

REVIEW

A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles

BY J. M. MURPHY*, B. B. B. BOOTH, M. COLLINS, G. R. HARRIS,
D. M. H. SEXTON AND M. J. WEBB

*Hadley Centre for Climate Prediction and Research, Met Office, Fitzroy Road,
Exeter EX1 3PB, UK*

A methodology is described for probabilistic predictions of future climate. This is based on a set of ensemble simulations of equilibrium and time-dependent changes, carried out by perturbing poorly constrained parameters controlling key physical and biogeochemical processes in the HadCM3 coupled ocean–atmosphere global climate model. These (ongoing) experiments allow quantification of the effects of earth system modelling uncertainties and internal climate variability on feedbacks likely to exert a significant influence on twenty-first century climate at large regional scales. A further ensemble of regional climate simulations at 25 km resolution is being produced for Europe, allowing the specification of probabilistic predictions at spatial scales required for studies of climate impacts. The ensemble simulations are processed using a set of statistical procedures, the centrepiece of which is a Bayesian statistical framework designed for use with complex but imperfect models. This supports the generation of probabilities constrained by a wide range of observational metrics, and also by expert-specified prior distributions defining the model parameter space. The Bayesian framework also accounts for additional uncertainty introduced by structural modelling errors, which are estimated using our ensembles to predict the results of alternative climate models containing different structural assumptions. This facilitates the generation of probabilistic predictions combining information from perturbed physics and multi-model ensemble simulations. The methodology makes extensive use of emulation and scaling techniques trained on climate model results. These are used to sample the equilibrium response to doubled carbon dioxide at any required point in the parameter space of surface and atmospheric processes, to sample time-dependent changes by combining this information with ensembles sampling uncertainties in the transient response of a wider set of earth system processes, and to sample changes at local scales.

The methodology is necessarily dependent on a number of expert choices, which are highlighted throughout the paper.

Keywords: perturbed physics ensembles; Earth system modelling;
Bayesian inference; pattern scaling; downscaling; HadCM3

* Author for correspondence (james.murphy@metoffice.gov.uk).

One contribution of 13 to a Theme Issue ‘Ensembles and probabilities: a new era in the prediction of climate change’.

1. Introduction

Complex three-dimensional global climate models (GCMs) are required to simulate the interactions between earth system processes which will determine the characteristics of future climate change (e.g. [Bony *et al.* 2006](#); [Friedlingstein *et al.* 2006](#)), including the multivariate information at fine spatial scales required for assessment of climate impacts. Policy makers and planners also need credible estimates of the uncertainty associated with model projections, in order to identify appropriate strategies for adaptation. A natural method of providing such estimates is to construct ensembles of simulations, which sample uncertainties arising from future emissions of radiative forcing agents (greenhouse gases (GHG) and aerosols), internal climate variability and modelling errors (these being an inevitable consequence of the discretization of the equations of motion on a grid of finite resolution, and of the need to parametrize the effects of unresolved processes). All three sources of uncertainty contribute significantly to the spread of possible changes, especially at the regional level ([Cubasch *et al.* 2001](#); [Räisänen 2001](#)).

Here we focus on the quantification of modelling errors and internal variability for an assumed pathway of future anthropogenic emissions ([Nakicenovic & Swart 2000](#)). The effects of internal variability can be estimated by varying the initial conditions in multiple simulations of a single climate model (e.g. [Selten *et al.* 2004](#)). In order to sample the effects of model error, it is necessary to construct ensembles which sample plausible alternative representations of earth system processes. Different climate models differ in both basic structural choices (such as choices of grid, resolution and numerical integration scheme, the set of processes included, fundamental assumptions informing the representation of physical processes, etc.), and the values of poorly constrained parameters controlling the outputs of their subgrid-scale parametrizations.

One approach to sampling modelling uncertainties consists of forming a multi-model ensemble (MME) from GCMs developed at different modelling centres around the world, whose results can be accessed from a central archive (e.g. [Cubasch *et al.* 2001](#); [Meehl *et al.* 2005a, 2007](#)). Another approach consists of exploring modelling uncertainties systematically within a single GCM, referred to hereafter as the ‘perturbed physics ensemble’¹ (PPE). One method of implementing a PPE is to identify a set of model variants which samples the settings of multiple uncertain parameters according to prior distributions estimated by experts based on their knowledge of the relevant physical processes ([Murphy *et al.* 2004](#); [Stainforth *et al.* 2005](#)). [Annan *et al.* \(2005a\)](#) describe another method in which an ensemble Kalman filter technique is used to converge on observationally constrained distributions for model parameters through repeat iterations of a cycle of short ensemble integrations and assimilation of observed datasets.

We focus on the PPE project currently in progress at the Hadley Centre, using a family of model configurations based on the HadCM3 coupled ocean–atmosphere GCM ([Gordon *et al.* 2000](#)). [Stainforth *et al.* \(2005\)](#) describe further PPE experiments using the same model family (see <http://www.climateprediction.net>). Here, we

¹We retain the term ‘PPE’ to refer to our approach, consistent with earlier papers. However, we stress that this paper describes extensions of the methodology to sample uncertainties in biogeochemical as well as physical earth system processes.

summarize developments to date in the Hadley Centre project and describe future plans to extend its scope to develop a methodology for the derivation of observationally constrained probability density functions (PDFs) of regional climate change. The aim is to produce a system appropriate for the estimation of joint PDFs of key climate variables at spatial scales of approximately 25 km, for use in assessments of regional climate impacts during the twenty-first century. Section 2 briefly compares the perturbed physics and the multi-model methods of sampling modelling uncertainties. In §3, we describe our methodology. This consists of a set of PPE experiments sampling key parameters in the earth system modules considered likely to contribute the largest uncertainties to twenty-first century climate change (atmosphere, ocean, land surface, terrestrial carbon cycle and atmospheric sulphur cycle), a further PPE of regional climate model simulations for downscaling and a set of statistical procedures required to derive PDFs from the PPE results. Detailed descriptions of these elements are left to supporting papers published already (Murphy *et al.* 2004; Collins *et al.* 2006; Harris *et al.* 2006; Webb *et al.* 2006; Rougier 2007) or forthcoming (e.g. Sexton *et al.* in preparation *a, b*). A summary is provided in §4 including a discussion of prospects for reducing uncertainties in future ensemble prediction exercises.

2. Multi-model and perturbed physics approaches to the quantification of modelling uncertainties

Results from MMEs have been used extensively to characterize the spread of GCM climate change projections for a number of different climate variables (e.g. Räisänen 2001, 2002, 2003, 2005; Palmer & Räisänen 2002; Weisheimer & Palmer 2005; Good & Lowe 2006; Meehl *et al.* 2005*b*; Tebaldi *et al.* 2006). 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 climate simulation (e.g. Blackmon *et al.* 2001; Anderson *et al.* 2004; Johns *et al.* 2006). 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 (typically 10–20 members), and are not designed to sample modelling uncertainties in a systematic fashion, being assembled on an opportunity basis from available models. Specifically, it is not clear how to define a space of possible model configurations of which the MME members are a sample. This creates the need to make substantial assumptions in order to obtain probabilistic predictions from their results (e.g. Dessai *et al.* 2005; Greene *et al.* 2006; Räisänen & Ruokolainen 2006). Tebaldi *et al.* (2005) and Furrer *et al.* (in press) describe Bayesian methods of converting MME results into probabilistic estimates of future changes in which each model is assumed to differ independently and randomly from the true climate. However, this assumption can be questioned on the basis that models show common systematic biases in their simulations of recent climate (Lambert & Boer 2001). Some studies assume all ensemble members to be equally plausible, while other studies exclude or down-weight MME members giving extreme changes (Räisänen 2001; Giorgi & Mearns 2003; Dessai *et al.* 2005; Tebaldi *et al.* 2005). In the latter case, the premise is that outliers are likely to be less credible

and would lead to overdispersive uncertainty estimates if included. However, [Allen & Ingram \(2002\)](#) argue that MMEs could be underdispersive rather than overdispersive, because they are not explicitly designed to sample the range of future responses consistent with recent historical observations. Such issues are inevitably open to debate, but it is clear that probabilistic estimates derived from MMEs are significantly dependent on methodological choices necessitated by the nature of the ensembles ([Lopez et al. 2006](#)).

The PPE method represents an attempt to develop a systematic method of sampling modelling uncertainties within a single model framework ([Murphy et al. 2004](#); [Stainforth et al. 2005](#); [Collins et al. 2006](#); [Webb et al. 2006](#)). Its main advantage is that it allows greater control over the design of experiments to sample process uncertainties. For example, it is possible to define a space of possible model configurations, by asking parametrization experts to estimate prior distributions for the set of model parameters to be perturbed. The definition of the model parameter space is subjective, and can therefore be questioned (for example, [Allen et al. \(2006\)](#) point out the risk that expert priors might double count some of the observational evidence later used to constrain the PPE simulations). However, the sensitivity to alternative choices of prior can be explored ([Murphy et al. 2004](#); [Rougier et al. submitted](#)). Although PPEs and MMEs are both characterized as methods of sampling modelling uncertainty, the spread of responses found in them also contains a component arising from internal climate variability. In experiments carried out to date, model uncertainty is the dominant driver of spread in simulated surface temperature changes (especially at continental to global scales), whereas internal variability plays a larger and sometimes comparable role at regional scales, particularly for circulation and precipitation changes ([Räisänen 2001](#); [Murphy et al. 2004](#)).

The main limitation of PPEs is that they do not sample major structural perturbations to the model physics. For example, in PPE simulations carried out using the atmosphere/mixed layer ocean configuration of HadCM3, some variables, such as cloud types in the middle troposphere, show large systematic biases which cannot be resolved by varying its uncertain parameters (see fig. 4 of [Murphy et al. 2004](#)). The PDFs obtained to date from PPE experiments ([Murphy et al. 2004](#); [Piani et al. 2005](#); [Knutti et al. 2006](#)) would be wider, were structural model errors to be accounted for. The methodology of §3 therefore includes plans to inflate PDFs obtained from PPE experiments to account for the estimated impact of structural modelling errors. For this purpose, we adopt a specific definition of structural error as the difference between simulated and observed climate which remains once that error has been minimized by finding the optimum location in the model parameter space.

3. Methodology for probabilistic prediction of climate change

In this section, we describe our methodology, summarize key results obtained to date and emphasize the assumptions and caveats associated with our approach.

Briefly, a large PPE of simulations of the equilibrium response to doubled CO₂ (Box A in [figure 1](#)) is used in conjunction with a statistical emulation technique (Box B) to generate a distribution of responses consistent with expert-specified prior distributions for uncertain parameters controlling the simulation of surface

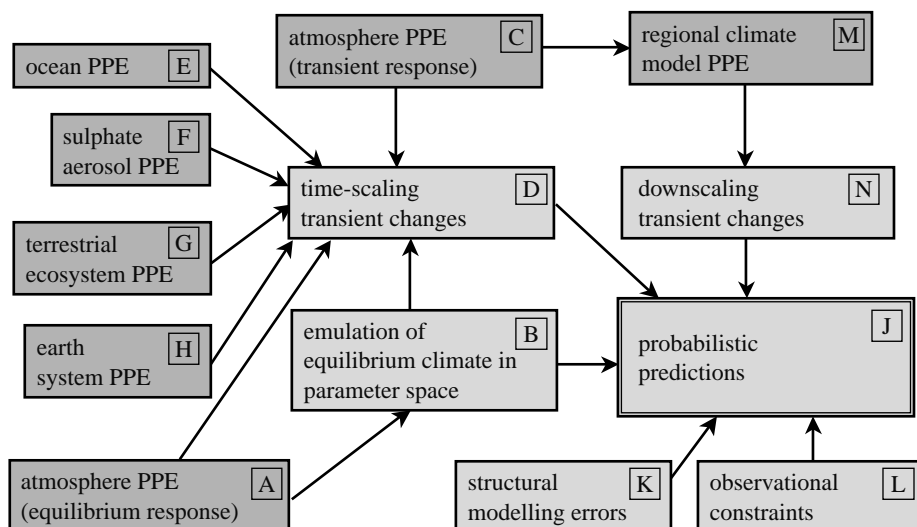


Figure 1. Constituents of a methodology for the production of probabilistic predictions of future climate change from PPE simulations using various configurations of the HadCM3 climate model. Boxes with a dark grey background denote ensemble simulation experiments, and boxes with a light grey background denote constituents of a suite of statistical techniques required to convert the model simulations into probabilistic predictions. The ‘atmosphere PPE (equilibrium response)’ box denotes simulations of the equilibrium response to doubled CO_2 carried out using the coupled atmosphere/mixed layer ocean (HadSM3) configuration of the model. The other dark grey boxes all denote simulations of transient climate change in which the atmosphere is coupled to a three-dimensional dynamical ocean component, sampling uncertain parameters in the earth system modules indicated. The ‘earth system PPE’ denotes an ensemble in which parameters in the atmosphere, ocean, sulphate aerosol and terrestrial ecosystem modules are perturbed simultaneously. The arrows indicate major linkages between the different elements, which are explained in more detail in §3.

and atmospheric processes in HadCM3. These results are then combined with a smaller PPE of simulations of the transient response (Box C) using a statistical methodology based on a pattern scaling approach (Box D), producing distributions of the transient response sampling the effects of internal climate variability and uncertainties in parameters. We plan to generalize these distributions to sample uncertainties in a wider range of earth system processes, using further PPE experiments designed to quantify the impacts of uncertain parameters in the ocean, sulphate aerosol and terrestrial carbon cycle modules available as part of the family of HadCM3 configurations (Boxes E–G). This may include a PPE in which parameters in all modules are perturbed simultaneously, in order to sample nonlinear interactions between feedbacks in different earth system components (Box H). These results will then be converted into probabilistic predictions using a Bayesian statistical framework (Box J). This framework (Rougier 2007) is based on the notion of a grand PDF representing the joint probabilities of all the uncertain elements of the problem, including climate model parameters, simulated historical and future values of model variables, structural modelling errors (Box K) and observed historical values of some chosen subset of variables (Box L). In order to apply the framework in practice, it will be necessary to supply quantitative estimates for each of these

elements and their associated uncertainties. Once this is done, it will become possible to sample this grand PDF to produce marginal PDFs of future values for individual climate variables or joint PDFs for several future variables, weighted according to the relative likelihood of alternative model versions within parameter space. Typical examples could include PDFs of multi-year means of surface temperature or precipitation change spatially averaged over sub-continental regions (such as those considered by [Harris *et al.* \(2006\)](#)) or joint PDFs of changes in (say) temperature and precipitation, or precipitation and circulation, to facilitate studies of the risks and impacts associated with twenty-first century climate change.

Boxes A–L constitute a methodology for the production of PDFs of variables at spatial scales skilfully resolved by HadCM3, which possesses a horizontal grid spacing of approximately 300 km. For Europe, predictions will be ‘downscaled’ to a resolution of 25 km by running a PPE of simulations with the regional configuration of HadCM3 (Box M), driven by time series of surface and lateral boundary conditions saved from the transient simulations of Box C. These simulations will be used to build statistical relationships between the climate changes simulated by the global and regional configurations of the model (Box N), which can then be used to estimate changes at high resolution from parts of the parameter space for which no regional simulation is available, and hence provide PDFs of changes at spatial scales required for more detailed studies of climate impacts (§3*k*).

The constituents of [figure 1](#) are described in more detail below. Most of the work in §3*a–d* is either published or completed and awaiting publication. The material in §3*e–k* is work in progress.

(*a*) *Simulations of the equilibrium response to doubled CO₂ (Box A)*

Given sufficient computing capacity, the ideal PPE experiment would consist of very large ensemble simulations of time-dependent climate change, using very high resolution coupled ocean–atmosphere GCMs in which values of multiple model parameters are simultaneously varied to quantify the interactions between uncertain earth system processes. This is not feasible, even at the limited resolution of current GCMs, so our large ensembles consist of simulations of present-day and doubled CO₂ climates, run to equilibrium using the cheaper coupled atmosphere-mixed layer ocean version of HadCM3, hereafter HadSM3. Each simulation contains alternative settings for uncertain parameters, controlling key processes in the HadSM3 parametrizations of large-scale cloud, convection, radiative transfer, sea ice, surface and boundary layer processes and dynamical transports ([Barnett *et al.* 2006](#)). Most parameters were continuous variables for which plausible minimum, standard and maximum values were obtained by consultation with experts, and others were factors taking a discrete set of values (usually two, representing a choice between activating or not activating an additional process, or an alternative method of representing a process). In total, the set of parameters constitutes a 31-dimensional space, sampling of which can provide, in principle, quantification of the range of climate changes consistent with *a priori* expert understanding of the physical processes included in the model. [Murphy *et al.* \(2004\)](#) carried out an initial ensemble of 53 members in which one parameter was perturbed at a time. This has since been

augmented by a second ensemble of 128 members containing multiple parameter perturbations chosen to sample a wide range of climate sensitivities, achieve skilful simulations of present climate and maximize coverage of parameter space (details in [Webb *et al.* \(2006\)](#)). This ensemble accounts for nonlinear interactions between parameters, the importance of which was demonstrated by the *climateprediction.net* PPE of [Stainforth *et al.* \(2005\)](#). These studies yielded key early results supporting the use of PPE simulations in climate prediction.

- The ensembles give ranges of global climate sensitivity ([Murphy *et al.* 2004](#); [Stainforth *et al.* 2005](#)) and cloud feedback ([Webb *et al.* 2006](#)), which either match or exceed those obtained from MME results. While such comparisons should not be regarded as formal performance criteria (since MME ranges are themselves only estimates of uncertainty; e.g. [Allen & Ingram 2002](#)), they do provide evidence that perturbing parameters in a single model samples uncertainties in large-scale climate system processes to a level consistent with current modelling approaches available worldwide.
- The PPE results simulate a wider range of regional changes than can be inferred by scaling a single pattern of change by the global mean temperature response ([Murphy *et al.* 2004](#)), and also simulate a substantial range of changes in climate extremes ([Barnett *et al.* 2006](#); [Clark *et al.* 2006](#)), demonstrating the importance of using large GCM-based ensembles to obtain realistic estimates of uncertainty at regional scales.

The ensembles of [Murphy *et al.* \(2004\)](#) and [Webb *et al.* \(2006\)](#) are being augmented by further HadSM3 simulations designed to improve the sampling of parameter space. [Rougier *et al.* \(submitted\)](#) report climate sensitivity values from a total of 297 simulations, and more are in progress. This set of experiments forms the bedrock of our probabilistic prediction methodology.

(b) *Emulation of the equilibrium response to doubled CO₂ (Box B)*

Given the large dimensionality of the HadSM3 parameter space, it is not feasible to sample it comprehensively through GCM simulations alone, even if the numbers of simulations run into hundreds (as above) or even thousands ([Stainforth *et al.* 2005](#)). A method of estimating the results of the climate model for any desired combination of parameter settings is therefore required ([Rougier & Sexton 2007](#)), in order to build up distributions of estimated climate states consistent with specified prior distributions for the input parameters. We plan to achieve this by developing a multivariate statistical emulator of the climate model, trained on the available ensemble runs. The emulator, described in detail by [Sexton *et al.* \(in preparation a\)](#), will provide best estimate values and associated errors for a range of climate variables expressed as spatial fields of present-day climate means (in order to facilitate the calculation of relative likelihood for different locations in parameter space) and the equilibrium response to doubled CO₂. A stepwise linear algorithm will be used to select potential regressors consisting of functions of the GCM parameters, including terms to capture nonlinear interactions between pairs of parameters ([Stainforth *et al.* 2005](#); [Rougier *et al.* submitted](#)). The role of the emulator in our Bayesian statistical framework is outlined in §3f.

(c) *Simulations of transient climate change sampling uncertainties in surface and atmospheric processes (Box C)*

Collins *et al.* (2006) describe the extension of the PPE approach to the simulation of transient climate change, using perturbed versions of HadCM3 in which the atmosphere is coupled to a dynamical, three-dimensional ocean module. They considered an idealized forcing scenario of CO₂ increasing at 1% per year, documenting the properties of a 17-member PPE sampling multiple perturbations to the set of surface and atmospheric parameters described above. One member used standard parameter settings, with a further 16 members sub-selected from the 128 perturbed members of Webb *et al.* (2006) using a procedure designed to yield the best simulations of present climate while sampling a wide range of climate sensitivity values. The results demonstrated that multiple locations in parameter space can be identified which yield simulations of present climate of comparable quality to the standard published version of HadCM3 (Gordon *et al.* 2000), while predicting a range of transient climate responses similar to that of the latest MME experiment, generated for the fourth assessment report of the Intergovernmental Panel on Climate Change (Meehl *et al.* 2007). Collins *et al.* used flux adjustments to limit simulation biases in sea surface temperature and salinity in their PPE. This was done partly to reduce regional systematic errors (and hence increase the credibility of predicted regional climate changes) and partly to allow the effects of parameter perturbations on the transient response to be explored without being excessively constrained by the need to achieve precise balance in the planetary radiation budget (see further discussion in Collins *et al.*, also §3*h*).

A new 17-member PPE of HadCM3 transient simulations (denoted by PPE_A1B) has recently been produced, in order to provide user-relevant predictions of twenty-first century climate change. These are driven from 1860 to 2000 by historical time series of well-mixed GHGs, ozone, anthropogenic sulphur emissions and reconstructions of variations in solar activity and volcanic aerosol, and from 2000 to 2100 by future GHG concentrations and sulphur emissions from the SRES A1B scenario (Nakicenovic & Swart 2000). The configuration of HadCM3 for these experiments is as described by Gordon *et al.* (2000), except that the representation of the atmospheric sulphur cycle has been upgraded to use the fully interactive module of Jones *et al.* (2001), thus avoiding the need to approximate the effect of sulphate aerosol on cloud albedo using an off-line calculation (Johns *et al.* 2003). One ensemble member uses standard parameter settings, with 16 perturbed members chosen from the parameter combinations used by Webb *et al.* (2006). These were selected by choosing the member with the best simulation of present-day climate, followed by 15 further members designed to maximize a non-dimensional measure of the average distance between members in terms of both climate sensitivity and model parameter values. This procedure differs in detail from that used by Collins *et al.* (2006) and gives a different selection of parameter combinations sampling a somewhat wider range of climate sensitivities. Flux adjustment terms were calculated from preliminary simulations in which sea surface temperatures and salinities were relaxed towards observed climatological values, using a longer relaxation time constant than Collins *et al.* (2006). This change significantly reduced the Northern Hemisphere sea surface temperature and sea ice biases found in the flux-adjusted simulations of Collins *et al.*

(d) 'Time scaling' of transient changes (Box D)

In order to quantify uncertainties in transient climate change, a method is required for combining information from the equilibrium and transient simulations described above. Harris *et al.* (2006) achieve this by scaling equilibrium response patterns from the 128-member Webb *et al.* (2006) ensemble to augment the 17 Collins *et al.* (2006) transient simulations and hence obtain frequency distributions of transient regional surface temperature and precipitation changes. Here we refer to their technique as 'time scaling'. It is achieved by scaling the equilibrium spatial response patterns ($\Delta P_{\text{eqm}}(s)$) of each Webb *et al.* ensemble member according to $\Delta T(t)$, the transient response of global mean surface temperature predicted by a simple climate model (SCM) tuned to the climate sensitivity of the relevant ensemble member. The SCM is based on that of Rowntree (1998) and predicts land and ocean surface temperatures in response to imposed radiative forcing anomalies, representing vertical heat transfer in the ocean via a globally averaged heat diffusion equation, modified to include upwelling and downwelling following Schlesinger *et al.* (1997). The equation used by Harris *et al.* can be written as

$$\Delta P(s, t) = \Delta T(t)[\Delta P_{\text{eqm}}(s) + \text{correction}(s)] + \text{error}(s, t), \quad (3.1)$$

where $\Delta P(s, t)$ is the inferred transient regional response for some region s at time t , and $\text{correction}(s)$ is a pattern representing the difference between scaled equilibrium and transient response patterns arising from the spatially varying impacts of oceanic thermal inertia and circulation changes, captured in the HadCM3 transient simulations but not in the HadSM3 equilibrium simulations. The $\text{error}(s, t)$ term captures the distribution of pattern scaling error, arising from uncertainty in the true response patterns introduced by internal variability, plus errors associated with the assumption in equation (3.1) that the transient response is linearly related to global temperature, and errors in SCM predictions of the GCM global temperature response. The $\text{correction}(s)$ term, and the mean and variance² of $\text{error}(s, t)$, are quantified by comparing the responses of the 17 Collins *et al.* transient simulations against the scaled equilibrium responses of corresponding members of the Webb *et al.* ensemble. Owing to the limited number of verifying transient simulations, Harris *et al.* assumed $\text{correction}(s)$ to be invariant across the model parameter space,³ which introduces a further contribution to $\text{error}(s, t)$. Harris *et al.* construct frequency distributions of 20-year mean regional changes in response to a 1% per year increase in CO₂ by predicting 128 realizations of $\Delta P(s, t)$ from equation (3.1) and adding noise by sampling $\text{error}(s, t)$. For surface temperature, they find that the mean and spread of the regional frequency distributions are explained mainly by the distribution of PPE responses in $\Delta P_{\text{eqm}}(s)$ plus the impact of internal climate variability. For precipitation, larger differences between the transient and equilibrium response

²The methodology described in this paper will require the specification of distributions for terms representing errors of various kinds, or values of SCM parameters. For the sake of tractability, it will be assumed that the error populations are Gaussian. This assumption will be tested, and variables will be transformed where necessary (and feasible) to ensure that it is valid.

³Subsequent investigations using our latest experiments suggest that $\text{correction}(s)$ may be significantly related to climate sensitivity. This will be incorporated by modifying $\text{correction}(s)$ according to sensitivity in the time-scaling equation.

patterns, enhanced by nonlinear dependencies on $\Delta T(t)$ in some regions, lead to a larger role for the pattern scaling error terms, consistent with previous work (e.g. Mitchell 2003).

In order to obtain *probability* distributions of transient changes (as opposed to the *frequency* distributions described by Harris *et al.*), it will be necessary to obtain time-scaled responses sampling the entire parameter space of surface and atmospheric processes, consistent with the prior distributions specified by experts. This will be achieved by applying the time-scaling technique to a large Monte Carlo sample of patterns $\Delta P_{\text{eqm}}(s)$ generated by statistical emulation of the equilibrium response of HadSM3 throughout its parameter space (§3*b*). The time-scaling technique will then be calibrated using the PPE_A1B ensemble (§3*c*), in order to generate a large sample of transient responses from 1860 to 2100.

In PPE_A1B, spatial heterogeneities in the patterns of external forcing arising from agents other than well-mixed GHGs will give rise to response patterns different from $\Delta P_{\text{eqm}}(s)$. The Harris *et al.* time-scaling technique will therefore be modified to include additional terms, $\Delta T_{\text{nohgh}}(t)$ and $\Delta P_{\text{nohgh}}(s, t)$, representing the global mean and spatial pattern of the net response to these forcings, the largest component of which is likely to arise from the effects of anthropogenic sulphate aerosols (Tett *et al.* 2002). The time-scaling equation then becomes

$$\begin{aligned} \Delta P(s, t) = & \Delta T_{\text{ghg}}(t) \cdot \Delta P_{\text{eqm}}(s) + \Delta T_{\text{nohgh}}(t) \cdot \Delta P_{\text{nohgh}}(s, t) \\ & + \Delta T_{\text{all}}(t) \cdot \text{correction}(s) + \text{error}(s, t). \end{aligned} \quad (3.2)$$

The ΔP_{nohgh} term includes a time dependence to account for changes in the spatial patterns expected to arise from changes in the distribution of sulphur emissions between the historical and future periods of the simulations. This term will be estimated from a further ensemble (PPE_A1B_NOGHG), identical to PPE_A1B except that concentrations of well-mixed GHGs are held fixed at pre-industrial levels, allowing the transient response to the additional forcings to be isolated.

When predicting the transient climate response consistent with some set of HadCM3 parameter settings, it will be necessary to run the SCM once to predict $\Delta T_{\text{ghg}}(t)$ in response to changes in the forcing from well-mixed GHGs, a second time to predict $\Delta T_{\text{nohgh}}(t)$ in response to changes in the other forcings, and then a third time to predict $\Delta T_{\text{all}}(t)$, the response to the full set of forcings.

(e) *Sampling uncertainties in a wider range of earth system processes*
(Boxes D–H)

Uncertainties in surface and atmospheric physical processes (Bony *et al.* 2006; Soden & Held 2006; Webb *et al.* 2006) are the major source of uncertainty in equilibrium climate sensitivity and can therefore be expected to contribute substantially to the uncertainty in the transient climate response (Raper *et al.* 2002). The PPE simulations described above sample uncertainties in these processes, as represented in HadCM3. Here we identify a set of additional earth system processes likely to contribute significantly to the uncertainty in twenty-first century climate change, and which can be investigated within the HadCM3 model framework. These include ocean heat uptake (e.g. Raper *et al.* 2002), terrestrial carbon cycle feedbacks (Friedlingstein *et al.* 2006) and radiative forcing by aerosols (Anderson *et al.* 2003). Additional PPE experiments to

sample uncertainties in these processes are described below, by perturbing parameters in component modules of HadCM3 representing ocean transport processes (Gordon *et al.* 2000), terrestrial ecosystem processes (Cox 2001) and the radiative forcing due to sulphate aerosols (Jones *et al.* 2001). It will also be necessary to modify further the time-scaling technique, as outlined in §3e(i), so that uncertainty estimates diagnosed from these additional ensembles can be included in our PDFs of transient changes.

(i) *Generalization of time-scaling method to include earth system uncertainties (Box D)*

In principle, the range of responses simulated in the earth system ensemble experiments will each increase the total uncertainty associated with both global mean and regional climate change. In order to include the additional uncertainty in global mean changes, the SCM used in the time-scaling technique will be augmented to include simple globally averaged parametrizations of processes associated with the conversion of sulphur emissions into sulphate aerosol forcing, and the effects of terrestrial carbon cycle feedbacks on the atmospheric CO₂ concentration. The SCM coefficient (or set of coefficients) controlling these processes, denoted by *aer* and *carb*, respectively, then join those determining the globally averaged climate sensitivity over ocean and land (λ_o and λ_l), and the efficiency of heat uptake by the ocean (κ), to form a collection $\zeta = \{\lambda_o, \lambda_l, \kappa, aer, carb\}$ which can be varied to sample uncertainties in global mean transient changes implied by the various PPE experiments carried out with the GCM.

It is worth emphasizing a key difference between the methods which will be used to sample uncertainties in the components of ζ . The availability of a large PPE of simulations allows us to estimate the equilibrium response to doubled CO₂ in any part of the model parameter space using statistical emulation (§3b). It will therefore be possible to estimate distributions of λ_o and λ_l consistent with expert-specified distributions of climate model parameters and the detailed interactions between the processes they control. Ideally the same approach would be used to obtain distributions for κ , *aer* and *carb*; however, the relevant PPE ensembles (see below) will be too small to support the generation of emulated responses in regions of parameter space not sampled by a model simulation. We will therefore take a simpler approach of using the spread of PPE responses to estimate the mean and variance of κ , *aer* and *carb*. Ensemble predictions of time-dependent global mean temperature rise will then be built up by the following procedure.

- (i) Sample the HadSM3 parameter space to draw a large ensemble of $O(10^4)$ paired values for λ_o and λ_l , obtained by statistical emulation of the climate model response (Box B).
- (ii) For each draw of λ_o and λ_l , run the SCM several times, sampling alternative values of κ , *aer* and *carb*, and thus increasing the spread of time-scaled global temperature responses.

In step (ii), the distributions of the coefficients κ , *aer* and *carb* will be assumed independent of one another, and independent of λ_o and λ_l , because the design of our ensemble simulations (Boxes E–G) will not allow us to sample covariations

between them.⁴ In the case of carbon cycle feedbacks, the functional form of the SCM parametrization will nevertheless contain explicit temperature dependences, allowing the effects of variations in the global temperature response on the global mean carbon cycle response to be captured (Andreae *et al.* 2005). This will be achieved using globally averaged calculations of changes to the vegetation and soil carbon stores consistent with the main features of the corresponding calculations used in the terrestrial ecosystem module of HadCM3 (Jones *et al.* 2003a), which contains temperature-dependent parametrizations of photosynthesis and plant and soil respiration.

The PPE experiments sampling ocean, sulphur cycle and terrestrial ecosystem feedbacks will be examined to quantify their impact on spatial patterns of climate change. Where the uncertainties in these feedbacks are found to indicate a significant impact on the mean and spread of regional responses, exceeding that consistent with internal climate variability, an additional contribution $\text{extra}(s, t)$ will be added to the time-scaling equation, the full version of which can now be defined as

$$\begin{aligned} \Delta P(s, t) = & \Delta T_{\text{ghg}}(t; \zeta) \cdot \Delta P_{\text{eqm}}(s) + \Delta T_{\text{noghg}}(t; \zeta) \cdot \Delta P_{\text{noghg}}(s, t) \\ & + \Delta T_{\text{all}}(t; \zeta) \cdot \text{correction}(s) + \text{extra}(s, t) + \text{error}(s, t). \end{aligned} \quad (3.3)$$

The time- and space-dependent means and variances of $\text{extra}(s, t)$ will be estimated by adding separate contributions from the ocean, sulphur cycle and terrestrial ecosystem PPE experiments (assumed independent of one another). When running the SCM to predict changes for some specific region s , alternative values of $\text{extra}(s, t)$ will be sampled by random draws from the relevant time-dependent distributions, assumed independent of the SCM parameters controlling its global mean temperature (ζ).

If resources allow, an additional ensemble PPE_A1B_ESYS will be run in which uncertain parameters in the atmosphere, ocean, sulphur cycle and terrestrial ecosystem components are perturbed together (Box H, described in §3e(v)). If produced, PPE_A1B_ESYS will be used to estimate the time-dependent mean and variance of the $\text{error}(s, t)$ term in equation (3.3) through a validation exercise similar to Harris *et al.* (2006), in which time-scaled estimates of regional transient responses $\Delta P(s, t)$ (obtained from the first four terms of equation (3.3)) are compared to responses simulated by members of PPE_A1B_ESYS. This would allow the distribution of $\text{error}(s, t)$ to include contributions from nonlinear interactions between feedbacks in different earth system modules not accounted for in the other terms of equation (3.3). If resource limitations prevent production of PPE_A1B_ESYS, then the validation exercise will be done using the PPE_A1B ensemble instead, in which case $\text{error}(s, t)$ would not account for these missing nonlinear interactions, though other sources of time-scaling error (§3d) would still be accounted for.

(ii) *Ocean processes (Box E)*

Preliminary simulations have been run with perturbations to parameters controlling three ocean processes which influence the vertical transfer of heat (diffusion in the vertical, diffusion along isopycnal surfaces and the depth profile

⁴One exception is that the ensemble PPE_A1B_NOGHG has allowed us to diagnose a significant relationship between sulphate aerosol forcing and climate sensitivity, which will be included in step (ii).

of wind mixing energy in the surface mixed layer). These show a limited sensitivity of the rate of global mean transient climate change to those processes (Collins *et al.* 2007). A more comprehensive study is now underway, consisting of a 16-member ensemble (PPE_A1B_OCEAN), in which simultaneous perturbations are made to a wider range of parameters in the ocean module of HadCM3, controlling various aspects of the resolved and subgrid-scale transports of heat, salt and momentum in both the horizontal and vertical. In this ensemble, atmosphere module parameters are held fixed at the values used in the standard version of HadCM3 (Gordon *et al.* 2000), flux adjustments are applied and the same external forcings as PPE_A1B are used to simulate transient climate change from 1860 to 2100. Parameter values for the ensemble runs were determined using a Latin Hypercube methodology (McKay *et al.* 1979) designed to sample parameter space as efficiently as possible, given the limited number of ensemble members. This ensemble will be used to estimate uncertainties in both global mean ocean heat uptake, and the effects of ocean transport on regional changes (Boer & Yu 2003), which will be introduced into the final PDFs of transient climate response via equation (3.3) (§3*e*(i)).

(iii) *Radiative forcing due to sulphate aerosols (Box F)*

The interactive sulphate aerosol module of HadCM3 is similar to that described by Jones *et al.* (2001). It simulates sulphate aerosol concentrations from prescribed emission fields of anthropogenic sulphur dioxide (SO₂), natural dimethyl sulphide and tropospheric sulphur arising from quasi-regular volcanic eruptions. Three ‘modes’ of aerosol are represented, comprising sulphate dissolved in cloud droplets plus two free particle modes assumed to possess log-normal size distributions: the Aitken mode is produced by gas-phase oxidation of SO₂ and the larger accumulation mode results mainly from cloud processing via the dissolved mode. The model simulates production of sulphate by oxidation of SO₂; transport via horizontal and vertical advection, convection and turbulent mixing; dry deposition in the boundary layer; wet scavenging (rain out); transfers between the sulphate modes resulting from evaporation of dissolved sulphate to form accumulation mode, nucleation of cloud droplets on accumulation mode sulphate to form dissolved sulphate and diffusion of Aitken mode sulphate into cloud droplets to form dissolved sulphate. The atmospheric sulphur burden affects radiation via the direct (cooling) influence of scattering and absorption of incoming solar radiation, and through increases in cloud albedo resulting from the action of sulphate aerosols as cloud condensation nuclei (generally referred to as ‘the first indirect effect’). The ‘second indirect effect’, in which reductions in cloud droplet size reduce precipitation efficiency and increase cloud lifetime, is not included since the calculation of precipitation in HadCM3 does not allow for any dependence on cloud droplet number concentrations.

A 16-member ensemble of HadCM3 simulations (PPE_A1B_SULPHATE) has recently been completed, in which simultaneous perturbations are made to parameters controlling some of the processes outlined above, and to emissions of aerosol precursors. Parameter ranges are based on expert advice and are sampled using a Latin Hypercube design. Ensemble members are forced from 1860 to 2100, using the same specifications of forcing agents as PPE_A1B and PPE_A1B_OCEAN. Flux adjustments are applied, and all ensemble members use the settings

for atmosphere and ocean module parameters employed in the standard version of HadCM3. These experiments will be used to estimate uncertainties in the globally averaged forcing due to sulphate aerosols, and in the regional climate response resulting from uncertainties in both global and regional forcing. These will be introduced into the time-scaling methodology as described in §3*e*(i).

An important role of the PPE_A1B_SULPHATE ensemble is to sample the uncertainty in the historical response of climate from the pre-industrial period to present day. This is because the uncertainties in past aerosol forcing are substantial, implying that model versions with relatively high climate sensitivities can be shown to be consistent with the observed record, hence increasing the range of future responses consistent with past changes (Andreae *et al.* 2005). This is an important consideration, given that historical trends in large-scale averages of surface temperature and ocean heat uptake provide a key observational constraint on future changes (e.g. Allen *et al.* 2000; Stott & Kettleborough 2002; Frame *et al.* 2005; Stott *et al.* 2006*a*). The use of observational constraints in the present study (Box L) is discussed below.

(iv) *Terrestrial ecosystem processes (Box G)*

Uncertainties in terrestrial ecosystem processes will be sampled by adding the TRIFFID dynamic vegetation module of Cox (2001) to HadCM3. TRIFFID simulates soil carbon, and the growth and replacement of five functional types of vegetation (broadleaf tree, needleleaf tree, C3 grass, C4 grass and shrubs). The functional types vary according to the net available carbon and competition between plant types, parametrized using empirical relationships. Soil carbon can be increased by litterfall and is returned to the atmosphere by microbial respiration, which depends on temperature and soil moisture. CO₂ fluxes at the land–atmosphere interface are determined by photosynthesis and plant and microbial respiration. Grid box surface fluxes are obtained by aggregating values from tiles representing the time-varying fraction of each vegetation type within each model grid box (Essery *et al.* 2003),⁵ using values of surface characteristics (albedo, roughness length, etc.) determined from the state of the vegetation. In order to simulate carbon fluxes at the ocean–atmosphere interface, an ocean carbon cycle module (Cox *et al.* 2001) is also included in HadCM3. This simulates exchange of gaseous CO₂ with the atmosphere, the transport of dissolved inorganic carbon and cycling of carbon by marine biota via a nutrient–phytoplankton–zooplankton–detritus ecosystem module (Palmer & Totterdell 2001) that simulates the effects of light penetration, alkalinity and nutrient availability on biological carbon uptake. In previous carbon cycle experiments using HadCM3 (e.g. Cox *et al.* 2000; Jones *et al.* 2003*a*), the horizontal resolution of the ocean module was reduced; however, here we shall maintain the standard resolution of 1.25° × 1.25° in order to ensure that our carbon cycle simulations are physically consistent with the other coupled ocean–atmosphere ensembles included in our methodology.

⁵This differs from the earlier MOSES land surface module of Cox *et al.* (1999) used in the PPE_A1B, PPE_A1B_OCEAN and PPE_A1B_SULPHATE ensembles, in that MOSES calculates grid box surface fluxes from prescribed and aggregated vegetation characteristics. Any systematic offset between these alternative configurations of HadCM3 will be accounted for when combining results from ensembles with and without interactive vegetation processes in equation (3.3).

In order to estimate the effects of uncertainties in terrestrial ecosystem feedbacks on climate change, a 16-member ensemble (PPE_A1B_VEG) will be produced, sampling simultaneous perturbations to TRIFFID parameters controlling soil carbon and the five vegetation functional types using a Latin Hypercube design. Notable among these is the ‘q10’ parameter controlling the microbial respiration, assumed to accelerate with temperature in TRIFFID and other terrestrial carbon cycle models (Friedlingstein *et al.* 2006), based on laboratory and field measurements (Raich & Schlesinger 1992). Uncertainty in q10 implies a significant uncertainty in future levels of atmospheric CO₂, and hence global temperature rise (Jones *et al.* 2003b). There is also potential for interactions between climate and terrestrial ecosystems to influence future changes at a regional level (e.g. Betts *et al.* 2004; Cox *et al.* 2004), hence we anticipate that the PPE_A1B_VEG ensemble will make a significant contribution to the spread of future changes expressed in our probabilistic predictions.

Before starting the climate change simulations, it is necessary to perform a long preliminary ‘spin-up’ integration to achieve equilibrium in ocean–atmosphere carbon fluxes (Cox *et al.* 2001). Such an integration has been completed, using standard settings for parameters controlling ocean transports and ecosystem processes. Resource limitations will prevent sampling of the effects of uncertainties in these processes on the ocean carbon sink in PPE_A1B_VEG. While this is recognized as a missing element in our sampling of uncertainties in carbon cycle feedbacks, we note that most current models suggest that climate-induced changes in carbon storage are explained mainly by the land (Friedlingstein *et al.* 2006).

(v) *Earth system perturbed physics ensemble (Box H)*

The purpose of this ensemble, hereafter PPE_A1B_ESYS, would be to estimate the impact of nonlinear interactions between the process uncertainties sampled in PPE_A1B, PPE_A1B_OCEAN, PPE_A1B_SULPHATE and PPE_A1B_VEG, beyond those already accounted for in the simplified modelling of global temperature responses discussed in §3e(i). Members of PPE_A1B_ESYS would contain multiple simultaneous perturbations to parameters in the atmosphere, ocean, sulphur cycle and terrestrial ecosystem modules. Such an ensemble would be an ambitious and unprecedented undertaking. If produced, its results will be used in the time-scaling procedure as described in §3e(i).

Our ensemble runs will not sample other possible forcings and feedbacks which may affect future climate change, such as those associated with ocean carbon cycle processes (§3e(iv)), carbonaceous aerosols (e.g. Jones *et al.* 2005), non-aerosol atmospheric chemistry (e.g. Johnson *et al.* 2001), future land use change or methane release from permafrost (Christensen *et al.* 2004) and ocean hydrates (Archer & Buffett 2005). Nevertheless, we believe that our set of experiments samples uncertainties in the processes most likely to exert a major influence on climate change during the twenty-first century, based on current knowledge.

(f) *Conversion of ensemble simulations into probabilistic predictions (Box J)*

Recent progress in the statistical field of computer experiments has resulted in the development of a general Bayesian framework for the inference of observationally constrained probabilistic predictions from historical and forecast

ensemble runs of complex but imperfect models (e.g. Kennedy & O’Hagan 2001; Goldstein & Rougier 2004). Rougier (2007) describes this approach in the context of climate prediction. Here we summarize its key ingredients and describe briefly how we plan to apply this approach to our PPE experiments. More detailed descriptions will appear in forthcoming papers (Sexton *et al.* in preparation *a, b*).

The Bayesian framework arises from the definition of relationships between uncertain objects in the climate prediction problem. These elements include a vector of variables y representing the true climate, observations of a set of historical variables o and a climate model whose outputs are a function f of inputs x . In our case, x consists of our set of perturbed model parameters.⁶ The true climate, and the model simulation of it, consists of historical and future components (y_h, y_f) and ($f_h(x), f_f(x)$). A key step is the ‘best input assumption’ (Goldstein & Rougier 2004), which asserts that somewhere in the model parameter space there exists a single location x^* that provides the best simulation of true climate that the model is capable of providing. This assertion makes sense because we envisage the true climate y as a large vector consisting of many different variables specified at many different times and spatial locations. It would be less defensible if we were to consider only a small subset of climate variables, since it might not then be possible to distinguish between multiple locations in parameter space which provided historical simulations of similar quality, through differing and fortuitous compensations of errors. Nevertheless, it is recognized that x^* will not give rise to a perfect simulation, due to the effects of missing or inadequately parametrized processes in the model. The resulting mismatch between the true climate and that simulated at x^* is captured in an uncertain vector ϵ , termed ‘discrepancy’ (Goldstein & Rougier 2004), and defined by

$$y = f(x^*) + \epsilon. \quad (3.4)$$

The discrepancy can be recognized as an expression of the impact of structural modelling errors, to use terminology more familiar to climate scientists. Hereafter, we use these terms interchangeably, this amounting to a specific definition of ‘structural modelling errors’ in the present application. The discrepancy vector contains historical and future components (ϵ_h, ϵ_f) and would be expected to be different for different variables, dependent on the varying significance of structural inadequacies in simulating different climate processes.

The historical component of the true climate is related to the observations through $o = y_h + e$, where e represents observing errors. Hence, the observations and the historical component of the model simulation are related through

$$o = f_h(x^*) + \epsilon_h + e. \quad (3.5)$$

We do not know the location of x^* , but its prior distribution $\Pr(x=x^*)$ can be estimated by consulting experts (e.g. Murphy *et al.* 2004). In order to obtain observationally constrained probabilistic predictions of future climate, $\Pr(y_f|o)$, it

⁶In general, x can also include initial conditions and boundary conditions, such as the values of external radiative forcing agents or their precursors; however, the PPE experiments to which we apply the Bayesian framework in this study (Box A in figure 1) consider only perturbations to parameters controlling model processes.

is necessary to apply Bayes' theorem in an integration over parameter space, i.e.

$$\Pr(y_f|o) = \int \Pr(y_f|x) \cdot \Pr(o|x) \cdot \Pr(x = x^*) \cdot dx, \quad (3.6)$$

where $\Pr(y_f|x)$ is the *prior predictive probability* of simulating y_f at x before observational constraints have been applied, and $\Pr(o|x)$ is the likelihood of simulating the observed values at x .

Since it is not possible to run a GCM simulation for every location in the model parameter space, we replace the climate model $f(x)$ by $\mu(x) + u(x)$, where $\mu(x)$ is an estimate of $f(x)$ produced by a statistical emulator of the GCM output (§3b), and $u(x)$ is emulation error. We therefore obtain the following relationships:

$$y_f = \mu_f(x) + u_f(x) + \varepsilon_f, \quad (3.7)$$

$$o = \mu_h(x) + u_h(x) + \varepsilon_h + e. \quad (3.8)$$

In order to estimate $\Pr(y_f|o)$ from equation (3.6), it will be necessary to sample the joint prior distribution of all uncertain objects in equations (3.7) and (3.8) using, for example, Monte Carlo techniques (Sexton *et al.* in preparation a). In principle, this is a massive task, since the joint prior distribution is constructed from multidimensional vectors of model parameters, observational errors, emulation errors (obtained by calibrating the emulator against available PPE simulations) and discrepancy values (§3g). In practice, certain assumptions are necessary to make the calculation tractable, in particular that the distributions of parameter values, emulator errors, observation errors and discrepancy values are mutually independent (Rougier 2007). These assumptions are difficult to test. For example, we currently have no method of estimating how discrepancy might vary within parameter space, and no way of knowing how observational errors might vary with parameter values or emulator errors; however, we believe that they are reasonable expert judgements, in the sense that errors in them would only have a second-order impact on the results. For example, covariances between emulator errors in different parts of parameter space are likely to be relatively unimportant, provided the emulator captures the bulk of the variation found in GCM outputs (Rougier *et al.* submitted).

The approach we intend to use for the calculation of posterior probabilities is described elsewhere (Sexton *et al.* in preparation a,b); however, we note that the likelihood $\Pr(o|x)$ for location x in parameter space (given the above assumptions) is given by

$$\log\{\Pr(o|x)\} = -c - \left(\frac{N_o}{2}\right) \times \log|V| - \frac{1}{2}(\mu_h(x) - o)^T V^{-1}(\mu_h(x) - o), \quad (3.9)$$

where V is the covariance matrix of the sum of historical discrepancy (ε_h), emulator error and observational error; N_o is the number of observed variables; and c is a normalizing constant. The historical discrepancy will be largest for those variables which are modelled with the least skill (Murphy *et al.* 2004) and will therefore prevent such variables from exerting undue influence on the calculation of likelihood in equation (3.9). The role of discrepancy for *future* variables is to broaden the prior predictive distribution $\Pr(y_f|x)$ by increasing the uncertainty associated with linking emulated model predictions to the real future climate (equation (3.7)). Both of these contributions have the impact of broadening the posterior predictive distribution $\Pr(y_f|o)$.

Murphy *et al.* (2004) estimated probabilities for global climate sensitivity, using a methodology similar to that outlined above, but with a number of simplifying assumptions, including neglect of covariances between observables used to calculate the likelihood (Piani *et al.* 2005), and (in common with all previous probabilistic climate studies) neglect of discrepancy. Adoption of the more comprehensive and rigorous framework of Rougier (2007) allows us to avoid these simplifications and also confers the considerable advantage of supporting the construction of a *joint* posterior distribution ($\Pr(y_f|o)$) of a potentially large set of future variables. This can then be sampled to provide marginal joint PDFs of smaller subsets of variables, say changes in surface temperature and precipitation over some region, which are physically and statistically consistent with PDFs of individual variables sampled from $\Pr(y_f|o)$ one at a time. Such consistency could not be guaranteed if we were to approach the task by attempting to define an optimal set of constraints for each individual future variable (e.g. Piani *et al.* 2005; Knutti *et al.* 2006). Our method accounts for correlations between prediction errors in different parts of parameter space (e.g. equation (3.9)), avoiding the assumption of independent errors made in recent studies using MME results (Tebaldi *et al.* 2005; Furrer *et al.* in press). Such an assumption would not be appropriate in PPEs, where we expect structural deficiencies to contribute a common component to prediction error regardless of location in parameter space.

(g) *Structural modelling errors (Box K)*

The Bayesian framework provides an opportunity to include the effects of structural model deficiencies on historical and future climate variables, by specifying the multivariate distribution of discrepancy ϵ , assumed Gaussian. We therefore require estimates of mean values of discrepancy for the $N_o + N_p$ elements of ϵ , where N_p is the number of future variables we wish to predict. We must also specify a matrix of dimension $N_o + N_p$, describing covariances of these elements about their mean values. This is a demanding task, given the inherent difficulty of anticipating the effects on particular climate variables of missing or inadequately understood processes, and their complex interactions. The task is made even more difficult by the large set of variables to be considered, which would typically consist of a number of different climate elements (surface temperature, precipitation, pressure at mean sea level, radiative fluxes, etc.), each measured at a number of different locations (e.g. at each model grid point) in different seasons of the year. Therefore, we do not believe discrepancy could credibly be estimated by expert elicitation. In addition, we emphasize that discrepancy is a *prior input* to the Bayesian framework described in §3f, and should therefore be specified assuming (as far as possible) no knowledge of the observations used to determine the posterior distribution $\Pr(y_f|o)$.

Given these constraints, we propose to estimate discrepancy using results from our PPE simulations to predict results from an alternative MME, whose members consist of coupled atmosphere–mixed layer ocean models of similar complexity and credibility as HadSM3, but employing different basic assumptions in some of their parametrizations of physical processes. Note that this exercise must be carried out using ensembles of atmosphere–mixed layer ocean simulations, rather than ensembles of coupled models using a full

dynamical ocean (e.g. Meehl *et al.* 2005a), because our PPE experiments with HadCM3 are too small to support a direct application of the Bayesian framework to their results (§3j).

Our proposed approach for the specification of discrepancy is described in detail by Sexton *et al.* (in preparation a, b). It is based on the judgement that relationships between *model errors* for different climate variables can reasonably be expected to follow relationships between *inter-model differences* for different variables. We take a given MME member as a proxy for the true climate, and use our emulator of HadSM3 (§3b) to locate a point in the HadSM3 parameter space which achieves the best multivariate fit between HadSM3 and the multi-model member. The difference represents one estimate of discrepancy, under the above judgement. This process is repeated for each multi-model member, in turn, building up an ensemble of difference estimates. The ensemble mean of these is taken as our estimate of the mean value of discrepancy, and the covariances of the differences about the ensemble mean serve as our estimate of the discrepancy covariance matrix, after allowing for a component due to internal climate variability.

It is important to stress that our approach to the specification of discrepancy can only be expected to capture a subset of possible structural modelling errors and should be regarded as a lower bound. This is because models tend to share certain common systematic biases, which can be found in diverse elements of climate including multiannual means of basic quantities such as surface temperature, precipitation and pressure at mean sea level (e.g. Lambert & Boer 2001), distributions of daily rainfall events (Dai 2006), surface absorption of solar radiation (Wild 2005), cyclone frequencies (Lambert & Fyfe 2006), stationary waves in the upper atmosphere (Boyle 2006), observed low-frequency variations in tropical cloudiness (Wielicki *et al.* 2002), etc. In general, such biases could arise from either missing processes or common limitations such as insufficient resolution or the widespread adoption of a deficient parametrization scheme. They introduce an important general caveat on the confidence that can be placed in ensemble predictions (Smith 2002). Nevertheless, different climate models simulate changes in response to increases in GHGs which show a number of robust features (e.g. Räisänen 2001). While these are uncertain in magnitude, they are unlikely to be fundamentally compromised by such shortcomings. We therefore believe our climate model (albeit imperfect) to be *informative* about the real climate, and the discrepancy quantifies the extent of this belief. We believe it is preferable to include a lower bound for the effects of structural modelling errors than to ignore them altogether, since this will reduce the risk of providing policy makers with over-confident climate predictions. Our specification method also has the attractive feature of allowing us to provide predictions which combine results from perturbed physics and multi-model ensembles, thus making maximum use of the database of simulations available to the international community. Sensitivity tests can also be carried out by perturbing our discrepancy estimates in order to identify those aspects of our PDFs which are robust to reasonable variations in our assumptions.

(h) *Observational constraints (Box L)*

Implementation of the Bayesian framework of §3f requires the ability to estimate values of a set of observables at any point in the model parameter space (equation (3.6)). However, our choice of observational constraints is restricted by

resource-imposed limitations in our experimental design, because only our ensemble of HadSM3 simulations is large enough to support credible emulation of unsampled parts of parameter space. We can therefore use only variables available from these coupled atmosphere–mixed layer ocean simulations (Box A in figure 1), or which can be inferred with acceptable accuracy via the time-scaling procedure of Box D. This precludes, for example, the use of observations of ocean properties such as salinity or (sub-surface) temperature, or of coupled ocean–atmosphere modes of variability in which ocean transport plays a role, such as the El Niño–Southern Oscillation. In the main, therefore, we will be restricted to the use of spatial fields of multiannual seasonal means of physical variables describing surface and atmospheric characteristics of recent historical climate. Nevertheless, this still constitutes a substantial subset of the metrics typically used to assess climate simulations (e.g. Lambert & Boer 2001; Taylor 2001; Covey *et al.* 2003; Stainforth *et al.* 2005), or to produce probabilistic predictions of climate sensitivity (Murphy *et al.* 2004; Piani *et al.* 2005; Knutti *et al.* 2006). Specifically, we will use observed fields relating to temperature, humidity, cloud, wind, surface heat and moisture fluxes and radiative transfer. The set of fields will be somewhat smaller than that used by Murphy *et al.* (2004), because we are restricted by the set of variables which can be retrieved from simulations of the alternative models used to estimate discrepancy (§3g).

The HadSM3 simulations require the use of a prescribed heat convergence term (Hewitt *et al.* 2001), added to its mixed layer ocean as a function of position and season, and calibrated separately for each ensemble member. This term not only accounts for the effects of ocean heat transport, but also corrects for errors in simulated atmosphere–ocean heat fluxes in a manner akin to the use of flux adjustments in transient coupled simulations (§3c). The global mean value of the heat convergence, therefore, offsets any error in the global mean radiation balance and prevents such an error from increasing systematic biases in other model variables through the global cooling or warming which would otherwise have resulted. We therefore plan to include the global and annual mean of the heat convergence as an additional observational constraint, in order to weight model variants according to the magnitude of the planetary energy balance correction implied by this term.

In addition, we will also constrain our predictions using changes in large-scale features of surface temperature patterns observed during the twentieth century. This is because our time-scaling technique should allow us to infer this aspect of transient climate change with reasonable precision, for points in the atmosphere model parameter space for which no transient simulation is available (Harris *et al.* 2006). We therefore plan to include historical changes in several indices identified by Braganza *et al.* (2003), which capture features of the response to increasing GHGs found to be robust across GCM ensembles, these being the global mean, land–ocean and interhemispheric temperature contrasts and the zonally averaged meridional temperature gradient in Northern Hemisphere mid-latitudes. Stott *et al.* (2006a) show that these indices capture most of the information obtained from comprehensive spatiotemporal analyses of the past warming attributable to forcing from GHGs, aerosols and natural forcing agents, and should therefore provide an important constraint on future temperature changes at continental to global scales (Allen *et al.* 2000; Stott & Kettleborough 2002; Stott *et al.* 2006a,b; Kettleborough *et al.* 2007).

It may also be possible to use historical changes in the atmospheric CO₂ concentration as a further constraint (Jones & Cox 2001), dependent on the precision with which our SCM is able to replicate time series simulated in HadCM3 simulations sampling uncertainties in carbon cycle processes. With this exception, however, it will not be possible to use observations specifically relating to sulphur cycle or terrestrial ecosystem processes. This is because our atmosphere–mixed layer ocean PPE does not sample uncertainties in these processes.

Our set of observables, while incomplete, constitutes a multivariate collection covering a variety of physical climate characteristics. This should substantially reduce the risk of erroneously assigning a high weight to a location in parameter space which achieves a good fit to observations through a fortuitous compensation of errors. Studies which use a smaller number of observables to weight ensemble members (e.g. Giorgi & Mearns 2003; Tebaldi *et al.* 2005; Greene *et al.* 2006; Knutti *et al.* 2006) are more heavily reliant on the assumption that error compensations in the simulation of historical variables will remain operative in their predictions of future changes. Such an assumption is questionable, and limits the effectiveness of the applied constraint. For example, Forest *et al.* (2006) show that the record of surface temperature changes during the twentieth century may be explained by many combinations of climate sensitivity and ocean heat uptake efficiency, yet different combinations of these variables give rise to very different predictions of future changes (Knutti *et al.* 2005).

Finally, we emphasize the importance of obtaining reliable estimates of the errors with which our set of observations represents the true climate (the term e in equation (3.5)). These errors arise from internal climate variability, plus additional uncertainties associated with instrument errors, errors of representivity (when point observations from several locations are used to estimate a spatial mean over a GCM grid box; e.g. Osborn & Hulme 1997), or from the use of a model integration to assimilate observations, and hence produce spatially complete estimates of observed fields (e.g. Kalnay *et al.* 1996; Uppala *et al.* 2005). Estimates of observational uncertainty are themselves uncertain. For example, it is not generally possible to specify accurately the level of internal (unforced) climate variability, either because the observed records are not long enough, or because the record also contains variations arising from changes in external forcing which cannot be reliably disentangled from unforced variability. Internal variability simulated in climate model simulations is therefore often used as a proxy for observed variability (e.g. Allen *et al.* 2000; Stott & Kettleborough 2002; Murphy *et al.* 2004; Piani *et al.* 2005). The dependence of our probabilistic predictions on uncertainties in observing errors will be tested.

(i) *Downscaling transient changes for impact assessments (Boxes M, N)*

Boxes A–L in figure 1 constitute a methodology for the production of probabilistic climate predictions at spatial scales resolved by GCMs (approx. 300 km in the atmospheric module of HadCM3). However, this resolution is generally insufficient to serve climate impact applications (Hostetler 1994; Mearns *et al.* 2001), many of which require information at much finer scales to predict impacts in, say, river flow (Bell *et al.* 2006) or crop growth (Mearns *et al.* 1999). A method is therefore needed of obtaining realizations of fine-scale climate physically consistent with the simulations of large-scale climate obtained from

GCMs. This has been addressed by ‘downscaling’ GCM results using either statistical methods (e.g. Huth 1999; Wilby *et al.* 2002) or dynamical techniques (e.g. Jones *et al.* 1995; Christensen *et al.* 2001; Laprise *et al.* 2003). Here we adopt a commonly used dynamical downscaling technique in which a high-resolution regional climate model (RCM) is driven by time series of lateral boundary conditions (atmospheric surface pressure, wind, temperature and moisture plus chemical species required for the calculation of sulphate aerosol concentrations) and surface boundary conditions (sea surface temperatures and sea ice extents) saved from a previous GCM simulation. Nested RCMs have been shown to generate skilful fine-scale information in idealized predictability studies (Denis *et al.* 2002), and in verification of simulations driven by observed or GCM boundary conditions, for both time averaged climate means (e.g. Christensen *et al.* 1998; Noguer *et al.* 1998) and extreme events (e.g. Frei *et al.* 2003, 2006; Fowler *et al.* 2005; Buonomo *et al.* 2007).

We are currently producing a 17-member PPE of RCM simulations of transient climate change over Europe, at a horizontal resolution of 25 km (Box M in figure 1). Boundary conditions are supplied from data saved from each member of the PPE_A1B ensemble of HadCM3 simulations (§3c) for the period 1950–2100. We choose an RCM domain which is large enough to avoid the risk that relaxation to GCM data at the lateral boundaries will damp or distort the simulation of fine-scale detail over the interior region of interest (Jones *et al.* 1995), yet small enough to minimize the risk that inconsistencies will develop at larger scales between the climates simulated by the RCM and its driving model (e.g. Jones *et al.* 1995; Rinke & Dethloff 2000). In European simulations, such inconsistencies are most likely to develop over southeastern parts of the region in summer, where the climate simulated by RCMs is typically warmer and drier than that of the driving circulation (Jacob *et al.* 2007). This can arise either from differences in regional dynamical transports (Hagemann *et al.* 2004), or from the regional influence of the physical parametrizations (Vidale *et al.* 2003). We seek to minimize the latter as a source of divergence, by choosing parameter settings in each RCM to be consistent with the values used in the relevant driving simulation.⁷ However, the impacts of higher resolution (25 km cf. approx. 300 km) may still cause some degree of divergence between members of our RCM ensemble and their driving simulations, especially in regions where the advection of information from the lateral boundaries is weak. In such regions, it may not be credible to interpret the RCM results as a realization of local climate consistent with the driving large-scale circulation.

Probabilistic predictions at fine scales will be obtained by developing statistical relationships between changes simulated at individual RCM land points, and changes simulated at nearby points in the driving GCM (Box N in figure 1). This is somewhat analogous to a traditional statistical downscaling approach, in which a set of large-scale ‘predictor’ variables is used to obtain values of localized ‘predictand’ variables, using relationships trained on historical observations (Wilby *et al.* 2004). However, in our case, the relationships are trained using *future changes* in the predictor and predictand variables simulated by the GCM and RCM, since their purpose is to allow us to infer fine-scale changes for parts of the model parameter space for which no RCM simulation is available.

⁷This is achieved using the same parameter values in corresponding regional and global model ensemble members, with the exception of a few parameters for which values in the RCM are adjusted to allow for known dependencies on horizontal resolution.

Initially, we will use the simplest possible approach of predicting the response of a given RCM variable, say $\Delta P^R(s, t)$, from a univariate linear regression on the response of the corresponding variable at the nearest land point in the driving GCM ($\Delta P(s, t)$, following the notation of equation (3.3)), i.e.

$$\Delta P^R(s, t) = \alpha \Delta P(s, t) + \beta. \quad (3.10)$$

The regression coefficient α , and the variance of the residual β , will be obtained by training equation (3.10) using changes simulated in our 17-member RCM ensemble and the driving HadCM3 simulations. Owing to the limited number of simulations available for this purpose, it will probably be necessary to assume that the relationship between $\Delta P^R(s, t)$ and $\Delta P(s, t)$ is invariant across the model parameter space. Errors arising from this assumption will contribute to the downscaling uncertainty captured in the variance of β , which will also contain a component arising from the effects of internal climate variability generated by the RCM within its domain (Vidale *et al.* 2003). The viability of this approach rests on the assumption that the GCM changes in equation (3.10) will be found to be reasonably strongly related to their RCM counterparts. Results from previous experiments suggest that this is likely to be the case (e.g. Durman *et al.* 2001).

We close this section by comparing our experimental design with that of PRUDENCE, a recent experiment in which boundary data from two GCMs were used to drive several European RCMs with different representations of dynamical and physical processes (Christensen *et al.* 2007). This matrix design was chosen to maximize the extent to which uncertainties arising from regional physical processes can be sampled when only a limited ensemble of driving simulations is available. However, such a design would not be appropriate in our experiment, since we wish to isolate the additional contribution to regional uncertainty arising specifically from downscaling. This contribution is likely to be significant (e.g. Wilby & Harris 2006); however, we would expect it to be smaller than the total contribution to uncertainty from regional processes, which is, in turn, likely to be smaller than the contribution arising from processes external to the region of interest, based on the analysis of PRUDENCE results by Rowell (2006). However, our use of one specific downscaling approach does neglect the possibility that alternative downscaling methods could be found which provide regional detail different from those provided by our RCM simulations, but equally consistent with the driving GCM changes. We assume that any additional spread arising from a hypothetical set of alternative approaches would be relatively small compared with other sources of uncertainty.

(j) Production of probabilistic predictions

This section outlines the computational procedure that we will use to generate PDFs of transient climate change from the elements described in the preceding sections.

- *Step (i)*. Produce a large Monte Carlo sample of the model parameter space, using our emulator (§3b) to estimate values of multiannual mean values of a set of climate variables, including the observables identified in the first two paragraphs of §3h. Values will be obtained for present-day climate, and for equilibrium changes in response to doubled CO₂.

- *Step (ii)*. Obtain realizations of transient climate changes by applying the generalized time-scaling technique (see equation (3.3) and §3e(i)) to each member of the Monte Carlo sample, using emulated values of equilibrium response patterns ($\Delta P_{\text{eqm}}(s)$) and land and ocean climate sensitivities (λ_o and λ_l). For each Monte Carlo member, the SCM required to predict the transient response of global mean temperature will be run several times, sampling alternative values of uncertain parameters ζ from their pre-calibrated distributions. Equation (3.3) will then be applied several times to each realization of the global temperature response, using alternative values of $\text{extra}(s, t)$ and $\text{error}(s, t)$ sampled from their distributions, in order to account for additional uncertainties in the transient regional response arising from pattern scaling errors and earth system feedbacks. The end result will be an enlarged Monte Carlo sample for use in step (iv).
- *Step (iii)*. Calculate the weight to be assigned to each point in parameter space, given by the likelihood $\Pr(o|x)$ (see equation (3.9)). The weight will be calculated from the emulated values of present-day climate observables from step (i), plus the Braganza *et al.* (2003) indices measuring changes in surface temperature patterns observed since the industrial revolution (see §3h for details). Hindcasts of the Braganza *et al.* indices for points in parameter space will be obtained from step (ii). The weight will depend on the multivariate distance between the observations and our emulated and time-scaled estimates of the GCM values, normalized by our estimates of the uncertainties introduced by discrepancy, emulator error and observational error via equation (3.9), and accounting for covariances between different observables.
- *Step (iv)*. Calculate posterior PDFs of changes in individual climate variables, or joint PDFs of changes in two or more variables, by integrating over the Monte Carlo sample resulting from steps (i) and (ii). This procedure provides a numerical approximation to the integration over parameter space required by equation (3.9), in which each sampled point receives the weight calculated in step (iii).

The above procedure will allow us to generate PDFs for predictions at spatial scales resolved by HadCM3. For Europe, PDFs at finer scales will be obtained by adding the downscaling step of §3i. The generation of PDFs at RCM points will be achieved by applying equation (3.10) several times to each predicted response produced by step (ii), incorporating downscaling uncertainty by sampling alternative values of the residual β . Steps (iii) and (iv) will then be applied to the resulting sample of downscaled changes. The weight applied in step (iii) will be determined purely from the set of global model variables described above, and will not be modified to account for variations in the skill with which the RCM simulates fine-scale detail in its simulation of historical variables. In addition, no attempt will be made to account for structural downscaling errors in the discrepancy term, which will be based solely on the estimates of structural simulation errors at scales resolved by the global model (§3g). These limitations are imposed because our RCM ensemble is too small to support estimation of variations in downscaling performance across the model parameter space.

(k) UKCIP08 climate scenarios

The UK Climate Impacts Programme (UKCIP) issues future projections of national climate (Hulme *et al.* 2002), which are used by UK stakeholders to plan responses to changes expected during the twenty-first century. The next update (UKCIP08) is planned for 2008, and will for the first time be presented as PDFs, using most of the methodology described in this paper.⁸ Thirty-year averages of monthly, seasonal and annual changes will be predicted for a set of variables at 25 km resolution, for different periods during the twenty-first century. The set of variables will include diurnal mean, maximum and minimum temperature, total precipitation, snowfall, near-surface wind speed, mean sea level pressure, soil moisture content, specific and relative humidity, total cloud and surface flux components, plus certain additional variables derived from these, such as frost, gale and rain day frequencies. For surface temperature and precipitation, PDFs will also be supplied for changes in the 30-year mean values of several different percentiles of their daily distributions, thus providing users with information on changes in extreme events. Output from the Monte Carlo sampling used to generate the predictions (§3*j*) will also be provided, allowing users to quantify joint changes in combinations of variables. Users requiring information in the form of time series, rather than PDFs of changes in multi-year mean values, will have the choice of using data directly from our RCM ensemble, or of using a statistical weather generator tool (Kilsby *et al.* in press) to generate time series conditioned on the PDFs.

Results will be provided for the SRES A1B, B1 and A1FI emissions scenarios. Production of results for the B1 and A1FI scenarios will be achieved by running additional HadCM3 ensemble simulations parallel to those described in §3*c* (Box C of figure 1), and extending the time-scaling methodology of §3*e*(i). Details will be provided in a future paper.

4. Summary and concluding remarks

We have described a methodology for probabilistic prediction of time-dependent future climate change at regional scales, through ensembles of simulations designed to quantify the effects of uncertainties in climate system processes. This is achieved using a systematic approach in which ensemble members sample multiple perturbations to poorly constrained parameters in various configurations of the HadCM3 coupled ocean–atmosphere GCM. A large ensemble of (currently) approximately 300 members quantifies the effects of uncertainties in surface and atmospheric processes on the equilibrium response to doubled CO₂. This is augmented by a series of smaller (17 member) ensembles of historical and future time-dependent changes (in progress), exploring uncertainties in atmospheric, oceanic, sulphur cycle and terrestrial ecosystem processes, respectively. A further ensemble is planned in which uncertain parameters in all these modules are perturbed simultaneously, in order to estimate the effects of nonlinear interactions between feedbacks in different earth system components. This set of experiments samples uncertainties in the feedback processes most likely to exert a major influence on climate change during the twenty-first century.

⁸ It may not be possible to include results from all of the PPE experiments involving perturbations to ocean, sulphur cycle and terrestrial ecosystem parameters in time for UKCIP08, in which case the sampling of earth system process uncertainties would be somewhat less complete than shown in figure 1.

The ensemble runs are to be converted into probabilistic predictions using a Bayesian statistical framework designed for use with complex but imperfect models (Kennedy & O'Hagan 2001; Goldstein & Rougier 2004; Rougier 2007). This framework will allow us to estimate predictive distributions of future climate change, given expert-specified prior distributions for uncertain model parameters. In order to achieve this, a statistical emulator will be used to estimate GCM results in parts of parameter space for which no model simulation is available. The Bayesian framework also includes a discrepancy term to represent the additional uncertainty associated with structural modelling errors, defined as differences between the simulated and true climate which cannot be resolved by varying model parameters. The presence of such errors is indicated by the existence of systematic simulation biases across our set of ensemble members. We will estimate discrepancy using results from one of our PPEs to predict the results of alternative climate models, developed at different modelling centres and containing structural assumptions and errors partially independent of HadCM3. This will allow us to generate probabilistic estimates which combine information from perturbed physics and multi-model GCM ensembles. However, MMEs also share common structural limitations, potentially arising from a combination of missing processes, shared parametrization errors or limited resolution, so our method of estimating discrepancy should be recognized as providing only a lower bound.

The Bayesian framework will be applied to our equilibrium climate change ensemble, sampling changes at a large number of points in its parameter space using a Monte Carlo technique, and attaching a weight to each point representing its relative likelihood, estimated from the fit to observations of an emulated set of historical model variables. In order to obtain user-relevant predictions of twenty-first century climate change, we shall employ a time-scaling technique to convert the Monte Carlo sample members into transient response patterns, using a simple globally averaged climate model allied to a pattern scaling technique trained on a subset of equilibrium and transient simulations using common parameter combinations. Results from the ocean, sulphur cycle, terrestrial ecosystem and earth system ensembles will allow us to estimate uncertainties in global and regional changes additional to those explained by surface and atmospheric processes, and hence broaden the range of time-scaled transient responses. From these, observationally constrained probabilistic predictions will be derived for regional scales resolved by HadCM3, by integrating over our weighted Monte Carlo sample of parameter space.

For Europe, we are also producing a 17-member PPE of the regional climate model configuration of HadCM3, run at 25 km horizontal resolution in order to provide downscaled information required for studies of climate impacts. This ensemble will be used to establish simple statistical relationships between the transient responses of the global and regional versions of the model. By applying these relationships to the time-scaled transient responses described in the previous paragraph, it will become feasible to produce probabilistic predictions at spatial scales resolved by the regional model.

We believe that the approach described in this paper will provide a credible basis for the assessment of regional climate risks during the coming century. However, it necessarily involves a number of assumptions and limitations, which have been emphasized throughout the paper. The main assumptions are summarized in [table 1](#). It will be important to assess the robustness of the results

Table 1. Key assumptions in probabilistic climate prediction from PPEs.

assumption/caveat	justification	limitations	action/comments
perturbed physics ensembles are an effective method of sampling uncertainties in processes affecting future climate change	results to date show a spread of large-scale responses consistent with alternative multi-model ensemble approaches	structural model errors may bias aspects of the response and/or limit effective sampling of the associated uncertainties	include estimated impact of structural modelling errors via discrepancy term of our Bayesian statistical framework
discrepancy can be estimated using HadSM3(the atmosphere–mixed layer ocean configuration of HadCM3) to predict the results of other climate models	other models use different structural assumptions to HadSM3, and this method avoids risk of double counting the observations in our Bayesian inference	some structural errors likely to be common to all climate models, so this approach likely to provide a lower bound on discrepancy	sensitivity of results to discrepancy estimates will be tested
results depend on expert-specified prior distributions for climate model parameters	allows results to be constrained by expert understanding of detailed model processes	shape and extremes of distributions uncertain, because parameters do not typically correspond to directly observable quantities, and different experts may disagree	sensitivity of results to alternative priors can be tested and is likely to be reduced by the application of observational constraints
predictions can be effectively constrained using observations derived from spatial fields of present-day climate means, plus twentieth century trend in large-scale changes in surface temperature	constitutes a large multivariate set of variables likely to penalize heavily model variants giving rise to unrealistic simulations	potential additional constraints not accounted for	refine the method to include additional constraints, by running additional ensemble simulations when resources allow
transient climate responses can credibly be estimated by time scaling the equilibrium response to doubled CO ₂	transient and equilibrium responses are sufficiently closely related that scaling errors do not dominate the spread of transient changes	scaling error introduces an additional source of uncertainty	future increases in computing capacity may allow us to avoid this assumption by running larger ensembles of transient simulations

(Continued.)

Table 1. (Continued.)

assumption/caveat	justification	limitations	action/comments
use of flux adjustments in coupled ocean–atmosphere simulations	reduces regional simulation biases and prevents simulations of transient responses from being restricted to a small region of parameter space which achieves planetary radiation balance through a compensation of errors between structural model biases and parameter values	model variants with unrealistic radiation balances will give misleadingly credible simulations	use the global mean value of the heat convergence term in HadSM3 (which offsets radiation balance errors) as an additional observational constraint
uncertainties in earth system feedbacks likely to contribute to twenty-first century climate change (in addition to those arising from physical surface and atmospheric processes) are effectively sampled	available evidence suggests that ocean transport, sulphate aerosol forcing and terrestrial carbon cycle feedbacks are the main known uncertainties, and these are sampled in our ensemble experiments	we can only run small ensembles to sample these uncertainties, which will limit the precision with which their effects can be included. Structural errors in the representation of these feedbacks in HadCM3 will not be accounted for	a more comprehensive approach would require greater computing resource
model parameter values, observational errors, emulator errors and discrepancy values assumed mutually independent in Bayesian framework	no obvious reason why these elements should be strongly related	basic assumption required to make the Bayesian calculation tractable: difficult to test	effects of errors in these assumptions are judged likely to be of second-order importance
uncertainty in downscaling can be effectively captured by running an ensemble of regional climate model variants with parameter perturbations corresponding to their driving global model simulations	likely to ensure maximum physical consistency between the down-scaled regional detail and the driving large-scale changes, thus avoiding the creation of spurious uncertainty at regional scales	in principle, alternative downscaling strategies could be found which give regional signals different from the regional climate model, but equally consistent with the driving large-scale patterns of change	alternative downscaling strategies should be compared when feasible

to these assumptions and compare them to alternative approaches based on different methodological choices (e.g. Allen *et al.* 2006). These would include methods to derive probabilistic information constrained by the spread of MME results (Tebaldi *et al.* 2005; Räisänen & Ruokolainen 2006; Furrer *et al.* in press), and methods which use GCM projections, but seek to provide probabilities determined principally by uncertainties in constraining historical observations, and relatively independent of the GCM(s) used (Allen *et al.* 2000; Stott & Kettleborough 2002; Stott *et al.* 2006a).

A number of developments may be identified for improving our methodology in future. The bedrock of our present approach consists of large ensembles of relatively cheap equilibrium simulations of the coupled atmosphere–mixed layer ocean configuration of our GCM. An increase in computing resources would enable larger ensemble simulations of transient climate change to be produced. This would remove the need to infer transient responses from equilibrium responses (hence removing the uncertainty associated with that element of the procedure), and would also allow improved sampling of uncertainties associated with ocean transports and biogeochemical processes. In the present framework, these will be included via simple adjustments to the mean and variance of global and regional transient responses, whereas larger ensembles would enable emulation of variations in their effects across the relevant parameter spaces, leading to improved specifications of their impacts. Large ensembles of transient simulations would also expand the range of historical observables available to constrain the predictions: at present, these are restricted to variables which can reliably be emulated from our equilibrium simulations (present-day long-term mean fields and historical changes in a few large-scale variables), thus excluding sub-surface ocean diagnostics, or metrics of coupled ocean–atmosphere modes of variability. Potential additional constraints include palaeoclimatic changes (Annan *et al.* 2005b; Schneider von Deimling *et al.* 2006), temperature changes over the past millennium (e.g. Hegerl *et al.* 2006), the response to major volcanic eruptions (Wigley *et al.* 2005; Yokohata *et al.* 2005), measures of cloud feedback derived from climate variability (Bony *et al.* 2004; Bony & Dufresne 2005; Williams *et al.* 2005), short-range tendencies in weather forecasts (Rodwell & Palmer 2007) and the verification of seasonal to interannual ensemble hindcasts initialized from observations (Palmer *et al.* submitted). Expanding the ensemble prediction methodology to include such constraints, either through additional integrations or extraction of the required data, could potentially narrow the range of future predictions (Annan & Hargreaves 2006), and reduce the extent to which they are dependent on prior assumptions such as uncertainty ranges for model parameters.

There is also, of course, scope to improve the methodology by improving the climate model itself. In one sense, improving the model could tend to increase the range of the predictions, by incorporating additional earth system feedbacks, or better resolving regional processes driving climate variability, such as atmospheric blocking (Palmer *et al.* submitted). Ultimately, however, model improvements should narrow the range of predicted changes, by reducing structural modelling errors and hence the contribution of the discrepancy term in our Bayesian framework.

This work is 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. © Crown Copyright 2007, the Met Office, UK.

References

- Allen, M. R. & Ingram, W. J. 2002 Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232. (doi:10.1038/nature01092)
- Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. & Delworth, T. L. 2000 Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* **407**, 617–620. (doi:10.1038/35036559)
- Allen, M. R., Frame, D. J., Kettleborough, J. A. & Stainforth, D. A. 2006 Model error in weather and climate forecasting. In *Predictability of weather and climate* (eds T. Palmer & R. Hagedorn), pp. 391–427. Cambridge, UK: Cambridge University Press.
- Anderson, T. L., Charlson, R. J., Schwartz, S. E., Knutti, R., Boucher, O., Rodhe, H. & Heitzberg, J. 2003 Climate forcing by aerosols—a hazy picture. *Science* **300**, 1103–1104. (doi:10.1126/science.1084777)
- Anderson, J. L. *et al.* 2004 The new GFDL global atmosphere and land model AM2-LM2: evaluation with prescribed SST simulations. *J. Clim.* **17**, 4641–4673. (doi:10.1175/1520-0442(2004)017<4089:IOALMO>2.0.CO;2)
- Andreae, M. O., Jones, C. D. & Cox, P. M. 2005 Strong present-day aerosol cooling implies a hot future. *Nature* **435**, 1187–1190. (doi:10.1038/nature03671)
- Annan, J. D. & Hargreaves, J. C. 2006 Using multiple observationally-based constraints to estimate climate sensitivity. *Geophys. Res. Lett.* **33**, L06704. (doi:10.1029/2005GL025259)
- Annan, J. D., Hargreaves, J. C., Edwards, N. R. & Marsh, R. 2005a Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Model.* **8**, 135–154. (doi:10.1016/j.ocemod.2003.12.004)
- Annan, J. D., Hargreaves, J. C., Ohgaito, R., Abe-Ouchi, A. & Emori, S. 2005b Efficiently constraining climate sensitivity with ensembles of paleoclimate simulations. *SOLA* **1**, 181–184. (doi:10.2151/sola.2005-047)
- Archer, D. & Buffett, B. 2005 Time-dependent response of the global ocean clathrate reservoir to climatic and anthropogenic forcing. *Geochem. Geophys. Geosyst.* **6**, Q03002. (doi:10.1029/2004GC000854)
- Barnett, D. N., Brown, S. J., Murphy, J. M., Sexton, D. M. H. & Webb, M. J. 2006 Quantifying uncertainty in changes in extreme event frequency in response to doubled CO₂ using a large ensemble of GCM simulations. *Clim. Dynam.* **26**, 489–511. (doi:10.1007/s00382-005-0097-1)
- Bell, V. A., Kay, A. L., Jones, R. G. & Moore, R. J. 2006 Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrol. Earth Syst. Sci.* **11**, 532–549.
- Betts, R. A., Cox, P. M., Collins, M., Harris, P., Huntingford, C. & Jones, C. D. 2004 The role of ecosystem-atmosphere interactions in simulated Amazonian precipitation decrease and forest dieback under global climate warming. *Theor. Appl. Clim.* **78**, 157–175.
- Blackmon, M. *et al.* 2001 The community climate system model. *Bull. Am. Met. Soc.* **82**, 2357–2376. (doi:10.1175/1520-0477(2001)082<2357:TCCSM>2.3.CO;2)
- Boer, G. J. & Yu, B. 2003 Dynamical aspects of climate sensitivity. *Geophys. Res. Lett.* **30**, 1135. (doi:10.1029/2002GL016549)
- Bony, S. & Dufresne, J.-L. 2005 Marine boundary layer clouds at the heart of cloud feedback uncertainties in climate models. *Geophys. Res. Lett.* **32**, L20 806. (doi:10.1029/2005GL023851)
- Bony, S., Dufresne, J.-L., Le Treut, H., Morcrette, J.-J. & Senior, C. A. 2004 On dynamic and thermodynamic components of cloud changes. *Clim. Dynam.* **22**, 71–86. (doi:10.1007/s00382-003-0369-6)
- Bony, S. *et al.* 2006 How well do we understand and evaluate climate change feedback processes? *J. Clim.* **19**, 3445–3482. (doi:10.1175/JCLI3819.1)
- Boyle, J. S. 2006 Upper level atmospheric stationary waves in the twentieth century climate of the Intergovernmental Panel on Climate Change simulations. *J. Geophys. Res.* **111**, D14 101. (doi:10.1029/2005JD006612)

- Braganza, K., Karoly, D. J., Hirst, A. C., Mann, M. E., Stott, P., Stouffer, R. J. & Tett, S. F. B. 2003 Simple indices of global climate variability and change: part I—variability and correlation structure. *Clim. Dynam.* **20**, 491–502.
- Buonomo, E., Jones, R., Huntingford, C. & Hannaford, J. 2007 On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations. *Q. J. R. Meteorol. Soc.* **133**, 65–81. (doi:10.1002/qj.13)
- Christensen, O. B., Christensen, J. H., Machenhauer, B. & Botzet, M. 1998 Very high-resolution regional climate simulations over Scandinavia—present climate. *J. Clim.* **11**, 3204–3229. (doi:10.1175/1520-0442(1998)011<3204:VHRRCSS>2.0.CO;2)
- Christensen, J. H., Räisänen, J., Iversen, T., Bjørge, D., Christensen, O. B. & Rummukainen, M. 2001 A synthesis of regional climate change simulations—a Scandinavian perspective. *Geophys. Res. Lett.* **28**, 1003–1006. (doi:10.1029/2000GL012196)
- Christensen, T. R. *et al.* 2004 Thawing sub-arctic permafrost: effects on vegetation and methane emissions. *Geophys. Res. Lett.* **31**, L04501. (doi:10.1029/2003GL018680)
- Christensen, J. H., Carter, T. R., Rummukainen, M. & Amanatidis, G. 2007 Evaluating the performance and utility of regional climate models: the PRUDENCE project. *Clim. Change* **81**, 1–6. (doi:10.1007/S10584-006-9211-6)
- Clark, R., Brown, S. & Murphy, J. 2006 Modelling northern hemisphere summer heat extreme changes and their uncertainties using a physics ensemble of climate sensitivity experiments. *J. Clim.* **19**, 4418–4435. (doi:10.1175/JCLI3877.1)
- Collins, M., Booth, B. B. B., Harris, G., Murphy, J. M., Sexton, D. M. H. & Webb, M. J. 2006 Towards quantifying uncertainty in transient climate change. *Clim. Dynam.* **27**, 127–147. (doi:10.1007/s00382-006-0121-0)
- Collins, M., Brierley, C. M., MacVean, M., Booth, B. B. B. & Harris, G. R. 2007 The sensitivity of the rate of transient climate change to ocean physics perturbations. *J. Clim.* **20**, 2315–2320.
- Cox, P. M. 2001 Description of the TRIFFID dynamic global vegetation model. Technical Note 24, Hadley Centre, Met Office, UK.
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. & Smith, J. 1999 The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dynam.* **15**, 183–203. (doi:10.1007/s003820050276)
- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. 2000 Acceleration of global warming due to carbon cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187. (doi:10.1038/35041539)
- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. 2001 Modelling vegetation and the carbon cycle as interactive elements of the climate system. In *Meteorology at the millennium* (ed. R. Pearce), pp. 259–279. New York, NY: Academic Press.
- Cox, P. M., Betts, R. A., Collins, M., Harris, P., Huntingford, C. & Jones, C. D. 2004 Amazon dieback under climate-carbon cycle projections for the 21st century. *Theor. Appl. Clim.* **78**, 137–156.
- Covey, C., AchutaRao, K. M., Cubasch, U., Jones, P., Lambert, S. J., Mann, M. E., Phillips, T. J. & Taylor, K. E. 2003 An overview of results from the coupled model intercomparison project. *Global Planet. Change* **37**, 103–133. (doi:10.1016/S0921-8181(02)00193-5)
- Cubasch, U., Meehl, G. A., Boer, G. J., Stouffer, R. J., Dix, M., Noda, A., Senior, C. A., Raper, S. & Yap, K. S. 2001 Projections of future climate change. In *Climate change 2001: the scientific basis* (eds J. T. Houghton *et al.*). Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change, pp. 525–582. Cambridge, UK: Cambridge University Press.
- Dai, A. 2006 Precipitation characteristics in eighteen coupled climate models. *J. Clim.* **18**, 4605–4630. (doi:10.1175/JCLI3884.1)
- Denis, B., Laprise, R., Caya, D. & Côté, J. 2002 Downscaling ability of one-way-nested regional climate models: the big-brother experiment. *Clim. Dynam.* **18**, 627–646. (doi:10.1007/s00382-001-0201-0)

- Dessai, S., Lu, X. & Hulme, M. 2005 Limited sensitivity analysis of regional climate change probabilities for the 21st century. *J. Geophys. Res.* **110**, D19 108. (doi:10.1029/2005JD005919)
- Durman, C. F., Gregory, J. M., Hassell, D. C., Jones, R. G. & Murphy, J. M. 2001 A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Q. J. R. Meteorol. Soc.* **127**, 1005–1015. (doi:10.1002/qj.49712757316)
- Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M. & Taylor, C. M. 2003 Explicit representation of subgrid heterogeneity in a GCM land–surface scheme. *J. Hydromet.* **4**, 530–543. (doi:10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2)
- Forest, C. E., Stone, P. H. & Sokolov, A. P. 2006 Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.* **33**, L01705. (doi:10.1029/2005GL023977)
- Fowler, H. J., Ekström, M., Kilsby, C. G. & Jones, P. D. 2005 New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. *J. Hydrol.* **300**, 212–233. (doi:10.1016/j.jhydrol.2004.06.017)
- Frame, D. J., Booth, B. B. B., Kettleborough, J. A., Stainforth, D. A., Gregory, J. M., Collins, M. & Allen, M. R. 2005 Constraining climate forecasts: the role of prior assumptions. *Geophys. Res. Lett.* **32**, L09702. (doi:10.1029/2004GL022241)
- Frei, C., Christensen, J. H., Déqué, M., Jacob, D., Jones, R. G. & Vidale, P. L. 2003 Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. *J. Geophys. Res.* **108**, D34 124.
- Frei, C. *et al.* 2006 Future change of precipitation extremes in Europe: intercomparison of scenarios from regional climate models. *J. Geophys. Res.* **111**, D06105. (doi:10.1029/2005JD005965)
- Friedlingstein, P., Bopp, L., Schnur, R. & Zeng, N. 2006 Climate-carbon cycle feedback analysis: results from the C⁴MIP model intercomparison. *J. Clim.* **19**, 3337–3353. (doi:10.1175/JCLI3800.1)
- Furrer, R., Sain, S. R., Nychka, D., & Meehl, G. A. In press. Multivariate Bayesian analysis of atmosphere–ocean general circulation models. *Environ. Ecol. Stat.*
- Giorgi, F. & Mearns, L. O. 2003 Probability of regional climate change based on the reliability ensemble averaging (REA) method. *Geophys. Res. Lett.* **30**, 1629. (doi:10.1029/2003GL017130)
- Goldstein, M. & Rougier, J. C. 2004 Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM J. Sci. Comput.* **26**, 467–487. (doi:10.1137/S106482750342670X)
- Good, P. & Lowe, J. 2006 Emergent behaviour and uncertainty in multi-model climate projections of precipitation trends at small scales. *J. Clim.* **27**, 357–375.
- Gordon, C., Cooper, C., Senior, C. A., Banks, H. T., Gregory, J. M., Johns, T. C., Mitchell, J. F. B. & Wood, R. A. 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. (doi:10.1007/s003820050010)
- Greene, A. M., Goddard, L. & Lall, U. 2006 Probabilistic multimodel regional temperature change projections. *J. Clim.* **19**, 4326–4346. (doi:10.1175/JCLI3864.1)
- Hagemann, S., Machenhauer, B., Jones, R., Christensen, O., Déqué, M., Jacob, D. & Vidale, P. 2004 Evaluation of water and energy budgets in regional climate models applied over Europe. *Clim. Dynam.* **23**, 547–607. (doi:10.1007/s00382-004-0444-7)
- Harris, G., Sexton, D. M. H., Booth, B. B. B., Collins, M., Murphy, J. M. & Webb, M. J. 2006 Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. *Clim. Dynam.* **27**, 357–375. (doi:10.1007/s00382-006-0142-8)
- Hegerl, G. C., Crowley, T. J., Hyde, W. T. & Frame, D. J. 2006 Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440**, 1029–1032. (doi:10.1038/nature04679)
- Hewitt, C. D., Senior, C. A. & Mitchell, J. F. B. 2001 The impact of dynamic sea ice on the climatology and climate sensitivity of a GCM: a study of past, present and future climates. *Clim. Dynam.* **17**, 655–668. (doi:10.1007/s003820000140)

- Hostetler, S. 1994 Hydrologic and atmospheric models: the problem of discordant scales: an editorial comment. *Clim. Change* **27**, 345–350. (doi:10.1007/BF01096266)
- Hulme, M. *et al.* 2002 Climate-change scenarios for the United Kingdom: the UKCIP02 scientific report. Tyndall Centre for Climate Change Research. School of Environmental Sciences. University of East Anglia, Norwich, UK, pp. 120.
- Huth, R. 1999 Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Clim. Res.* **13**, 91–101.
- Jacob, D. *et al.* 2007 An intercomparison of regional climate models for Europe: model performance in present-day climate. *Clim. Change* **81**, 31–52. (doi:10.1007/S10584-006-9213-4)
- Johns, T. C. *et al.* 2003 Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim. Dynam.* **20**, 583–612.
- Johns, T. C. *et al.* 2006 The new Hadley Centre climate model HadGEM1: evaluation of coupled simulations. *J. Clim.* **19**, 1327–1353. (doi:10.1175/JCLI3712.1)
- Johnson, C. E., Stevenson, D. S., Collins, W. J. & Derwent, R. G. 2001 Role of climate feedback on methane and ozone studied with a coupled ocean–atmosphere–chemistry model. *Geophys. Res. Lett.* **28**, 1723–1726. (doi:10.1029/2000GL011996)
- Jones, C. & Cox, P. 2001 Constraints on the temperature sensitivity of global soil respiration from the observed interannual variability in atmospheric CO₂. *Atmos. Sci. Lett.* **2**, 166–172. (doi:10.1006/asle.2001.0041)
- Jones, R. G., Murphy, J. M. & Noguer, M. 1995 Simulation of climate change over Europe using a nested regional climate model. I: assessment of control climate, including sensitivity to location of lateral boundaries. *Q. J. R. Meteorol. Soc.* **121**, 1413–1449.
- Jones, A., Roberts, D. L., Woodage, M. J. & Johnson, C. E. 2001 Indirect sulphate aerosol forcing in a climate model with an interactive sulphur cycle. *J. Geophys. Res.* **106**, 20 293–20 310. (doi:10.1029/2000JD000089)
- Jones, C. D., Cox, P. M., Essery, R. L. H., Roberts, D. L. & Woodage, M. A. 2003a Strong carbon cycle feedbacks in a climate model with interactive CO₂ and sulphate aerosols. *Geophys. Res. Lett.* **30**, 1479. (doi:10.1029/2003GL016867)
- Jones, C. D., Cox, P. & Huntingford, C. 2003b Uncertainty in climate-carbon cycle projections associated with the sensitivity of soil respiration to temperature. *Tellus B* **55**, 642–648. (doi:10.1034/j.1600-0889.2003.01440.x)
- Jones, G. S., Jones, A., Roberts, D. L., Stott, P. A. & Williams, K. D. 2005 Sensitivity of global-scale climate change attribution results to inclusion of fossil fuel black carbon aerosol. *Geophys. Res. Lett.* **32**, L14 701. (doi:10.1029/2005GL023370)
- Kalnay, E. *et al.* 1996 The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471. (doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kennedy, M. & O'Hagan, A. 2001 Bayesian calibration of computer models. *J. Roy. Stat. Soc. B* **63**, 425–464. (doi:10.1111/1467-9868.00294)
- Kettleborough, J. A., Booth, B. B. B., Stott, P. A. & Allen, M. R. 2007 Estimates of uncertainty in predictions of global mean surface temperature. *J. Clim.* **20**, 843–855. (doi:10.1175/JCLI4012.1)
- Kilsby, C. G., Jones, P. D., Burton, A., Ford, A. C., Fowler, H. J., Harpham, C., James, P., Smith, A. & Wilby, R. L. In press. A daily weather generator for use in climate change studies. *Environ. Model. Software*.
- Knutti, R., Joos, F., Müller, S. A., Plattner, G.-K. & Stocker, T. F. 2005 Probabilistic climate change projections for CO₂ stabilization profiles. *Geophys. Res. Lett.* **32**, L20 707. (doi:10.1029/2005GL023294)
- Knutti, R., Meehl, G. A., Allen, M. R. & Stainforth, D. A. 2006 Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Clim.* **19**, 4224–4233. (doi:10.1175/JCLI3865.1)
- Lambert, S. J. & Boer, G. J. 2001 CMIP1 evaluation and intercomparison of coupled climate models. *Clim. Dynam.* **17**, 83–106. (doi:10.1007/PL00013736)

- Lambert, S. J. & Fyfe, J. C. 2006 Changes in winter cyclone frequencies and strengths simulated in enhanced greenhouse warming experiments: results from the models participating in the IPCC diagnostic exercise. *Clim. Dynam.* **26**, 713–728. (doi:10.1007/s00382-006-0110-3)
- Laprise, R., Caya, D., Frigon, A. & Paquin, D. 2003 Current and perturbed climate as simulated by the second-generation Canadian Regional Climate Model (CRCM-II) over northwestern North America. *Clim. Dynam.* **21**, 405–421. (doi:10.1007/s00382-003-0342-4)
- Lopez, A., Tebaldi, C., New, M., Stainforth, D. A., Allen, M. R. & Kettleborough, J. A. 2006 Two approaches to quantifying uncertainty in global temperature changes. *J. Clim.* **19**, 4785–4796. (doi:10.1175/JCLI3895.1)
- McKay, M. D., Beckman, R. J. & Conover, W. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245. (doi:10.2307/1268522)
- Mearns, L. O., Mavromatis, E., Tsvetsinskaya, E., Hays, C. & Easterling, W. 1999 Comparative responses of EPIC and CERES crop models to high and low resolution climate change scenario. *J. Geophys. Res.* **104**, 6623–6646. (doi:10.1029/1998JD200061)
- Mearns, L. O., Hulme, M., Carter, T. R., Leemans, R., Lal, M. & Whetton, P. 2001 Climate scenario development. In *Climate change 2001: the scientific basis* (eds J. T. Houghton *et al.*). Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change, pp. 739–768. Cambridge, UK: Cambridge University Press.
- Meehl, G. A., Covey, C., McAvaney, B., Latif, M. & Stouffer, R. J. 2005a Overview of the coupled model intercomparison project. *Bull. Am. Meteorol. Soc.* **86**, 89–93. (doi:10.1175/BAMS-86-1-89)
- Meehl, G. A., Arblaster, J. M. & Tebaldi, C. 2005b Understanding future patterns of precipitation extremes in climate model simulations. *Geophys. Res. Lett.* **32**, L18 719. (doi:10.1029/2005GL023680)
- Meehl, G. A. *et al.* 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* (eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor & H. L. Miller). Cambridge, UK and New York, NY: Cambridge University Press.
- Mitchell, J. F. B. 2003 Pattern scaling. *Clim. Change* **60**, 217–242. (doi:10.1023/A:1026035305597)
- Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M. & Stainforth, D. A. 2004 Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **429**, 768–772. (doi:10.1038/nature02771)
- Nakicenovic, N. & Swart, R. 2000 *Special report on emissions scenarios*. Cambridge, UK: Cambridge University Press.
- Noguer, M., Jones, R. G. & Murphy, J. M. 1998 Effect of systematic errors in the lateral boundary forcing on the climatology of a nested regional climate model over Europe. *Clim. Dynam.* **14**, 691–712. (doi:10.1007/s003820050249)
- Osborn, T. J. & Hulme, M. 1997 Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Clim.* **10**, 1885–1908. (doi:10.1175/1520-0442(1997)010<1885:DOARBS>2.0.CO;2)
- Palmer, T. N. & Räisänen, J. 2002 Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature* **415**, 514–517. (doi:10.1038/415512a)
- Palmer, J. R. & Totterdell, I. J. 2001 Production and export in a global ocean ecosystem model. *Deep Sea Res.* **48**, 1169–1198. (doi:10.1016/S0967-0637(00)00080-7)
- Palmer, T. N., Doblas-Reyes, A., Weisheimer, A., Rodwell, M. & Jung, T. Submitted. Reliability of climate-change projections of precipitation: towards “seamless” climate prediction.
- Piani, C., Frame, D. J., Stainforth, D. A. & Allen, M. R. 2005 Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Lett.* **32**, L23 825. (doi:10.1029/2005GL024452)
- Raich, J. & Schlesinger, W. 1992 The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B* **44**, 81–99. (doi:10.1034/j.1600-0889.1992.t01-1-00001.x)

- Räisänen, J. 2001 CO₂-induced climate change in CMIP2 experiments: quantification of agreement and role of internal variability. *J. Clim.* **14**, 2088–2104. (doi:10.1175/1520-0442(2001)014<2088:CICCIC>2.0.CO;2)
- Räisänen, J. 2002 CO₂-induced changes in interannual temperature and precipitation variability in 19 CMIP2 experiments. *J. Clim.* **15**, 2395–2411. (doi:10.1175/1520-0442(2002)015<2395:CICII-T>2.0.CO;2)
- Räisänen, J. 2003 CO₂-induced changes in angular momentum in CMIP2 experiments. *J. Clim.* **16**, 132–143. (doi:10.1175/1520-0442(2003)016<0132:CICIAA>2.0.CO;2)
- Räisänen, J. 2005 Impact of increasing CO₂ on monthly-to-annual precipitation extremes: analysis of the CMIP2 experiments. *Clim. Dynam.* **24**, 309–323. (doi:10.1007/s00382-004-0510-1)
- Räisänen, J. & Ruokolainen, L. 2006 Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A* **58**, 461–472.
- Raper, S. C. B., Gregory, J. M. & Stouffer, R. J. 2002 The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *J. Clim.* **15**, 124–130. (doi:10.1175/1520-0442(2002)015<0124:TROCSA>2.0.CO;2)
- Rinke, A. & Dethloff, K. 2000 On the sensitivity of a regional Arctic climate model to initial and boundary conditions. *Clim. Res.* **14**, 101–113.
- Rodwell, M. J. & Palmer, T. N. 2007 Using NWP to assess climate models. *Q. J. R. Meteorol. Soc.* **133**, 129–146. (doi:10.1002/qj.23)
- Rougier, J. C. 2007 Probabilistic inference for future climate using an ensemble of climate model evaluations. *Clim. Change* **81**, 247–264. (doi:10.1007/s10584-006-9156-9)
- Rougier, J. & Sexton, D. M. H. 2007 Inference in ensemble experiments. *Phil. Trans. R. Soc. A* **365**, 2133–2143. (doi:10.1098/rsta.2007.2071)
- Rougier, J. C., Sexton, D. M. H., Murphy, J. M. & Stainforth, D. A. Submitted. Emulating the sensitivity of the HadAM3 climate model using different but related experiments.
- Rowell, D. P. 2006 A demonstration of the uncertainty in projections of UK climate change resulting from regional model formulation. *Clim. Change* **79**, 243–256. (doi:10.1007/s10584-006-9100-z)
- Rowntree, P. R. 1998 Global average climate forcing and response since 1750. *Int. J. Climatol.* **18**, 355–377. (doi:10.1002/(SICI)1097-0088(19980330)18:4<355::AID-JOC241>3.0.CO;2-B)
- Schlesinger, M. E., Andronova, N. G., Entwistle, B., Ghanem, A., Ramankutty, N., Wang, W. & Yang, F. 1997 Modelling and simulation of climate and climate change. In *Past and present variability of the solar-terrestrial system*, vol. 133 (eds G. C. Catagnoli & A. Provenzale). International School of Physics Enrico Fermi, pp. 389–429. Amsterdam, The Netherlands: IOS Press.
- Schneider von Deimling, T., Held, H., Ganopolski, A. & Rahmstorf, S. 2006 Climate sensitivity estimated from ensemble simulations of glacial climate. *Clim. Dynam.* **27**, 149–163. (doi:10.1007/s00382-006-0126-8)
- Selten, F. M., Branstator, G. W., Kliphuis, M. & Dijkstra, H. A. 2004 Tropical origins for recent and future northern hemisphere climate change. *Geophys. Res. Lett.* **31**, L21 205. (doi:10.1029/2004GL020739)
- Sexton, D. M. H., Murphy, J. M. & Collins, M. In preparation *a*. Multivariate prediction using imperfect climate models. Part I. Outline of methodology.
- Sexton, D. M. H., Murphy, J. M., Collins, M. & Webb, M. J. In preparation *b*. Multivariate prediction using imperfect climate models. Part II. Using a large number of observations to constrain a climate prediction.
- Smith, L. A. 2002 What might we learn from climate forecasts? *Proc. Natl Acad. Sci. USA* **99**, 2487–2492. (doi:10.1073/pnas.012580599)
- Soden, B. J. & Held, I. M. 2006 An assessment of climate feedbacks in coupled ocean–atmosphere models. *J. Clim.* **19**, 3354–3360. (doi:10.1175/JCLI3799.1)
- Stainforth, D. A. *et al.* 2005 Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403–406. (doi:10.1038/nature03301)
- Stott, P. A. & Kettleborough, J. A. 2002 Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**, 723–726. (doi:10.1038/416723a)

- Stott, P. A., Mitchell, J. F. B., Allen, M. R., Delworth, T. L., Gregory, J. M., Meehl, G. A. & Santer, B. D. 2006a Observational constraints on past attributable warming and predictions of future global warming. *J. Clim.* **19**, 3055–3069. (doi:10.1175/JCLI3802.1)
- Stott, P. A., Kettleborough, J. A. & Allen, M. R. 2006b Uncertainty in continental-scale temperature predictions. *Geophys. Res. Lett.* **33**, L02708. (doi:10.1029/2005GL024423)
- Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**, 7183–7192. (doi:10.1029/2000JD900719)
- Tebaldi, C., Smith, R. W., Nychka, D. & Mearns, L. O. 2005 Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. *J. Clim.* **18**, 1524–1540. (doi:10.1175/JCLI3363.1)
- Tebaldi, C., Hayhoe, K., Arblaster, J. M. & Meehl, G. A. 2006 Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events. *Clim. Change* **79**, 185–211. (doi:10.1007/s10584-006-9051-4)
- Tett, S. F. B. *et al.* 2002 Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.* **107**, 4306. (doi:10.1029/2000JD000028)
- Uppala, S. M. *et al.* 2005 The ERA-40 reanalysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012. (doi:10.1256/qj.04.176)
- Vidale, L., Lüthi, D., Frei, C., Seneviratne, S. I. & Schär, C. 2003 Predictability and uncertainty in a regional climate model. *J. Geophys. Res.* **108**, 4586. (doi:10.1029/2002JD002810)
- Webb, M. J. *et al.* 2006 On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Clim. Dynam.* **27**, 17–38. (doi:10.1007/s00382-006-0111-2)
- Weisheimer, A. & Palmer, T. N. 2005 Changing frequency of occurrence of extreme seasonal-mean temperatures under global warming. *Geophys. Res. Lett.* **32**, L20 721. (doi:10.1029/2005GL023365)
- Wielicki, B. A. *et al.* 2002 Evidence for large decadal variability in the tropical mean radiative energy budget. *Science* **295**, 841–844. (doi:10.1126/science.1065837)
- Wigley, T. M. L., Ammann, C. M., Santer, B. D. & Raper, S. C. B. 2005 Effect of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.* **110**, D09107. (doi:10.1029/2004JD005557)
- Wilby, R. L. & Harris, I. 2006 A framework for assessing uncertainties in climate change impacts: low flow scenarios for the River Thames, UK. *Water Resour. Res.* **42**, W02419. (doi:10.1029/2005WR004065)
- Wilby, R. L., Dawson, C. W. & Barrow, E. M. 2002 SDSM—a decision support tool for the assessment of regional climate change impacts. *Environ. Model. Software* **17**, 147–159. (doi:10.1016/S1364-8152(01)00060-3)
- Wilby, R. L. *et al.* 2004 Guidelines for use of climate scenarios developed from statistical downscaling methods, IPCC task group on data and scenario support for impact and climate analysis (TGICA). See http://ipcc-ddc.cru.uea.ac.uk/guidelines/StatDown_Guide.pdf.
- Wild, M. 2005 Solar radiation budgets in atmospheric model intercomparisons from a surface perspective. *Geophys. Res. Lett.* **32**, L07704. (doi:10.1029/2005GL022421)
- Williams, K. D. *et al.* 2005 Evaluation of a component of the cloud response to climate change in an intercomparison of climate models. *Clim. Dynam.* **26**, 145–165. (doi:10.1007/s00382-005-0067-7)
- Yokohata, T., Emori, S., Nozawa, T., Tsushima, Y., Ogura, T. & Kimoto, M. 2005 Climate response to volcanic forcing: validation of climate sensitivity of a coupled atmosphere–ocean general circulation model. *Geophys. Res. Lett.* **32**, L21 710. (doi:10.1029/2005GL023542)