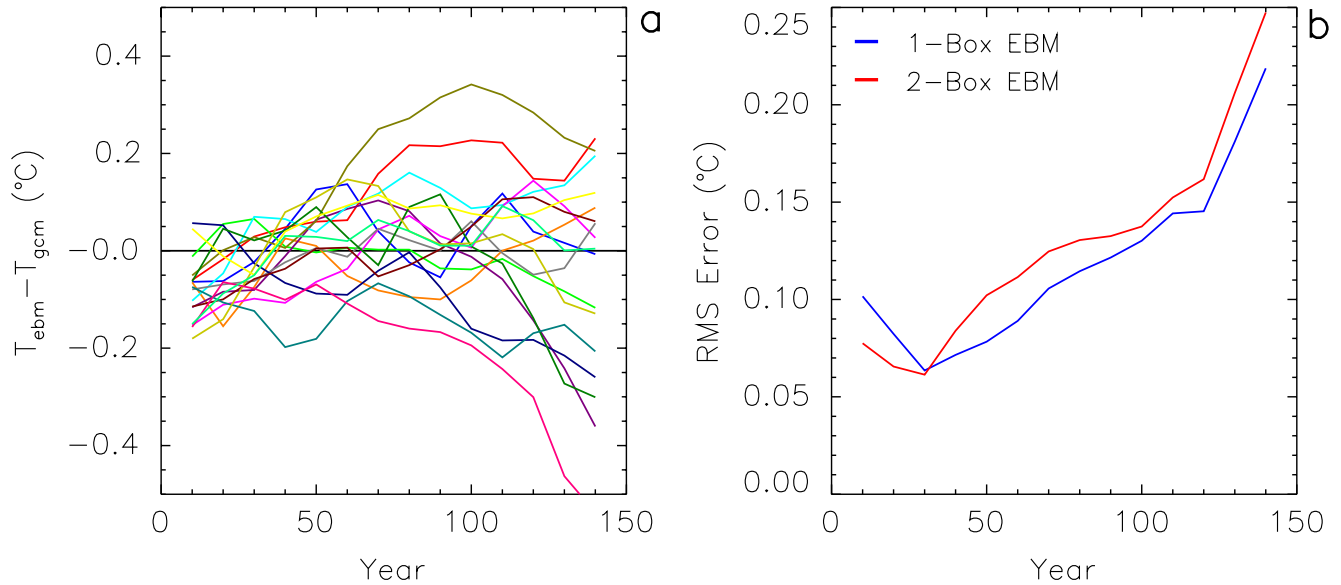


Fig. 5 **a** Differences between the one-box EBM and AOGCM predictions of 20-year mean global surface temperature anomalies under a 1% per annum CO_2 increase, for the 17 members of the cross-validation ensemble. **b** Ensemble average of the RMS error between the one-box (blue) and two-box (red) EBM versions and AOGCM projections of global surface temperature change

Errors in emulated values for a specified region k can be measured by three simple statistics: the ensemble mean bias, the ensemble mean RMS error, and the standard deviation of emulation error:

$$\begin{aligned} \bar{e}_k(t) &= \frac{1}{M} \sum_{j=1}^M e_{j,k}(t), \quad \text{rmse}_k(t) = \sqrt{\frac{1}{M} \sum_{j=1}^M e_{j,k}^2(t)}, \\ \sigma_k(t) &= \sqrt{\frac{1}{M-1} \sum_{j=1}^M (e_{j,k}(t) - \bar{e}_k(t))^2}, \end{aligned} \quad (8)$$

where

$$e_{j,k}(t) = \left\langle \Delta F_j^{\text{pred}}(x, t) - \Delta F_j^{\text{gcm}}(x, t) \right\rangle_k. \quad (9)$$

For example, the mean RMS error between scaled patterns and the 17 AOGCM simulations of annual surface temperature change over all land is shown in Fig. 6. The four curves here correspond to four different methods of obtaining emulated anomalies $\Delta F_j^{\text{pred}}(x, t)$: (1) slab pattern scaled by ΔT^{ebm} , (2) slab pattern scaled by the ‘true’ anomaly ΔT^{gcm} , (3) slab pattern plus ensemble mean correction field scaled by ΔT^{ebm} , (4) slab pattern plus ensemble mean correction field scaled by ΔT^{gcm} . This comparison allows errors due to the scaling assumptions to be separated from errors in the emulation of globally averaged temperature by the EBM, since scaling by the AOGCM anomaly ΔT^{gcm} is equivalent to using a perfect EBM. However, when we apply pattern scaling to the full slab ensemble in section 5, case (3) is used (green line). Several conclusions can be drawn from Fig. 6. Firstly, emulation errors increase in time, although since the warming also increases with time, the fractional error relative to the simulated warming re-

mains small. For example, following the first 30 years (when warming is small) the relative error for case (3) never exceeds 7%. Secondly, as expected, scaling by ΔT^{ebm} rather than ΔT^{gcm} leads to an increase in RMS error. The simulated land–sea contrast in warming exceeds the ratio specified in the EBM during the first two decades (see Fig. 4c), leading to a small increase in error for this period. However, the dominant contribution to

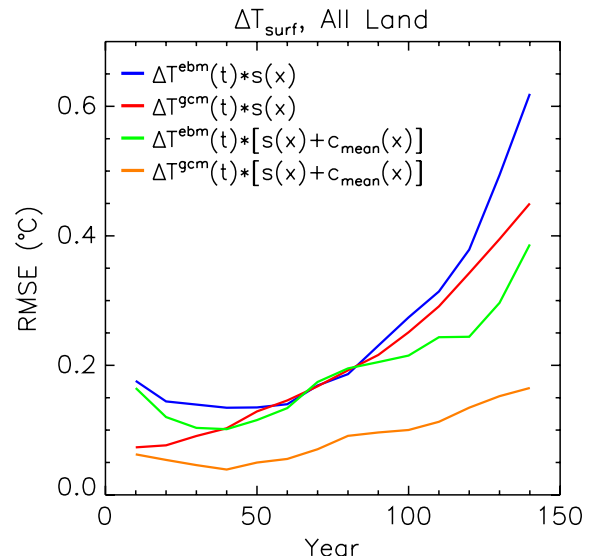
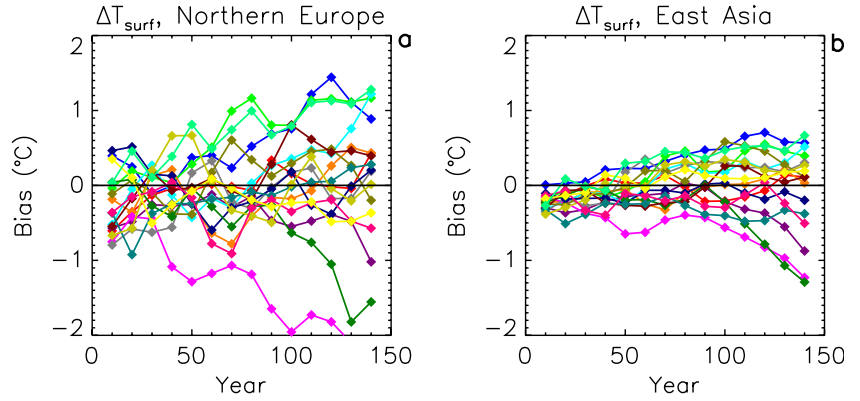


Fig. 6 Comparison of the mean RMS emulation error for annual land surface temperature for the 17 member AOGCM ensemble when scaling the slab pattern, or the slab pattern plus ensemble mean correction field, by either the AOGCM or EBM prediction of transient global surface temperature anomaly, for a 1% per annum increase in CO_2 concentration

Fig. 7 Errors between the 17 AOGCM simulations of annual surface temperature and the emulated response obtained by using the EBM to scale the slab plus ensemble mean correction field pattern, for **a** Northern Europe, **b** East Asia



emulation error during early decades (when the forcing is small) is provided by internal variability, which affects both the slab model predictor pattern, the verifying AOGCM response and (to a lesser degree) the ensemble mean correction field. Thirdly, inclusion of the ensemble mean correction field leads to a reduction in prediction error when compared to scaling the slab patterns.

Figure 7 plots regional emulation error defined by Eq. 9 for all 17 AOGCM ensemble members, for the Northern Europe and East Asia GF regions. The emulated anomaly here is given by

$$\Delta F_j^{\text{pred}}(x, t) = \Delta T_j^{\text{ebm}}(t) \left[s_j^{\text{slab}}(x) + c_{\text{mean}}(x) \right]. \quad (10)$$

The spread in uncertainty for East Asia is a lot smaller than for Northern Europe. This is partly due to the higher regional internal variability in Northern Europe (evidenced here by the higher spread at early times when the forcing is small). Pattern scaling error is also larger for Northern Europe, since there is more variability across the ensemble in the optimum correction fields obtained for each member to correct the slab pattern (recall from Eq. 5 that $c_{\text{mean}}(x)$ is the ensemble average of the optimum fields, used because we lack sufficient members to predict variations across the model parameter space). For example, those members in Fig. 7a with a large discrepancy between the AOGCM and the scaled

response (e.g. 2°C by 140 years) correspond to cases of large difference between the optimum correction field for that member and the mean correction field. One can conclude that for temperature the equilibrium response is less able to explain the pattern of the transient response in Northern European than in East Asia. This is not surprising, since weakening of the North Atlantic thermohaline circulation in the AOGCM simulations, and its influence on future climate in this region, will not be represented in the slab model patterns. Also, variations in the weakening across the AOGCM ensemble will not be picked up by use of an ensemble mean correction field. Nevertheless, errors of 2°C by 140 years are still only about 20% of the simulated warming, so fractional emulation errors remain reasonably small, even in Northern Europe.

The errors in the emulated responses shown in Fig. 7 were obtained by cross validation between the AOGCM and slab model ensembles, removing one member at a time from the ensemble and emulating its response. Cross validation is an essential part of the process of producing pseudo-AOGCM ensembles of the transient response, as it allows us to quantify emulation uncertainties and hence account for them in the widths of our frequency distributions of future climate change (see Sect. 5). As an illustration, Fig. 8a shows the ensemble mean transient responses simulated by the AOGCM for

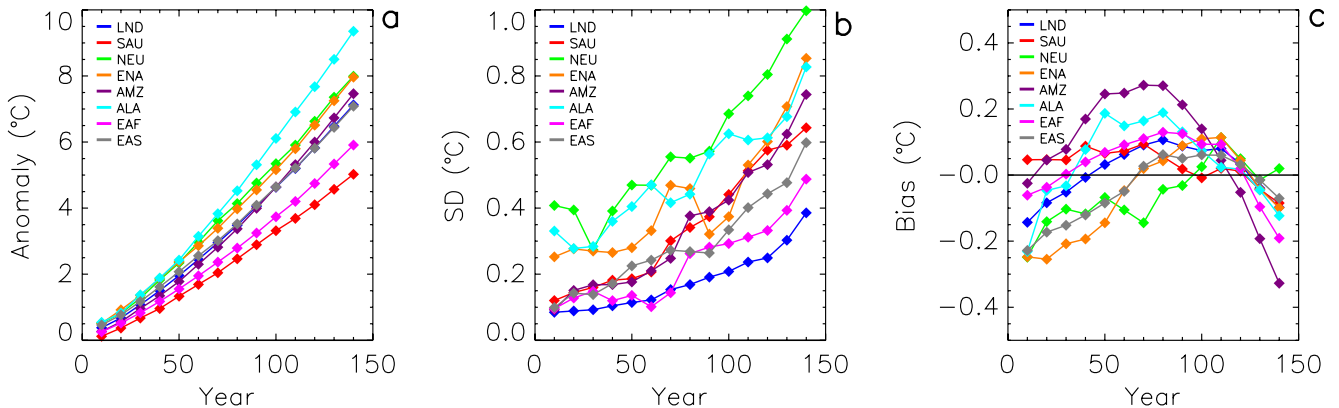


Fig. 8 Evolution of **a** the mean anomaly of the AOGCM ensemble of simulated changes in annual temperature, and the standard deviation **(b)**, and bias **(c)** of error in the corresponding emulated responses, in response to a 1% increase in CO₂ concentration over 150 years, for eight selected Giorgi and Francisco regions

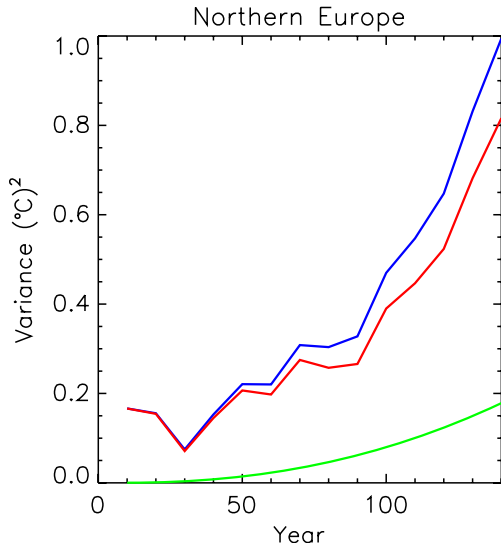


Fig. 9 Evolution of the variance σ^2 of error (blue curve) in the emulated annual surface temperature for Northern Europe, in response to a 1% increase in CO_2 concentration over 150 years. The green curve shows an independent estimate of the extra variance $2\overline{\varepsilon^2_{\text{int}}} \overline{\Delta T^2}$ for this region due to scaling of random internal variability in the slab patterns by the global surface temperature response. The difference between these two quantities (red curve) corrects for this unwanted noise, and represents the variance δ^2 of the distribution of random uncertainty assumed to be associated with each emulation

eight representative GF regions, with quantitative estimates of the standard deviation and bias of the emulation error (Eq. 8) shown in Fig. 8b, c, respectively. Emulation error is largest for GF regions in the high northern latitudes.

In the next section we emulate the transient response for each of the 129 slab ensemble members, to create frequency distributions of transient regional climate change. Each emulation is assumed to be sampled from

a normal distribution with mean value predicted by the emulation, and variance δ^2 in prediction error given by Eq. 25 (see Sect. 8):

$$\overline{\delta^2} = \sigma^2 - \overline{\varepsilon^2 \Delta T^2}. \quad (11)$$

The cross validation variance σ^2 contains a contribution from the variance ε^2 of random internal variability in the slab patterns, scaled by the global temperature response. To correct for this extra variance, we independently estimate $\overline{\varepsilon^2}$ from thirty 20-year means obtained from two 600-year simulations with the HadSM3 standard slab model configuration with pre-industrial and doubled CO_2 concentrations:

$$\overline{\varepsilon^2}(x) = \frac{1}{29} \sum_{l=1}^{30} [s_l(x) - \bar{s}(x)]^2. \quad (12)$$

Equation 2 is used to calculate slab patterns $s_l(x)$ from 20-year samples in these experiments. The estimate for $\overline{\varepsilon^2}$ in Eq. 12 assumes the same internal variability for all physics parameter perturbations. The blue curve in Fig. 9 shows the variance σ^2 of emulation error for Northern Europe obtained from the cross validation. The green curve shows $\overline{\varepsilon^2}$ for this region scaled by $\overline{\Delta T^2}$, while the red curve corresponds to δ^2 (the difference between the blue and green curves).

5 Frequency distributions of transient regional climate change

We are now in a position to emulate the transient regional responses which would be obtained by taking each atmospheric model version from the full 129 member slab ensemble, and coupling it to the dynamic ocean of HadCM3. The frequency distribution of equilibrium climate sensitivity for the full slab ensemble is

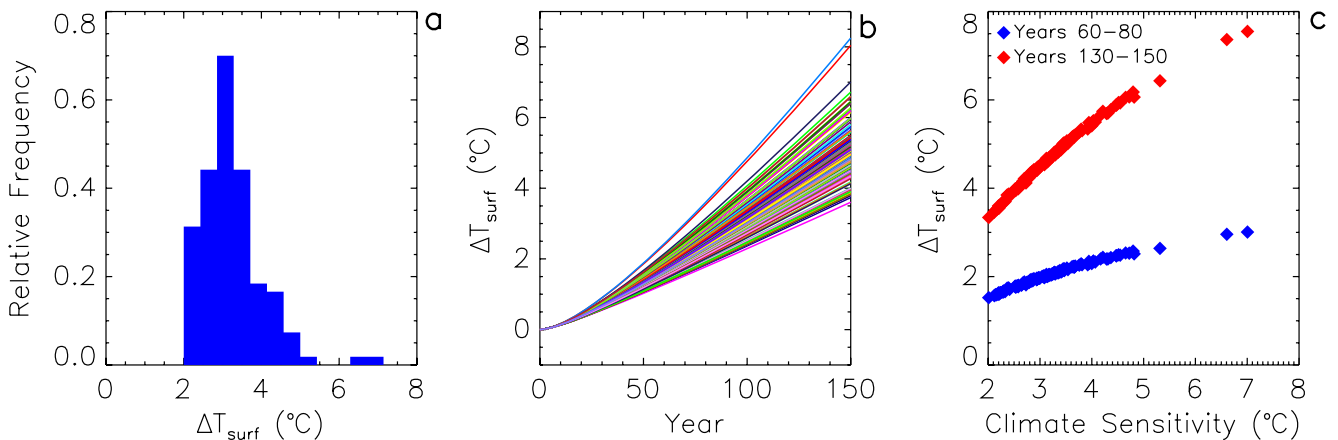


Fig. 10 a The frequency distribution of $2 \times \text{CO}_2$ equilibrium climate sensitivity, diagnosed from the 129 member perturbed physics slab model ensemble. **b** Plume of global surface temperature responses to a 1% increase in CO_2 concentration predicted by the EBM, using feedbacks diagnosed from **a**. **c** EBM transfer

function between climate sensitivity and global surface temperature warming, for decades 60–80 (blue curve) and 130–150 (red curve), for a 1% increase in CO_2 concentration (i.e., two times and four times pre-industrial CO_2 concentrations, respectively)

shown in Fig. 10a. Values range from 2.0 to 7.0°C across the ensemble (equivalent to a range of 0.53–1.86 W m⁻² K⁻¹ for global climate feedback parameter). Ocean climate feedbacks λ_o were likewise diagnosed for each slab member, and used to drive the EBM. Fixing other ocean parameters of the EBM to those obtained in Sect. 3, the global surface temperature responses of 129 variants of HadCM3 to a 1% per annum increase in CO₂ for 150 years are shown in Fig. 10b. Figure 10c shows the resulting transfer function between equilibrium climate sensitivity and mean global surface temperature response for years 60–80 and 130–150. This warming for years 60–80 (the time of CO₂ doubling) is commonly referred to as the transient climate response (TCR), and the EBM estimates a range of between 1.5 and 3.0°C for the pseudo-ensemble of HadCM3 versions.

Using Eq. 10, each EBM projection in Fig. 10b is used to scale the sum of the corresponding slab model response pattern plus the ensemble mean correction field, and is then meaned over 20 year periods and averaged over the GF regions. For each region k , member j and time t , we assume the resulting emulation of the transient response to be normally distributed with unknown mean and variance. The mean can be estimated from the emulated response $\langle \Delta F_j^{\text{pred}}(x,t) \rangle_k$, minus the mean bias $\bar{e}_k(t)$ obtained from cross validation using the AOGCM ensemble. The variance $\delta_k(t)$ can be estimated from cross validation using the 17 member AOGCM ensemble, with the variance in emulation error

reduced to remove variance due to scaled noise in the slab patterns (Eq. 11). As the variance has been estimated, we sum cumulative t -distribution functions $t(z; M-1)$ with $M-1 = 16$ degrees of freedom:

$$D_k(\Delta F, t) = \sum_{j=1}^{129} t \left(\frac{\Delta F - \langle \Delta F_j^{\text{pred}}(x,t) \rangle_k - \bar{e}_k(t)}{\delta_k(t)}; M-1 \right). \quad (13)$$

to obtain cumulative frequency distributions $D_k(\Delta F, t)$ for surface climate anomaly ΔF as a function of region and time. The derivative with respect to climate anomaly gives the frequency distribution. The blue curve in Fig. 11 shows the frequency distribution obtained using Eq. 14 for the TCR for the All Land region. Also shown here for comparison is the frequency histogram (grey shading) obtained from the EBM projections alone. Contributions to the uncertainty $\delta_k(t)$ in Eq. 13 include (see Sect. 8):

- Discrepancy between the emulated and simulated patterns of transient climate change
- Error due to the assumption that emulated patterns scale linearly with global temperature
- Error in the global surface temperature response obtained from the EBM
- Internal regional variability in the transient simulation.

For the All Land region, $\delta_{\text{LND}}(70 \text{ years}) = 0.14^\circ\text{C}$, which is small relative to the spread observed in Fig. 11. The main source of uncertainty in this frequency distribution is therefore due to the uncertainty in the physical parameters in the climate model rather than emulation uncertainty. The shape of the distribution is also affected by incomplete sampling of the model parameter space; e.g., the second minor peak at 4.2°C corresponds to the two highest sensitivity members of the ensemble.

Frequency distributions for regional transient climate change can be derived in an identical manner for smaller sub-regions. For example, Fig. 12a plots frequency distributions for the regional TCR for eight selected GF regions, while Fig. 12b shows the evolution of the frequency distribution for the Northern Europe surface temperature response through five different 20 year periods. The results illustrate wide spatial and temporal variations in the uncertainty of the response. South Australia and Alaska are the two GF regions which give the smallest and largest median response respectively, in our ensemble. At early times the widths of the distributions are dominated by internal variability, although there is some contribution from pattern scaling error (e.g., failure to capture the early transient response in land–sea temperature contrast). At later times the widths of the distributions for surface temperature response are dominated by uncertainty in the physical parameters in the climate model. It is worthwhile noting that the

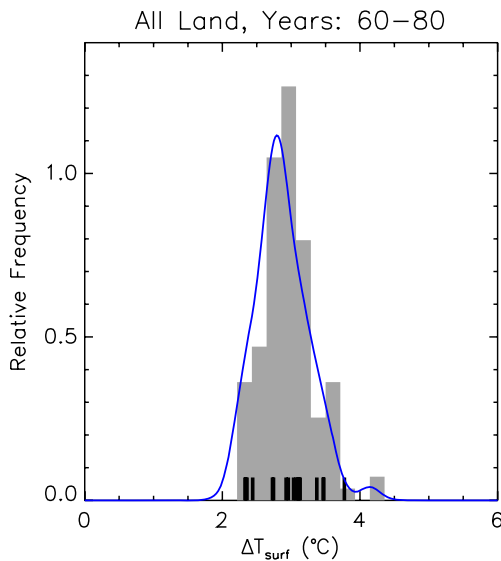


Fig. 11 The frequency distribution (grey shading) for the transient climate response (TCR) in surface temperature, for all land surface points for the 60–80-year interval during a 1% per annum CO₂ increase. The grey distribution includes uncertainty due to the perturbations of uncertain climate model parameters and error in emulated values of global surface temperature from the EBM. Including (in addition) pattern scaling error and pattern bias correction gives the frequency distribution in blue. Black vertical bars on the horizontal axis show the TCR for this decade for the 17 member AOGCM ensemble

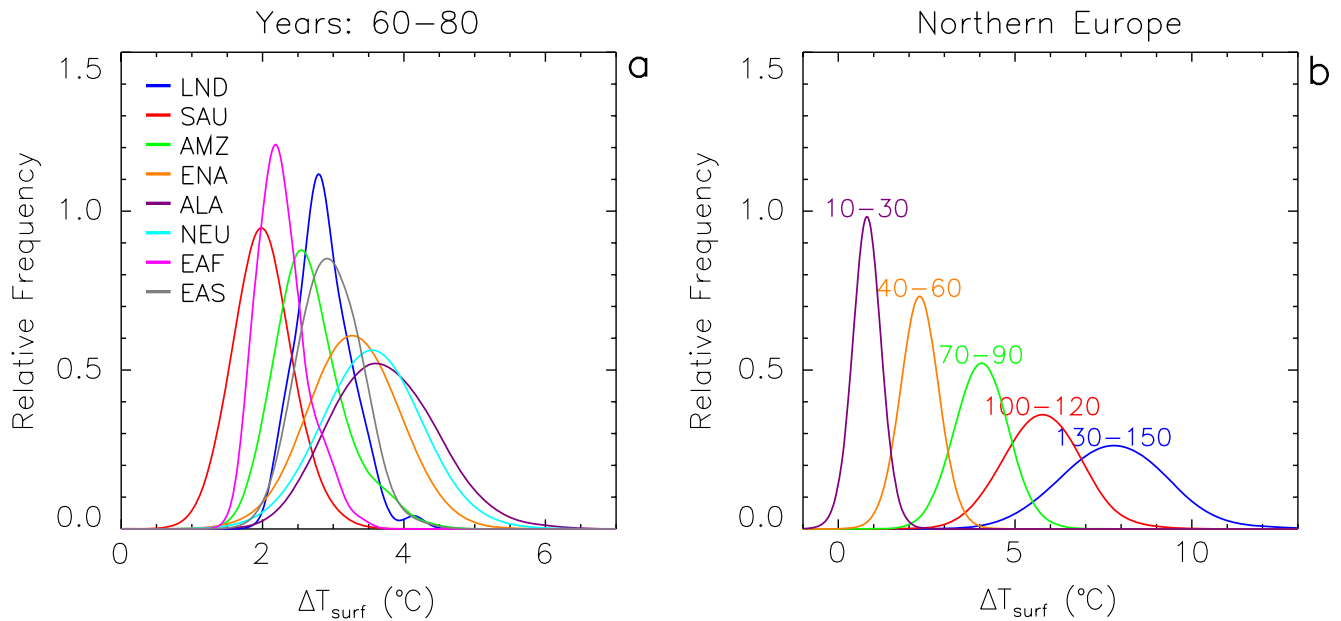


Fig. 12 a Frequency distributions of surface temperature response for the 60–80-year interval during a 1% increase in CO_2 concentration, for eight selected Giorgi and Francisco regions. Results were obtained by scaling the equilibrium responses of the

perturbed physics slab model ensemble. **b** Evolution of the frequency distribution of annual surface temperature change in the Northern Europe GF region, for five selected 20-year periods during a 1% per annum increase in CO_2 concentration

existence of pattern scaling error does not invalidate the emulation technique, since any uncertainty introduced is quantified by the cross validation ensemble, and included as an additional broadening of the final emulated frequency distribution.

A full description of the evolution in the predicted response and its uncertainty for a particular region can be given by looking at the evolution of specified confidence ranges of the frequency distributions. In Fig. 13 for each of the 24 GF regions, the median (black line) and 2.5, 5, 10, 90, 95 and 97.5% percentiles are chosen, with shading corresponding to the 80, 90 and 95% confidence ranges, respectively. Both the median response and the width of the distributions are a strong function of region, with higher response and uncertainty in the high northern latitudes than in the tropics and southern hemisphere. The confidence ranges increase in time for all regions, in a similar way to Fig. 12b. This is due both to the increasing spread of the response due to variations in climate model parameters, and also to trends in emulation error.

6 Emulation of the precipitation response

Scaling patterns of precipitation change (measured in $\text{mm day}^{-1} \text{K}^{-1}$) is a more difficult proposition than temperature. Mitchell et al. (1999) using the earlier HadCM2 version of the Hadley Centre GCM, concluded it was hard to distinguish between errors in the emulation and the effect of natural variability, although some regions did show a genuine scaleable climate change response. Mitchell (2003) concluded that by

redefining anomalies with respect to a control simulation of pre-industrial climate rather than the recent past, and increasing the averaging period from 10 to 30 years, then the simulated signal-to-noise ratio could be raised sufficiently to obtain robust response patterns for precipitation change. For HadCM3, Huntingford and Cox (2000) explained 68% of the global precipitation change in response to increasing CO_2 by scaling the optimum pattern (see Eq. 3). Another potential issue for the emulation is whether the distributions for the errors in predicted anomalies are normally distributed. Kolmogorov-Smirnov tests for the AOGCM cross validation ensemble produce no evidence to suggest these distributions for emulated precipitation are anything other than Gaussian.

The technique described here uses a 20-year averaging period, and anomalies are defined with respect to a control simulation, so we should expect scaling of precipitation patterns to be skilful wherever the signal is significant. Unlike Mitchell (2003), we scale slab patterns augmented by the ensemble mean correction field (see Eq. 10) rather than the optimum response pattern. Different atmospheric parameterisations with different response patterns are also scaled here. It is therefore important to demonstrate the validity of scaling for emulation of a pseudo-ensemble of AOGCM precipitation changes.

As an example of the precipitation anomalies we expect the scaling technique to emulate, Fig. 14 plots the DJF (boreal winter) precipitation responses of our 17 member AOGCM ensemble for two GF regions: Northern Europe and North Australia. In Northern Europe there is a clear common response across the

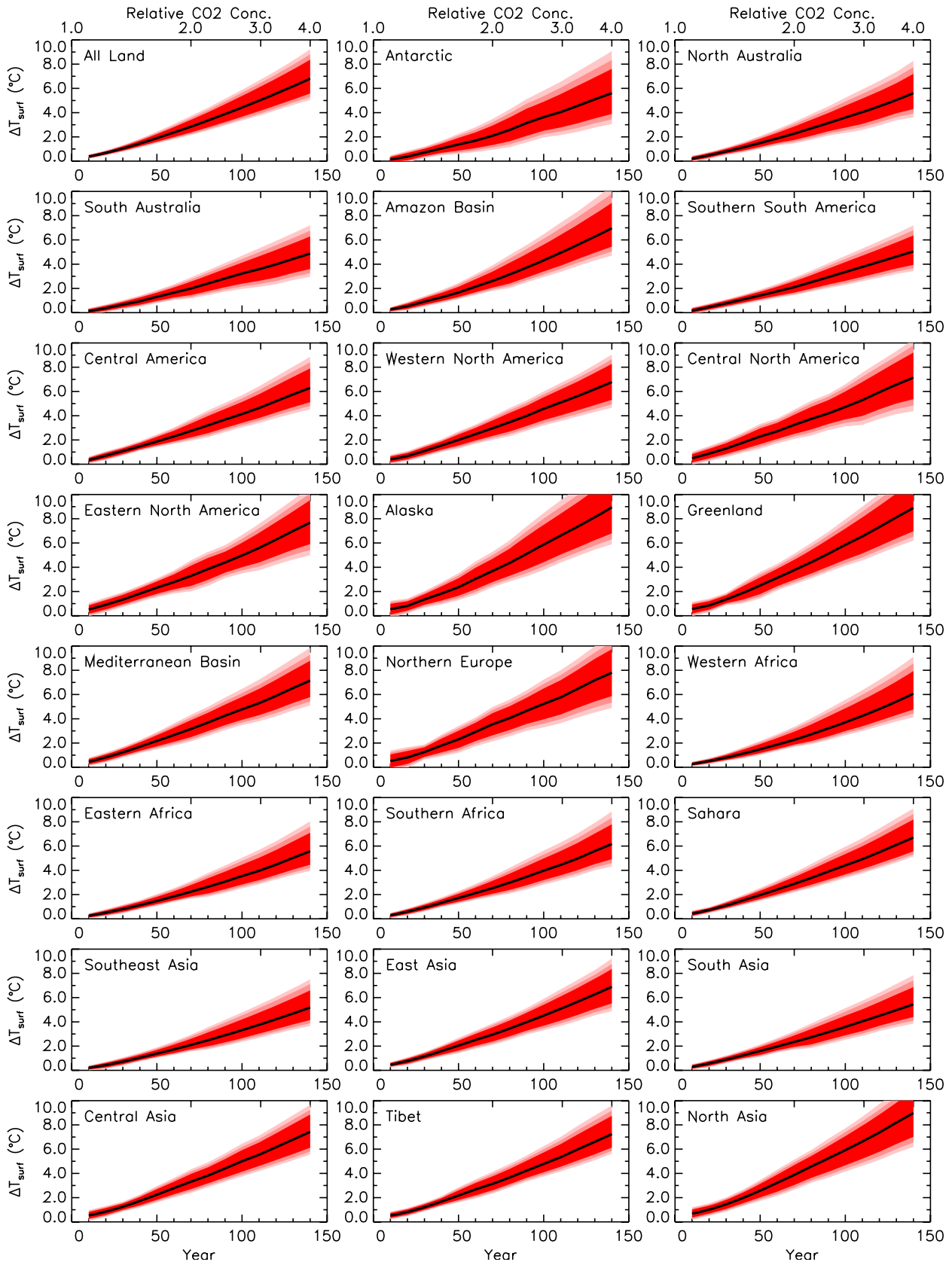
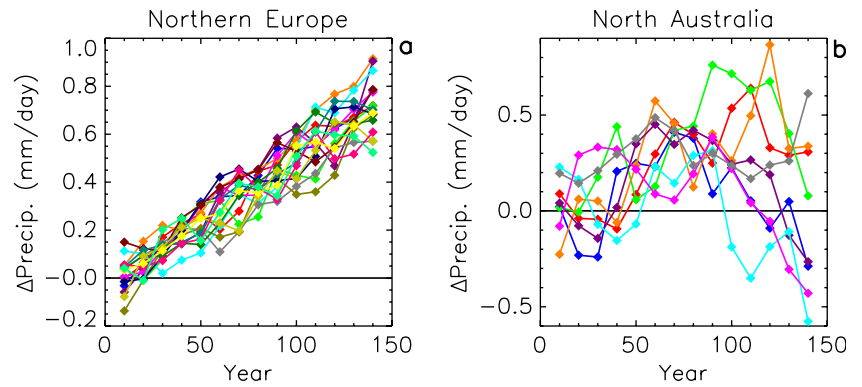


Fig. 13 Evolution in the median, and 80, 90, and 95% confidence ranges for annual surface temperature change, for a 1% per annum increase in CO₂ concentration for 150 years, for all 24 Giorgi and Francisco regions

Fig. 14 Regional response of **a** Northern Europe and **b** North Australia (eight members shown for clarity), for precipitation changes for the AOGCM ensemble, under a 1% increase in CO₂ for 150 years



ensemble for an increase in precipitation rate with increasing radiative forcing, although compared to annual surface temperature we observe more variability in the climate change signal.

North Australia is the region whose precipitation response is the most difficult to emulate. Internal variability for any given ensemble member is large in this region. The response for a majority of members is for an initial increase in precipitation rate, although some members show no clear signal or predict a decrease in precipitation. Furthermore, in some cases the precipitation anomaly depends non-linearly on global surface temperature anomaly, with an initial increase in precipitation followed by a reduction. The median North Australian response of the ensemble is similarly non-linear (red curve in Fig. 16a), with an initial increase in DJF precipitation followed by a decrease. Sub-regional trends (Good and Lowe 2006) are implicated in this response, although further analysis is required.

We now look at the ability of pattern scaling to emulate regional responses typified in Fig. 14. For precipitation changes over land, Fig. 15a, b show the slab and optimum correction field patterns respectively for one member of the ensemble, while Fig. 15c shows the ensemble mean correction field. In some areas the correction field is no longer small compared to the slab pattern of precipitation change (the average value of the absolute ratio of the two fields in Fig. 15a, b is 0.62). This implies that for some regions, the equilibrium anomaly pattern does not represent the transient response well. In North-East Australia for example, the slab pattern implies a reduction in rainfall, while the correction field is similar in magnitude, but opposite in sign, suggesting the equilibrium response is drier than the transient response. The discrepancy in this region could be due to the effects of ocean dynamics, represented in the transient response but not the equilibrium response. However, the correction field is derived from a single optimum pattern calibrated from the whole period of the transient simulation, and is unable to represent non-linearities in the dependence of the response on global temperature. Such non-linearities are seen in Fig. 14b, and will also contribute to discrepancy between the emulated and AOGCM

responses. A further contribution arises from use of an ensemble mean correction field in the emulation technique, which can differ substantially from the (usually unknown) optimum correction field for a given location in the model parameter space (cf. Fig. 15b, c). Cross validation therefore gives large uncertainty for emulated DJF precipitation anomalies in areas where these factors make emulation difficult, such as North Australia.

In addition, for the ensemble member in Fig. 15 there is a signal for a reduction in precipitation in the western half, and an increase in precipitation in the eastern half of both the North Australia and Amazon GF regions. By averaging over the full GF region, we reduce the climate change signal and make pattern scaling in these regions less robust. There is a good case for a redefinition of some regions to permit more accurate scaling predictions, but as the GF regions have wide recognition, we continue at present to use them.

Similarly to Fig. 8 for surface temperature, Fig. 16 shows the results of cross validating emulated changes obtained from scaled precipitation patterns with simulated changes from the 17 member AOGCM ensemble. Simulated AOGCM anomalies, and the standard deviation and bias of errors in the emulated precipitation changes, are shown for eight selected GF regions. For the reasons just outlined, North Australia is difficult to emulate, and the standard deviation of error for this region is the largest of all the GF regions (for the DJF season). The non-linear time dependence of the median response for North Australia and the Amazon is reflected in changes in sign of the bias in Fig. 16c, the effects of which are included in the emulated responses via Eq. 10. This further underlines the importance of cross validation to our technique, both in optimising the skill of the emulated changes, and in quantifying those regions, seasons and climate variables for which emulation uncertainty is large.

Emulating the transient precipitation response for the full 129 member pseudo-ensemble of HadCM3 versions leads to the plumes of evolving uncertainty shown in Fig. 17 for the 24 GF regions. The median response for most regions is for an increase in DJF precipitation. In the high Arctic and many temperate

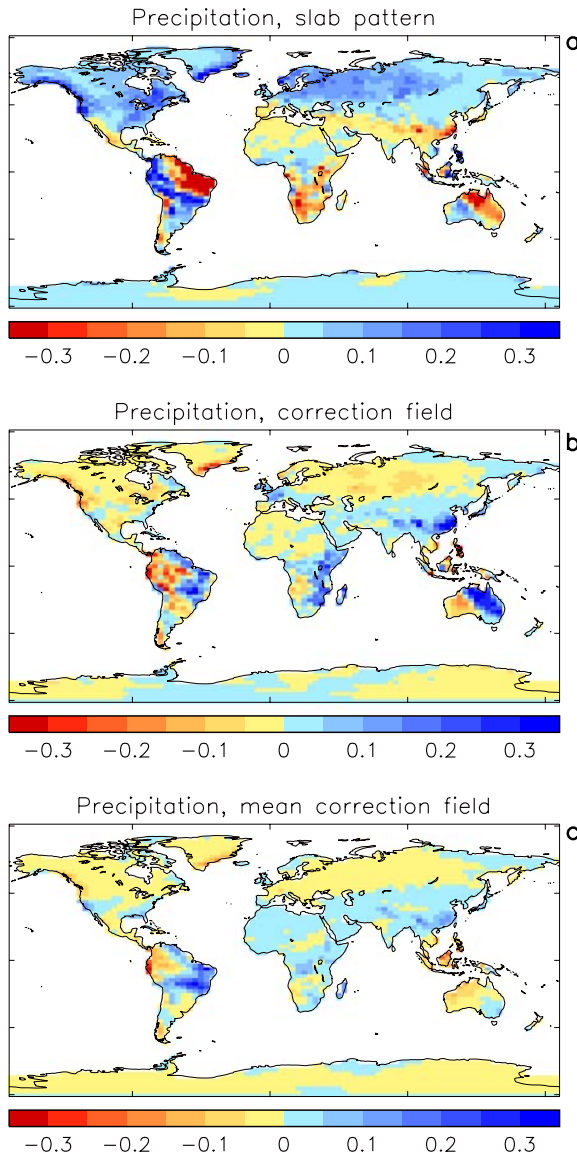


Fig. 15 **a** Pattern of the equilibrium response to doubled CO_2 of boreal winter (DJF) precipitation change ($\text{mm day}^{-1} \text{K}^{-1}$) simulated by one perturbed physics slab model ensemble member. **b** Correction field that minimises the MSE between the scaled slab pattern in **a** and the response when the same perturbed version of the atmosphere model is coupled to a dynamic ocean, for a 1% increase in CO_2 for 150 years. **c** Ensemble mean correction field for DJF precipitation change for the 17 members of the AOGCM ensemble

Northern Hemisphere regions (e.g. NEU, ENA, NAS, GRL) the sign of this precipitation change is robust to modelling uncertainty as explored by our ensemble. In tropical and Southern Hemisphere regions, uncertainty is sufficient to preclude the conclusion that the predicted precipitation responses are significantly different from zero during the DJF season. Sample frequency distributions for the 60–80 year period (the time of CO_2 doubling) for eight selected GF regions are plotted in Fig. 18.

7 Summary and discussion

Perturbed physics ensembles of coupled atmosphere mixed layer ocean (“slab” model) simulations have been used to examine the effect of modelling uncertainty on the equilibrium response of climate to a doubling of CO_2 . However, to predict uncertainty in the transient response of regional climate we need large ensembles in which atmospheric general circulation models are coupled to a dynamical ocean model (AOGCMs). Collins et al. (2006) report an initial step towards this goal, consisting of an ensemble of 17 versions of HadCM3 with multiple perturbation of 29 key atmospheric, surface and sea-ice parameters. However, it is not yet computationally feasible to produce the larger ensembles required to explore fully the model parameter space. We have therefore developed a technique to scale equilibrium patterns of climate change derived from much cheaper ensembles of slab model simulations, in order to emulate the transient response of an equivalent ensemble of AOGCM simulations.

With this technique climate sensitivities are diagnosed for each member of a slab model ensemble, and used to drive an EBM for a specified forcing scenario to predict the time dependent global surface temperature response expected for the equivalent AOGCM version. The EBM projections are then used to scale normalised patterns of climate change for each slab member to emulate the transient response of a 129 member AOGCM ensemble with multiple parameter perturbations. In this paper, we have emulated the response for annual surface temperature and DJF precipitation anomalies for a 1% per annum increase in CO_2 concentration. The technique can be used however for any plausible forcing scenario realisable with an EBM, and for any climate surface variable that has a strong response to forcing whose dependence on global surface temperature anomaly is approximately linear. In the future we shall extend the technique to other seasons, and attempt emulation for other important climate impact variables such as soil moisture, surface wind speed, surface relative humidity, etc.

The emulation technique is validated by comparing estimations of the transient response based on scaling of patterns with corresponding output from the 17 member AOGCM ensemble referred to above. For these cases we can also derive the optimum single pattern which, when scaled, minimises the mean square error between the AOGCM and the emulated response. The optimum pattern offers substantial reduction in emulation error when added to the slab pattern for the purpose of scaling. The difference between the slab and optimum patterns (the correction field) is a measure of how well the equilibrium response represents the pattern of transient response. For surface temperature the correction fields are considerably smaller in magnitude than the slab pattern over much of the globe, confirming the assumption that the equilibrium pattern explains most

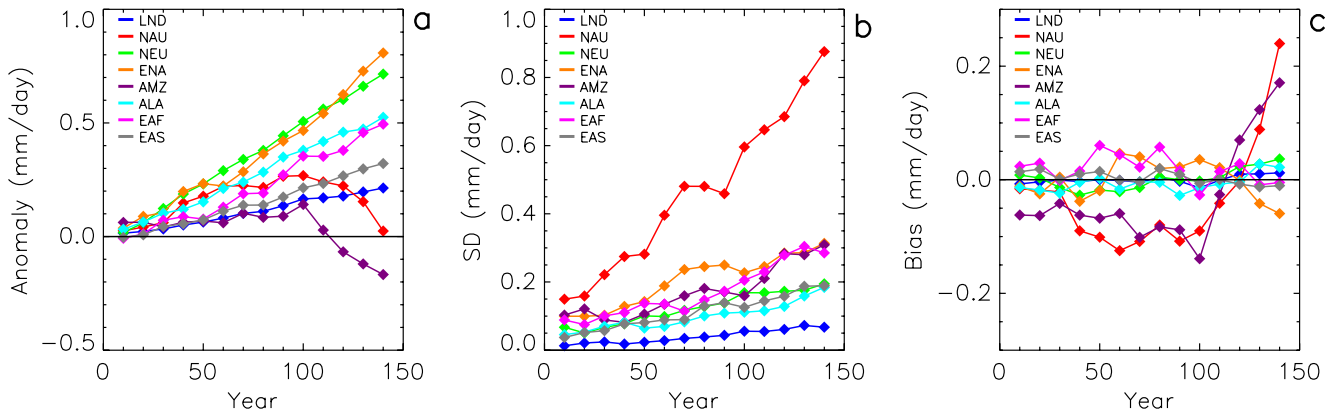


Fig. 16 Evolution of **a** the mean anomaly of the AOGCM ensemble of simulated changes in DJF precipitation, and the standard deviation (**b**), and bias (**c**) of error in the corresponding emulated responses, in response to a 1% increase in CO₂ concentration over 150 years, for eight selected Giorgi and Francisco regions

of the variation in regional response. Locally, the largest values of correction field relative to the slab pattern are located over the oceans in regions where the slab model fails to capture a dynamical oceanic response to greenhouse gas forcing.

Precipitation is more difficult to emulate. The correction field can be of comparable magnitude to the slab pattern in some regions, implying the equilibrium response does not correspond well to the transient response. Also, in some regions, such as North Australia and the Amazon, we find sub-regions of increasing and decreasing precipitation (Good and Lowe 2006). Averaging over these regions can reduce the climate change signal, leading to less robust emulation. Thirdly, the precipitation response for some AOGCM simulations was found in some regions to possess a non-linear relationship to global temperature, with initial precipitation increases followed by a decrease (perhaps a reflection of sub-regional trends). Development of non-linear pattern scaling techniques could therefore potentially reduce emulation error in such regions.

Information from the correction fields can be used to reduce the errors associated with scaling slab ensemble members for which no corresponding AOGCM simulation is available. The correction fields vary as uncertain model parameters are perturbed, and with just 17 verifying AOGCM simulations spanning a 29-parameter phase space, we cannot hope to derive statistical relationships between the parameter perturbations and the correction fields. However, it is possible to calculate a mean correction field (assumed invariant across parameter space) by averaging the fields obtained from the 17 members of the verifying AOGCM ensemble. Addition of this mean correction field to the slab pattern is shown to reduce emulation error.

Cross-validation also permits quantification of the mean bias and standard deviation of emulation error as a function of time and region, and is a key component of the emulation method. For example, correcting for a time-dependent emulation bias allows the technique to

capture non-linearities in the transient response (see above), at least to the extent that departures from linearity are independent of variations in AOGCM physics. The estimates of emulation error include error due to the scaling assumption that climate response patterns scale linearly with global temperature, error in the EBM predictions of global surface temperature change, and discrepancy between the emulated and simulated response patterns. The presence of pattern scaling error does not invalidate the emulation technique, since any uncertainty introduced is quantified by the cross validation ensemble, and is included as an additional broadening of the final emulated frequency distributions. Of course it is desirable to reduce emulation error wherever possible. Assuming the net emulation error to be normally distributed, we obtain regional frequency distributions for transient climate change whose widths are determined both by the divergence of regional response for the different AOGCM parameter combinations, and by regional emulation error. For example, after 70 years (at the time of CO₂ doubling) the median changes in annual surface temperature and precipitation rate for Northern Europe are 3.5°C and 0.34 mm/day, while the 90% confidence ranges for the frequency distributions are 2.4–4.7°C, and 0.12–0.57 mm/day respectively.

In this study we considered the spread of transient climate changes consistent with perturbations to uncertain parameters relating to atmospheric, surface and sea-ice processes in one climate model. We neglected structural uncertainty (e.g., variations in model resolution or in the basic physical assumptions used in the parameterisation of sub grid scale processes), and we also neglected modelling uncertainties in other key components of the Earth system, such as ocean physics and the carbon and sulphur cycles. Work is under way at the Hadley Centre to address some of these additional sources of model uncertainty with an enlarged AOGCM ensemble, the results of which can be expected to broaden the uncertainty ranges associated with predicted future changes. We also plan to estimate distributions of the

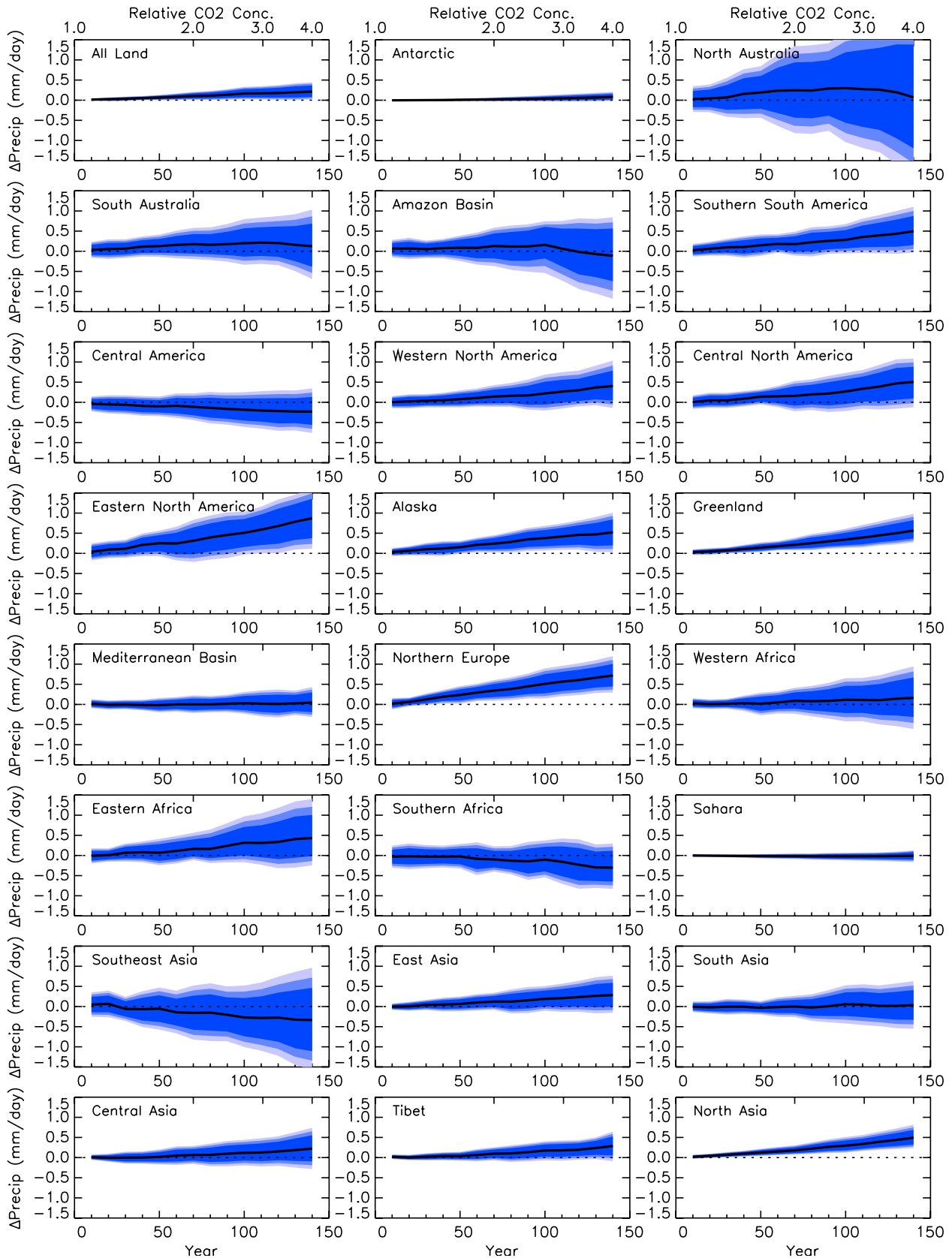


Fig. 17 Evolution in the median, and 80, 90, and 95% confidence ranges for DJF precipitation changes, for a 1% per annum increase in CO_2 concentration for 150 years, for all 24 Giorgi and Francisco regions

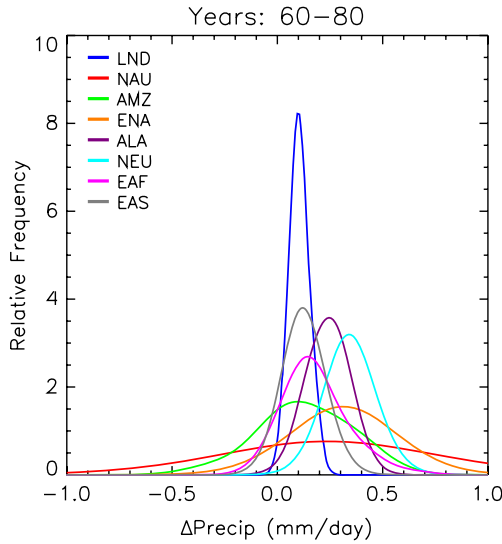


Fig. 18 Frequency distributions of DJF precipitation change for the 60–80-year interval during a 1% increase in CO₂ concentration, for eight selected Giorgi and Francisco regions. Results were obtained by scaling the equilibrium responses of the perturbed physics slab model ensemble

response to policy-relevant forcing scenarios using the techniques described in this paper. For this purpose we are creating a new 17 member AOGCM ensemble forced by historical forcings from 1860–2000, followed by the A1B SRES scenario to 2100 (Nakicenovic et al. 2000).

Our frequency distributions for future climate change depend on the 129 member slab model ensemble (Webb et al. 2006) from which the transient responses are emulated. Although this is a large ensemble by climate modelling standards, it was not designed, and indeed is not large enough, to provide a comprehensive, unbiased sample of the model parameter space according to some specified prior distribution of values. As it is not computationally feasible to run sufficient slab model versions to sample fully the 29 parameter space, we are constructing a new statistical emulator to estimate the non-linear response for parameter combinations not sampled by an actual slab model simulation. The new emulator (which predicts equilibrium responses), in conjunction with the emulator described here (which predicts transient changes from the equilibrium response), will enable us to utilise Bayesian methods to obtain distributions of predicted equilibrium and transient changes consistent with specified prior distributions for uncertain quantities. Ultimately, we aim to produce probability distributions of future transient regional changes accounting for the relative likelihood of different model versions based on comparison with observations.

Acknowledgments This work could not have been possible without input from many dedicated Hadley Centre staff. The work was supported by the UK Department for Environment, Food and Rural Affairs under Contract PECD/7/12/37 and by the European Community ENSEMBLES project (GOCE-CT-2003–505539) under the Sixth Framework Programme.

8 Appendix

This section provides an estimate of the variance in the frequency distribution of the transient regional response emulated from the slab ensemble. For a given region and time, the emulated response P_j in Eq. 10 is assumed to be equivalent to:

$$P_j = \Delta T_j (s_j + \varepsilon_j). \quad (14)$$

where s_j is the signal in the combined slab and correction field pattern for member j , and ε_j is random internal variability in the normalised slab pattern uncorrelated with s_j . Internal variability in the slab pattern arises since we have used 20 year means to estimate the response. The variance of the emulated frequency distribution is:

$$V = \frac{1}{N-1} \sum_{j=1}^N \left[(P_j + \delta_j) - \overline{(P_j + \delta_j)} \right]^2 \quad (15)$$

where to each emulated response P_j we add random noise δ_j , sampled from a normal distribution with some yet to be determined variance δ^2 . In the limit $(N-1)/N \rightarrow 1$, and with $\overline{\delta_j} = 0$ this becomes

$$V = \overline{P^2} - \overline{P}^2 + \overline{\delta^2}. \quad (16)$$

Equation 14 for P_j is used in Eq. 16 to give

$$V = \overline{\delta^2} + \overline{s^2 \Delta T^2} - \overline{s \Delta T}^2 + \overline{\varepsilon^2 \Delta T^2} + 2\overline{s \varepsilon \Delta T^2} - \overline{\varepsilon \Delta T}^2 - 2\overline{\varepsilon \Delta T} \overline{s \Delta T}. \quad (17)$$

Random internal variability in the slab pattern is assumed to be uncorrelated with EBM projections for ΔT , so the last three terms in Eq. 17 equal zero. Defining the variance V_s in scaled prediction of the true slab pattern signal:

$$V_s = \frac{1}{N} \sum_{j=1}^N [s_j \Delta T_j - \overline{s_j \Delta T_j}]^2 = \overline{s^2 \Delta T^2} - \overline{s \Delta T}^2 \quad (18)$$

the variance in the emulated frequency distribution becomes

$$V = V_s + \overline{\varepsilon^2 \Delta T^2} + \overline{\delta^2}. \quad (19)$$

The second term in Eq. 19, with internal variability scaled by the EBM global temperature response ΔT^2 , represents variance due to our uncertain knowledge of the true slab signal. The remaining variance $\overline{\delta^2}$ will include a contribution V_{em} arising from uncertainty in the emulation, and variance $\overline{\eta^2}$ due to random internal variability in the transient response being emulated:

$$\overline{\delta^2} = V_{em} + \overline{\eta^2}. \quad (20)$$

The variance $\overline{\delta^2}$ can be estimated from the cross validation ensemble. Similarly to Eq. 14, we assume the AOGCM simulated response M_j for a given region and time is equal to a model signal m_j , plus random internal

variability η_j uncorrelated with m_j :

$$M_j = m_j + \eta_j. \quad (21)$$

Substituting $e_j = P_j - M_j$ into the definition for σ^2 in Eq. 8, expanding, and setting the means of uncorrelated terms such as $\overline{\eta_j}$ to zero, the variance in cross validation error is given by

$$\begin{aligned} \sigma^2 = & \frac{1}{M-1} \sum_{j=1}^M \left[(\Delta T_j s_j - m_j) - \overline{(\Delta T_j s_j - m_j)} \right]^2 \\ & + \frac{1}{M-1} \sum_{j=1}^M \eta_j^2 + \frac{1}{M-1} \sum_{j=1}^M \varepsilon_j^2 \Delta T_j^2. \end{aligned} \quad (22)$$

The first term in Eq. 22 is the variance of the emulation error for the true signal, which we identify as V_{em} , i.e.:

$$V_{\text{em}} = \frac{1}{M-1} \sum_{j=1}^M \left[(\Delta T_j s_j - m_j) - \overline{(\Delta T_j s_j - m_j)} \right]^2. \quad (23)$$

V_{em} includes error due to non-linearities in the simulated AOGCM signal m_j , and error due to discrepancy between m_j and the slab signal scaled by global temperature from the EBM. The second term in Eq. 22 represents uncertainty in the validation due to random internal variability in the AOGCM transient simulations. The third term represents emulation uncertainty due to scaling of random internal variability in the slab patterns by the temperature response. A quadratic dependence on ΔT is indeed demonstrated in Fig. 9, where the variance in emulation error σ^2 for Northern Europe surface temperature is shown (blue curve).

Equation 22 can be rewritten:

$$\sigma^2 = V_{\text{em}} + \overline{\eta^2} + \overline{\varepsilon^2 \Delta T^2} \quad (24)$$

and combining this with Eq. 20, we arrive at an expression that relates δ^2 to the variance of cross validation error:

$$\overline{\delta^2} = \sigma^2 - \overline{\varepsilon^2 \Delta T^2} \quad (25)$$

assuming ε_j to be uncorrelated with ΔT_j . The required variance in random uncertainty associated with each emulation P_j is therefore equal to the variance in cross validation error, minus the variance associated with scaling of random internal variability in the slab patterns by the global temperature response. An estimate for $\overline{\varepsilon^2}$ is obtained from thirty 20 year means from a 600 year control simulation with the HadSM3 standard slab model configuration.

References

- Annan JD, Hargreaves JC, Ohgaito R, Abe-Ouchi A, Emori S (2005) Efficiently constraining climate sensitivity with paleoclimate simulations. *SOLA* 1:181–184, <http://www.jstage.jst.go.jp/browse/sola>
- Collins M, Booth BBB, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2006) Towards quantifying uncertainty in transient climate change. *Clim Dynam* (in press). DOI 10.1007/s00382-006-0121-0
- 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: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds) *Climate change 2001: the scientific basis. Contribution of Working Group I to the 3rd assessment report of the IPCC*, pp 525–582. Cambridge University Press, Cambridge, pp 881
- Currin C, Mitchell TJ, Morris M, Ylvisaker D (1991) Bayesian prediction of deterministic functions, with application to the design and analysis of computer experiments. *J Am Stat Assoc* 86:953–963
- Furrer R, Sain SR, Nychka D, Meehl GA (2006) Multivariate Bayesian analysis of atmosphere–ocean general circulation models. *Environ Ecol Stat* (in press)
- Giorgi F, Francisco R (2000a) Uncertainties in regional climate change predictions. A regional analysis of ensemble simulations with the HadCM2 GCM. *Clim Dynam* 16:169–182
- Giorgi F, Francisco R (2000b) Evaluating uncertainties in prediction of regional climate change. *Geophys Res Lett* 27:1295–1298
- Good P, Lowe J (2006) Emergent behaviour and uncertainty in multi-model climate projections of precipitation trends at small spatial scales. *J Clim* (in press)
- Gordon C, Cooper CA, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim Dynam* 16:147–168
- Hoffert MI, Callegari AJ, Hsieh CT (1980) The role of deep sea heat storage in the secular response to climatic forcing. *J Geophys Res* 85:6667–6679
- Huntingford C, Cox PM (2000) An analogue model to derive additional climate change scenarios from existing GCM simulations. *Clim Dynam* 16:575–586
- Johns TC, Gregory JM, Ingram WJ, Johnson CE, Jones A, Lowe JA, Mitchell JFB, Roberts DL, Sexton DMH, Stevenson DS, Tett SFB, Woodage MJ (2003) Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim Dynam* 20:583–612
- Mitchell T (2003) Pattern scaling. *Climatic Change* 60:217–242
- Mitchell JFB, Johns TC, Eagles M, Ingram WJ, Davis RA (1999) Towards the construction of climate change scenarios. *Climatic Change* 41:547–581
- 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 Climate* 8:496–514
- Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DJ (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430:768–772
- Nakicenovic NJ, Alcamo J, Davis G, de Vries B, Fenham J, Gaffin S, Gregory K, Grübler A, Jung TY, Kram T, La Rovere EL, Michaelis L, Mori S, Morita T, Pepper W, Pitcher H, Price L, Raihi K, Roehrl A, Rogner H-H, Sankovski A, Schlesinger M, Shukla P, Smith S, Swart R, van Rooijen S, Victor N, Dadi Z, (2000) Emissions scenarios. A special report of working group III of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 599
- Pittock AB, Jones RN, Mitchell CD (2001) Probabilities will help us plan for climate change. *Nature* 413:249–249
- Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. *Clim Dynam* 16:123–146
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical recipes in Fortran 77: the art of scientific computing*, vol 1. Cambridge University Press, Cambridge, pp 402

- Raper SCB, Gregory JM, Osborn TJ (2001) Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. *Clim Dynam* 17:601–613
- Santer BD, Wigley TML, Schlesinger ME, Mitchell JFB (1990) Developing climate scenarios from equilibrium GCM results. Max-Planck Institute for Meteorology Report Number 47, Hamburg
- Schlesinger ME, Andronova NG, Entwistle B, Ghanem A, Ramankutty N, Wang W, Yang F (1997) Modelling and simulation of climate and climate change. In: Catagnoli GC, Provenzale A (eds) Past and present variability of the solar-terrestrial system. International School of Physics Enrico Fermi, vol 133. IOS Press, Amsterdam pp 389–429
- Shine K, Derwent R, Wuebbles D, Morcette JJ (1990) Radiative forcing of climate. In: Houghton J, Jenkins G, Ephraums J (eds) Climate change. The IPCC scientific assessment. Cambridge University Press, Cambridge, pp 45–68
- Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton DMH, 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 Climate* 18:1524–1540
- Webb MJ, Senior CA, Sexton DMH, Ingram WJ, Williams KD, Ringer MA, McAvaney BJ, Colman R, Soden BJ, Gudgel R, Knutson T, Emori S, Ogura T, Tsushima Y, Andronova N, Li B, Musat I, Bony S, Taylor KE (2006) On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Clim Dynam* (in press). DOI 10.1007/s000382-006-0111-2
- Williams KD, Senior CA, Mitchell JFB (2001) Transient climate change in the Hadley Centre models: the role of physical processes. *J Climate* 14:2659–2674