Perturbed parameter ensembles as a tool for sampling model uncertainties and making climate projections

J. Murphy[1], R. Clark[1], M. Collins[2], C. Jackson[3], M. Rodwell[4], J. Rougier[5], B. Sanderson[6], D. Sexton[1] and T. Yokohata[7]

[1]*Met Office Hadley Centre, Exeter, UK,* [2]*College of Engineering, Mathematics and Physical Sciences, Univ. of Exeter, UK,* [3]*Institute for Geophysics, Univ. of Texas, USA,* [4]*ECMWF, Reading, UK,* [5]*Dept. of Mathematics, Univ. of Bristol, UK,* [6]*NCAR, Boulder, Colorado, USA,* [7]*National Institute for Environmental Studies, Tsukuba, Japan*

1. Introduction

Uncertainties in the formulation of weather forecasting or climate models are manifested in alternative structural choices (for example, choices of resolution, numerical integration scheme and the basic physical assumptions on which parameterisations of sub-grid scale processes are based), and in the values of poorly constrained parameters controlling the representation of earth system processes or properties within a given model structure. Multi-model ensembles (MMEs), constructed from models available at different centres, have been extensively used in weather and seasonal climate forecasting applications (e.g. Krishnamurti et al., 1999; Palmer et al., 2004; Doblas-Reyes et al., 2009; Johnson and Swinbank, 2009), and for climate change projections (e.g. Meehl et al., 2007). An important strength of the MME approach is that each model undergoes extensive testing, including validation against a wide range of observables to establish its credibility as a tool for weather prediction or climate simulation. Another strength is that the models are constructed from a large pool of alternative components, hence sampling, to some extent, the effects of variations in model structure. The main limitations are that MMEs are rather small (ranging from a handful of models to a maximum of around 20 members), and are not designed to sample modelling uncertainties in a systematic fashion, being assembled on an opportunistic basis. Specifically, it is not clear how to define a space of possible model configurations of which the MME members are a sample. In climate change applications, for example, this creates the need to make substantial assumptions in order to interpret their results in a probabilistic form suitable for risk assessments (Tebaldi and Knutti, 2007).

Alternative approaches consist of exploring modelling uncertainties more systematically within a single model framework. In numerical weather prediction (NWP), attempts have been made to account for structural uncertainties in “multi-parameterisation” ensembles in which several versions of a model are used, distinguished by different choices for some of its parameterisation schemes (e.g. Houtekamer et al., 1996; Charron et al., 2010). This report focuses on perturbed parameter ensembles (PPEs), another technique applicable in a single model context. In PPEs, alternative model variants are created by sampling the settings of multiple uncertain parameters within specified ranges, either estimated *a priori* by experts based on their knowledge of the relevant physical processes (Murphy et al, 2004; Stainforth et al, 2005), or by using objective techniques to identify observationally constrained distributions, from which model variants are then selected (Annan et al., 2005a). While some investigation of PPEs has been pursued in NWP (Marsigli, 2009), usage has to date been more widespread in climate modelling. In section 2 we summarise worldwide PPE activities in climate simulation and prediction, and then discuss selected properties and applications from these in section 3, with a brief future outlook provided in section 4.

In all the studies presented here, the outputs from model parameterisation schemes are assumed to be deterministic, and are coupled to the resolved flow exclusively at the grid scale. They do not sample the potential stochastic effects of alternative unresolved sub-grid scale organisations (Palmer et al., 2005), now represented in some medium-range and seasonal forecasting systems via methods such as perturbations to the parameterised physical tendency terms (e.g. Buizza et al., 1999), or energy backscatter to the resolved scales (e.g. Berner et al., 2008). The potential to combine stochastic parameterisation and perturbed parameter techniques in ensemble prediction systems is a subject for future study.

2. Perturbed parameter experiments worldwide

Two comprehensive PPE projects based on the HadCM3 family of model configurations have been pursued during the past decade. The climate*prediction*.net (cpdn) project (Allen, 1999; Allen and Stainforth, 2002) has exploited spare capacity on personal computers worldwide to run large PPEs amounting to thousands of members, using either HadSM3 (the atmospheric component coupled to a mixed-layer (slab) ocean (e.g. Stainforth et al., 2005)), or HadCM3L, in which a three-dimensional dynamic ocean component with reduced horizontal resolution is used (e.g. Frame et al., 2009). Results have been used to investigate uncertainties in global and regional climate change feedbacks, how these are driven by the effects of model parameters both individually and in combination, and how they relate to errors in the simulation of historical observables that might be used to constrain future projections (Stainforth et al., 2005; Piani et al., 2005; Sanderson et al 2008a,b; Sanderson, 2012). A number of new experiments are currently in progress, including studies of past climate epochs, possible ocean circulation changes in the North Atlantic, and attribution studies in which the effects of climate change on past events are investigated using atmosphere model simulations driven by prescribed sea surface temperatures (see http://www.climateprediction.net).

The Met Office Hadley Centre QUMP (Quantifying Uncertainty in Model Predictions) project has also pursued a substantial programme of experiments using HadSM3 and HadCM3, deployed in studies of climate sensitivity, cloud feedbacks and observational constraints (Murphy et al., 2004; Webb et al., 2006; Collins et al., 2011). In the QUMP case, the experimental design has been more explicitly focused on the aim of providing future climate scenarios for the UK and other world regions, in which uncertainties are quantified more comprehensively than would be possible using MMEs in isolation. Some of this work is summarised in section 3.4. QUMP ensembles were run on Met Office supercomputers, being smaller in size than corresponding cpdn experiments, but involving a larger set of output diagnostics to support, for example, studies of future changes in climate extremes (e.g. Barnett et al., 2006; Clark et al., 2010). A study combining QUMP and cpdn simulations of the equilibrium response to doubled CO_2 has also been performed (Rougier et al., 2009).

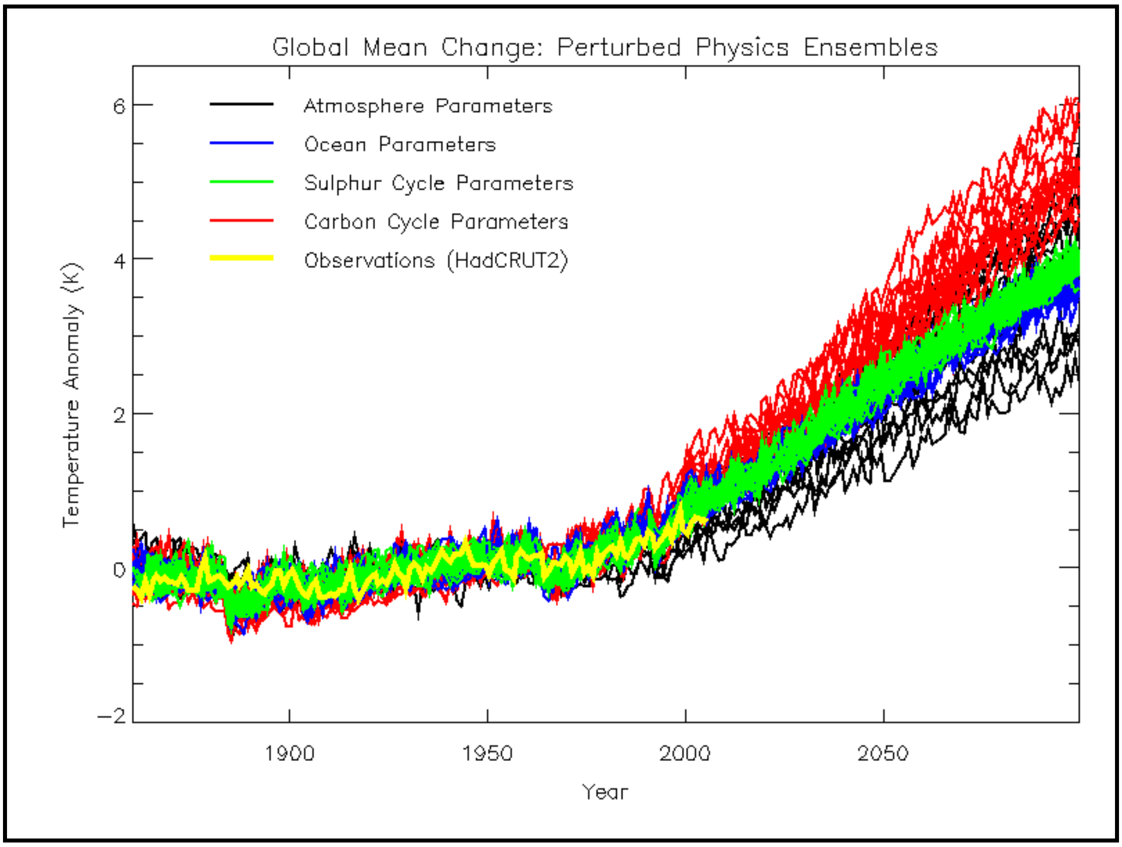

Figure 1. Global-mean temperature anomalies from 1860-2100 in perturbed parameter HadCM3 experiments forced by historically observed changes in anthropogenic and natural forcing agents and future greenhouse gas and sulphate aerosol emissions under the SRES A1B scenario, compared with observations to 2000. The different colours indicate ensembles with perturbations to parameters in different model components (as indicated in the legend) while keeping parameters in the other components fixed. Reproduced from Murphy et al (2009b).

Although much of the HadCM3-based work has focused on the effects of atmospheric model parameters, ensembles exploring separately the effects of perturbing ocean model parameters (Collins et al., 2007; Brierley et al., 2010), sulphur cycle parameters (Ackerley et al, 2009; Murphy et al., 2009a) and carbon cycle parameters (Booth and Jones, 2011) have also been produced. As an example, Figure 1 shows projected changes in global mean surface air temperature (SAT) during the 21st century, from 17-member QUMP ensembles sampling separately the effects of plausible perturbations to atmosphere, ocean, sulphur cycle and terrestrial carbon cycle processes. Uncertainties in the response of the carbon cycle drive a spread of changes similar to that derived from uncertainties in atmospheric feedback processes, and also comparable to uncertainties in the carbon cycle response found in a multimodel ensemble (Friedlingstein et al., 2006). Ensembles have also been produced in which parameters in multiple model components are perturbed simultaneously (Frame et al., 2009; Hugo Lambert, pers. comm.), facilitating study of the effects of interactions between uncertainties in different earth system modules.

Several PPE studies have also been carried out using the NCAR Community Atmosphere Model (CAM). Jackson et al. (2008) used Bayesian inference to study the uncertainty in parameter optimisation exercises, and implications for the response to doubled CO_2, using CAM version 3.1 in simulations with prescribed sea surface temperatures (SSTs) and with coupling to a slab ocean. Sanderson (2011) performed atmosphere-slab ocean simulations with CAM version 3.5 to study the effects of selected parameters on the global and regional responses to doubled CO_2, in comparison to their impacts in corresponding cpdn simulations. Fischer et al (2011) also used this version of CAM to explore how parameters controlling some key land surface properties affect CO_2-driven changes in regional temperature-related extremes. Covey et al. (2011) have used versions 3.6 and 4.0 to perform an extensive set of prescribed-SST simulations in order to assess the effects of 28 parameters on a number of standard measures of historical model performance. They plan to extend their experiment to coupled ocean-atmosphere climate change simulations.

In the Japan Uncertainty Modelling Project (JUMP), Yokohata et al. (2010) used an ensemble of 32 PPE simulations of pre-industrial and doubled CO_2 climates, based on a low resolution configuration of the MIROC3.2 atmospheric model coupled to a slab ocean. Each simulation sampled multiple perturbations to thirteen model parameters. These were determined by using an ensemble Kalman Filter to assimilate observational data into the model (following Annan et al., 2005a,b), generating a set of perturbed model variants which gave the best overall reproduction of 15 present-day climatological variables. The shortwave cloud feedbacks simulated by this ensemble were compared against those found in a corresponding 128 member ensemble of simulations using HadSM3. Yoshimori et al (2011) compared the above ensemble of doubled CO_2 simulations to a corresponding PPE of MIROC3.2 simulations of the response to the negative forcing experienced at the Last Glacial Maximum (LGM). They found smaller short-wave cloud feedbacks and a smaller spread in climate sensitivity in the LGM experiment, and suggested that this may be associated with rapid tropospheric adjustments to the imposed forcing. However, cross-ensemble variations between the total feedback, and its components, are found to be related in the LGM and doubled CO_2 experiments, providing support for the use of past changes at the LGM as an observational constraint on future climate change. Shiogama et al (2011) have recently constructed a new PPE using the more recent MIROC5 coupled ocean-atmosphere GCM, using results from short simulations of the atmospheric component to identify a range of parameter combinations for which the coupled configuration would provide plausible simulations, without recourse to the flux adjustment strategies used in some other PPE studies (e.g. Collins et al., 2006, 2011). Jackson et al. (2011) also constructed a PPE of coupled ocean-atmosphere simulations without flux adjustments (based in their case on HadCM3), to study the sensitivity of the ocean meridional overturning circulation to modelling uncertainties.

Klocke et al (2011) used a large set of 1 year prescribed-SST simulations of the ECHAM5 atmosphere model to explore perturbations to twelve parameters, and then chose a subset of 50 perturbed variants to perform doubled CO_2 simulations with coupling to a slab ocean. The results were used to explore relationships between historical cloud, radiation and precipitation-related model errors and climate sensitivity, comparing the relationships against those obtained from the multi-model CMIP3 ensemble of simulations assessed in IPCC AR4 (Meehl et al., 2007). Implications for the use of observational constraints to weight model projections were also discussed. The effects of uncertainties in seven cloud parameters in ECHAM5 were found by Haerter et al (2009) to exert a substantial influence on the present-day radiative forcing due to direct and first indirect (cloud albedo) effects of sulphate aerosols, particularly when multiple perturbations to the seven parameters were considered. In these simulations, aerosol mass mixing ratios were prescribed. Lohmann and Ferrachat (2010) used ECHAM5 coupled interactively to the HAM aerosol module to investigate the total anthropogenic aerosol forcing due to sulphates and black and organic carbon, accounting for fast aerosol-climate feedbacks such as the cloud lifetime effect, as well as the direct and first indirect effects. They considered uncertainties due to four cloud parameters typically used to tune the model's radiation balance, in simulations constrained to follow the observed atmospheric circulation for the year 2000. Lohmann and Ferrachat found an impact of about 25% on the total anthropogenic forcing, reducing to ~10% when only simulations close to global radiative balance are considered.

Niehörster (2009) compared the effects of varying convective entrainment, the rainout efficiency of cloud droplets and cloud ice fall speed on simulated global mean SAT changes in climate change experiments with the EGMAM atmosphere model coupled to a dynamic ocean. The impacts were found to be positively correlated with the effects of corresponding parameter variations in HadSM3 simulations. Neelin et al (2010) used prescribed-SST and coupled atmosphere-slab ocean simulations of the ICTP model to explore the effects of four parameters influencing the modelled hydrological cycle. They assessed the utility of low-order polynomial fits in determining optimal parameter settings for the skilful simulation of observed fields of climatological variables, and in quantifying the parameter-dependence of the simulated patterns of precipitation response to doubled CO_2.

3. Understanding and quantifying uncertainties in climate change

Here we provide a few examples from the studies summarised in section 2, drawn mainly from PPEs investigating the effects of uncertainties in surface and atmosphere processes. These illustrate some of the key properties and uses of PPE studies in understanding climate feedbacks and their uncertainties, diagnosing structural model limitations, using observations to identify credible regions of parameter space, and in developing approaches to the provision of observationally-constrained projections.

3.1. Understanding uncertainties by comparing different ensembles

In general, results from PPE experiments depend both on the model chosen for the experiment, and on the scope and design of the parameter perturbations. If the aim is to use a PPE for climate projections supporting decision making in the real world (see section 3.4), then the ensemble must be capable of generating a spread of outcomes consistent, at least to leading order, with other current sources of understanding and knowledge. In order to achieve this, it is important to design PPEs to sample key process uncertainties in as comprehensive a fashion as possible. In the QUMP experiments, for example, 31 parameters in the atmospheric component of HadCM3 were perturbed with this in mind, covering all major areas of sub-grid scale parameterisation in the model: large scale cloud and precipitation, convection, radiation, surface fluxes, boundary layer transports, land surface, sea ice, gravity wave drag and diffusion of heat, moisture and momentum. Yokohata et al (2011) examined several model ensembles from the standpoint of whether their distributions of simulated outcomes (for several mean climate variables) suggested that historical observations could be regarded as a possible sample drawn from the model ensemble. They found that contemporary MMEs, and the above QUMP PPEs, satisfied this condition to a greater degree than PPEs designed to treat a smaller number of uncertain parameters. Hargreaves et al. (2011) applied a similar reliability test to simulations of SSTs at the LGM, finding that reconstructions of observations were much more consistent with MME simulations from the Paleoclimate Model Intercomparison Project, than with a PPE of MIROC 3.2. simulations perturbing 13 model parameters (see also Yokohata et al., 2010). Masson and Knutti (2011) show that members of a 17-strong QUMP ensemble are more similar to each other (based on simulated monthly fields of historical surface temperature and precipitation) than they are to simulations of alternative CMIP3 models, reflecting the lack of sampling of the structural component of model error in PPEs. However, Collins et al. (2011) show that a larger QUMP ensemble, designed with a greater emphasis on sampling parameter space rather than optimising model skill, can produce distributions of error correlations between pairs of members similar to those found in a corresponding CMIP3 multi-model ensemble, for a number of climatological variables.

Even if a model ensemble passes some test of its sampling strategy based on historical performance, this does not guarantee that it will provide a realistic spread of plausible future outcomes, because the relationship between historical simulation errors and projected future changes is not necessarily strong (e.g. Knutti et al., 2010a). Given the widespread use of MME results to derive information on future changes (e.g. Meehl et al., 2007), one relevant test is whether a PPE generates a spread of changes similar to corresponding MME results. Collins et al (2011) showed that this is the case for global mean atmospheric feedbacks, in the case of the QUMP experiments (Figure 2). However, the PPE spread does depend on the design of the parameter perturbations (see top three rows of Figure 2): The S-PPE-E and S-PPE-M experiments both use (different sets of) multiple parameter perturbations, whereas S-PPE-S members each sample a perturbation to an individual parameter.

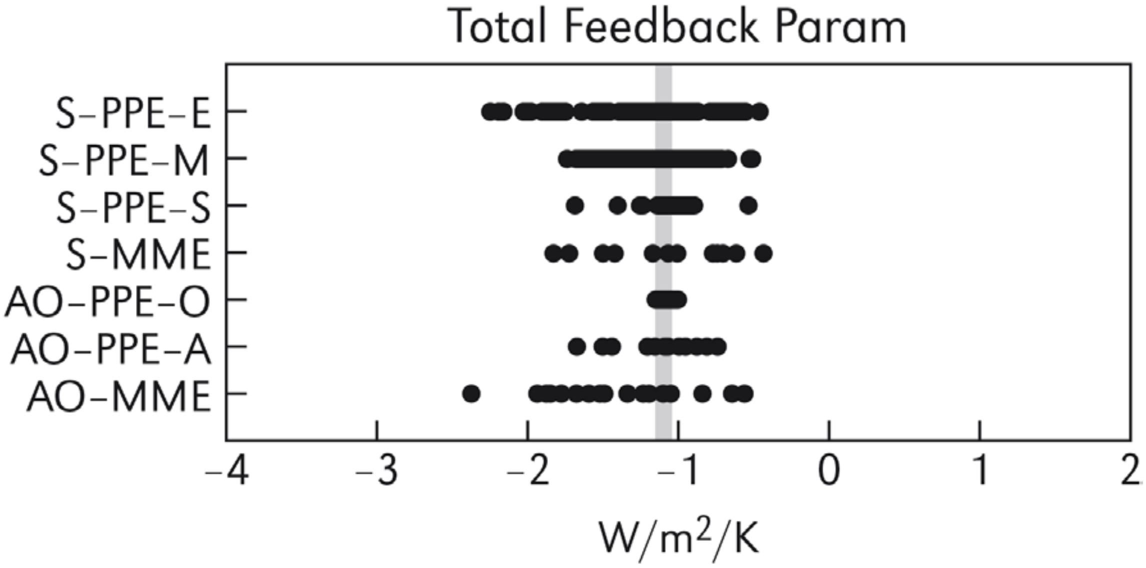

Figure 2. Global surface and atmospheric climate feedback parameters in $Wm^{-2}K^{-1}$ at time of doubling CO_2 (experiments prefixed AO-), or at $2xCO_2$ equilibrium (experiments prefixed S-), in various PPE and MME climate change simulations (Collins et al., 2011). Each circle is an ensemble member, while the grey shading is an estimate of the uncertainty due to natural variability. The PPE experiments (based on the HadCM3 model) give comparable ranges to the MMEs (run for IPCC AR4) when atmosphere parameters are perturbed, noting that when ocean transport processes are perturbed with atmosphere parameters held fixed (the AO-PPE-O experiment), the spread is much smaller than when atmosphere parameters are perturbed with the ocean parameters held fixed (AO-PPE-A) .

Sanderson (2011) studied the effects of four key parameters: convective entrainment coefficient, critical relative humidity for cloud formation, accretion constant for cloud droplets and cloud ice fall speed. These were chosen in view of their large impact on global climate sensitivity in cpdn experiments. On perturbing related parameters in the NCAR CAM3.5 model, Sanderson found an ensemble range of 2.2-3.2 K for climate sensitivity, compared to a larger range of 1.7-9.9K in cpdn experiments with a parallel perturbation strategy. Differences in the clear-sky longwave feedback played a major role in explaining the different ranges in the two ensembles: In their control simulations, the cpdn experiments simulated a wide range of values for specific humidity in the upper troposphere and lower stratosphere, and a strong positive correlation with the clear sky feedback on doubling CO_2, whereas the CAM3.5 ensemble exhibited neither of these features, showing that the effects of perturbing similar processes show structural differences between the two models.

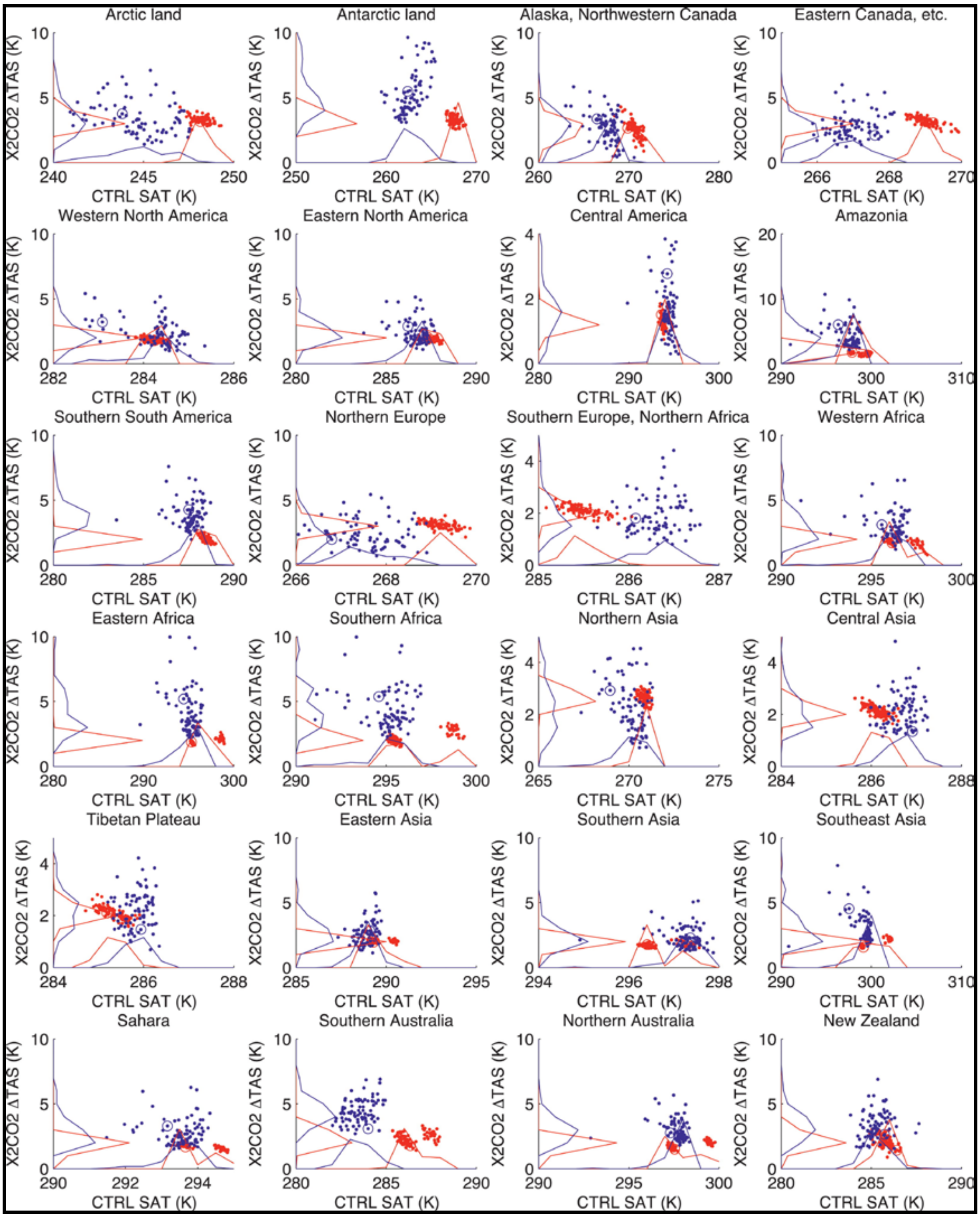

Figure 3. Scatter plots showing the joint distribution of regional surface temperature mean state and response to doubled CO_2 in cpdn (blue) and CAM3.5 (red) atmosphere-slab ocean PPE simulations (from Sanderson, 2011). In each case, the horizontal axis represents the preindustrial annual mean temperature for the region, while the vertical axis shows the regional equilibrium temperature response to CO_2 doubling. Histograms on each axis represent the distribution of that quantity in the ensemble. Circled points show the unperturbed, default configuration of each model.

However, both ensembles showed substantial diversity in their simulations of regional surface temperatures, both in control simulations and in the response to doubled CO_2 (Figure 3). In many cases, the regional ranges found in the CAM3.5 ensemble were comparable to those in the cpdn experiment, despite the narrower spread found for the global mean response. This illustrates the potential to use PPEs to study the processes driving regional systematic errors and climate feedbacks, as well as the aforementioned applications to global climate sensitivity. This is supported by the results of Harris et al. (2010), who found that QUMP simulations generate a spread of projected warming for northern Europe consistent with that from a corresponding CMIP3 MME of atmosphere-slab ocean simulations. Rowell (2011) uses the CMIP3 coupled ocean-atmosphere MME simulations (Meehl et al., 2007), and three QUMP PPEs, to assess the contributions of modelling uncertainty and internal variability to the uncertainty in future regional 20 year-mean precipitation changes. The largest of the QUMP ensembles (280 members) shows rich spatial patterns in the varying contribution of modelling uncertainty, with similar features found in the MME. The relative contribution from modelling uncertainties is highest in the deep tropics, and is also high over large parts of the northern continents in summer, central and eastern Asia in winter, and over the Arctic in winter. The QUMP results also demonstrate that uncertainties in modelling the terrestrial carbon cycle add significant uncertainty to projected rainfall changes over tropical land.

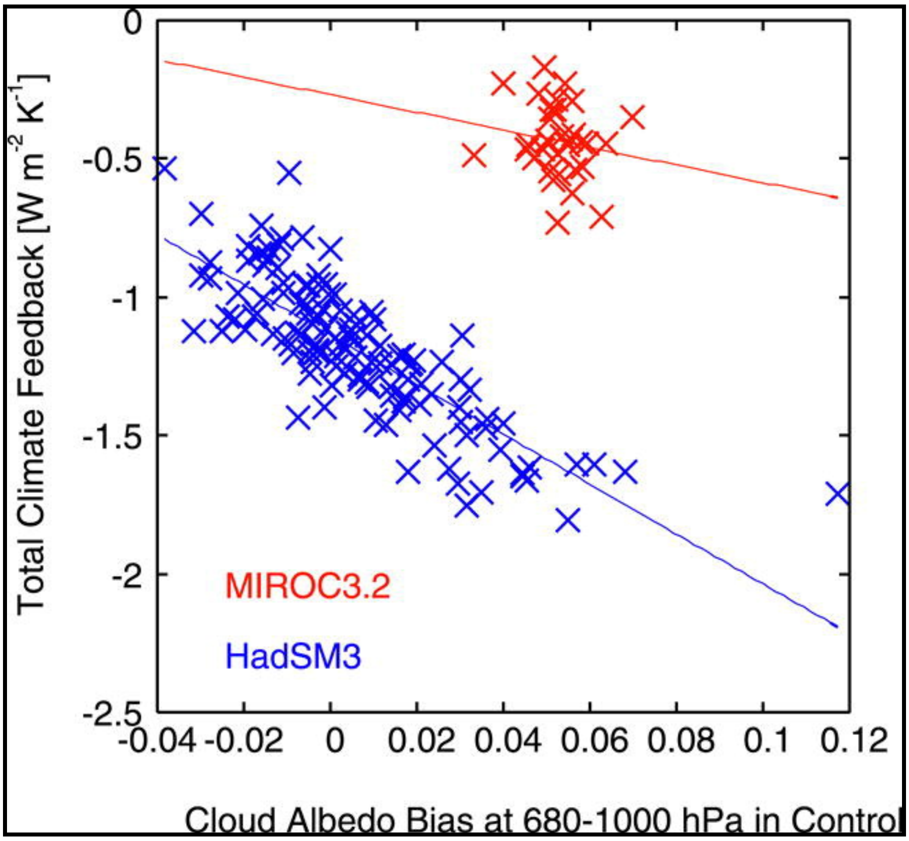

Figure 4. Scatterplot of global mean total climate feedback in response to doubled CO_2 versus the global mean cloud albedo in the lower troposphere (680-1000 hPa) in control ($1xCO_2$) simulations. Each point represents a member of the MIROC 3.2 (red) and HadSM3 (blue) perturbed parameter ensembles of Yokohata et al. (2010). Cloud albedo values are shown as biases relative to observations from ISCCP-D2 (Rossow and Schiffer, 1999).

Yokohata et al (2010) also compared PPEs generated using different atmosphere-slab ocean models, derived in their case from MIROC3.2 and HadSM3 (see section 2). They found that uncertainties in the shortwave cloud feedback (arising principally from changes in low cloud) provide the largest contribution to the spread of climate sensitivity values in each ensemble, while uncertainties in clear-sky longwave feedback are also significant in both PPEs. In this sense the two ensembles show a degree of structural similarity, however they also show a key structural difference: The ensemble-mean climate sensitivity in the MIROC ensemble is considerably higher than that of the HadSM3 ensemble, with only limited overlap between the two uncertainty ranges. Yokohata et al. find that the climate feedback in the HadSM3 ensemble is strongly related to variations in the albedo of low clouds in the control simulations (Figure 4). Ensemble members with high albedo possess more extensive cloud cover than model variants with low albedo; on doubling CO_2, these variants produce a stronger negative component to cloud feedback associated with optical thickening of this cloud cover, leading to lower climate sensitivities. In MIROC3.2, no relationship is found between the control cloud albedo and climate sensitivity: all the members possess clouds which are somewhat less extensive but optically thicker than observed, and the larger ensemble-mean response to doubled CO_2 compared to HadSM3 is explained by larger reductions in cloud cover (particularly at low and middle latitudes). This leads to a net positive low cloud feedback in the MIROC3.2 ensemble, and hence relatively high climate sensitivities. In a new PPE based on the MIROC5 coupled ocean-atmosphere GCM, Shiogama et al (2011) find that shortwave cloud feedbacks relating to changes in middle level cloud dominate variations in climate sensitivity, in contrast to Yokohata et al. (2010).

Watanabe et al (2011) investigate structural differences between the response of MIROC5 and MIROC3.2 by constructing a multiparameterisation ensemble (MPE) in which one or more of the convection, large scale cloud and boundary layer mixing schemes from MIROC3.2 is transplanted into MIROC5. The MPE generates a spread of climate sensitivity values which bridges the gap between the ranges sampled by the two PPEs. The MPE range is explained mainly by variations in the shortwave feedback associated with changes in tropical clouds, which arises from the non-linear coupling of two pairs of parameterization changes. In subsidence regions, the control simulations and future response of low cloud in MIROC5 is significantly altered when the MIROC3 cloud and boundary layer schemes are deployed together. In convective regions, combining the MIROC3.2 convection and cloud schemes has a large influence on both low and middle level clouds. Both of these coupled effects enhance the positive shortwave cloud feedback in MIROC5, reducing the contrast in behaviour between MIROC5 and MIROC3.2.

3.2. Identifying processes driving uncertainties

A key advantage of PPEs is that the explicit experimental design opens up possibilities to identify the effects of specific processes on model biases and projected changes, by quantifying the effects of different model parameters. In MMEs, differences between ensemble members arise from a complex combination of alternative choices for grid resolution, numerical integration scheme, basic parameterisation assumptions and model parameters, making it very difficult to trace differences in the emergent behaviour of simulated outputs back to detailed process-level choices in the model formulation.

Clark et al (2010) documented uncertainties in simulated regional changes in heat extremes in response to doubled CO_2, using a HadSM3 PPE constructed from members containing multiple parameter perturbations. Figure 5 shows the fraction of northern hemisphere land points over which individual model parameters drive an uncertainty range greater than that expected from internal climate variability, for changes in the temperature associated with the typical hottest day of summer. Several parameters directly controlling aspects of surface moisture or radiation balance appear among the leading influences, including forest roughness length, stomatal conductance, the boundary layer cloud fraction in a saturated grid box, and vegetation root depth. However, Figure 5 suggests that perturbations in all 31 perturbed parameters contribute to the spread of responses beyond the level attributable to internal variability, often through indirect influences on summer temperatures. For example, uncertainties in sea-ice albedo significantly affect the spread of changes over north–eastern Russia (not shown), possibly through effects on the seasonal cycle of soil moisture driven by the impact of variations in winter and spring temperatures on the timing of snowmelt. The large regional ranges therefore reflect uncertainties arising from an array of feedback processes, some of which operate directly at a regional level, while others operate indirectly via effects remote in space and/or season.

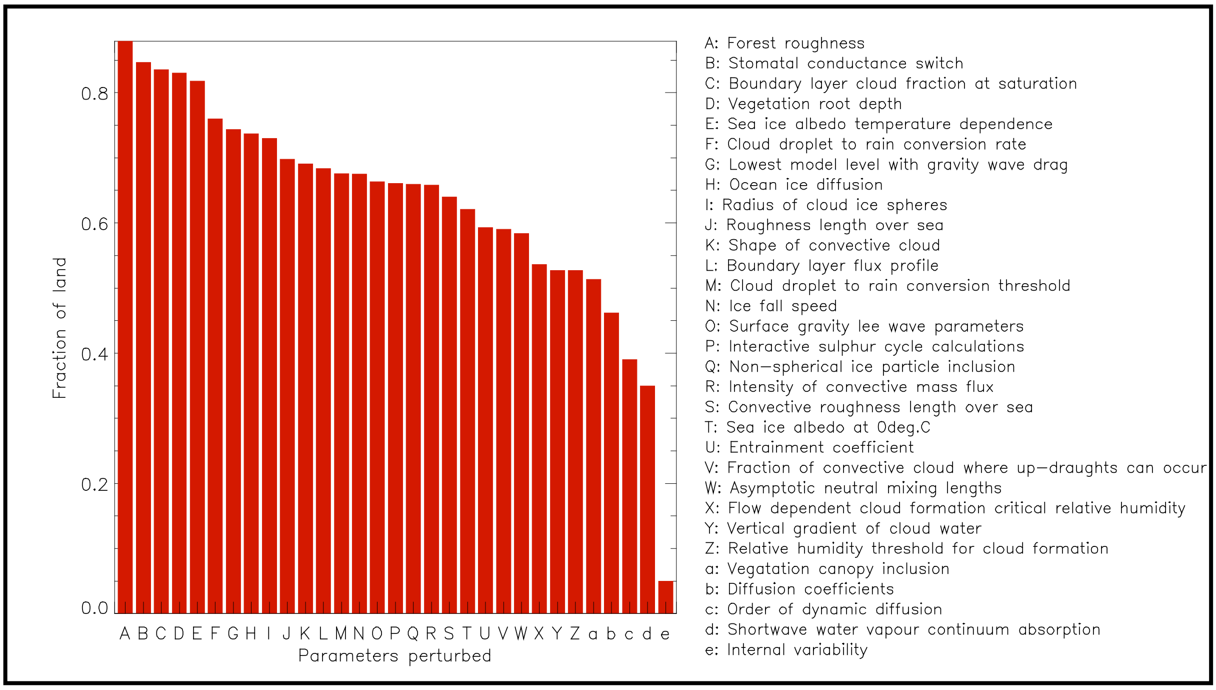

Figure 5. Areal fraction of northern hemisphere land for which perturbed parameters in the HadSM3 PPE of Clark et al (2010) drive uncertainties greater than internal variability in the equilibrium response to doubled CO_2 of the 99th percentile of daily maximum surface air temperature values simulated during June to August. The results are drawn from 44 ensemble members simulating global climate sensitivity in the range 1.5-2.5K. Sampling uncertainty in internal variability (rightmost bar) is expected to give a false signal at 5% of points.

In the above study, the effects of individual parameters were estimated by differencing the average changes found in subsets of ensemble members containing values of the relevant parameter in the upper and lower terciles of its continuous range, or by averaging differences of the most extreme settings for discrete parameters consisting of 2, 3 or 4-level switches. In this simple method, the estimated impacts for any given target parameter may be partially confounded by uneven sampling of other parameters in the binned subsets. This can be addressed by building statistical emulators, trained on the available ensemble results to predict the behaviour of parts of parameter space for which no climate model simulation exists (Sanderson et al, 2008a; Rougier et al., 2009; Sexton et al., 2011). Such emulators express the values of some set of model outputs (typically a mixture of historical observables and projected changes in some chosen variables) as functions of model parameters, allowing the effects of individual parameters, and key interactions, to be isolated.

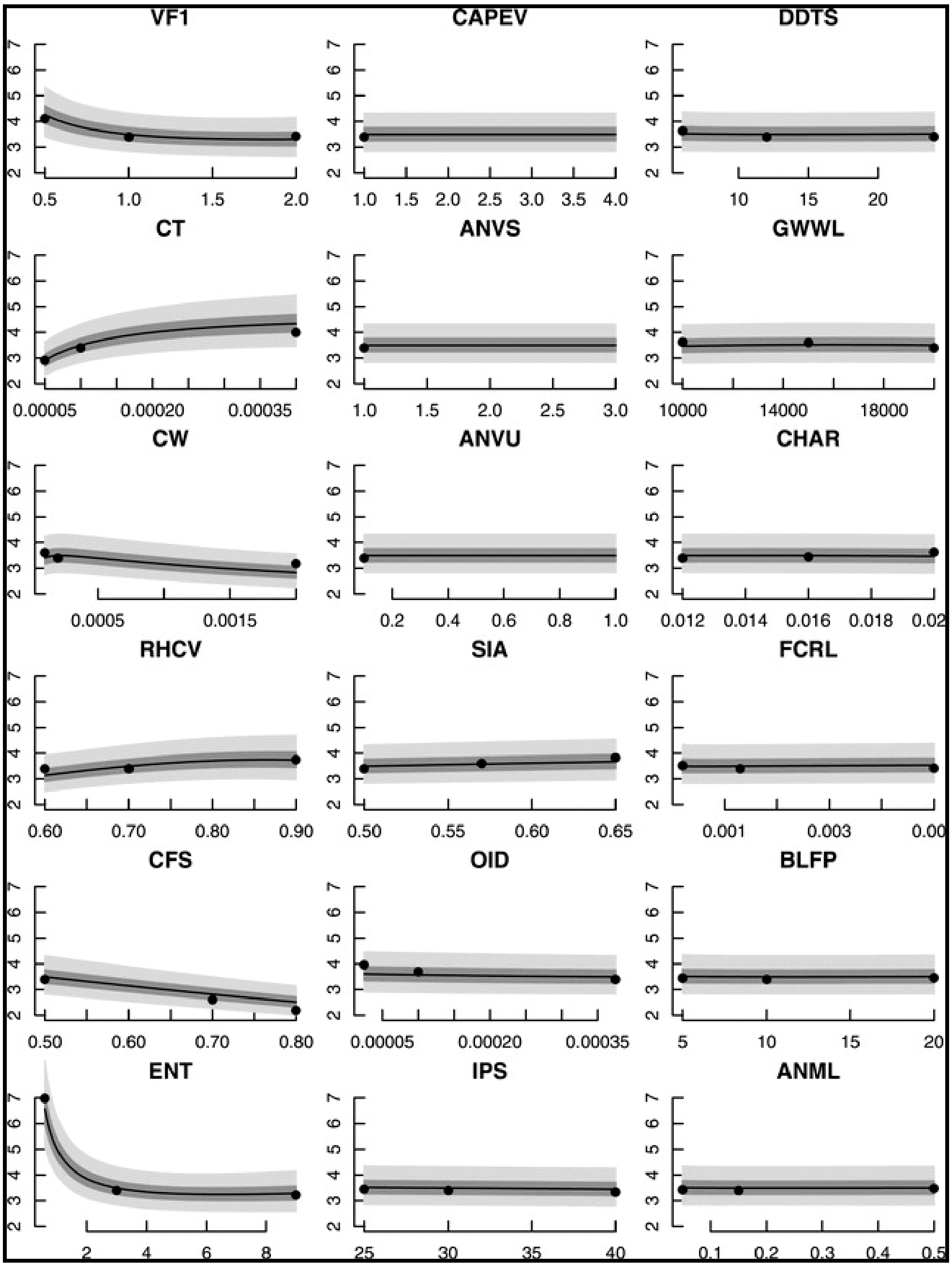

Figure 6. The effect on climate sensitivity of varying each of the continuous HadSM3 parameters considered by Rougier et al (2009), based on an emulator trained on cpdn and QUMP PPEs. The line shows the median emulated value, the darker and lighter grey shading denoting associated 50% and 95% credible intervals. The dots show values from HadSM3 simulations in which parameters were perturbed individually relative to their standard values. See Rougier et al for definitions of the full set of parameters shown. Those showing the largest effects are explained in the text below.

Sanderson et al (2008a) developed an emulator based on neural network techniques, and trained on a multi-thousand member cpdn ensemble of HadSM3 simulations. The emulator was used to predict climate sensitivity, and also a set of present day climate mean observables based on empirical orthogonal functions of surface temperature, precipitation and radiative fluxes in a set of worldwide regions. The performance of the emulator was verified against a set of cpdn simulations withheld from the training step, and used to estimate results of the whole parameter space of the cpdn ensemble, via a large Monte Carlo sample. Results were used to map relationships between the skill of the emulated observables and climate sensitivity (the smallest model errors were found for sensitivities in the range 3-5K when considering all the observables together), and to estimate the dependence of climate sensitivity on individual model parameters.

Rougier et al (2009) built a Bayesian emulator trained by combining cpdn and QUMP simulations of HadSM3, based on a set of regression functions capturing linear and quadratic dependencies on each continuously-variable parameter, augmented by a number of key interaction terms capturing non-linear relationships between parameters. Discrete parameters (sampled in the QUMP simulations only) were treated as factors. The results (Figure 6) show that climate sensitivity increases rapidly as low values of convective entrainment rate (ENT) are approached, and also increases at low values of cloud ice fall speed (VF1) and the threshold for conversion of cloud droplets to precipitation (CW), while sensitivity reduces for low values of the cloud droplet accretion rate (CT), and the critical relative humidity for cloud formation (RHCV). These features are qualitatively consistent with the parameter dependencies found by Sanderson et al (2008a), although there are quantitative differences between the two studies, reflecting differences in the sets of climate simulations used for training, and different approaches to emulation. Other parameters in Figure 6 affect climate sensitivity to a smaller or (in some cases) negligible degree. We note also that some of the parameters exerting the strongest influence on regional heat extremes (Figure 5) do not significantly affect global climate sensitivity, underlining the importance of a comprehensive approach to the sampling of process uncertainties for applications in which both global and regional contributions to climate change are important (see section 3.4).

3.3. Constraining parameter space using observations

Many studies have demonstrated that the historical simulation skill of PPE members can vary substantially between different parts of parameter space (see Collins et al (2011) for a recent example). During the development of “best effort” models contributed to IPCC assessments of future climate change, a “tuning” process is usually performed (Randall et al., 2007), in which a few parameters are adjusted to ensure acceptable replication of certain key observables, such as planetary radiation balance. Such exercises are based on the judgement of modelling experts, but are usually limited in scope, somewhat subjective, and hard to document. Jackson et al (2008) demonstrate an application of Bayesian inference in which PPE simulations are used to create a formal, objective and repeatable procedure for identifying model configurations with optimised parameter settings. They performed simulations of the CAM3.1 atmosphere model using sea surface temperatures prescribed from observations. Six optimisation exercises were carried out, each starting from a different set of values for six parameters representing important sources of uncertainty in the simulation of large scale clouds and convection. The purpose was to study the uncertainty in the model development process, assuming an aim of creating a single model variant the best represents climate. In each optimisation experiment, a multivariate stochastic importance sampling algorithm was used to determine successive steps in parameter space at which new model simulations would be carried out, seeking convergence towards a location minimising modelling errors. These errors were diagnosed from seasonal spatial fields of a variety of variables measuring humidity, cloud cover, radiative fluxes, surface exchanges, surface temperature, precipitation and circulation. Jackson et al found that each of their six optimisation exercises led to an overall improvement in model performance relative to the default model configuration (Figure 7, top left panel). The change in skill varied between observables, and a few variables (notably mid- and high-level clouds) were less skilful in the optimised versions compared to the default. Also, the levels of skill for particular variables varied somewhat between the six different optimisation exercises. These results illustrate the alternative multivariate error balances explored in PPEs. A related finding was that the parameter settings found in the six optimised model variants of Figure 7 differed considerably, despite the similarity in the overall level of skill achieved.

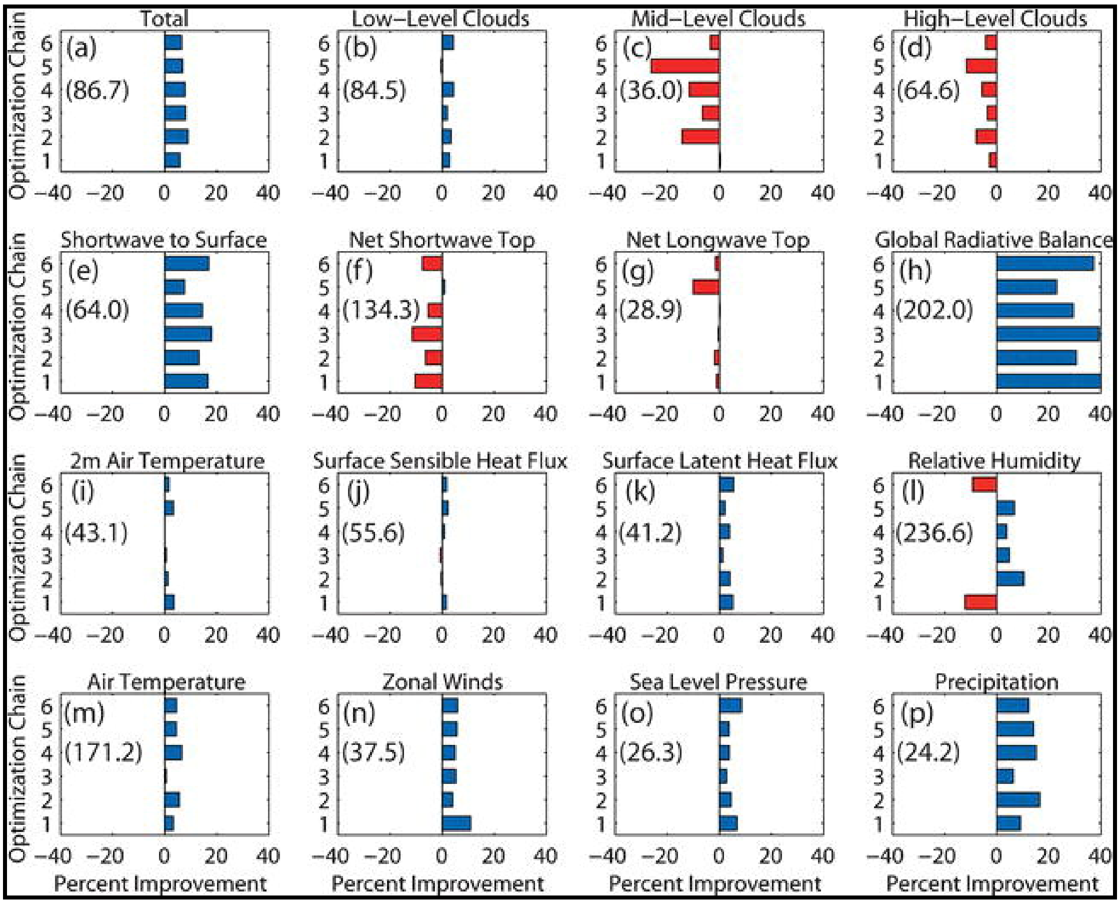

Figure 7. Impact of stochastic parameter optimisation on the CAM3.1 model, from Jackson et al (2008). Bars show skill in predicting seasonal and regional climatologies of 15 observable fields commonly used in model development. Each bar represents the change in skill in one of six optimised configurations, as compared to the default configuration. The size of each bar represents the percent improvement (blue) or degradation (red) in a skill score based on mean squared differences, for (a) a weighted average, and (b)-(p) the 15 individual observational targets. The skill score for the default configuration is given in parentheses.

While searching for optimally tuned parameter sets is one application of PPEs, another is to seek an ensemble of projections which represent uncertainties as fully as possible, by considering all parts of parameter space judged capable of yielding useful (while imperfect) information. Alternative projections will in general be characterised by different error balances and simulation characteristics (see above), and may in principle be weighted according to relevant metrics of historical model performance.

Sexton et al (2011) describe one method of approaching this, based on a Bayesian statistical framework (Goldstein and Rougier, 2004; Rougier, 2007). Seasonal climatological fields of several present-day observables (similar in scope to Jackson et al (2008)) are expressed as a set of multivariate spatial eigenvectors which capture most of the variability in simulated values across 280 members of a HadSM3 PPE. An emulator is trained on the 280 simulations to predict the modelled values of the associated principal components at any point in HadSM3 parameter space, and these are then compared against the observed principal components to estimate the relative likelihood of alternative parameter combinations given the observations, accounting for uncertainties due to emulator error (including internal climate variability), observational error, and a prior estimate of the uncertainty in structural model error. The structural error represents a systematic component to the model-data misfit that cannot be resolved by varying uncertain parameters in HadSM3, and is estimated by comparing HadSM3 against other international climate models, as outlined in section 3.4.

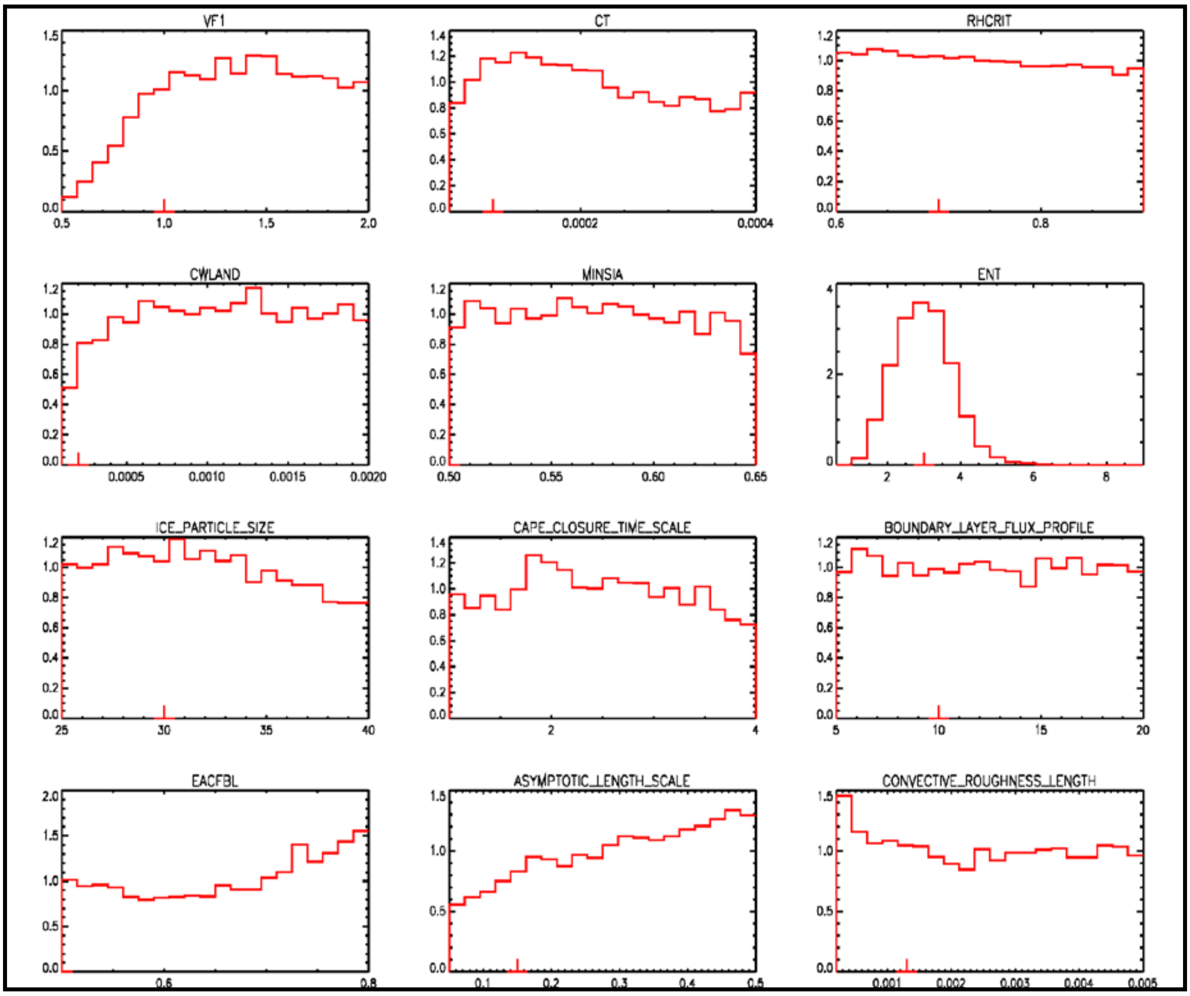

Figure 8. The ratio of posterior to prior probability for twelve continuously-variable HadSM3 parameters from Sexton et al (2011), based on likelihood estimates derived from modelled and observed projections onto six multivariate spatial eigenvectors of a set of seasonal climatological fields commonly used to evaluate climate model simulations. The prior probabilities are assumed constant within the middle 75% of the expert-specified range, dropping linearly to zero at the extreme values. Parameter values used in the standard version of HadSM3 are shown by the red dash on the x-axis. See Sexton et al for definitions of the full set of parameters shown.

Figure 8 shows the impact of this procedure in constraining the HadSM3 parameter space. The curves express the posterior probability for each value of continuously variable parameters, relative to a prior probability based on expert judgement. The extent to which the observational constraint modifies the prior distributions varies between the parameters. Low values of VF1 and ENT are substantially downweighted, implying that the high climate sensitivities obtained from these settings (see Figure 6) are less plausible than realisations of climate sensitivity (regardless of value) obtained from other parts of parameter space. In the case of ENT, this is supported by Joshi et al (2010), who find that HadSM3 variants with low values simulate excessive levels of stratospheric water vapour, because convective plumes transport too much moisture into the upper tropical troposphere, some of which is then transported into the stratosphere. The same mechanism operates in the doubled CO_2 simulation, causing an unrealistically large increase in stratospheric water vapour, and a large positive longwave feedback (see also Sanderson, 2011). These results demonstrate that while knowledge uncertainty in the behaviour of specific processes might justify exploring a wide range of perturbed values to parameters like ENT or VF1, some of these values actually lead to unrealistic simulations of the climate system as a whole. This underlines the importance of constraining parameter space *a posteriori* in PPEs, using metrics of key emergent properties.

In cases where Figure 8 shows a flat distribution, this indicates that the set of mean climate observables used by Sexton et al (2011) is not effective in discriminating between different values of the relevant parameter. This implies (not surprisingly) that a more comprehensive set of observational constraints is needed to refine estimates of the relative credibility of different model variants. Many candidates have been suggested, including: measures based on observed climate change over the past few decades, since the industrial revolution, during the past millennium, or from earlier palaeoclimatic epochs; the observed response to major volcanic eruptions; detailed process-based metrics relating directly to the outputs from specific parameterisation schemes; metrics based on the seasonal cycle, or aspects of intraseasonal or interannual climate variability, that might be related to climate change feedbacks. See, for example, Meehl et al (2007) and Knutti et al (2010b), and references therein.

One approach is to evaluate ensemble prediction systems used for multidecadal climate projections by using verification data from hindcasts of shorter lead times, potentially from weather forecasting time scales to a season, year or decade ahead. This reflects the seamless prediction concept of building more unified forecasting systems to exploit commonalities in the mechanisms determining the evolution of the climate system on different time scales (e.g. Palmer et al., 2008; Hurrell et al., 2009).

Rodwell and Palmer (2007) used a version of the ECMWF atmospheric weather forecasting model to show how temperature tendencies found in the first few forecast time steps could be used to assess the credibility of parameter perturbations similar to some of those deployed in cpdn and QUMP climate simulations. This approach allows the effects of parameters affecting physical processes on short time scales ("fast physics errors") to be assessed cheaply, and independently of systematic errors arising from interactions between multiple process errors in longer simulations ("slow physics errors"). Figure 9 (top panel) shows vertical profiles of the total temperature tendency (red curve) over the Amazon region, from a set of forecasts for January 2005 using the unperturbed version of the model. The total tendency arises essentially from a balance between convective warming (blue curve) balanced by dynamical cooling due to ascent (orange curve), with radiative destabilisation (dark green curve) playing a key role in triggering the convection, and also contributing to the overall balance. When convective entrainment is reduced by a factor of five, the convective heating increases substantially (Figure 9, lower panel), consistent with reduced loss of moisture and buoyancy from convective plumes. These changes lead to larger magnitudes for the total tendency (red curve), indicating a lack of overall balance compared to the unperturbed model, and leading to larger systematic forecast errors after 5 days (cf black curves in upper and lower panels). The low entrainment perturbation creates similar imbalances in the total temperature (and moisture) tendencies in many parts of the world, leading Rodwell and Palmer to conclude that this perturbation degrades the model physics, and should be rejected. This conclusion is consistent with the low weight of credibility attached to low entrainment based on the analysis of longer climate simulations (see Figure 8 and related discussion).

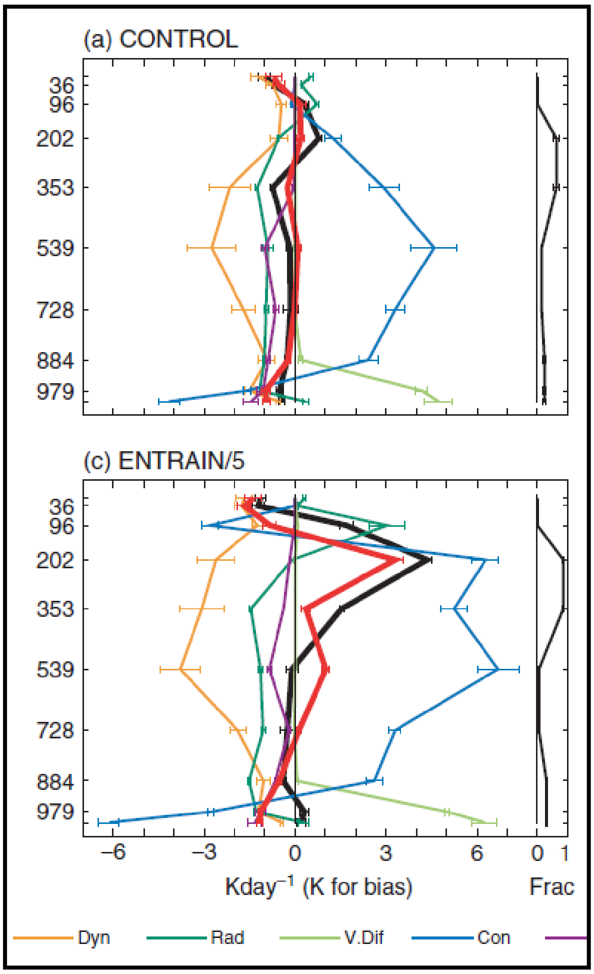

Figure 9. Vertical profile of initial temperature tendencies in K per day for various pressure levels in the Amazon/Brazil region. Results are from January 2005 forecasts using the ECMWF operational forecast model, for the unperturbed model (top panel), and a version with reduced convective entrainment (lower panel). The initial tendencies correspond to the model's dynamical tendencies, those from major parts of the model physics, and the total tendency (red curve). Also shown are the profiles of systematic error after five days (black curve, in K), and the mean initial cloud fraction (right-hand profile in each panel). Bars indicate 70% confidence intervals. The vertical coordinate is linear in pressure. See Rodwell and Palmer (2007) for details.

3.4. Deriving observationally-constrained climate projections from PPEs

A number of studies address the question of how to obtain observationally-constrained projections of future climate change from ensembles of climate model simulations. These fall essentially into two classes (Knutti et al., 2010b). The first is based on the idea of finding a direct relationship between one or more observables and a prediction variable of interest (e.g. global climate sensitivity), and the second on the application of Bayesian statistical methods in which a joint analysis of multiple uncertain objects in the prediction problem (observed climate, modelled historical and future outputs, and factors controlling the relationships between them) is performed, based on some set of specified prior distributions updated with evidence from model simulations and available observations.

The first approach includes several studies based on direct relationships between past and future climate changes (see Stott and Forest (2007) for a review). Some of these use large ensembles of simple or intermediate complexity climate model simulations to explore the spread of future projections in global quantities constrained by the fit of past changes to observations (e.g. Forest et al., 2002; Knutti et al., 2002), while others use small ensembles of simulations of comprehensive climate models in which a linear relationship is assumed between fractional errors in simulating past and future changes (ASK methods, based on Allen et al. (2000) and subsequent papers). Other studies consider links between present day climate variables and future change: for example, Hall and Qu (2006) find a strong relationship between future changes in surface albedo per unit warming, and the current seasonal cycle. In the PPE context, Piani et al (2005) used multivariate linear relationships with spatial fields of several climatological mean observables to obtain an observationally-constrained probability distribution for climate sensitivity from cpdn simulations using HadSM3, while Knutti et al. (2006) used a relationship with the amplitude of the seasonal cycle in surface temperature for a similar purpose.

Figure 10 demonstrates schematically how this approach works: (i) An ensemble of climate model simulations (black dots) is used to find a regression relationship between an observable variable (noting that in practice a multivariate set of observables can be used) and a predictand; (ii) a probability distribution for the observed value (accounting for uncertainties due to internal variability, and ideally measurement biases) is found (blue pdf on the x-axis), which can be converted into a probability distribution for the future prediction variable from the regression relationship (blue pdf on y-axis); (iii) uncertainty in the relationship diagnosed from the scatter between the observable and the predictand is then added, leading to a wider probability distribution for the predictand (red probability distribution). In this approach, uncertainties in the observations, and in the relationship between the observable(s) and the predictand diagnosed from the ensemble, play key roles in determining the nature and robustness of the results. Sanderson et al (2008b) illustrate these points in a study in which observationally-constrained distributions for climate sensitivity are obtained from around 6,000 members of the cpdn ensemble of HadSM3 simulations of present-day and doubled CO_2 climates. Their distributions are derived from linear relationships between spatial eigenvectors capturing variations across their ensemble in climate feedback strength and simulated historical values of a set of key regional climate variables, following Piani et al (2005). Importantly, they demonstrate the potential dependence of their results on key assumptions, by using alternative observational datasets to constrain the future changes, and also by comparing the relationship between the historical observables and climate sensitivity diagnosed from their PPE against results from other climate models with different structural assumptions. Sanderson (2012) shows that a methodology of this type overestimates the climate sensitivities of alternative CMIP3 models, unless the set of PPE simulations is pre-filtered to exclude unrealistic model variants with large biases in their planetary radiation balance. A modified regression approach is also demonstrated, in which the CMIP3 models are used to constrain the regression parameters linking eigenvectors of present climate and climate sensitivity, focusing the PPE-based transfer function on relationships which validate well against other models. This yields a probability distribution for climate sensitivity with a reduced upper bound of 4.3K.

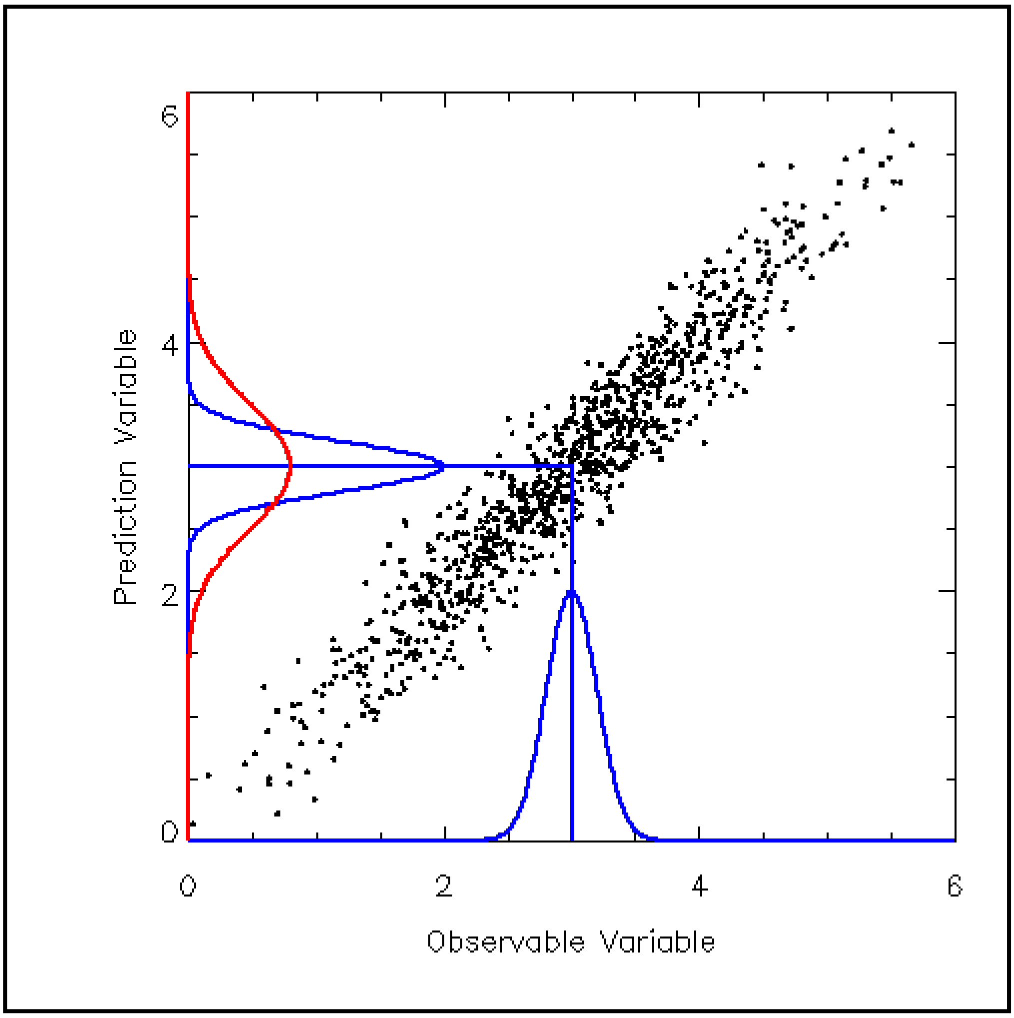

Figure 10. Schematic illustration of future climate projections based on relationships between a historical climate observable and a future variable of interest derived from an ensemble of climate model simulations. Contents explained in text below.

A number of Bayesian approaches based on MMEs have been published (Tebaldi et al., 2005; Greene et al., 2006; Lopez et al., 2006; Furrer et al., 2007; Buser et al., 2009), in which various multivariate statistical frameworks and prior assumptions are used to convert the model simulations into probabilistic estimates of future changes. Most of these are reviewed by Tebaldi and Knutti (2007). In the recent UK climate projections (UKCP09, see Murphy et al., 2009a), a Bayesian method was applied to PPEs to generate probabilistic projections designed to inform climate adaptation activities. The UKCP09 projections involved a relatively comprehensive attempt to account for current known sources of uncertainty in 21st century climate change. The underpinning methodology combined information from global climate model ensembles sampling key process uncertainties associated with surface and atmospheric feedbacks, ocean transport, and the sulphur and carbon cycles, with a regional climate model ensemble sampling uncertainties in downscaling to sub-UK regions, and a set of observational constraints. At its core was a methodology for probabilistic projections of the equilibrium response to doubled CO_2, described in detail by Sexton et al. (2011), and summarised briefly in the discussion of Figure 8 above. These equilibrium projections were then converted into projections of time-dependent change using a pattern-scaling approach (Harris et al., 2006, updated), adding in uncertainties associated with other earth system components, diagnosed from relevant ensemble simulations (see Figure 1). Wherever feasible, information from corresponding MME experiments was combined with HadCM3-based PPEs to provide uncertainty estimates combining the complementary strengths of both types of ensemble (see Introduction).

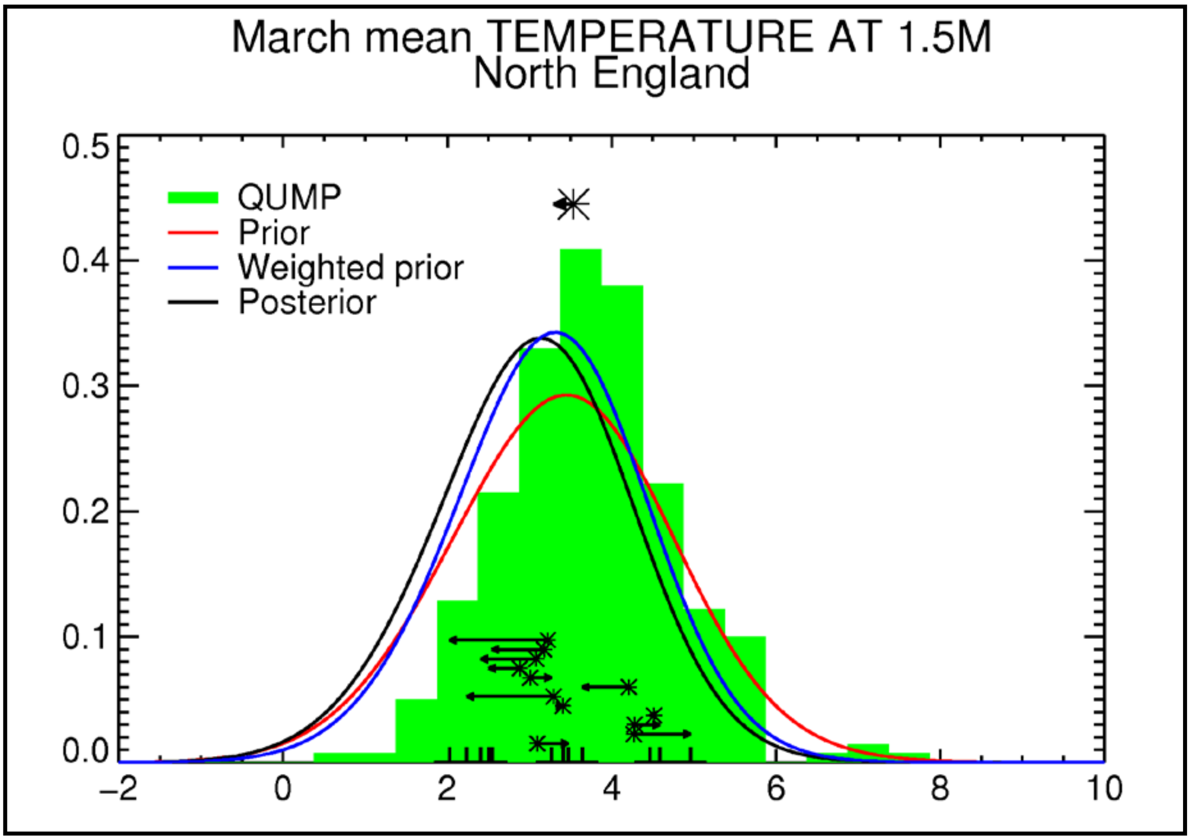

Figure 11. Probability distributions for the equilibrium surface temperature change over Northern England in March, in response to doubled CO_2, using the methodology of Sexton et al (2011). See text for explanation of contents.

Figure 11 illustrates how this was done for the core calculations of equilibrium climate change. The green histogram (labelled QUMP) shows the response simulated by 280 perturbed parameter variants of the HadSM3 model. The red curve (prior) shows the distribution of changes obtained using an emulator trained on the HadSM3 simulations to sample changes from a space of 31 HadSM3 parameters, defined from expert-specified distributions. The blue curve (weighted prior) shows the effect of applying constraints derived from eigenvectors of a set of present day mean climate observables. The black dashes on the x-axis denote changes simulated by a multi-model ensemble of 12 alternative models assessed in IPCC AR4, and the arrows show the discrepancy between these, and predictions of their response found by locating best analogues of their properties in the HadSM3 parameter space (asterisks). The 12 values of discrepancy are converted into a Gaussian distribution assumed to represent the future effects of structural model error. A similar calculation is used to estimate the effects of structural error on the historical simulations. The blue curve includes the broadening effect of the discrepancy variance on the future projections, and also the effect of the historical component of discrepancy on the weights applied to different parts of parameter space. The black curve (the final posterior distribution) differs from the blue curve by the mean effect of the future discrepancy, shown also by the large asterisk and arrow. Sexton and Murphy (2011) find that such projections are reasonably robust to plausible variations in key assumptions in this methodology, including the assumed prior distributions for model parameters, the amount of information used in the observational constraint, and the specification of the discrepancy term. A key caveat in this approach, and that of Sanderson (2012) discussed above, is that the use of alternative models to estimate the effects of structural model error inevitably neglects the effects of systematic errors common to all models.

4. Summary and outlook

The use of perturbed parameter ensembles (PPEs) to understand and quantify uncertainties in climate simulations has grown during the past decade, and now involves a community of projects at several modelling institutes. The technique provides a systematic approach to the study of modelling uncertainties within a single model framework, facilitating understanding of the effects of errors in specific processes on emergent behaviour in historical simulations and future projections. In this sense PPEs are complementary to multi-model ensembles of opportunity (MMEs), which are not designed to support traceability of this sort, but do possess the advantages of sampling structural variations in model formulation which cannot be covered in PPEs. The developing literature in PPEs includes applications in understanding drivers of uncertainty, model optimisation, finding relationships between historical observables and future projections, and identifying structural model limitations. The properties of a PPE depend significantly on the model used to generate them, and on the design and scope of the applied set of parameter perturbations. Perturbed parameter ensembles have also been used to make national climate change projections providing a basis for adaptation decisions (in combination with multi-model information and the application of observational constraints), and as a tool for promoting public awareness of climate change and its uncertainties, through the climate*prediction*.net project.

In short-range weather forecasting and seasonal prediction, the use of PPEs is less widespread, compared to the use of MMEs and stochastic parameterisation techniques. This is despite evidence that perturbing model parameters can significantly affect model forecasts on NWP time scales (Rodwell and Palmer, 2007). One practical reason for this may be the developmental and computational overhead of maintaining and running operational ensemble prediction systems consisting of multiple model variants with differing systematic error characteristics (e.g. Houtekamer et al., 1996). The Met Office medium-range forecasting system MOGREPS avoids this by deploying a technique in which several model parameters vary in time during the forecast integration of each of a set of ensemble members, according to a stochastic first order regressive process (Bowler et al., 2008). This contrasts with the climate simulation applications described in this report, in which alternative sets of parameter combinations are held fixed during the simulations of each ensemble member, creating systematically different realisations of climate feedbacks and other emergent characteristics.

The relative benefits of multi-model, stochastic parameterisation and perturbed parameter techniques may vary with different forecasting lead times, due, for example, to the changing balance between the effects of internal variability and forced climate change (Hawkins and Sutton, 2009), but systematic comparisons of alternative methods are needed to assess this. Also, more work is needed to assess the benefits of combining the different methods, where these can be shown to be sampling independent contributions to model uncertainty (e.g. Figure 10 and related discussion).

From a forecasting perspective, the question of whether and how to weight model projections of climate change depends on: (i) the potential benefits of reducing the influence of poorly performing ensemble members; (ii) the risk of reducing the degree of error cancellation which might be achieved by combining partially independent representations of the real climate system; (iii) the challenge of finding relevant observational constraints (Knutti et al., 2010a; Weigel et al, 2010). This applies to PPEs just as to MMEs, however the balance between factors (i) and (ii) is likely to be different, given that some parts of a model's parameter space will inevitably generate model simulations that would not pass the criteria normally applied when creating a tuned "best effort" simulation typical of MME members. The evaluation of climate prediction models in directly-verifiable shorter range forecasts (see section 3.3) is one of several ways in which the basis could be improved for weighting PPE members (e.g. Sexton et al., 2011), or for choosing parts of parameter space from which ensemble members of approximately equal credibility could be obtained (e.g. Jackson et al., 2008).

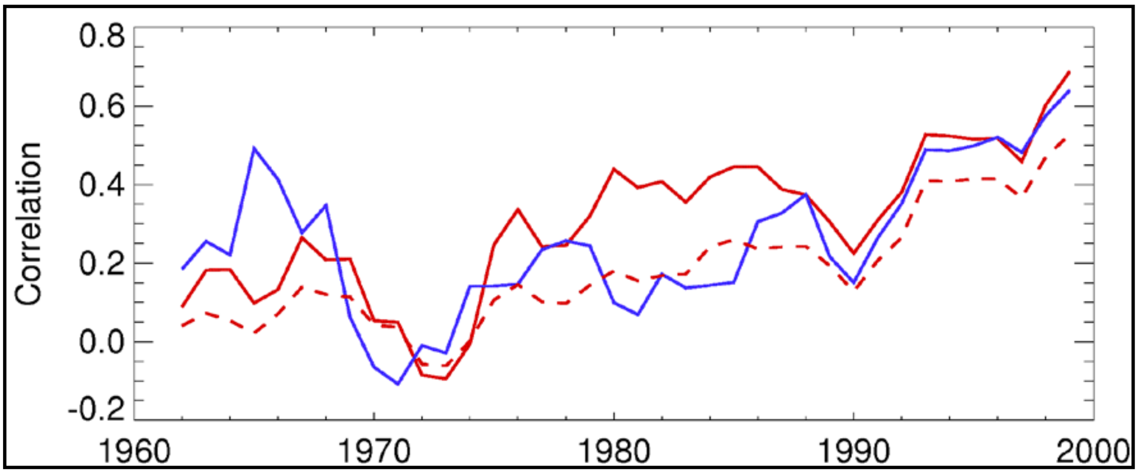

Figure 12. Time series of correlation between hindcast and observed global patterns of near-surface temperature anomalies, for hindcasts of nine-year means. Red dashed curve shows average scores for individual perturbed parameter variants of HadCM3 initialised from analyses of atmosphere and ocean observations for 1 November 1960, 1965,…, 2005. Solid red line shows scores for the ensemble mean of the nine constituent variants. Blue curve shows scores for the ensemble mean of a corresponding perturbed parameter ensemble in which hindcasts are driven by the same time-dependent specification of external radiative forcing anomalies, but lacking initialisation from recent observations. Time-dependent forcing anomalies arise from man-made greenhouse gases and sulphate aerosols from the SRES A1B scenario, and projected natural forcing from volcanoes and solar variations, assuming no prior knowledge of volcanic eruptions after the initialisation date. From Murphy et al (2009b). The ENSEMBLES work reproduced here is from the EU-funded FP6 Integrated Project ENSEMBLES (Contract number 505539).

In this seamless prediction context (Hurrell et al., 2009), another interesting development could be to assess the performance of PPEs initialised with recent observations in predictions on decadal and longer time scales, reflecting the developing interest in this area in the IPCC modelling community (Meehl et al., 2009). Figure 12 shows an early example of a PPE of nine variants of HadCM3, used to make initialised decadal hindcasts started from November 1960, 1965, 1970,.., 2005 (Murphy et al., 2009b), as part of the EU Framework Programme 6 ENSEMBLES project. The results show a time series of global pattern correlations for nine year average hindcasts of surface temperature. The average skill of individual members is consistently smaller than that of the PPE ensemble-mean (compare dashed and solid red curves), showing a benefit from combining the nine different realisations of modelling error and internal variability. On average, the skill of the initialised hindcasts is also slightly higher than that of a parallel ensemble containing the same external forcing from greenhouse gases, sulphate aerosols, volcanoes and solar variations, but initialised from randomly-selected model states rather than analyses of observations (compare solid red and blue curves).

The ENSEMBLES project also contained an MME of decadal hindcast experiments (Doblas-Reyes et al., 2010), as does the current Coupled Model Intercomparison Project phase 5 (CMIP5) protocol (Meehl et al., 2009). These experiments will provide an opportunity to extend comparison and combination of MME and PPE approaches, reported here in the context of historical simulations and long-term future projections, to initialised near-term climate forecasting.

References

Ackerley, D., E.J. Highwood and D.J. Frame, 2009: Quantifying the effects of perturbing the physics of an interactive sulfur scheme using an ensemble of GCMs on the climateprediction.net platform. J. Geophys.Res. 114: D01203.

Allen, M.R., 1999: Do-it-yourself climate prediction. Nature 401: 627.

Allen, M.R., P.A. Stott, J.F.B. Mitchell, R. Schnur and T.L. Delworth, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. Nature 407: 617-620.

Allen, M.R., and D.A. Stainforth, 2002: Towards objective probabilistic climate forecasting. Nature 419: 228.

Annan, J.D., J.C. Hargreaves, N.R. Edwards and R. Marsh, 2005a: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. Oc. Model. 8: 135-154.

Annan, J.D., J.C. Hargreaves, R. Ohgaito, A. Abe-Ouchi, and S. Emori, 2005b: Efficiently constraining climate sensitivity with ensembles of paleoclimate simulations. SOLA 1: 181-184.

Barnett, D.N., S.J. Brown, J.M. Murphy, D.M.H. Sexton and M.J. Webb, 2006: Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations. Clim. Dyn. 26: 489-511.

Berner, J., F. J. Doblas-Reyes, T. N. Palmer, G. Shutts, and A. Weisheimer, 2008: Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. Phil. Trans. R. Soc A 366: 2561–2579.

Booth, B.B.B., and C.D. Jones, 2011: Terrestrial response of QUMPC ensemble. Hadley Centre Technical Note No. 89, available from Met Office, FitzRoy Road, Exeter, UK.

Bowler N.E., A. Arribas, K.R. Mylne, K.B. Robertson and S.E. Beare, 2008: The MOGREPS short-range ensemble prediction system. Quart. J. Roy. Met. Soc. 134: 703-722.

Brierley, C.M., M. Collins and A.J. Thorpe, 2010: The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. Clim. Dyn. 34: 325-343.

Buizza, R., M. Miller and T.N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quart. J. Roy. Met. Soc. 125: 2887-2908.

Buser, C.M., H.R. Künsch, D. Lüthi, M. Wild and C. Schär, 2009: Bayesian multi-model projection of climate: bias assumptions and interannual variability. Clim. Dyn. 33: 849-868.

Charron, M., G. Pellerin, L. Spacek, P.L. Houtekamer, N. Gagnon, H.L. Mitchell and L. Michelin, 2010: Toward random sampling of model error in the Canadian ensemble prediction system. Mon. Wea. Rev. 138: 1877-1901.

Clark, R.T., J.M. Murphy and S.J. Brown, 2010: Do global warming targets limit heatwave risk ? Geophys. Res. Lett. 37: L17703.

Collins M., B.B.B. Booth, G.R. Harris, J.M. Murphy, D.M.H. Sexton and M.J. Webb, 2006: Towards quantifying uncertainty in transient climate change. Clim. Dyn. 27: 127-147.

Collins, M., C.M. Brierley, M. MacVean, B.B.B. Booth and G.R. Harris, 2007: The sensitivity of the rate of transient climate change to ocean physics perturbations. J. Clim. 20: 2315-2320.

Collins, M., B.B.B. Booth, B. Bhaskaran, G.R. Harris, J.M. Murphy, D.M.H. Sexton and M.J. Webb, 2011: Climate model errors, feedbacks and forcings. A comparison of perturbed physics and multi-model ensembles, Clim. Dyn. 36: 1737-1766.

Covey, C., S. Brandon, P.-T. Bremer, D. Domyancic, X. Garaizar, G. Johannesson, R. Klein, S.A. Klein, D.D. Lucas, J. Tannahill and Y. Zhange, 2011: Quantifying the uncertainties of climate predictions. Bull. Am. Met. Soc., submitted.

Doblas-Reyes, F.J., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J.M. Murphy, P. Rogel, D. Smith and T.N. Palmer, 2009: Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. Quart. J. Roy. Met. Soc. 135: 1538-1559.

Doblas-Reyes, F. J., A. Weisheimer, T.N. Palmer, J.M. Murphy and D. Smith, 2010: Forecast quality assessment of the ENSEMBLES seasonal-to-decadal Stream 2 hindcasts. ECMWF Tech Memo 621, 45pp, Reading UK

Fischer, E.M., D.M. Lawrence and B.M. Sanderson, 2011: Quantifying uncertainties in projections of extremes – a perturbed land surface experiment. Clim. Dyn. 37:1381-1398..

Forest, C.E., P.H. Stone, A.P. Sokolov, M.R. Allen and M.D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. Science 295: 113-117.

Friedlingstein, P., et al., 2006: Climate-carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. J Clim 19: 3337-3353.

Frame, D.J., T. Aina, C.M. Christensen, N.E. Faull, S.H.E. Knight, C. Piani, S.M. Rosier, K. Yamazaki, Y. Yamazaki and M.R. Allen, 2009: The climate*prediction*.net BBC climate change experiment: Design of the coupled model ensemble. Phil. Trans. R. Soc. A 367: 855-870.

Furrer, R., S.R. Sain, D. Nychka, and G.A. Meehl, 2007: Multivariate Bayesian analysis of atmosphere-ocean general circulation models. Env. Ecol. Stat. 14: 249-266.

Goldstein, M., and J. C. Rougier, 2004: Probabilistic formulations for transferring inferences from mathematical models to physical systems. SIAM J. Sci. Comput. 26: 467-487.

Greene, A.M., L. Goddard and U. Lall, 2006: Probabilistic multimodel regional temperature change projections. J. Clim. 19: 4326-4346.

Haerter, J.O., E. Roeckner, L. Tomassini and J.-S. von Storch, 2009: Parametric uncertainty effects on aerosol radiative forcing. Geophys. Res. Lett. 36: L15707.

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

Harris, G.R., D.M.H. Sexton, B.B.B. Booth, M. Collins, J.M. Murphy and M.J. Webb, 2006: Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. Clim. Dyn. 27: 357-375.

Harris, G.R., M. Collins, D.M.H. Sexton, J.M. Murphy and B.B.B. Booth, 2010: Probabilistic projections for 21^{st} century European climate. Nat. Hazards Earth Syst. Sci. 10: 2009-2020.

Hawkins, E. and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. Bull. Amer. Met. Soc. 90: 1095-1107.

Houtekamer, P.L., L. Lefaivre, J. Derome, H. Ritchie and H.L. Mitchell, 1996: A system simulation approach to ensemble prediction. Mon. Wea. Rev 124: 1225-1242.

Hurrell, J.H., G.A. Meehl, D. Bader, T.L. Delworth, B. Kirtman and B. Wielicki, 2009: A unified modelling approach to climate system prediction. Bull. Am. Met. Soc. 90: 1819-1832.

Jackson, C.S., M.K. Sen, G. Huerta, Y. Deng and K. Bowman, 2008: Error reduction and convergence in climate prediction. J. Clim. 21: 6698-6709.

Jackson, L.C., M. Vellinga and G.R. Harris, 2011: The sensitivity of the meridional overturning circulation to modelling uncertainty in a perturbed physics ensemble without flux adjustment. Clim. Dyn., Online first, doi:10.1007/s00382-011-1110-5.

Johnson, C., and R. Swinbank, 2009: Medium-range ensemble combination and calibration. Quart. J. Roy. Met. Soc. 135: 777-794.

Joshi, M.M., M.J. Webb, A.C. Maycock and M. Collins, 2010: Stratospheric water vapour and high climate sensitivity in a version of the HadSM3 climate model. Atmos. Chem. Phys. 10: 7161-7167.

Klocke, D., R. Pincus and J. Quaas, 2011: On constraining estimates of climate sensitivity with present-day observations through model weighting. J. Clim., early online release, doi:10.1175/2011JCLI4193.1

Knutti, R., T.F. Stocker, F. Joos and G. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature 416: 719-723.

Knutti, R., G.A. Meehl, M.R. Allen and D.A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. J. Clim. 19: 4224-4233.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak and G.A. Meehl, 2010a: Challenges in combining projections from multiple climate models. J. Clim. 23: 2739-2758.

Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P.J. Gleckler, B. Hewitson and L. Mearns, 2010b: Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections. In: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, and P.M. Midgley (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland.

Krishnamurti, T., C. Kishtawal, T. LaRow, D.R. Bachiochi, C. Williford, S. Gadgil and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. Science 13: 4196–4216.

Lohmann, U., and S. Ferrachat, 2010: Impact of parametric uncertainties on the present-day climate and on the anthropogenic aerosol effect. Atmos. Chem. Phys. 10: 11373-11383.

Lopez, A., C. Tebaldi, M. New, D.A. Stainforth, M.R. Allen and J.A. Kettleborough, 2006: Two approaches to quantifying uncertainty in global temperature changes. J. Clim. 19: 4785-4796.

Marsigli, C., 2009: COSMO Priority Project. Short-range ensemble prediction system (SREPS): Final report. COSMO Technical Report No. 13 (Ed. M. Milelli). Available from http://www.cosmo-model.org.

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

Meehl, G.A., T.F. Stocker, W.D. Collins, P. Friedlingstein, A.T. Gaye, J.M. Gregory, A. Kitoh, R. Knutti, J.M. Murphy, A. Noda, S.C.B. Raper, I.G. Watterson, A.J. Weaver and Z.-C. Zhao, 2007: Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Meehl, G. A., L. Goddard, J. Murphy, R. J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, and T. Stockdale, 2009: Decadal Prediction: Can it be skillful? Bull. Am. Meteorol. Soc. 90: 1467-1485.

Murphy, J.M., D.M.H. Sexton, D.N. Barnett, G.S. Jones, M.J. Webb, M. Collins and D.A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 429: 768-772.

Murphy, J.M., B.B.B. Booth, M. Collins, G.R. Harris, D.M.H. Sexton and M.J. Webb, 2007: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Phil. Trans. R. Soc. A 365: 1993-2028.

Murphy, J.M., D.M.H. Sexton, G.J. Jenkins, P.M. Boorman, B.B.B. Booth, C.C. Brown, R.T. Clark, M. Collins, G.R. Harris, E.J. Kendon, R.A. Betts, S.J. Brown, T.P. Howard, K.A. Humphrey, M.P. McCarthy, R.E. McDonald, A. Stephens, C. Wallace, R. Warren, R. Wilby and R.A. Wood, 2009a: UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter.

Murphy, J.M., M. Collins, F. Doblas-Reyes and T. Palmer, 2009b: Development of ensemble prediction systems [Research Theme 1]. In: van der Linden, P., and J.F.B. Mitchell (eds). 2009: ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK. 160pp.

Neelin, J.D., A. Bracco, H. Luo, J.C. McWilliams and J.E. Meyerson, 2010: Considerations for parameter optimization and sensitivity in climate models. PNAS 107: 21349-21354.

Niehörster, F., 2009: Uncertainty in feedback mechanisms in climate change projections. PhD Thesis, Freie Universität Berlin.

Palmer, T.N. and co-authors, 2004: Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER). Bull. Amer. Met. Soc. 85: 853-872.

Palmer, T.N. and co-authors, 2005: Representing model uncertainty in weather and climate prediction. Annu. Rev. Earth. Planet. Sci. 33: 4.1-4.31.

Palmer, T.N., F. Doblas-Reyes, A. Weisheimer, and M. Rodwell, 2008: Reliability of climate change projections of precipitation: Towards "seamless" climate predictions. Bull. Amer. Meteor. Soc. 89: 459-470, DOI: 10.1175/BAMS-89-4-459.

Piani, C., D.J. Frame, D.A. Stainforth, and M.R. Allen, 2005: Constraints on climate change from a multi-thousand member ensemble of simulations. Geophys. Res. Lett. 32: L23825.

Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, 2007: Climate models and their evaluation. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Rougier, J.C., 2007: Probabilistic inference for future climate using an ensemble of climate model evaluations. Clim. Change 81:247-264.

Rougier, J., D.M.H. Sexton, J.M. Murphy and D.A. Stainforth, 2009: Analyzing the climate sensitivity of the HadSM3 climate model using different but related experiments. J. Clim. 22: 3540-3557.

Rossow, W.B., and R.A. Schiffer, 1999: Advances in understanding clouds from ISCCP. Bull. Am. Met. Soc. 80: 2261-2287.

Rowell, D.P., 2011: Sources of uncertainty in future changes in local precipitation. Clim. Dyn., in press, doi: 10.1007/s00382-011-1210-2.

Sanderson, B.M., and co-authors, 2008a: Constraints on model response to greenhouse gas forcing and the role of sub-grid scale processes. J. Clim. 21: 2384-2400.

Sanderson, B.M., C. Piani, W. Ingram, D. Stone and M.R. Allen, 2008b: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. J. Clim. 21:2384-2400.

Sanderson, B.M., 2011: A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. J. Clim. 24: 1362-1377.

Sanderson, B.M., 2012: On the estimation of systematic error in regression-based predictions of climate sensitivity. Clim. Change, submitted.

Sexton, D. M. H., J. M. Murphy, M. Collins and M. J. Webb, 2011: Multivariate probabilistic projections using imperfect climate models Part I: Outline of methodology. Clim. Dyn., Online First, doi:10.1007/s00382-011-1208-9.

Sexton, D. M. H., and J.M. Murphy, 2011: Multivariate probabilistic projections using imperfect climate models Part II: Robustness of methodological choices and consequences for climate sensitivity. Clim. Dyn., Online First, doi:10.1007/s00382-011-1208-8.

Shiogama, H., M. Wanatabe, M. Yoshimori, T. Yokohata, T. Ogura, J. Annan, J. Hargreaves, M. Abe, Y. Kamae, R. O'ishi, R. Nobui, S. Emori, T. Nozawa, A. Abe-Ouchi and M. Kimoto, 2011: Physics parameter uncertainty and observational constraints of climate feedback: An ensemble of coupled atmosphere ocean GCMs without flux corrections. Clim. Dyn., submitted.

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R.A. Spicer, A. J. Thorpe and M. R. Allen, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433: 403-406.

Stott, P.A., and C.E. Forest, 2007: Ensemble climate predictions using climate models and observational constraints. Phil. Trans. R. Soc. A 365: 2029-2052.

Tebaldi, C., R.W. Smith, D. Nychka and L.O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multi-model ensembles. J. Clim. 18: 1524-1540.

Tebaldi, C. and R. Knutti, 2007: The use of multi-model ensembles in probabilistic climate projections. Phil. Trans. Roy. Soc. A 365: 2053-2076.

Watanabe, M., H. Shiogama, T. Yokohata, Y. Kamae, M. Yoshimori, T. Ogura, J.D. Annan, J.C. Hargreaves, S. Emori and M. Kimoto, 2011: Using a multi-physics ensemble for exploring diversity in cloud-shortwave feedback in GCMs. J. Clim., submitted.

Webb, M.J., et al., 2006: On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. Clim. Dyn 27: 17-38.

Weigel, A.P., R. Knutti, M.A. Liniger and C. Appenzeller, 2010: Risks of model weighting in multimodel climate projections. J. Clim. 23: 4175-4191.

Yokohata, T., M.J. Webb, M. Collins, K.D. Williams, M. Yoshimori, J.C. Hargreaves and J.D. Annan, 2010: Structural similarities and differences in climate responses to CO_2 increases between two perturbed physics ensembles. J. Clim. 23: 1392-1410.

Yokohata, T., J.D. Annan, M. Collins, C.S. Jackson, M. Tobis, M.J. Webb and J.C. Hargreaves, 2011: Reliability of multi-model and structurally different single-model ensembles. Clim. Dyn. Online first, doi 10.1007/s00382-011-1203-1.

Yoshimori, M., J.C. Hargreaves, J.D. Annan, and T. Yokohata, 2011: Dependency of feedbacks on forcing and climate state in physics parameter ensembles. J. Clim., Early online release, doi: 10.1175/2011JCLI3954.1.