Review

Ensemble climate predictions using climate models and observational constraints

By Peter A. Stott[1],[*] and Chris E. Forest[2]

[1]*Hadley Centre for Climate Change (Reading Unit), Meteorology Building, University of Reading, Reading RG6 6BB, UK*
[2]*Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Two different approaches are described for constraining climate predictions based on observations of past climate change. The first uses large ensembles of simulations from computationally efficient models and the second uses small ensembles from state-of-the-art coupled ocean–atmosphere general circulation models. Each approach is described and the advantages of each are discussed. When compared, the two approaches are shown to give consistent ranges for future temperature changes. The consistency of these results, when obtained using independent techniques, demonstrates that past observed climate changes provide robust constraints on probable future climate changes. Such probabilistic predictions are useful for communities seeking to adapt to future change as well as providing important information for devising strategies for mitigating climate change.

Keywords: climate change; attribution; prediction; ensembles; uncertainty; probability

1. Introduction

Traditionally, the information obtained from a climate model about future climate change has been presented as a projection, without any information about the likelihood of such a projection (e.g. Cubasch *et al.* 2001). However, adaptation to the consequences of climate change requires estimates of risks, and therefore of the likelihoods of different impacts. Likewise, devising mitigation strategies requires an understanding of the risks contingent on a chosen emissions path, with particular recent attention being given to the issue of avoiding dangerous climate change and minimizing the risk of reaching thresholds that could lead to irreversible climate changes (e.g. Schellnhuber *et al.* 2006). Consequently, there is an increasingly urgent requirement to provide probabilistic predictions of future climate for a wide variety of different users and applications.

Uncertainty in model predictions of future climate change arises from three main sources. First, future emissions of greenhouse gases (GHGs) and changes in other anthropogenic and natural factors that can affect climate

are uncertain. The second source of uncertainty results from natural internal variability, which can be partially estimated by running climate models many times from different initial conditions. The third source arises from modelling uncertainty in estimating the response of the climate system to a particular emission scenario.

The main focus here is on combining models and observations to provide constraints on probable future climate change and thereby quantify the second and third sources of uncertainty arising from modelling uncertainty. (There will be some brief discussion of a methodology for generating uncertainties in future emission scenarios and that has been incorporated in one approach described below, while other approaches assume a particular emission scenario and calculate probable future climate change contingent on that emissions pathway.) The common approach in all the work described here is that past climate change has been used to constrain future climate change. An alternative approach is simply to use ensembles of models to provide ranges of probable future climate without reference to observations (e.g. as was done in the third assessment report (TAR) of the IPCC to provide the range of future warming of 1.4–5.8°C; Cubasch *et al.* 2001). This has the drawback that unweighted ensembles of models, for example, from 'ensembles of opportunity' such as for the TAR, have no basis for assigning likelihoods and will not necessarily span the range of uncertainty (Allen & Ingram 2002). Much recent interest has been devoted to combining large perturbed physics ensembles of simulations with observational constraints based on current climate and variability, such as the model's representation of the seasonal cycle and other climate variables (Murphy *et al.* 2004; Stainforth *et al.* 2005; Knutti *et al.* 2006). In contrast, here we use past climate change (rather than the equilibrium state of current climate) to constrain future climate change. By quantifying the effects of past forcings, to the extent they are known, on past climate change, we aim to exploit information on the timescales and processes that are most relevant for future climate change.

The focus of §2 of this review paper is on using the observational record to provide probabilistic predictions of future global mean temperature change and to constrain basic properties of the climate system such as the climate sensitivity, rate of ocean heat uptake and the total aerosol forcing. Much progress has been made since the TAR in moving on from purely model-based projections to probabilistic assessments constrained by observations. Clearly, however, there is a requirement for information at regional as well as global scales, and for probabilistic predictions of variables other than temperature or mean values (e.g. variability or extremes). These are discussed in §3.

2. Constraints on future climate change based on past climate change

Within the framework of using past climate changes to constrain climate system properties and predictions of future climate change, two main approaches are discussed in this paper.

The first approach is to make large ensembles of model simulations and compare the fit between past observed changes and modelled changes. Those simulations that have better fits will be given higher likelihoods than those with

worse fits, and probability distributions of climate system properties (including the climate sensitivity, the rate of deep-ocean heat uptake and the net aerosol forcing) can then be derived based on the likelihood statistics.

This approach requires the computing capability to make many climate model simulations. Therefore, most studies have employed computationally efficient climate models such as an energy balance model (EBM) or an Earth system model of intermediate complexity (EMIC), although state-of-the-art (SOTA) coupled atmosphere–ocean general circulation models (AOGCMs) can be used, given very large computing resources, such as distributed computing projects as in Climate*Prediction*.net (Stainforth *et al.* 2005). The EBMs only require closure of the energy balance equation and neglect the balances (i.e. conservation laws) of other physical components in the climate system (e.g. mass, moisture, chemical constituents, momentum or angular momentum) that are included in higher complexity models. The EMICs also simplify the complexities of processes included in SOTA coupled AOGCMs, but aim to include the most relevant climate processes with the focus on the fidelity of the most important feedbacks between the components. The resulting models take a comprehensive approach to the Earth system rather than representing individual components with varying levels of complexity. As computing resources increase, these models will then increase their complexity equally across components. Like all climate models, EMICs are subject to uncertainties in modelling climate processes and feedbacks, uncertainties that derive from imperfect understanding and the approximations required in representing processes taking place on smaller spatial or temporal scales than the model's resolution. In addition, the need for SOTA models still remains debated as the predicted variables are transformed from global to regional scales and from decadal to shorter timescales.

The second approach is to use much smaller ensembles of simulations of past climate change by SOTA climate models in order to deduce probable future changes. SOTA climate models include a much wider range of processes than EBMs and EMICs and are able to resolve large-scale (regional) weather phenomena on regional scales. The likelihoods of future changes are estimated by scaling the response to historical climate forcings as simulated by a model and using the scaling factors to adjust the future predictions by the same model. The basic assumption is that if a climate model overestimates the response to past climate forcings as compared with observed climate changes, then it will also overestimate the response to future forcings provided the forcings remain similar. This further implies that a linear relation can be estimated between observed and modelled past climate change, and that the fractional errors from the historical period will apply to future scenarios. Such a method, which is based on optimal detection analyses, can be used to deduce the observed temperature changes attributable to different forcing factors. It has been dubbed the ASK method and was first described by Allen *et al.* (2000).

(*a*) *Large-ensemble approach to assessing uncertainty*

With current computational resources, EMICs are the tools that have been most commonly used for estimating large ensembles of future climate simulations including a range of different forcings and climate model properties. Two projects have approached the problem in two distinct manners.

The first project (Knutti *et al.* 2002) performed a direct Monte Carlo simulation in which a small set of parameters is perturbed, an historical simulation is calculated and these results are compared with observed surface and ocean temperature changes. If the simulation is not inconsistent with the observations, a simulation of future climate is calculated and joins the growing ensemble. This technique requires a highly efficient model (the model used, BERN2D, takes order minutes per 100 years of simulation time). We note that only single future emission scenarios can be considered in this fashion because only the climate system uncertainty is included in the analysis. Additional scenarios require additional Monte Carlo simulations with the entire system.

The second project (Webster *et al.* 2003) approached the task differently for two reasons. First, the model used (the MIT integrated global system model, MIT IGSM; Prinn *et al.* 1999) has higher complexity and requires more computational resources (requiring approx. 24 h for a 100-year simulation). Second, the method includes uncertainty in future climate forcings by using probabilistic emission scenarios (Webster *et al.* 2002).[1]

Since uncertainties in future emission scenarios are also considered, an estimate of total uncertainty in future climate change is generated without dependence on specific forcing scenarios. Uncertainties relating to future climate forcings are considered resulting from anthropogenic emissions of GHGs (carbon dioxide, CO_2; methane, CH_4; nitrous oxide, N_2O; hydrofluorocarbons, HFCs; perfluorocarbons, PFCs; sulphur hexafluoride, SF_6) and anthropogenic emissions of short-lived climate-relevant air pollutants (sulphur dioxide, SO_2; nitrogen oxides, NO_x; carbon monoxide, CO; ammonia, NH_3; black carbon, BC; organic carbon, OC; non-methane volatile organic compounds, NMVOCs). Using the MIT emissions prediction and policy analysis model (Babiker *et al.* 2001; Paltsev *et al.* 2006), uncertainties are estimated for anthropogenic emissions (Webster *et al.* 2002) of all relevant GHGs as well as aerosol and GHG precursors.

By including uncertainties in future emissions, this approach is able to provide the total likelihood of future climate change, which is of direct relevance to policy makers, especially if comparisons can be made between future predictions of future climate change that include emission mitigation measures and predictions of future climate change that do not include mitigation measures. The alternative approach (as followed by the large-ensemble (LE) approach of Knutti *et al.* (2002) and by the ASK small-ensemble approach described in §2*b*) is to estimate the uncertainty in the climate response under a single projection of emissions. This enables useful comparisons to be made between the consequences of different possible emissions pathways.

The century time-scale response of the climate system to changes in the radiative forcing is primarily controlled by two uncertain global properties of the climate system: the climate sensitivity and the rate of oceanic heat uptake

---

[1] The critical input data for uncertainty analyses are the probability density functions (PDFs) for the uncertain parameters and when possible, these should be based on objective analyses. A key error frequently made in assembling such PDFs is to use the distribution of point estimates drawn from the literature rather than from estimates of uncertainty (e.g. standard deviation). There is nothing inherently wrong with using literature estimates, but the point estimates of uncertain parameters should span the population of interest and not simply be derived from a distribution of mean estimates from different studies (see Webster *et al.* (2003) for further discussion).

(Sokolov & Stone 1998; Sokolov *et al.* 2003).[2] In coupled AOGCMs, these are emergent properties which depend on the model equations and parameters of the model. The sensitivity, *S*, of the MIT climate model, however, can be easily varied by changing the strength of the cloud feedback. (We note that this mimics structural differences in the AOGCMs and also that *S* can be varied in all EMICs, although it is often directly changed by the total feedback parameter rather than the cloud feedback alone.) Mixing of heat into the deep ocean is parameterized in the MIT model by an effective diffusion applied to a temperature difference from values in a present-day climate simulation. Therefore, the rate of the oceanic heat uptake is defined by the value of the globally averaged vertical diffusion coefficient, K_v, for diffusion of temperature anomalies below the mixed layer.[3] (NB: the rate of ocean heat uptake is proportional to $\sqrt{K_v}$ as discussed in Sokolov *et al.* 2003.) By varying these two parameters, the MIT climate model can reproduce the global-scale zonal-mean responses of different AOGCMs (Sokolov & Stone 1998; Sokolov *et al.* 2003). Significant uncertainty also exists in the historical forcing mainly associated with uncertainty in the radiative forcing in response to a given aerosol loading, F_{aer}. Thus, in the MIT IGSM, these three parameters (S, K_v and F_{aer}) characterize uncertainty both in the response of the climate system and in the historical climate forcing. Uncertainties in these basic climate system properties that govern the transient response of the climate system are estimated from constraints provided by recent observations of climate change (Forest *et al.* 2002, 2006).

Together, these uncertainty ranges for future forcing and for the basic climate system properties provide the input distributions that are used for the Earth system components of the MIT IGSM (Prinn *et al.* 1999; Reilly *et al.* 2001). To generate large ensembles of future climate simulations, each of these sources requires input probability distributions for the relevant parameters in the model components.

One crucial aspect of the Webster *et al.* (2003) work is using the joint probability density functions (PDFs) for the climate model parameters controlling S, K_v and F_{aer} from Forest *et al.* (2002). The estimation method uses observations of upper air, surface and deep-ocean temperatures for the twentieth century to jointly constrain these climate parameters ($\theta = (S, K_v, F_{aer})$), while including unforced climate variability as a source of uncertainty (Forest *et al.* 2002). The method for estimating PDFs relies on estimating goodness-of-fit statistics, r^2 (Forest *et al.* 2000, 2001), obtained from an optimal fingerprint detection algorithm (Allen & Tett 1999). We compute r^2 by taking the difference in the modelled and observed patterns of climate change, $T(\theta)$, T_{obs} and weighting the difference by the inverse of the unforced variability for the pattern, C_N^{-1}:

$$r^2(\theta, T_{obs}) = (T(\theta) - T_{obs})' C_N^{-1} (T(\theta) - T_{obs}). \qquad (2.1)$$

[2] Climate sensitivity, *S*, is defined as the equilibrium surface temperature change in response to doubling the CO_2 concentration. One alternative to this definition is the effective climate sensitivity which is defined in Murphy (1995) as the climate sensitivity that satisfies the transient energy balance equation at the time of CO_2 doubling. In most AOGCMs, the equilibrium and effective climate sensitivities are very similar, although significant differences can exist (Houghton *et al.* 2001).

[3] Each EMIC has its own method to vary the rate of oceanic heat uptake and the inter-comparison of this quantify can only be done via comparison of the transient response to forcings in each model. Thus, the effective K_v is unique to the MIT model and cannot be compared directly with diffusion in other climate models.

Differences in r^2 provide a statistic for hypothesis testing, and thereby provide probability estimates for parameter combinations (Forest *et al.* 2000, 2001)

$$\Delta r^2 = r^2(\theta) - r^2_{\min} \sim mF_{m,\nu}, \qquad (2.2)$$

with $r^2_{\min}$ being the minimum r^2 value in the θ-parameter space for an individual diagnostic. Thus, Δr^2 follows an F distribution with m and ν degrees of freedom and can be used to estimate a likelihood function over the θ-parameter space. The likelihood, $p(\Delta T_i | \theta, C_N)$, from individual diagnostics for surface, upper air and deep-ocean temperature changes can then be combined via the Bayes theorem (1763) to estimate the posterior, $p(\theta | \Delta T, C_N)$. This method requires an estimate of the unforced variability (aka natural) for the climate system, C_N, over very long periods. Ideally, observed climate variability would be used but reconstructed data are not of sufficient accuracy. Our estimate was obtained from long control runs of particular AOGCMs (Forest *et al.* 2002). Estimates of the variability from other AOGCMs could potentially change the results although tests of this sort are found not to change the results qualitatively.

In Webster *et al.* (2003), expert priors for both S and K_v were used from Webster & Sokolov (2000). The latest PDF estimates calculated from Forest *et al.* (2006) are shown in figure 1. The three diagnostics are treated as independent observations and, therefore, weighted equally in the Bayesian updating procedure. The result is a joint PDF for these three parameters that contains correlation among the marginal PDFs (e.g. a high climate sensitivity is only consistent with observed temperature under some combination of rapid heat uptake by the ocean and a strong aerosol cooling effect). Forest *et al.* (2006) concluded from figure 1 that most AOGCMs are mixing heat into the deep ocean too efficiently. The implications of this for models' estimates of the transient climate response (TCR) are discussed further in §2*b*.

From these distributions of the basic climate system properties, an efficient Monte Carlo sample is generated using the Latin-Hypercube sampling algorithm (Iman & Helton 1988).[4] This is shown in figure 2. Given both PDFs for emission scenarios and $p(\theta | \Delta T, C_N)$, a LHS can be generated and an ensemble of future climate change simulations is calculated. A sample size of 250 is sufficient to estimate probability distributions for climate outcomes of interest. Further details can be found in Webster *et al.* (2003).

The Webster *et al.* (2003) results are shown in figure 3 for two 250-member ensembles: the 'policy' and 'no policy' cases. The policy case assumes Kyoto Protocol caps for those countries agreeing to them with additional 5% reductions in emissions below these caps every 15 years. For the remaining countries, they adhere to caps beginning in 2025 at 5% below their 2010

---

[4] To reduce the computational requirements for a Monte Carlo simulation, sampling from the probability distributions for the uncertainty analysis is performed using Latin-Hypercube sampling (LHS; Iman & Helton 1988). LHS divides each parameter distribution into n segments of equal probability, where n is the number of samples to be generated. Sampling without replacement is performed, so that with n samples, every segment is used once. Samples for the climate parameters are generated from the marginal PDFs, and the correlation structure among the three climate model parameters is imposed (Iman & Conover 1982). This ensures that the low probability combinations of parameters are not overrepresented, as would be the case if the correlations were neglected.
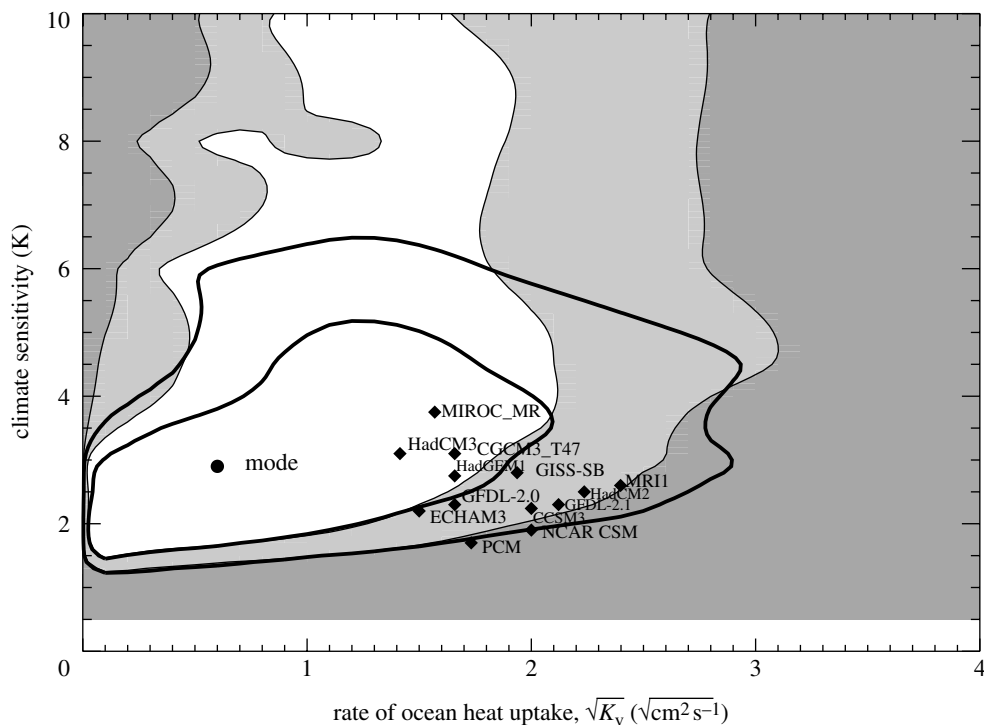
Figure 1. The marginal *posterior* PDF for GSOLSV results with uniform priors for the S–K_v parameter space. The light to dark shading denotes rejection regions for the 10 and 1% significance levels, respectively. The 10 and 1% boundaries for the posterior with expert prior on S are shown by thick black contours. The positions of AOGCMs represent the parameter values required in the MIT two-dimensional model to match the transient response in surface temperature and thermal expansion component of sea-level rise (following the method in Sokolov *et al.* (2003)). Lower K_v values imply less deep-ocean heat uptake and hence a smaller effective heat capacity of the ocean. Adapted from Forest *et al.* (2006).

emission levels and also reduce an additional 5% below these every 15 years. These targets achieve approximate stabilization at 550 ppm for CO_2 concentrations in 2100 (figure 4; additional details in Webster *et al.* (2003)). We stress that the no policy ensemble includes uncertainty in both the forcing and the response of the climate system, and thus provides an estimate of total uncertainty for assessing future impacts of climate change policies. Probable rates of temperature and sea-level increase are accordingly greater for the no policy case than the policy case, especially towards the upper tails of the probability distributions.

In addition, in figure 5, we estimate the uncertainty for the response to the SRES A1B scenario using a LH sample from the Forest *et al.* (2006) results. With the forcing scenario fixed, the probability for temperature changes can be compared with results later in this paper. As expected, the uncertainty ranges are smaller than the total uncertainty previously discussed. The future temperature rise is estimated to be 2.3–4.2 K in 2100 (5–95 percentile) relative to 2000.
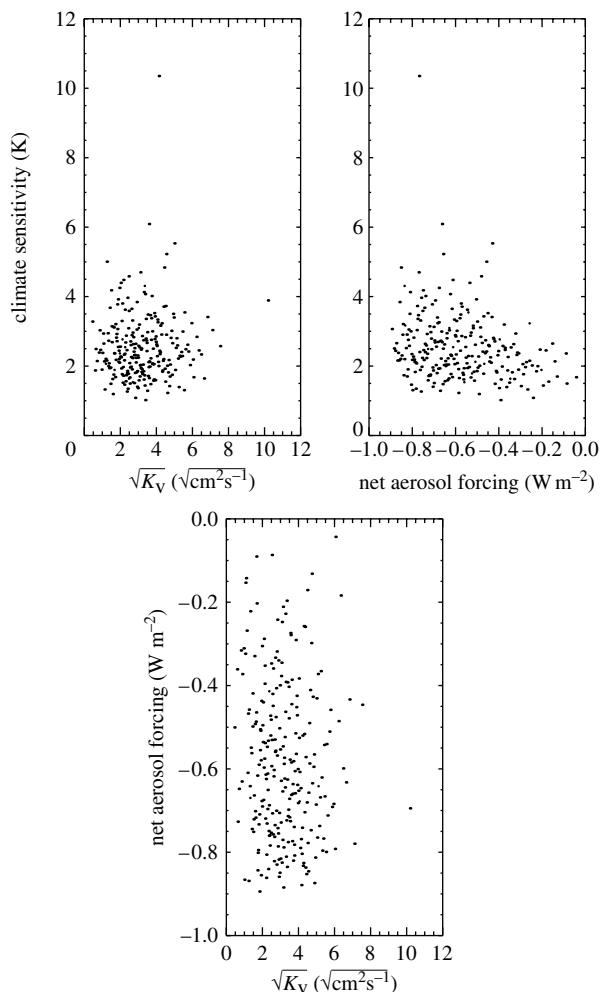
Figure 2. Monte Carlo sample of climate system properties generated using a 250-member Latin-Hypercube sampling. Adapted from Forest *et al.* (2006).

(*b*) *Small ensembles of coupled ocean–atmosphere climate models*

To complement the LE approach, a similar method to estimate uncertainty in climate predictions makes use of the small ensembles available from SOTA climate models. This has been dubbed the ASK method (Allen *et al.* 2000). The basic idea behind the ASK method is that knowledge of how the observational record constrains the probable contributions of GHGs and other forcing factors to past temperature change in turn provides observational constraints on probable future rates of warming. This is achieved by assuming that there is a linear relationship between fractional errors in simulating past climate change and in predicting future climate change. The robustness of such a relationship has been investigated in detail by Kettleborough *et al.* (2007) (see below). An important advantage of such an approach is that predictions are STAID (i.e. provide STAble Inferences from Data; Allen *et al.* (2006*a*)). STAID
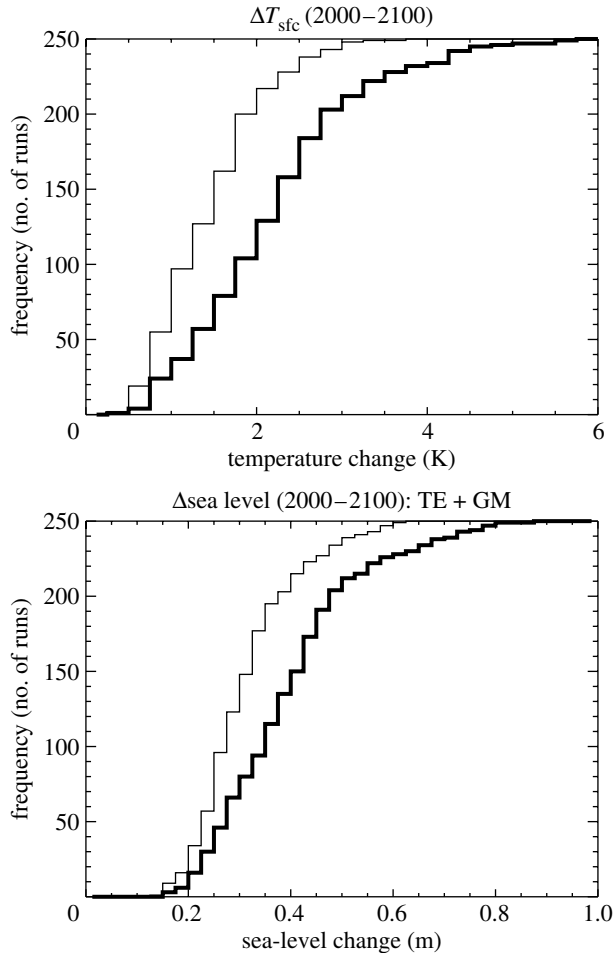
Figure 3. Probability distributions, shown as cumulative density functions (CDFs), of near-surface temperature change ($\Delta T_{\rm sfc}$) and sea-level rise (both thermal expansion and glacier melt) for policy (thin lines) and no policy (thick lines) ensembles.

estimates depend largely on observations and are relatively robust to changes in models. For example, introducing a new model into the analysis that has a lower sensitivity than other models has only a second-order effect on the result (Stott *et al.* 2006*b*). At the same time, uncertainty is reduced in a predictable way as the signal strengthens (Stott & Kettleborough 2002). For a STAID forecast, there should be a transfer function that links a forecast quantity to an observable quantity. In the case of the ASK method, the transfer function is a linear relationship linking past warming to future warming.

   The first component of the ASK procedure is to estimate the factors by which the model's response to different forcings can be scaled up or down while remaining consistent with the observed record. These scaling factors are determined from an optimal detection analysis of observed and coupled model's patterns of temperature change. An optimal detection analysis is a form of linear regression (Allen & Tett 1999) which seeks to explain the observed pattern of change in terms of the
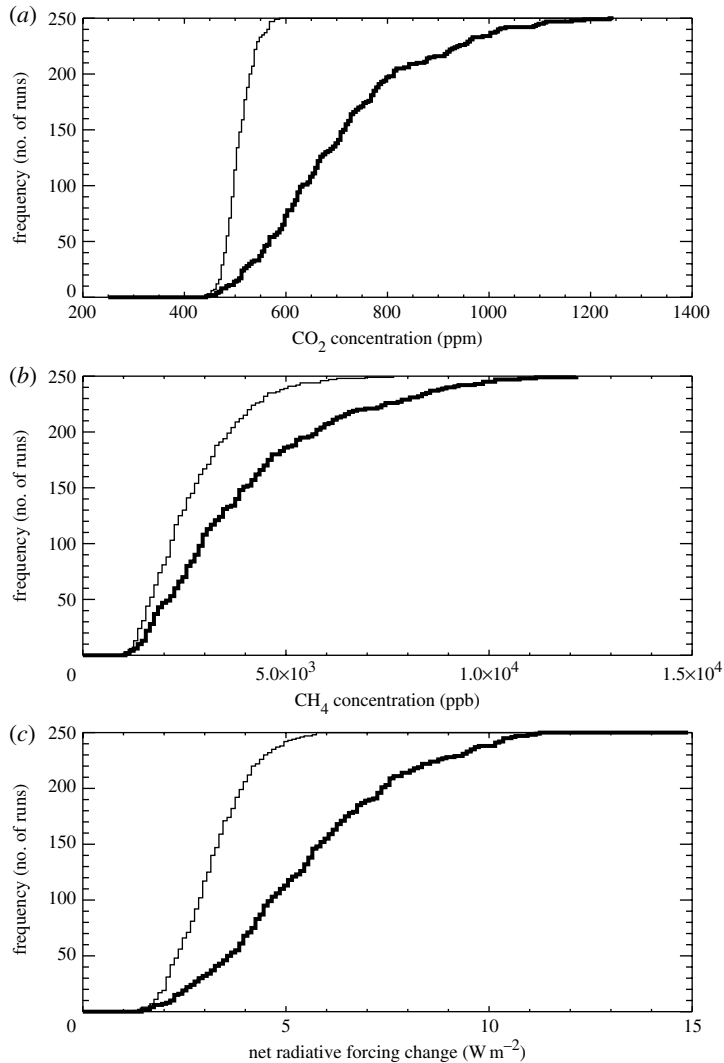
Figure 4. Net changes in major forcings: (*a*) CO_2 and (*b*) CH_4 concentrations in 2100 and (*c*) the total net radiative forcing changes (ΔF) from 2000 to 2100 (given as average of preceding decades) for policy (thin lines) and no policy (thick lines). Adapted from Webster *et al.* (2003).

contributions from different climate forcing agents, such as changes in well mixed GHGs or fluctuations in solar output. Such optimal detection analyses have been used extensively to determine the causes of observed changes in temperature and other climate variables such as changes in precipitation and sea-level pressure. They provided important evidence that led the fourth assessment report of the IPCC to conclude that 'most of the observed increase in global average temperatures since the mid-twentieth century is very likely due to the observed increase in anthropogenic GHG concentrations' (IPCC 2007).

   In an optimal detection analysis, observed temperature changes are expressed as a linear combination of model responses to specified forcings. A typical approach described here, and many of the ASK studies (Stott & Kettleborough 2002;
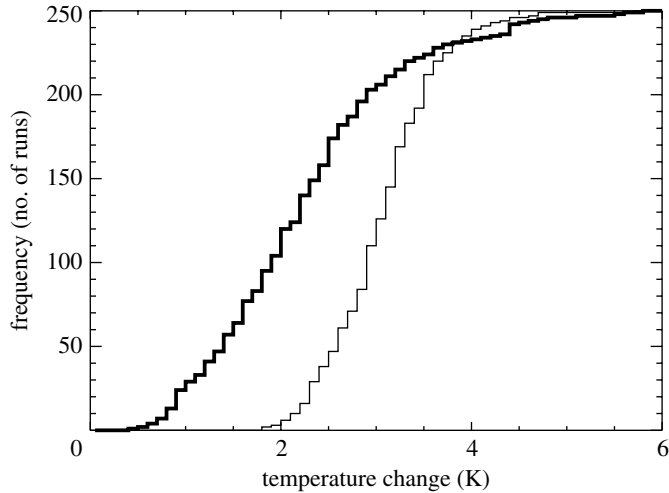
Figure 5. CDFs of temperature change (2000–2100) for no policy (thick lines) and 'SRES A1B' ensembles (thin lines).

Stott *et al.* 2006*a*,*b*; Kettleborough *et al.* 2007) is to partition the observed changes into contributions from the response to well-mixed GHGs, the combined effects of all anthropogenic forcings (ANTHRO) and the combined effects of all natural forcings (NAT). Anthropogenic forcings typically included in climate models are the effects of sulphate aerosols (both their direct effect through increased reflection of incoming solar radiation, and in some model simulations through their indirect effects by making clouds brighter or longer lasting) and the effects of tropospheric and stratospheric ozone changes. These are spatially hetero-geneous forcings and should have a different response pattern as compared with that from the well-mixed GHGs. Natural forcings often included in climate model simulations are changes in solar irradiance and stratospheric aerosols resulting from explosive volcanic eruptions. The pattern of response to a particular forcing is usually obtained from a multi-member initial condition ensemble, in which several simulations (typically three or four) of the climate model are made with an identical forcing combination but starting from different initial conditions taken from a long multi-century control run of the model. Averaging the responses from a multi-member initial condition ensemble reduces the noise contamination of the climate change signal because ensemble averaging filters out internal variability while keeping the common response to external forcings that is seen in all initial condition ensemble members.

Observational and model data are normally filtered first to retain the large spatial and temporal scales on which the signal of climate change is expected to be detected above the noise of unforced internal variability (Stott & Tett 1998). In the following example, observed decadal-mean near-surface temperature changes over the 1900–2000 period, y, are expressed as a linear sum of simulated changes from: GHG, x_1; ANTHRO, x_2; NAT, x_3; plus noise, v_0

$$y = \sum_{i=1}^{3} (x_i - v_i)\beta_i + v_0, \tag{2.3}$$

where $\boldsymbol{\beta}_i$ is the vector of unknown scaling factors to be estimated in the regression. If $\boldsymbol{\beta}_i$ is less than 1, this implies that the model overestimates the response to a particular forcing and so has to be scaled down, and if $\boldsymbol{\beta}_i$ is greater than 1, the model underestimates the response and has to be scaled up. Including the additional noise term $\boldsymbol{v}_i$ in the regression equation, Allen & Stott (2003) take into account the statistical uncertainty introduced by taking the model-simulated responses from the mean of a finite ensemble which differs from the underlying noise-free response that would be obtained from a hypothetical infinite ensemble.

The regression analysis is carried out in the reduced space spanned by the leading p empirical orthogonal functions of the covariance matrix of internal variability, C_{N_1}, which is obtained from the model-based estimates of internal variability (e.g. from long control simulations). To avoid bias in the estimate of the covariance of the scaling factors, $\boldsymbol{\beta}_i$ (Hegerl *et al.* 1996, 1997), a second statistically independent estimate of the covariance matrix is used to determine the uncertainty in the scaling factors (e.g. from an independent segment of the control simulation). Model-based estimates of internal variability are used because the instrumental record is too short to provide a reliable estimate of internal variability and is also affected by external forcing. Two methods of validation of model-based estimates of internal variability are made. As part of the optimal detection procedure, a consistency test (Allen & Tett 1999) is used to test whether the residuals of regression are consistent with internal variability. In addition, a further check is provided by comparing power spectra of models and data. Since the observational data contain forced as well as unforced variability, to compare directly with model-based estimates of internal variability, an estimate of the forced changes in the observational record is made, either by detrending the data in some way or by subtracting an independent model-based estimate of the externally forced response (Allen *et al.* 2006*b*). These tests indicate that for global mean and continental-scale near-surface temperatures, the estimated model variance is consistent with the observed variance although we have very limited observational data for such tests on the 50–100 year timescales of interest.

Having determined the distribution of scaling factors, $\boldsymbol{\beta}_i$, the next step is to apply these same scaling factors to climate model simulations of future temperature change on the assumption that a model which over- or under-estimates past temperature change will similarly over- or underestimate future temperature change. In the case of equation (2.1), where scaling factors for GHG, ANTHRO and NAT are calculated, the future modified forecast, $\boldsymbol{y}^{\text{for}}$ is given by

$$\boldsymbol{y}^{\text{for}} = \sum_{i=1}^{2} (\boldsymbol{x}_i^{\text{for}} - \boldsymbol{v}_i^{\text{for}})\boldsymbol{\beta}_i + \boldsymbol{v}_0^{\text{for}}, \tag{2.4}$$

where $\boldsymbol{x}_i^{\text{for}}$ is the model forecast response and $\boldsymbol{\beta}_i$ are the scaling factors for GHG and ANTHRO calculated in equation (2.3). The assumption being made here is that it is not possible to forecast deterministically future naturally forced changes (due to changes in output from the Sun and from explosive volcanic eruptions) and therefore the anthropogenically forced component is predicted according to a particular scenario of emissions. The noise term $\boldsymbol{v}_i^{\text{for}}$ represents the uncertainty in the model response pattern due to the modelled response pattern of future temperature change being based on small ensembles and therefore contaminated by internal variability. The additional noise term, $\boldsymbol{v}_0^{\text{for}}$ represents

departures of the future climate from a scaled version of the modelled forced response due to internal variability. Given knowledge of the probability distribution of $\boldsymbol{\beta}_i$, obtained from equation (2.3), the uncertainty in future temperature change can be calculated from equation (2.4) by estimating the statistics of $\boldsymbol{v}_i^{\text{for}}$ and $\boldsymbol{v}_0^{\text{for}}$ by Monte Carlo sampling of the distributions.

Using the same scaling factors for past and future temperature changes is equivalent to saying that the fractional error in the model's simulation of temperature change stays constant over time, a supposition that is supported by the evolution of global mean temperatures in GCMs that tend to evolve similarly over time in response to a given forcing despite differences in sensitivity and thus response to amplitude. Analysis of an intermediate complexity model (Forest *et al.* 2000), which includes similar nonlinear feedbacks as atmospheric GCMs, also supports a linear relationship. A further investigation of the robustness of the transfer function between past attributable temperature changes and future warming was carried out by Kettleborough *et al.* (2007) for a range of emissions scenarios using an EBM to sample the uncertainties introduced. Although an EBM is limited by a lack of nonlinear feedbacks, the transfer function was found to be sufficiently robust over a number of realistic forcing scenarios to introduce only small additional uncertainty. (Note that an EBM was used to test the linearity of the transfer function between past and future global mean temperature changes, whereas a fully coupled SOTA climate model was used to determine the probable spatial and temporal patterns of temperature change attributable to anthropogenic and natural forcings.) For the A1FI scenario, at 2100, they found an error of between -10 and $+10\%$ and, for the B1 scenario (which stabilizes to some extent), an error of between -30 and $+20\%$ was found depending on the climate sensitivity and ocean heat diffusivity of the model. The ASK method is best suited for casting forward uncertainties in future climate changes over the next few decades when forcings are likely to continue increasing linearly but is less well suited for scenarios with considerable stabilization. For these, the approach of running very large ensembles of models such as EMICs is better suited (§2*b*).

The first application of the ASK approach was by Allen *et al.* (2000), who calculated uncertainties in twenty-first century warming rates under the IS92A scenario scaling GHG and aerosol patterns deduced from the HadCM2 model separately (as in equations (2.1) and (2.2)). They also calculated future warming rates for a range of other coupled climate models using scaling factors on ANTHRO only (rather than both GHG and ANTHRO as in equations (2.3) and (2.4)), thereby assuming that the combined response to GHGs and aerosols could be represented by a single spatio-temporal pattern. They found a range of temperatures of 1–2.5 K warmer than in pre-industrial times in the decade 2036–2046 and that the range was relatively robust to errors in a model's climate sensitivity, rate of ocean heat uptake or global response to sulphate aerosols as long as these errors are persistent over time. Since across the range of models they assumed that the relative roles of GHGs and aerosols remain constant in the future, substantial changes in the balance of greenhouse warming and aerosol cooling would increase the uncertainty in their results.

Allen *et al.* (2000) demonstrated that the spread of observationally constrained predictions of future temperature estimated from different models is smaller than the spread of predictions from the raw unscaled models, thus
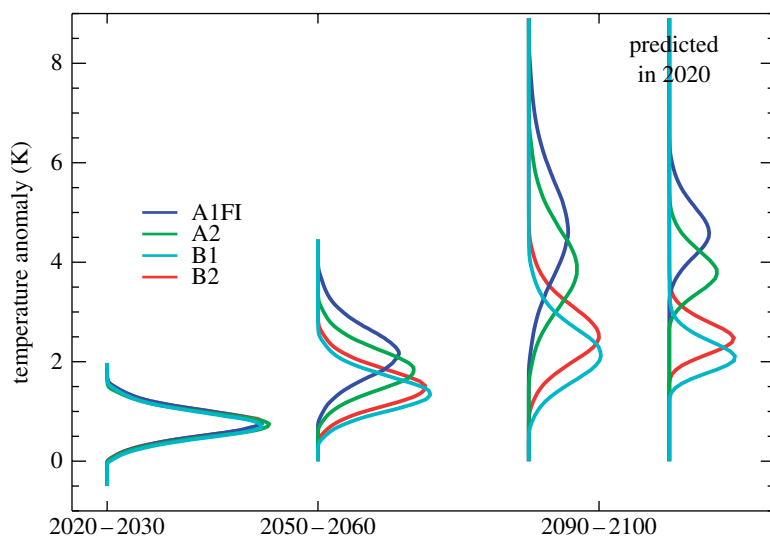
Figure 6. PDFs of temperature change. Shown are PDFs for four SRES scenarios (A1FI, A2, B1 and B2) for 2020–2030, 2050–2060 and 2090–2100 decades relative to 1990–2000 decade, calculated by constraining HadCM3 simulations to the observed temperature change over the 1900–1999 period. The PDFs at the far right are for the 2090–2100 decade calculated by constraining HadCM3 simulations to be consistent with the observed temperature change over the 1920–2019 period where the observations are assumed to follow a B2 scenario prediction after 1999. Adapted from Stott & Kettleborough (2002).

demonstrating the STAID nature of these forecasts. While this approach is conservative, since only one of a large number of possible observational constraints has been used to constrain the forecasts, it should also evolve in a more predictable way, since it is not likely to be affected to first order by introducing new models into the analysis. This could be an advantage for discussions with policy makers in providing a robust, if potentially sub-optimal, estimate of uncertainty (see discussion in Stott *et al.* (2006a)), and is in contrast to estimates of uncertainty based solely on ensembles of model simulations which will not necessarily span the range of uncertainty, as demonstrated by Allen & Ingram (2002).

The assumption that the relative roles of GHGs and aerosols should remain constant, as predicted by models was relaxed by Stott & Kettleborough (2002) who calculated separate scaling factors on the response patterns in HadCM3 to GHG and aerosols as well as including the response to natural forcings in the analysis. Probabilistic forecasts of global mean temperatures were obtained for four representative SRES emission scenarios. They also included uncertainty due to future natural forcings by adding variance to the distribution calculated from simulations of the response to past natural forcing. Global mean temperature rise was found to be insensitive to differences in emissions scenarios over the first few decades of the twenty-first century; a temperature rise of 0.3–1.3 K was predicted by the 2020s relative to the 1990s (figure 6). As discussed by Zwiers (2002), this range is consistent with an estimate using the alternative LE approach for estimating forecast uncertainty described in §2 (Knutti *et al.* 2002).
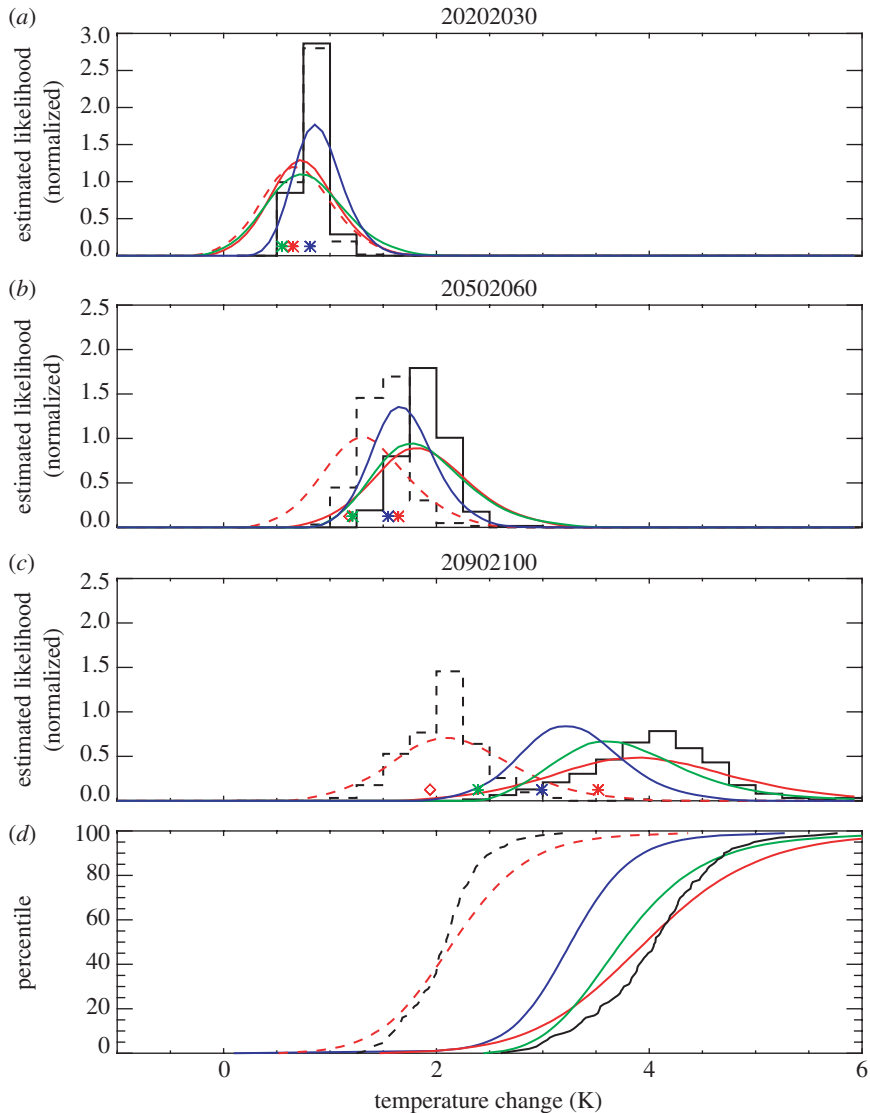
Figure 7. Probability density functions for global mean temperature for the A2 (solid lines) and B1 (dashed lines) scenarios for (*a*) 2020–2030, (*b*) 2050–2060 and (*c*) 2090–2100 relative to the 1980–2000 average as constrained by observed twentieth century temperature change and as calculated using HadCM3 model (red), PCM model (green) and GFDL R30 model (blue). In each panel, results are compared against histograms of distributions derived using MIT IGSM using Forest *et al.* (2006) data (black lines). Also shown as coloured stars are the raw model predictions for each model for the A2 scenario and as diamonds the raw HadCM3 prediction for the B1 scenario. (*d*) The same distributions for 2090–2100 plotted as cumulative distribution functions.

Although Stott & Kettleborough (2002) found little difference between scenarios in early decades, large differences emerge by the end of the century. Warming as high as 6.9 K (relative to 1990s temperatures) cannot be ruled out at the 95% confidence level in the SRES A1FI scenario (figure 6). Stott & Kettleborough (2002) also showed that as the signal of climate strengthens, the

uncertainty in future temperature rise is likely to reduce, potentially halving uncertainties in late twenty-first century warming by 2020 when compared with values estimated in 2000 (figure 6). The range of warming under the B2 scenario is 1.6–3.7 K by the 2090s decade relative to the 1990s decade. This is close to the estimate shown in figure 5 for the temperature range expected under the A1B scenario calculated according to the LE method and discussed in §2*a*. The A1B range is slightly higher reflecting its slightly higher emissions than B2. An examination of the sensitivity of the results to varying the size of the temperature variance assumed to result from future natural forcings was found by Kettleborough *et al.* (2007) to make relatively little difference to future warming rates, especially late in the century when the strengthening greenhouse signal and corresponding increase in the uncertainty in the transient response to the increasing GHGs dominates the overall uncertainty. The linear relationship between past and future warming assumed in this approach would not hold if there were large nonlinear feedbacks in future such as a shutdown of the THC or land biosphere switching from a weak sink for carbon to a strong source.

Whereas Stott & Kettleborough (2002) analysed only the HadCM3 model, Stott *et al.* (2006*b*) extended the analysis to three climate models (HadCM3, PCM and GFDL R30), all of which had ensembles of simulations including GHGs only (GHG), combined anthropogenic forcings (ANTHRO) and simulations from which the response to natural forcings could be estimated (either natural forcings for HadCM3 and PCM, or the response to all forcings for GFDLR30). The results shown in figure 7 show little difference between the A2 and B1 scenarios in the 2020s but an increasing separation of the two scenarios later in the century. Some structural uncertainty remains in the probable range of future temperature change according to which model is used (comparing red, blue and green curves in figure 7), although the lower sensitivity PCM predictions (green stars) are consistently scaled up more than predictions from the higher sensitivity GFDLR30 and HadCM3 models. The equivalent uncertainty ranges derived using the MIT IGSM (§2*a*) are consistent with those derived by the ASK methodology but are generally narrower, indicating that tighter constraints are derived using the EMIC LE-based approach.

The TCR is the global-mean surface temperature change that is realized at the time of CO_2 doubling under an idealized scenario in which CO_2 concentrations increase at 1% per year. TCR, like equilibrium climate sensitivity (ECS, the equilibrium temperature change resulting from CO_2 doubling), is a basic property of the climate system, but TCR is more relevant to determining the near-term climate change than ECS because it includes the delaying effects of mixing heat into the deep ocean. An ASK-based observationally constrained estimate of the TCR can be obtained in the same way as the transient response under SRES scenarios. In this case, the model's TCR is scaled by the scaling factor on GHG as obtained from the full ASK approach. Observationally constrained estimates of TCR are shown in figure 8 for estimates using the three models and are compared with TCRs (diamonds) from the multi-model ensemble archived at PCMDI for input into the IPCC fourth assessment report (AR4) and the range of TCR calculated by Forest *et al.* (2006). The unweighted average from the three models is also shown as was calculated by Stott *et al.* (2006*b*). We note that three cases is a small sample size to represent results accounting for both structural uncertainty in AOGCM models and the structural (or
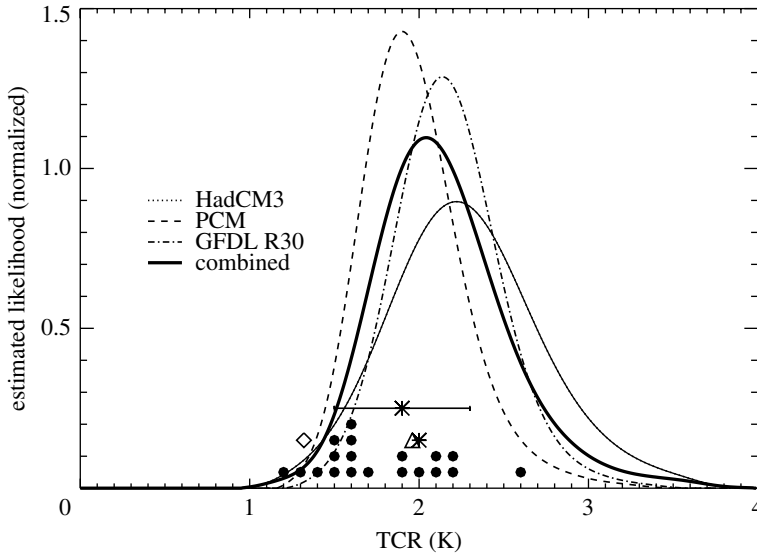
Figure 8. Probability distributions of TCR (expressed as warming rates over the century), as constrained by observed twentieth century temperature change, and as calculated using HadCM3 (dotted line), PCM (dashed line), GFDL (dot-dashed line) and from unweighted average of all three PDFs (solid line). The TCR of each individual model is shown as star (HadCM3), diamond (PCM) and triangle (GFDL). Also shown as filled circles are the TCRs of climate models forming part of the IPCC AR4 ensemble and the 5–95 percentile range derived by Forest *et al.* (2006), using a large ensemble of EMICs is shown as the black bar with the star showing the median.

methodological) uncertainty of the STAID approach. The similarity of the results indicates the method's robustness while the differences remain unexplored. As for the distribution of IPCC AR4 models (diamonds in figure 8), there are more in the lower half than the upper half of the observationally constrained distributions (solid curve and horizontal bar), indicating a tendency for models to overestimate ocean heat uptake (since increased transient heat uptake dampens the atmospheric warming response to radiative forcing requiring a stronger response to still match observations; Forest *et al.* 2002; Knutti *et al.* 2005). This is consistent with the result found by Forest *et al.* (2006) who showed, using the LE methodology described in §2*a*, that AOGCMs generally mix heat into the deep ocean too efficiently, as shown in figure 1.

The robustness of the relationship between warming attributable to GHGs and TCR was also investigated by Frame *et al.* (2006) using an EBM, finding a regression line of 2.22 between past attributable greenhouse warming and TCR. The relationship can also be deduced simply from the ratio of the forcing at doubling CO_2 (approx. $3.74 \mathrm{~W~m^{-2}}$ per 70 years) to the forcing over the twentieth century (approx. $1.66 \mathrm{~W~m^{-2}}$ per century) which at 2.25 is close to the estimate deduced by Frame *et al.* (2006) (M. R. Allen 2006, personal communication). A range in attributable greenhouse warming (5–95 percentiles) of 0.7–1.3 K for the range of models then translates into a range for TCR of 1.5–2.8 K. The 5–95 percentile range of TCR derived by Forest *et al.* (2006) is 1.5–2.3 K and is also shown in figure 8.

Splitting the observed response into contributions from the response to GHGs and to other anthropogenic forcing allows an estimate not only of the TCR but
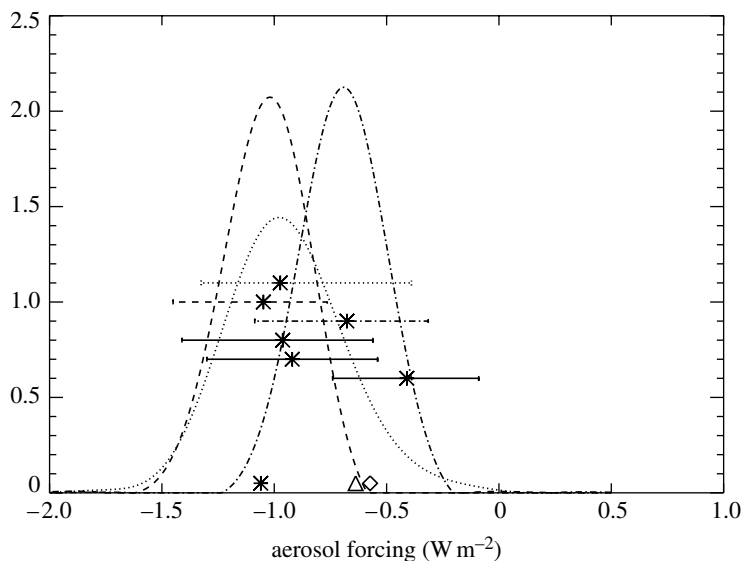
Figure 9. Estimates of net aerosol forcing (1990s relative to pre-industrial levels, plus any forcings not explicitly considered in the analysis) derived from optimal detection analyses carried out on HadCM3 (dotted line), PCM (dashed line) and GFDL R30 (dot-dashed line) models. The sulphate forcing of each individual model is shown as star (HadCM3), diamond (PCM) and triangle (GFDL). Discussion of the differences in the three results is contained in the text. Also shown as bars are the 5–95 percentile ranges and stars showing the medians for each of these analyses and as solid lines for inverse estimates of aerosol forcing from (reading from top to bottom on the figure): Knutti *et al.* (2003), Andronova & Schlesinger (2001), Forest *et al.* (2006).

also of the probable net forcing of the climate system, uncertainty in which is dominated by the forcing due to aerosols. From estimates of the scaling factors on GHG and ANTHRO, we can estimate the scaling factors on the GHG only signal (G) and the contribution from other anthropogenic factors (mainly aerosols, denoted S) by a linear transformation, assuming that the climate response to these forcings is linearly additive. This linearity assumption appears to hold reasonably well on large spatial scales (Gillett *et al.* 2004; Meehl *et al.* 2004). The probable range of aerosol cooling is inferred from these analyses by scaling the raw aerosol forcing in each model by the ratio of the scaling factors for S and G (the PDF of β_S/β_G). This takes account of observational constraints on the climate response by assuming that the GHG forcing is well known and that errors in the response of the model to different forcings scale equally. Note that the HadCM3 model-derived pattern also includes tropospheric and stratospheric ozone in addition to the direct and indirect effects of sulphate aerosols; the PCM model includes tropospheric and stratospheric ozone in addition to the direct effects of sulphate aerosols; and the GFDL R30 model includes only the direct effects of sulphate aerosols.

The estimates derived from the three models, shown in figure 9, are broadly consistent with other inverse estimate of aerosol forcing based on observational constraints (Andronova & Schlesinger 2001; Knutti *et al.* 2002, 2003; Forest *et al.* 2006) and appear to exclude larger magnitudes of aerosol forcing derived from forward calculations (Anderson *et al.* 2003; see also §2a). Note that all these

estimates of net aerosol forcing also implicitly include any forcings that are not explicitly included in the analysis. As an example, the GFDL model does not include the effects of tropospheric and stratospheric ozone. This means that the GFDL uncertainty estimate implicitly includes the missing ozone forcing, in addition to any other missing forcings.

Also, an estimate of the aerosol forcing based on the attribution analyses using a particular model will have unquantified uncertainties resulting from inadequacies in the treatment of aerosols in that model, such as, for example, omitting the indirect effects of sulphate aerosols, as is the case with the GFDLR30 and PCM models. The scaling procedure (equation (2.3)) will correct for gross model error. For example, to the extent that patterns of response to the direct and indirect effects of aerosols are similar, the procedure will correct for the omission of the indirect effects of aerosols in those models. However, differences between the real world's patterns of temperature response to the direct and indirect effects of aerosols will lead to errors in the aerosol forcing calculated by this method. By sampling different models, as is done in figure 9, the uncertainty introduced by these errors is sampled, but only to a limited extent where only three models are analysed.

3. Extension to regional scales and other climate variables

The above discussion has mainly concentrated on recent progress in developing probabilistic predictions of global mean temperature change. However, regional-scale probabilistic predictions are required in order to adapt to the probable effects of climate change, as well as plan mitigation strategies which reduce the risk of exceeding local thresholds of extreme weather. In addition to regional predictions of temperature, probabilistic predictions of other variables will be required such as precipitation and circulation changes and changes in extremes as well as mean quantities will be needed.

A first step in extending the ASK approach to sub-global scales was made by Stott *et al.* (2006*a*) who considered six continental-scale areas. An estimate was made of continental-scale temperature by carrying out an optimal detection analysis on each of the six continents separately, which showed that a significant human influence was detected in each of these six regions (Stott 2003). Then, by assuming a linear relationship between fractional errors in past and future temperature changes over these continental regions, probabilistic predictions of future continental mean temperature change were obtained in the same way as for global mean temperature. The results based on HadCM3 gave large uncertainty ranges for future continental mean temperatures if emissions are assumed to follow the SRES A2 scenario of 2–12 K and 2–11 K relative to 1990s values for Europe and North America, respectively.

By analysing each continent separately, this approach does not exploit the observational constraints provided by any possible relationships between warming rates in different continents. A lower limit on the possible uncertainty range was obtained by Stott *et al.* (2006*a*) by assuming that the spatial pattern of temperature change simulated by HadCM3 is correct and scaling future projections according to the fractional error in the global mean temperature. This approach gave much tighter uncertainty ranges for the SRES A2 emission

scenario of 4–8 K for North America and 4–7 K for Europe. However, this result is conditional as there being no additional uncertainty in the spatial patterns of response. This is not the case since different climate models do not have identical patterns of response and therefore climate models with different spatial patterns of response are likely to give different ranges of uncertainty when scaled by scaling factors derived from a global mean analysis of the model.

The standard optimal detection methodology has recently been extended to take account of the modelling uncertainty in the spatial patterns of response and this approach could be applied in future to estimate uncertainties in future climate change at sub-global scales. Huntingford et al. (2006) include an estimate of the inter-model covariance structure in the regression method and calculate attributable global mean temperature changes based on an analysis of three models simultaneously (the same three models analysed by Stott et al. (2006b), namely HadCM3, GFDLR30 and PCM). They found a tighter constraint of global mean greenhouse attributable warming than seen in individual models of 0.8–1.1 K, indicating that for global mean temperature, the reduction in uncertainty due to the EIV method accessing in effect a larger ensemble outweighs any increase in uncertainty due to the inter-model covariance structure. This range of attributable greenhouse warming translates into a range for TCR of 1.8–2.4 K (based on a ratio between attributable greenhouse warming and TCR of 2.22; see §2b). While this gives tighter constraints for attributable global mean warming rates, and therefore TCR, it also provides an estimate of attributable patterns of change that includes modelling uncertainty. Such an approach could therefore be used to provide a more accurate estimate of future uncertainties in continental-scale temperature changes that lies between the upper and lower limits provided by Stott et al. (2006a).

To apply such approaches, ensembles of coupled model simulations are required with different forcings for a range of different models. Only large multi-model multi-forcing ensembles of this sort are capable of characterizing the structural uncertainty resulting from model formulation (perturbing parameter choices in one particular model does not explore the structural uncertainty obtained when parameter choices from multiple different models are perturbed). Such structural uncertainty is likely to be a more important component of the uncertainty in regional-scale predictions than that of global-scale predictions. While many modelling centres have made simulations of past climate change including both anthropogenic and natural forcings combined (e.g. for the IPCC AR4) and a smaller number of modelling centres have also made simulations including only natural forcings, a relatively small number of models have been run with the full range of transient simulations required for an ASK analysis. Ideally, transient simulations of past climate change are required that can separate the effects of GHGs from other anthropogenic forcings, with these simulations continuing into the future with the same combinations of forcings (according to a particular emissions scenario). This enables a separation of the fractional error in TCR from the fractional error in forcings. Therefore, to fully characterize uncertainty in regional predictions requires multi-model multi-forcing ensembles of coupled models.

Further information about probable future climate changes can be found by searching coupled model ensembles for emergent constraints. Allen & Ingram (2002) find a consistent physically based relationship between global mean precipitation and

global mean temperature, which they use to provide a probabilistic prediction for precipitation based on a probabilistic distribution for future temperature changes based on the EMIC-based calculation of Forest *et al*. (2002). A similar type of approach has been used to derive probabilistic estimates of climate sensitivity based on model-derived relationships between climate sensitivity and observable variables such as the seasonal cycle (Piani *et al*. 2005; Knutti *et al*. 2006).

4. Summary and discussion

In this paper, we have shown that two different approaches for constraining climate predictions based on observations of past climate change produce consistent probabilistic predictions, indicating that the observed record contains robust information that provides important constraints on future climate changes.

The two approaches described in this paper, the EMIC-based approach and the ASK approach, differ most in their treatment of errors. While the ASK approach works well for scenarios of steadily increasing forcing, for stabilization scenarios, the relative fractions of forcing components will not be the same as in the twentieth century and so the system will be extrapolating beyond the calibration sample. In the LE approach, the model–observations comparison provides error estimates first for climate system properties (S, K_v, F_{aer}, etc.), or potentially feedbacks directly, and then the uncertainty in these properties are propagated with a dynamical model. As such the patterns of climate changes are allowed to evolve in the LE approach based on the model dynamics rather than being held fixed by a pattern-scaling approach.

While EMICs are well suited for exploring the large-scale effects of coupling between different components of the climate system, EMICs do not include the full range of processes incorporated in AOGCMs and their highly parametrized representation of climate processes and their coarse resolution mean that they are not well suited for quantifying uncertainty in regional climate change or extreme events. In addition, because ocean dynamics are important for determining regional feedbacks (Boer & Yu 2003), climate models require a full three-dimensional dynamic ocean component if they are going to represent future regional climate changes. Therefore, for informing regional adaptation policies, AOGCMs are required. Also, their high resolution and detailed parametrizations means that only AOGCMs are able to properly represent internal variability, and the EMIC-based analyses require AOGCM-based estimates of internal variability.

A potential shortcoming of both approaches is the inadequate sampling of the noise-covariance matrices by the control simulations. This becomes a more serious problem at smaller scales since a greater number of degrees of freedom is needed to describe regional patterns of climate change. In addition, the true number of degrees of freedom in the observational data is probably much smaller than is currently assumed and treatment of observational errors needs to be improved.

In summary, the evidence consistently provided by both approaches is that significant future warming is likely to follow from continued emissions of anthropogenic GHGs. Even targeted emissions reductions (to achieve approximate stabilization at 550 ppm for CO_2 concentrations in 2100) will very probably lead to future warming rates greater than 1 K by 2100 when compared with 2000,

which could imply significant impacts (Hansen *et al.* 2006). The inevitable necessity to adapt to some level of climate change underlines the importance of probabilistic predictions of future climate change.

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. 2003 Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Clim. Dyn.* **21**, 477–491. (doi:10.1007/s00382-003-0313-9)

Allen, M. R. & Tett, S. F. B. 1999 Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419–434. (doi:10.1007/s003820050291)

Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. & Delworth, T. L. 2000 Uncertainty in forecasts of anthropogenic climate change. *Nature* **407**, 617–620. (doi:10.1038/35036559)

Allen, M. R., Frame, D., Kettleborough, J. A. & Stainforth, D. A. 2006*a* Predictability in weather and climate. In *Model error in weather and climate forecasting. ECMWF* (eds T. Palmer & R. Hagedorn), ch. 15, pp. 391–428. Cambridge, UK: Cambridge University Press.

Allen, M. R. *et al.* 2006*b* Quantifying anthropogenic influence on recent near-surface temperature change. *Surv. Geophys.* **27**, 491–544. (doi:10.1007/s10712-006-9011-6)

Anderson, T. L., Charlson, R. J., Schwartz, S. E., Knutti, R., Boucher, O., Rodhe, H. & Heintzenberg, J. 2003 Climate forcings by aerosols—a hazy picture. *Science* **300**, 1103–1104. (doi:10.1126/science.1084777)

Andronova, N. G. & Schlesinger, M. E. 2001 Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.* **106**, 22 605–22 611. (doi:10.1029/2000JD000 259)

Babiker, M. H., Reilly, J. M., Mayer, M., Eckaus, R. S., Sue Wing, I. & Hyman, R. C. 2001 The MIT emissions prediction and policy analysis (EPPA) model: revisions, sensitivities, and comparison of results. Report no. 71. Technical report http://web.mit.edu/globalchange/www/Publications/, Massachusetts Institute of Technology, Joint Program on the Science and Policy of Global Change, Room E40-428, 77 Massachusetts Avenue, Cambridge, MA 02139.

Bayes, T. 1763 An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. A* **53**, 370–418. (doi:10.1098/rsbl.1763.0053)

Boer, G. J. & Yu, B. 2003 Climate sensitivity and response. *Clim. Dyn.* **20**, 415–429. (doi:10.1007/s00382-002-028-3)

Cubasch, U., Meehl, G. A., Boer, G. J., Stouffer, R. J., Dix, M., Noda, A., Senior, C. A., Raper, S. & Yap, K. S. 2001 Climate change 2001: the scientific basis. In *Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell & C. A. Johnsen). Projections of future climate change, ch. 9, pp. 525–582. Cambridge, UK: Cambridge University Press.

Forest, C. E., Allen, M. R., Stone, P. H. & Sokolov, A. P. 2000 Constraining uncertainties in climate models using climate change detection methods. *Geophys. Res. Lett.* **27**, 569–572. (doi:10.1029/1999GL010859)

Forest, C. E., Allen, M. R., Sokolov, A. P. & Stone, P. H. 2001 Constraining climate model properties using optimal fingerprint detection methods. *Clim. Dyn.* **18**, 277–295. (doi:10.1007/s003820100175)

Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. & Webster, M. D. 2002 Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117. (doi:10.1126/science.1064419)

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).

Frame, D. J., Stone, D. A., Stott, P. A. & Allen, M. R. 2006 Alternatives to stabilization scenarios. *Geophys. Res. Lett.* **33**, L14 707. (doi:10.1029/2006GL025801)

Gillett, N. P., Wehner, M. F., Tett, S. F. B. & Weaver, A. J. 2004 Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.* **31**, L14 201. (doi:10.1029/2004GL020111)

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W. & Medina-Elizade, M. 2006 Global temperature change. *Proc. Natl Acad. Sci. USA* **103**, 14 288–14 293. (doi:10.1073/pnas.0606291103)

Hegerl, G. C., Storch, H. V., Hasselman, K., Santer, B. D., Cubasch, U. & Jones, P. D. 1996 Detecting greenhouse gas-induced climate change with an optimal fingerprint method. *J. Clim.* **9**, 2281–2306. (doi:10.1175/1520-0442(1996)009<2281:DGGICC>2.0.CO;2)

Hegerl, G. C., Hasselman, K., Cubasch, U., Mitchell, J. F. B., Roeckner, E., Voss, R. & Waskewitz, J. 1997 Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.* **9**, 613–634. (doi:10.1007/s003820050186)

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K. & Johnson, C. A. (eds) 2001 *Climate change 2001: the scientific basis. Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change*, p. 881. Cambridge, UK: Cambridge University Press.

Huntingford, C., Stott, P. A., Allen, M. R. & Lambert, F. H. 2006 Incorporating model uncertainty into attribution of observed temperature change. *Geophys. Res. Lett.* **33**, L05710. (doi:10.1029/2005GL024831)

Iman, R. L. & Conover, W. J. 1982 A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. B* **11**, 311–334.

Iman, R. L. & Helton, J. C. 1988 An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal.* **8**, 71–90. (doi:10.1111/j.1539-6924.1988.tb01155.x)

IPCC 2007 *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.

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)

Knutti, R., Stocker, T. F., Joos, F. & Plattner, G. 2002 Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723. (doi:10.1038/416719a)

Knutti, R., Stocker, T. F., Joos, F. & Plattner, G.-K. 2003 Probabilistic climate change projections using neural networks. *Cim. Dyn.* **21**, 257–272. (doi:10.1007/s00382-003-0345-1)

Knutti, R., Joos, F., Muller, 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)

Meehl, G. A., Washington, W. M., Ammann, C. M., Arblaster, J. M., Wigley, T. M. L. & Tebaldi, C. 2004 Combinations of natural and anthropogenic forcings in 20th century climate. *J. Clim.* **17**, 3721–3727. (doi:10.1175/1520-0442(2004)017<3721:CONAAF>2.0.CO;2)

Murphy, J. D. 1995 Transient response of the Hadley centre coupled ocean–atmosphere model to increasing carbon dioxide. Part III: analysis of global-mean response using simple models. *J. Clim.* **8**, 496–514. (doi:10.1175/1520-0442(1995)008<0496:TROTHC>2.0.CO;2)

Murphy, J., Sexton, D., Barnett, G. J. D. M., Webb, M., Collins, M. & Stainforth, D. A. 2004 Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772. (doi:10.1038/nature02771)

Paltsev, S., Reilly, J., Jacoby, H., Eckaus, R., McFarland, J., Sarofim, M., Asadoorian, M. & Babiker, M. 2006 The MIT emissions prediction and policy analysis (EPPA) model: version 4. Technical report http://web.mit.edu/globalchange/www/Publications/, Massachusetts Institute of Technology, Joint Program on the Science and Policy of Global Change, Room E40-428, 77 Massachusetts Avenue, Cambridge, MA 02139.

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)

Prinn, R. *et al.* 1999 Integrated global system model for climate policy assessment: feedbacks and sensitivity studies. *Clim. Change* **41**, 469–546. (doi:10.1023/A:1005326126726)

Reilly, J. *et al.* 2001 Multi-gas assessment of the kyoto protocol. *Nature* **401**, 549–555. (doi:10.1038/44069)

Schellnhuber, H. J., Cramer, W., Nakicenovic, N., Wigley, T. & Yohe, G. 2006 *Avoding dangerous climate change*, p. 208. Cambridge, UK: Cambridge University Press.

Sokolov, A. P. & Stone, P. H. 1998 A flexible climate model for use in integrated assessments. *Clim. Dyn.* **14**, 291–303. (doi:10.1007/s003820050224)

Sokolov, A. P., Forest, C. E. & Stone, P. H. 2003 Comparing oceanic heat uptake in AOGCM transient climate change experiments. *J. Clim.* **16**, 1573–1582.

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. 2003 Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.* **30**, 1728. (doi:10.1029/2003GL017324)

Stott, P. A. & Kettleborough, J. A. 2002 Origins and estimates of uncertainty in twenty-first century temperature rise. *Nature* **416**, 723–726. (doi:10.1038/416723a)

Stott, P. A. & Tett, S. F. B. 1998 Scale-dependent detection of climate change. *J. Clim.* **11**, 3282–3294. (doi:10.1175/1520-0442(1998)011<3282:SDDOCC>2.0.CO;2)

Stott, P. A., Kettleborough, J. A. & Allen, M. R. 2006*a* Uncertainty in continental-scale temperature predictions. *Geophys. Res. Lett.* **30**, L02708. (doi: 10.1029/2003GL017324)

Stott, P. A., Mitchell, J. F. B., Allen, M. R., Delowrth, T. L., Gregory, J. M., Meehl, G. A. & Santer, B. D. 2006*b* Observational constraints on past attributable warming and predictions of future global warming. *J. Clim.* **19**, 3055–3069. (doi:10.1175/JCLI3802.1)

Webster, M. D. & Sokolov, A. P. 2000 A methodology for quantifying uncertainty in climate projections. *Clim. Change* **46**, 417–446. (doi:10.1023/A:1005685317358)

Webster, M. D., Babiker, M., Mayer, M., Reilly, J. M., Harnisch, J., Sarofim, M. C. & Wang, C. 2002 Uncertainty in emissions projections for climate models. *Atmos. Environ.* **36**, 3659–3670. (doi:10.1016/S1352-2310(02)00245-5)

Webster, M. *et al.* 2003 Uncertainty analysis of climate change and policy response. *Clim. Change* **61**, 295–320. (doi:10.1023/B:CLIM.0000004564.09961.9f)

Zwiers, F. W. 2002 Climate change: the 20-year forecast. *Nature* **416**, 690–691. (doi:10.1038/416690a)