

Estimates of Uncertainty in Predictions of Global Mean Surface Temperature

J. A. KETTLEBOROUGH

British Atmospheric Data Centre, Rutherford Appleton Laboratory, Chilton, United Kingdom

B. B. B. BOOTH

Hadley Centre for Climate Prediction and Research, Met Office, Exeter, United Kingdom

P. A. STOTT

Hadley Centre for Climate Prediction and Research (Reading Unit), Met Office, University of Reading, Reading, United Kingdom

M. R. ALLEN

Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, University of Oxford, Oxford, United Kingdom

(Manuscript received 21 July 2005, in final form 12 June 2006)

ABSTRACT

A method for estimating uncertainty in future climate change is discussed in detail and applied to predictions of global mean temperature change. The method uses optimal fingerprinting to make estimates of uncertainty in model simulations of twentieth-century warming. These estimates are then projected forward in time using a linear, compact relationship between twentieth-century warming and twenty-first-century warming. This relationship is established from a large ensemble of energy balance models.

By varying the energy balance model parameters an estimate is made of the error associated with using the linear relationship in forecasts of twentieth-century global mean temperature. Including this error has very little impact on the forecasts. There is a 50% chance that the global mean temperature change between 1995 and 2035 will be greater than 1.5 K for the Special Report on Emissions Scenarios (SRES) A1FI scenario. Under SRES B2 the same threshold is not exceeded until 2055. These results should be relatively robust to model developments for a given radiative forcing history.

1. Introduction

Forecasts of future climate change provide an essential source of information for policy makers. Their utility is increased considerably when the forecasts are accompanied by estimates of uncertainty. This has been recognized by the Intergovernmental Panel on Climate Change (IPCC; Houghton et al. 2001) and others (Reilly et al. 2001; Schneider 2001; Webster et al. 2003) and has motivated several estimates of uncertainty in forecasts of climate change (Allen et al. 2000; Wigley and Raper 2001; Forest et al. 2002; Knutti et al. 2002; Stott and Kettleborough 2002). These estimates pro-

vide information on the likely ranges of climate change: in particular they rule out certain futures as unlikely. This paper describes in detail a method for making estimates of the uncertainty in global mean temperature change introduced by Allen et al. (2000) and used by Stott and Kettleborough (2002), hereafter referred to as ASK.

Most estimates of uncertainty in climate change are based on weighted ensembles of climate models. Each member of the ensemble has different values for the key, uncertain, model inputs. The models are used to simulate a historical period, and usually a future period using one or more future radiative forcing time series. Individual model runs are then compared with observations to produce the relative weight for that ensemble member (Andronova and Schlesinger 2001; Forest et al. 2002; Harvey and Kaufmann 2002; Knutti et al. 2002). These weights are then used to build distributions of

Corresponding author address: J. A. Kettleborough, Hadley Centre for Climate Prediction and Research, Met Office, Fitzroy Road, Exeter, EX1 3PB, United Kingdom.
E-mail: jamie.kettleborough@metoffice.gov.uk

forecast quantities. In some cases the effective weight assigned to models is not determined by comparison with observations, but by expert judgment (Wigley and Raper 2001).

The results of all these model studies are, however, dependent on the way that the model inputs have been sampled. Sampling the parameters in a different way or introducing new models will change the forecast distribution (Frame et al. 2005; Stainforth et al. 2005). Allen et al. (2006) introduced the idea of Stable Inferences from Data (STAID). These are estimates that depend largely on observations, and are relatively robust to model changes such as the way that the model parameters are sampled or the introduction of new models into an ensemble. At the center of the STAID forecast is a transfer function that links a forecast quantity to an observable. In the case of ASK the transfer function is a relationship linking twentieth-century warming to future warming. The uncertainty in twentieth-century warming is assessed by comparing an atmosphere–ocean general circulation model (AOGCM) with observations. The transfer function is then used to project the uncertainty in twentieth-century warming into the future. This method is very similar in spirit to the Gregory et al. (2002) estimates of climate sensitivity. Gregory et al. (2002) identify a relationship between observable quantities and the forecast quantity, which in this case is climate sensitivity. They then use this relationship as a transfer function to project the uncertainty in the observations onto the uncertainty in climate sensitivity. In the case of Gregory et al. (2002) the transfer function is based on energy conservation and a simple representation of the climate feedbacks. For ASK the transfer function emerges from an ensemble of Energy Balance Model (EBM) integrations. The transfer function should be robust to known errors or uncertainty in models and is a key component of a STAID forecast.

The general method used by ASK is described in detail in section 2 of this paper. Section 3 discusses the particular application of the method to global mean temperature. An ensemble of EBM integrations is used to test the limitations of using the linear transfer function between twentieth-century and twenty-first-century warming (section 4). The impacts on forecasts of temperature change of errors associated with using an exact linear transfer function are determined in section 5. This section also presents new results concerning the probability of exceeding a temperature threshold.

2. General method

A summary of the general method used in ASK is given in Fig. 1. Two models are used: an EBM and an

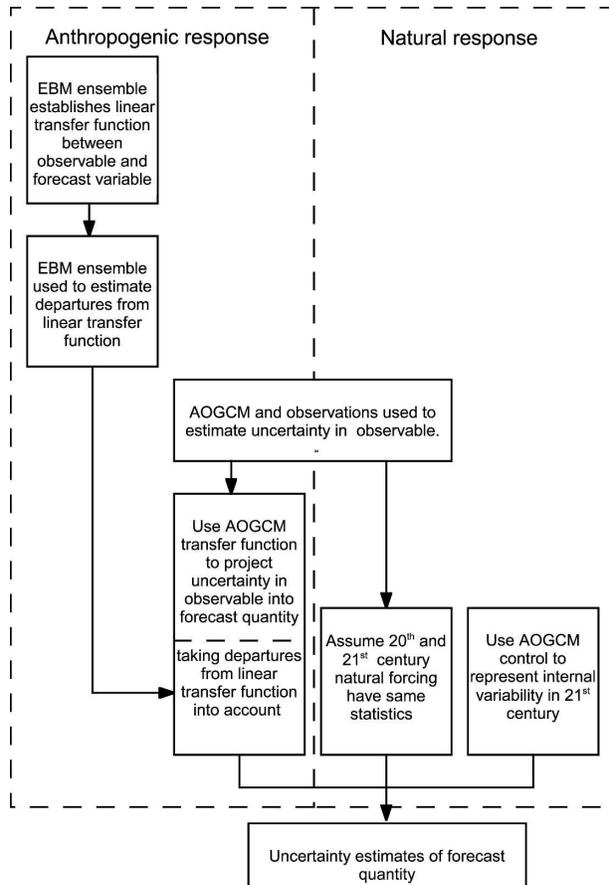


FIG. 1. Schematic of the method used to estimate uncertainty in future climate change.

AOGCM. The EBM is used to establish the existence of a linear transfer function for the anthropogenic component of the response, and to estimate the error associated with using a linear transfer function. This is discussed in detail in section 4. Here we begin with the use of the AOGCM in assessing the uncertainty in past climate change and, given the existence of a linear transfer function, discuss how a single AOGCM can be used to predict the uncertainty in future climate change.

a. Estimates of uncertainty in past climate change

An estimate of the uncertainty in past climate change is found by comparing model simulations with observations. The comparison is performed using optimal fingerprinting as used in the detection and attribution of climate change (Hasselmann 1997; Allen and Tett 1999). In essence the fingerprinting algorithm is a linear regression and can be thought of as scaling a linear combination of AOGCM response patterns to fit the observations:

$$y = \sum_{i=1}^m (x_i - v_i)\beta_i + v_0, \quad (1)$$

where y is the observations, x_i are the modeled response patterns, v_i are any errors or uncertainties in a response patterns, v_0 is any error or uncertainty that causes the observations to be different from the linear fit, and β_i are the scaling factors for the model response patterns. The observations and response patterns are written as vectors as they can be space–time patterns. The response patterns are calculated from integrations, or initial condition ensembles, forced with different radiative forcings. If any of the β_i are less than one, this implies that the model overestimates the response to a particular forcing and so the model’s pattern of response has to be scaled down. If any of the β_i are greater than one, the model underestimates the observed response and so its pattern of response has to be scaled up.

The method used to determine the β_i and their uncertainty estimates depends on the sources of noise terms, v_i and v_0 . Ordinary least squares (OLS) accounts for internal variability in the observations (v_0 ; Allen and Tett 1999). The observed forced response may differ from the scaled response patterns because of internal variability in the observations. Total least squares (TLS) allows both internal variability in the observations and the influence of internal variability on the sampling of the response patterns (v_i ; Allen and Stott 2001). When the response patterns are inferred from a small number of initial condition ensemble runs then the response patterns are contaminated by internal variability. In this study we use TLS rather than OLS to estimate the β_i and their uncertainty ranges. OLS suffers from low bias, particularly in the upper bound, when used with small ensemble sizes. TLS does not suffer from the same biases (Allen and Stott 2001; Stott et al. 2001). This is particularly important when estimating confidence limits on the scaling factors, especially if there are weak signals. If OLS was used in the current study, this would lead to an underestimate in the uncertainty.

TLS does not account for all factors that may cause the observations to differ from scaled versions of the model response patterns. Inadequacies in the model formulation or physical parameterizations and errors in the spatial distribution or temporal evolution of the forcing can all result in errors in the response patterns. By allowing the response patterns to be scaled up or down we are allowing for cases where these errors affect only the amplitude of the response. We do not allow for cases where the errors change the pattern of the response. This could, in principle, be relaxed by in-

cluding these errors in the term v_i (Huntingford et al. 2006). In practice a consistency test is applied to the residuals of the fit (Allen and Tett 1999). One of the reasons this consistency test may fail is if the errors in the patterns cannot be accounted for by internal variability alone. If the consistency test is passed, we can be reasonably confident that the response pattern predicted by the model is adequate.

TLS gives the best-guess estimates of the scaling factors and the probability that the true scaling factors are different from these best estimates (Allen and Stott 2001). This probability distribution, $P(\beta_i)$, gives an estimate of our uncertainty in past climate change.

b. Uncertainty estimates in future climate change

Having determined the distribution of the β_i using optimal fingerprinting, a natural way of predicting the future response of the climate system is to apply a modified version of Eq. (1), but at forecast times:

$$y^{\text{for}} = \sum_{i=1}^m (x_i^{\text{for}} - v_i^{\text{for}})\beta_i + v_0^{\text{for}} + e_i, \quad (2)$$

where x_i^{for} are the model forecast response patterns, β_i are the same scaling factors as calculated from the optimal fingerprinting of past climate change, and y^{for} is the modified forecast. The noise terms v_i^{for} are the uncertainty in the model response patterns. This uncertainty might be due to internal variability because only small ensembles are used to produce the forecast response, or due to uncertainties in the future forcing or model physics that result in errors in the forecast pattern. In this study we account for the uncertainty due to small ensemble size. The additional noise terms (in this case v_0^{for} and e_i) represent any reason why the future climate is different from the scaled versions of the modeled forced response. These include internal variability (v_0^{for}), and any error resulting from assuming the scaling factors calculated from optimal fingerprinting can be used to scale the future response (e_i).

Using the same scaling factors for past and future response is equivalent to saying that fractional error is constant over time. If a model underestimates the forced response by 20% in the twentieth century, then it will continue to underestimate the forecast-forced response by 20% through the twenty-first century. The errors do not change the shape of the response, only the magnitude (Fig. 2). It is worth putting this in the context of terminology used elsewhere. Following Allen et al. (2000) earlier in this paper we have referred to this as a linear transfer function. In the language of Allen and Ingram (2002) the linear transfer function is just an example of an emergent constraint. In this paper we

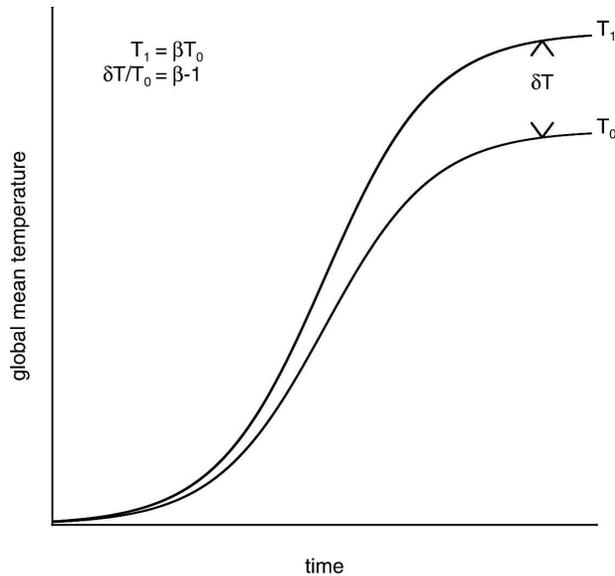


FIG. 2. Scaling a reference time series is equivalent to 1) allowing for errors in magnitude, but not shape, and 2) constant fractional error.

will also use the term linear constraint to indicate an emergent constraint that is associated with a linear transfer function.

The uncertainty in future climate change can be calculated from Eq. (2) given the $P(\beta_i)$ and the statistics of v_0^{for} , v_i^{for} , and e_i . The simplest way to do this is by Monte Carlo sampling of the distributions to generate a set of sample points for y^{for} .

In the introduction we mentioned that ensemble-based estimates of uncertainty are prone to a sampling problem, even if weighted by a likelihood based on observations. In the context of ASK we in effect generate an ensemble of model runs based on the scaling of a set of AOGCM response patterns. The weight on each ensemble member is then determined by the probability of the scale factors [$P(\beta_i)$]. In estimating $P(\beta_i)$ by optimal fingerprinting we do not assume any prior knowledge of how much the model response has to be scaled up or down to be made consistent with the observations. We assume that a model is equally likely to under- or overestimate the observed response. This is equivalent to saying that we start with a uniform sampling of scaling factors (and hence model ensemble members). The final distribution of scaling factors is then determined by weighting by the likelihood based on observations. Since we are using the same scaling factor for past and forecast temperature change, then a uniform prior in scaling factor is equivalent to a uniform prior in past and forecast temperatures. This uniform prior in future or forecast temperatures means

that, before accounting for model–observation differences, we are not assuming that any one forecast is any more likely than any other. This is exactly what most users of forecasts would expect (Allen et al. 2006; Frame et al. 2005; Stainforth et al. 2005).

If we were to apply the method used in other studies (Forest et al. 2002; Knutti et al. 2002; Murphy et al. 2004) we would find that, before weighting by the likelihood, the ensemble members would not be uniformly distributed in the forecast quantity. The distribution is biased by the sampling used to generate the ensemble. In practice the prior distribution will be weighted toward the base model since many of the perturbations to the models physics will have little impact on the modeled climate change. This will result in an underestimate of the uncertainty. This could, of course, have serious implications for any planning or policy based on these estimates of uncertainty.

3. Models and data

In this study we apply the method outlined in the previous section to forecasts of global mean surface temperature change. Here we use AOGCM simulations as outlined in Stott et al. (2000), Tett et al. (2002), and Stott and Kettleborough (2002). The model simulations, observations, and processing are all identical to that in Stott and Kettleborough (2002). The observations used to estimate the uncertainty in the response to recent climate change are surface temperatures taken from an updated version of Parker et al. (1994). The model used is the Third Hadley Centre Coupled AOGCM (HadCM3; Gordon et al. 2000; Pope et al. 2000). Three ensembles of simulations of the period 1869–2000 have been used to estimate the uncertainty in recent changes in surface temperature. Each ensemble includes different radiative forcing: greenhouse gas (GHG) only (G), greenhouse gas and anthropogenic sulphate (GS), and solar and volcanic forcing (NAT). For notational convenience, where we need to be specific about forcing, the subscript i will take values G, GS, or NAT in terms in Eqs. (1) and (2). Each of the forcings is described in more detail in Tett et al. (2002). The ensembles are four-member initial condition ensembles. The ensemble means are used to define the response patterns [x_i , or more specifically x_G , x_{GS} , and x_{NAT} , in Eq. (1)]. We use decadal mean data, and the observational mask is applied to the model patterns before the space patterns are filtered using spherical harmonics to scales above 5000 km (Tett et al. 2002). Segments of the HadCM3 control run are used to provide estimates of the internal variability, v_i . Optimal fingerprinting applied in this way gives β_G , β_{GS} , and

β_{NAT} , the factors by which the modeled responses have to be scaled in order to fit observations of the surface temperature.

The estimates of internal variability by the model are deficient at some scales (Stott and Tett 1998; Allen and Tett 1999). This is overcome by truncating the response patterns so that they include only modes of variability that are sufficiently well represented by the model. The level of truncation is determined by a consistency test (Allen and Tett 1999). Here we use a truncation of 15 throughout. The variation of results quoted in this paper is less than 0.1 K for truncations between 15 and 20.

The anthropogenic components of the forecast patterns [$x_{\text{G}}^{\text{for}}$, $x_{\text{GS}}^{\text{for}}$ in Eq. (2)] are taken from G and GS integrations of HadCM3 for the period 2000–2100. Simulations have been performed for each of the Special Report on Emissions Scenarios (SRES) scenarios SRES-A1FI, SRES-A2, SRES-B1, and SRES-B2. For most of the SRES scenarios and forcing histories only one simulation has been used (the exception is SRES-B2, which has two simulations for both G and GS). The resulting sampling error in the forecast response patterns is taken into account by adding an error term v_i^{for} . The estimate of the sampling error is made by taking the variance of the global mean temperature from the control, scaling this to account for the ensemble size, and adding it as Gaussian noise in Eq. (2).

Future changes in natural forcing due to volcanic eruptions and changing solar irradiance ($x_{\text{NAT}}^{\text{for}}$) are hard to determine. Volcanic eruptions with significant radiative forcing appear to be random (Hyde and Crowley 2000) and so can be modeled by a stochastic process. The solar output, at least over the next two decades, may be predictable using variations in the solar cycle (Lean 2001). Although predictions of longer-term solar output variations can be made, there is no established method for doing this. In principle, the forecasts of random volcanic eruptions and solar radiance output could be used in this analysis. Here, however, we use a simpler approach based on the modeled response to natural forcing during the period 1860–1990. We assume that the variation of solar output and effect of volcanic eruptions of the next century will not be very different from that in the last century. The uncertainty in the natural forcing in the next century is modeled by a random process with the same characteristics as the recent past response to natural forcing modeled in the HadCM3 NAT ensemble. This is achieved by adding a Gaussian process to $v_{\text{NAT}}^{\text{for}}$ in Eq. (2). The variance of this process is determined from the statistics of the NAT historical integrations. The forecast temperatures produced by this Gaussian process are based on the modeled response. They are scaled by β_{NAT} to make

them consistent with observations. (Fig. 1 puts this in context.)

For estimates of uncertainty in future climate change we concentrate only on the global mean surface temperature. The forecast responses ($x_{\text{G}}^{\text{for}}$ and $x_{\text{GS}}^{\text{for}}$) are the global mean, ensemble mean temperature from the forecast integrations. To maintain consistency with the data used in the optimal fingerprinting the forecast data have an observational mask applied before taking the global mean. The mask used is the same mask that is used in the last decade of the twentieth century. For most of the forecast period the difference between the global mean of the masked data and the true global mean is less than 0.1 K. Toward the end of the twenty-first century for high climate change scenarios the difference can be greater than 0.1 K, but remains less than 0.15 K. The decadal-scale internal variability to be added onto the forecast forced response [v_0^{for} in Eq. (2)] is represented by a first-order autoregressive process [AR(1)]. The process variance is 0.003 K^2 , and the lag-one correlation 0.251, estimated by fitting the control. All results presented here are based on decadal mean predictions.

4. The linear constraint

Having discussed the AOGCM component of Fig. 1 we now turn to the EBM and the linear transfer function. For a simple model of the climate system with a constant heat capacity and linear feedbacks there is a lagged response to a linearly increasing forcing (e.g., Hartmann 1994). The lag time is given by the response time of the system. This means that, after an initial adjustment time, uncertainties in the strength of the feedbacks or the heat capacity have little impact on the shape of the response, which is a linear function of time. The errors simply scale the response up or down. As discussed earlier this is equivalent to saying that the fractional error is constant, or that there is a linear transfer function between past and future warming. A similar derivation to that given in Hartmann (1994) also applies to an exponentially increasing forcing. For some idealized forcing profiles there is a linear transfer function between past and future warming. Allen et al. (2000) have demonstrated that a similar result holds for the IS92A scenario. Here we extend their analysis to SRES scenarios and estimate the error resulting from using the linear transfer function.

The error is estimated using an ensemble of EBM runs. The EBM consists of a mixed layer ocean with a linear feedback and heat transport to the deep ocean modeled by an effective diffusion (Hansen et al. 1985). The model has been run using a range of values of the

climate sensitivity (0.5–20 K) and ocean diffusivity (0.0–4.8 $\text{cm}^2 \text{s}^{-1}$). The model realizations sample uniformly in climate sensitivity and in the square root of ocean heat diffusivity following Forest et al. (2002). One hundred values of each parameter have been used. For each scenario the EBMs are run with both G only and GS forcing from 1860 to 2100. The radiative forcing time series used to force the EBMs are those diagnosed from corresponding runs of HadCM3. This helps ensure consistency between the EBM and HadCM3 results.

In this section we concentrate on two of the SRES scenarios. The forcing histories for A1FI and B1 are shown in Fig. 3. The A1FI forcing increases with time, with radiative forcing approaching 8 W m^{-2} by the end of the twenty-first century. The rate of increase of the forcing is not steady but accelerates, particularly in the middle of the century. This acceleration is especially noticeable for the GS forcing. In contrast, in the B1 scenario the G forcing is almost steady at 4 W m^{-2} by the end of the twenty-first century. The anthropogenic sulfate forcing is decreasing with time, so the GS forcing shows less stabilization. The decreasing G forcing is offset by the decrease in the magnitude of the negative anthropogenic sulfate forcing.

Figure 4 shows the error resulting from using a linear constraint as a function of the climate sensitivity and ocean heat diffusivity. Elsewhere in this study we scale HadCM3 projections to estimate future warming rates so the error has been calculated by scaling the EBM that most closely matches the temperature and heat content of HadCM3. For the A1FI scenario, at 2100, the additional error resulting from scaling the reference trajectory lies between -10% and 10% . Most of the models sampled have positive errors. For the B1 scenario, at 2100, the errors are larger and lie between -30% and 20% .

The source of the errors can be understood in terms of the relative time scales of the forcing and the model response time. For stabilization scenarios, such as B1, the time scale of the stabilization may be comparable to the model response time. Models with large ocean heat diffusion coefficients and climate sensitivities have longer response time scales than models with small ocean heat diffusion coefficients and climate sensitivities. Figure 4 shows that models with long time scales tend to have larger temperature response than their scaled analog because they have a large warming commitment. In other words representing the system as a scaled version of a model with less warming commitment leads to an underestimate of the system response. For models with very large warming commitment the

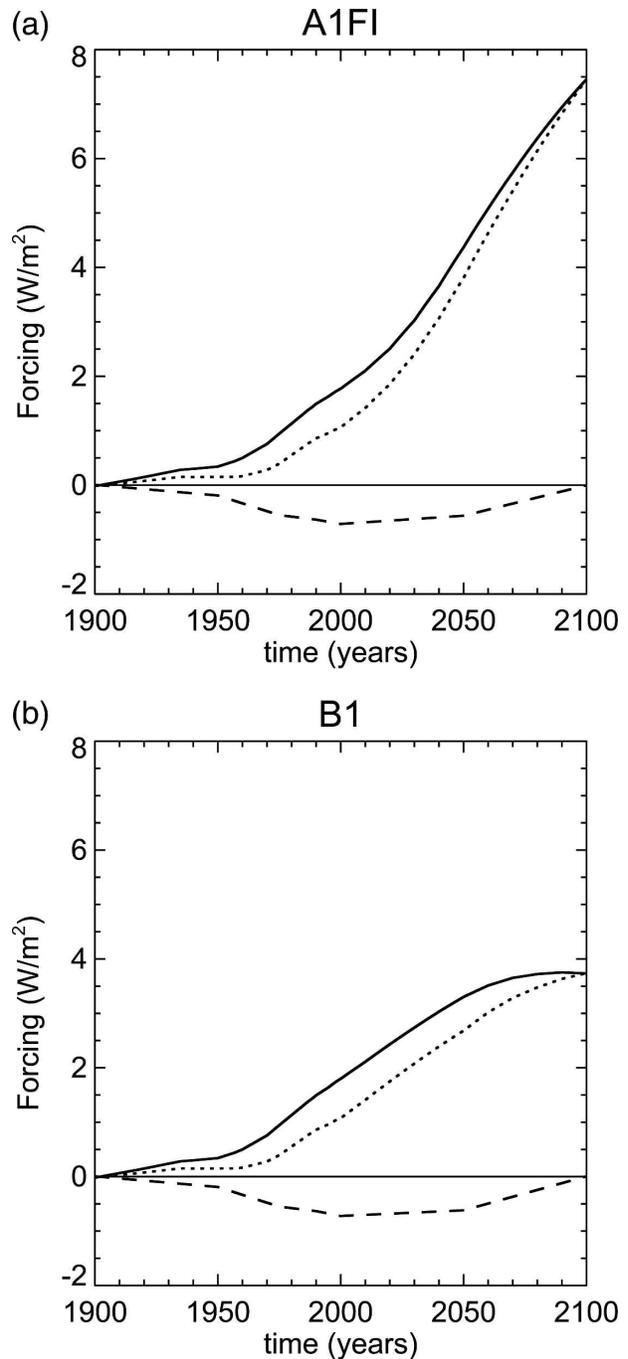


FIG. 3. Radiative forcing time series (a) A1FI and (b) B1. Solid lines: G; dotted lines: GS; dashed lines: anthropogenic sulfate.

response is relatively insensitive to any stabilization in the forcing. This explains why the error flattens out at high time scales.

Models with short response time scales have smaller temperature responses than their scaled analog. These errors can become large for very short response time

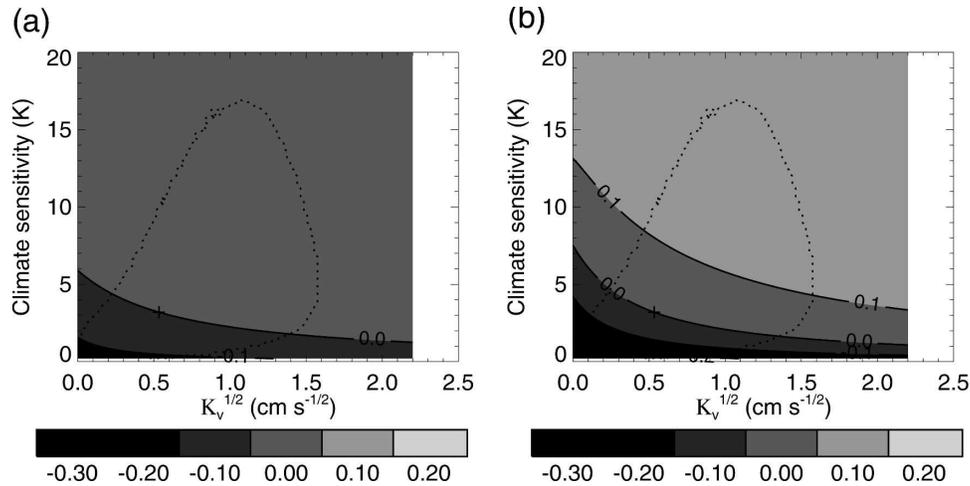


FIG. 4. Error resulting from using the linear transfer function shown as a function of EBM parameter K_v , designates the ocean heat diffusion coefficient. (a) 2100 error for A1FI, (b) 2100 error for B1. The overlaid dotted contour is the 95% confidence limit based on comparison with observations of the heat content and attributable GHG warming. The cross is the EBM parameter combination that best fits HadCM3 [EBM(HadCM3): climate sensitivity of 3.2K; ocean diffusion coefficient of $0.284\text{cm}^2\text{ s}^{-1}$]. The error has been calculated for each EBM by first regressing the twentieth century temperature response of the EBM onto the EBM(HadCM3) temperature. This gives the factor β by which the EBM(HadCM3) temperature needs to be scaled to give the actual EBM temperature. The error is then, $E = [T_{\text{EBM}} - \beta T_{\text{EBM(HadCM3)}}]/T_{\text{EBM}}$.

models. Models with short response times have very little warming commitment. If the forcing stabilizes, these models will respond quickly to the changes in forcing. Their temperature response will therefore be less than their scaled analog based on a model with larger warming commitment.

When the forcing accelerates rather than stabilizes, as it does in the A1FI scenario for the case of GS, then for large time-scale models the scaled analog tends to overestimate the true response. Again this can be understood in terms of the model response time. When the response is slow compared to the acceleration of the forcing the model tends to have a temperature response that is smaller than a scaled version of a shorter time scale model. This is reflected in a negative error. This behavior is seen in the EBMs, though the corresponding Fig. 4 for GS is not shown here.

Not all EBM models are equally realistic when compared with observations of the twentieth century. The dotted contour overlaid on Fig. 4 shows the 95% confidence limit based on comparing observations with the EBM outputs. The confidence limits have been calculated using two sets of observations. First, Levitus et al. (2000) observations of ocean heat content are used to infer the effective heat capacity over the second half of the twentieth century. Second, the optimal fingerprinting outlined in section 3 is used to infer the amount of twentieth-century warming attributable to GHG forc-

ing (Frame et al. 2005). The effective heat capacity and the GHG attributable warming can also be calculated from the G-forced runs of each EBM. The uncertainty estimates on the observed effective heat capacity and GHG attributable warming can then be used to assign each EBM a likelihood. The 95% contour on Fig. 4 indicates that EBMs with very high sensitivities ($>17\text{ K}$) and ocean diffusion coefficients ($>2.5\text{ cm}^2\text{ s}^{-1}$) can be excluded. Models with very large negative errors (those with very low sensitivities and heat uptake rates) can also be excluded.

The likelihoods of different EBMs can be used to weight the errors shown in Fig. 4 to give a distribution of error. This distribution, $P(E_i)$, gives an estimate of the probability that scaling the HadCM3 best fit EBM results in an error, E_i . This error is denoted E_i as it conceptually equivalent to the term e_i in Eq. (2). In this case i represents the forcings G and GS. Of course, $P(E_i)$ is dependent on the sampling strategy used to generate the ensemble of EBMs. The distribution would be different if we sampled evenly in the ocean heat diffusion coefficient as opposed to the square root of the ocean heat diffusion coefficient. Figure 5 shows $P(E_G)$ and $P(E_{GS})$ from the A1FI and B1 scenarios as a function of time. The errors clearly grow with time in all cases. The influence of down weighting the EBMs according to their likelihoods when compared to observations is clear. For instance for the G forcing in A1FI

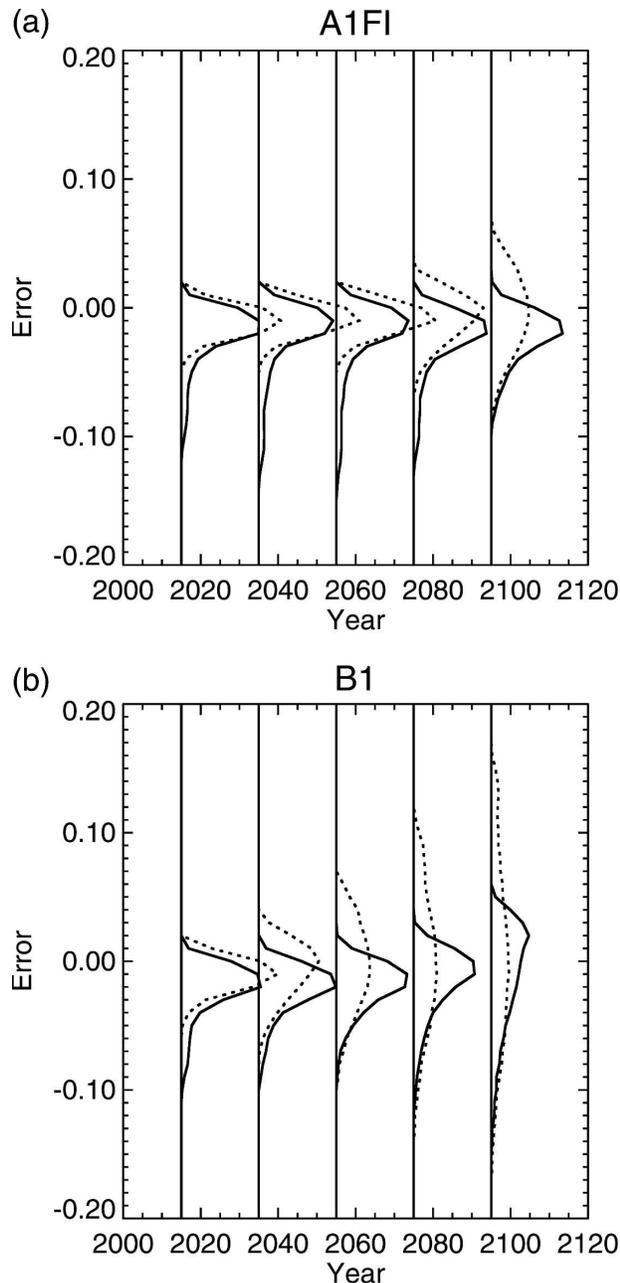


FIG. 5. Probability of an error, $p(E)$, resulting from scaling a reference trajectory relative to the unscaled trajectory. $E = [T_{\text{EBM}} - \beta T_{\text{EBM(HadCM3)}}]/T_{\text{EBM}}$. The probability density function is based on the EBM ensemble after weighting each model with its likelihood. (a) A1FI and (b) B1. Solid line: GS; dotted line: G.

(Fig. 4a) shows that there are large numbers of EBMs sampled that have positive errors. Most of these EBMs, however, have relatively low likelihoods, and so $P(E_G)$ and $P(E_{GS})$ do not reflect a predominance of positive values. The errors for A1FI are small, with most being less than 10%. There is also a small negative bias, par-

ticularly early in the century. For B1 the errors are larger than for A1FI, particularly for G. As discussed earlier this reflects the fact that the B1 scenario is stabilizing. Models with slow response times and large warming commitments warm more than their scaled analogs; models with fast response times and small warming commitments warm less than their scaled analogs. For B1 the errors associated with the slow response models become large and the error distribution shifts toward positive errors. These error distributions are sensitive to the reference model chosen as this sets the time scale of all the analog models. In this case the errors calculated will be applied to forecasts based on HadCM3. This means the EBM that best fits HadCM3 is a suitable choice for the reference.

The error $P(E_G)$ and $P(E_{GS})$ quantifies the linearity and compactness of the constraint. It can be used to estimate the statistics of the term e_i in Eq. (2) for the G and GS forcings. It is this error that joins the EBM and AOGCM parts of the analysis in Fig. 1. The errors are generally small, indicating that the use of a linear transfer function is valid.

5. Estimates of uncertainty

a. Accounting for uncertainty in the emergent constraint

In this section we use the estimates $P(E_i)$ to correct the forecasts of twenty-first-century temperature change made in Stott and Kettleborough (2002).

In this analysis we ignore the fact that there may be correlation between the errors in the β_i and e_i . In other words we ignore the fact that more severe errors in the linear relationship occur as β_i gets further away from 1. This will have most impact in the extremes of the probability density function (PDF), but little impact on the center of the PDF.

Figure 6 shows the uncertainty plumes for the A1FI and B1 scenarios. The plumes calculated taking the error in the linear constraint into account are very similar to the uncorrected plumes. The dominant error is our uncertainty in twentieth-century temperature change rather than the error associated with assuming a linear transfer function. This is particularly so in the first half of the century. For the A1FI scenario the correction tends to bring the 5th, 50th, and 95th percentiles down in temperature by less than 0.1 K in 2100. The lower values reflect the negative errors associated with the linear constraint, shown in Fig. 4. For the B1 scenario the influence of errors in the linear constraint is very small, and almost negligible. Given that the errors due to the linear constraint for G forcing are relatively large in the B1 scenario this is perhaps a surprising result. In

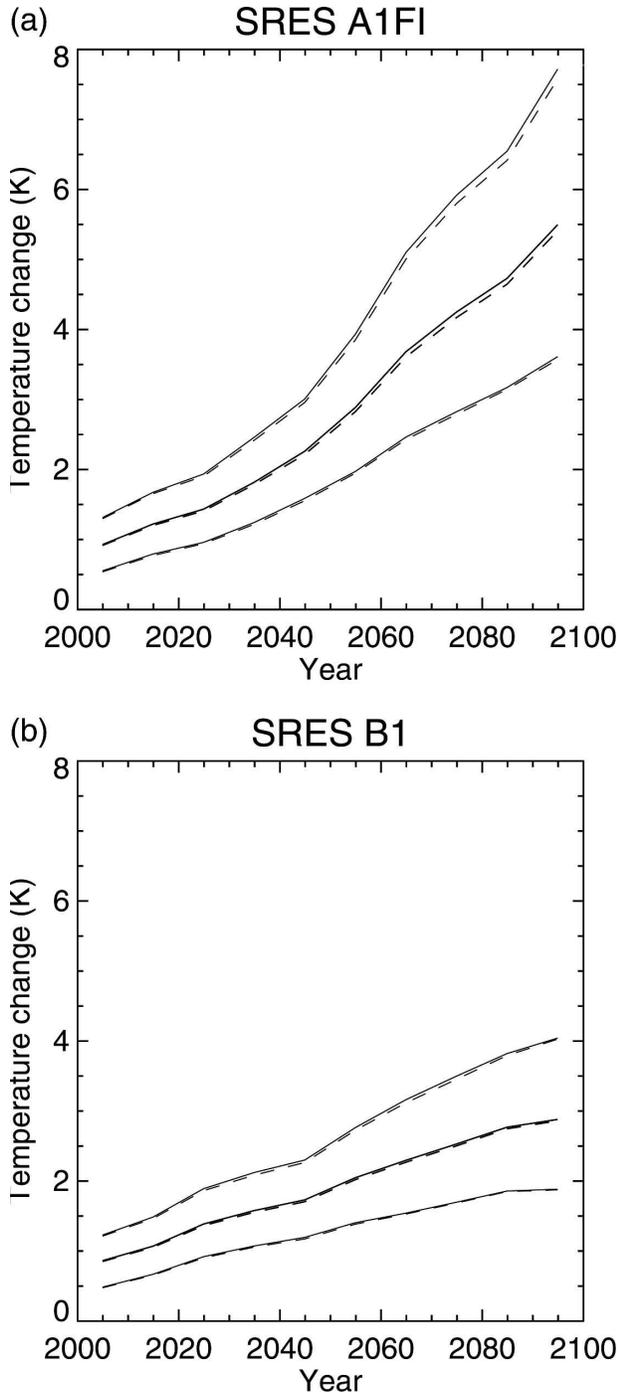


FIG. 6. Uncertainty estimates in future climate accounting for errors in the assumption of constant fractional error. (a) A1FI and (b) B1; solid lines are the uncorrected plumes, and dashed lines are the corrected plumes. For each case the midlines are the median estimate and the extremes are the 5% and 95% confidence limits. Warming is relative to preindustrial climate.

predictions made using HadCM3 based on the GS and G patterns the scaling on the G pattern alone is quite small, and so the influence of the errors associated with this pattern is small. This is because the scaling on the G pattern is the additional GHG warming needed to match the observations over and above that in GS (Allen and Tett 1999). If we applied the corrections to a model that had a different ratio of greenhouse gas to anthropogenic sulfate response in the twentieth century, then the deviation of GHG response from a linear transfer function could have more of an influence.

Accounting for uncertainty in the linear constraint due to uncertainty in climate sensitivity and ocean heat diffusivity has little effect on the main conclusions of Stott and Kettleborough (2002). Percentiles of the distribution of forecast decadal mean global mean temperature change relative to preindustrial climate are given in Table 1.

b. Sensitivity to the uncertainty in natural forcing

Given the uncertainty in the future natural forcing, it is worth understanding the sensitivity of results to changes in the assumed future natural forcing. In the context of this paper the future natural response is represented by the term $\beta_{\text{NAT}}x_{\text{NAT}}^{\text{for}}$ in Eq. (2). The uncertainty in the future natural response is therefore determined by the distributions of β_{NAT} and $x_{\text{NAT}}^{\text{for}}$. The distribution of β_{NAT} is determined by the optimal fingerprinting. The distribution of $x_{\text{NAT}}^{\text{for}}$ is represented by a Gaussian process. In this section we vary the standard deviation of this Gaussian distribution to determine the sensitivity of the forecasts of global mean temperature change to the uncertainty in the future natural forcing. Figure 7 demonstrates the sensitivity of the uncertainty in global mean temperature change to the uncertainty in the natural response. For reference the modeled global mean temperature standard deviation over the period 1860–2000, used as the estimate for much of this study and that of Stott and Kettleborough (2002), is shown by the vertical dashed line. The modeled global mean temperature variance over the period 1900–2000 is shown by the vertical dotted line. The period 1860–2000 includes the eruption of Krakatoa in 1883, which accounts for the larger natural variance over this extended period compared to the twentieth century. At both 2020 and 2060 there is some sensitivity to the assumed natural variance. This sensitivity, however, is relatively small compared to the total uncertainty in the temperature forecasts. This is especially the case at 2060. This reflects the strengthening greenhouse gas signal and corresponding increase in the uncertainty in the forecast due to the uncertainty in the greenhouse gas forcing.

TABLE 1. Percentiles of the estimated PDFs of future decadal mean temperature change relative to preindustrial climate.

Decade/ percentile	5%	10%	33%	50%	66%	90%	95%
SRES A1FI							
2000–10	0.8	0.8	1.0	1.1	1.2	1.4	1.6
2010–20	1.0	1.1	1.3	1.4	1.5	1.8	1.9
2020–30	1.2	1.3	1.5	1.6	1.7	2.0	2.2
2030–40	1.4	1.6	1.8	2.0	2.1	2.5	2.7
2040–50	1.8	1.9	2.2	2.4	2.6	3.0	3.2
2050–60	2.2	2.4	2.8	3.0	3.3	3.8	4.1
2060–70	2.6	2.9	3.5	3.8	4.1	4.9	5.3
2070–80	3.0	3.3	4.0	4.4	4.7	5.6	6.0
2080–90	3.4	3.7	4.5	4.8	5.2	6.2	6.7
2090–2100	3.8	4.2	5.1	5.6	6.1	7.2	7.8
SRES A2							
2000–10	0.7	0.8	1.0	1.1	1.1	1.4	1.5
2010–20	0.9	1.0	1.2	1.3	1.4	1.7	1.8
2020–30	1.1	1.2	1.5	1.6	1.7	2.0	2.1
2030–40	1.4	1.5	1.7	1.9	2.0	2.3	2.5
2040–50	1.7	1.8	2.1	2.2	2.4	2.8	2.9
2050–60	2.0	2.1	2.5	2.7	2.9	3.4	3.6
2060–70	2.3	2.5	2.9	3.2	3.4	4.0	4.2
2070–80	2.6	2.8	3.4	3.6	3.9	4.6	5.0
2080–90	3.0	3.2	3.9	4.2	4.5	5.3	5.7
2090–2100	3.4	3.7	4.4	4.8	5.2	6.1	6.6
SRES B1							
2000–10	0.7	0.8	0.9	1.0	1.1	1.4	1.5
2010–20	0.9	1.0	1.1	1.3	1.4	1.6	1.7
2020–30	1.1	1.2	1.4	1.6	1.7	2.0	2.1
2030–40	1.3	1.4	1.6	1.8	1.9	2.2	2.3
2040–50	1.4	1.5	1.8	1.9	2.0	2.4	2.5
2050–60	1.6	1.7	2.0	2.2	2.4	2.8	3.0
2060–70	1.8	1.9	2.3	2.5	2.7	3.1	3.4
2070–80	1.9	2.1	2.5	2.7	2.9	3.4	3.7
2080–90	2.1	2.3	2.7	2.9	3.2	3.7	4.0
2090–2100	2.1	2.3	2.8	3.1	3.3	4.0	4.3
SRES B2							
2000–10	0.8	0.9	1.0	1.1	1.2	1.5	1.6
2010–20	1.0	1.1	1.3	1.4	1.5	1.8	1.9
2020–30	1.2	1.3	1.5	1.6	1.7	2.0	2.2
2030–40	1.4	1.5	1.8	1.9	2.0	2.4	2.5
2040–50	1.5	1.7	2.0	2.1	2.3	2.6	2.8
2050–60	1.7	1.8	2.2	2.3	2.5	2.9	3.1
2060–70	1.9	2.1	2.5	2.7	3.0	3.5	3.8
2070–80	2.1	2.3	2.7	3.0	3.2	3.8	4.0
2080–90	2.3	2.5	3.0	3.3	3.5	4.2	4.5
2090–2100	2.5	2.7	3.2	3.5	3.7	4.3	4.7

c. Probability of global warming exceeding 1.5 K

An alternative way of viewing the uncertainties in twenty-first-century warming is to look at the probability of temperature exceeding a given threshold as a function of time. This shifts the emphasis from looking at whether a certain temperature change will occur to looking at when that temperature change occurs. This is particularly relevant when there are processes with

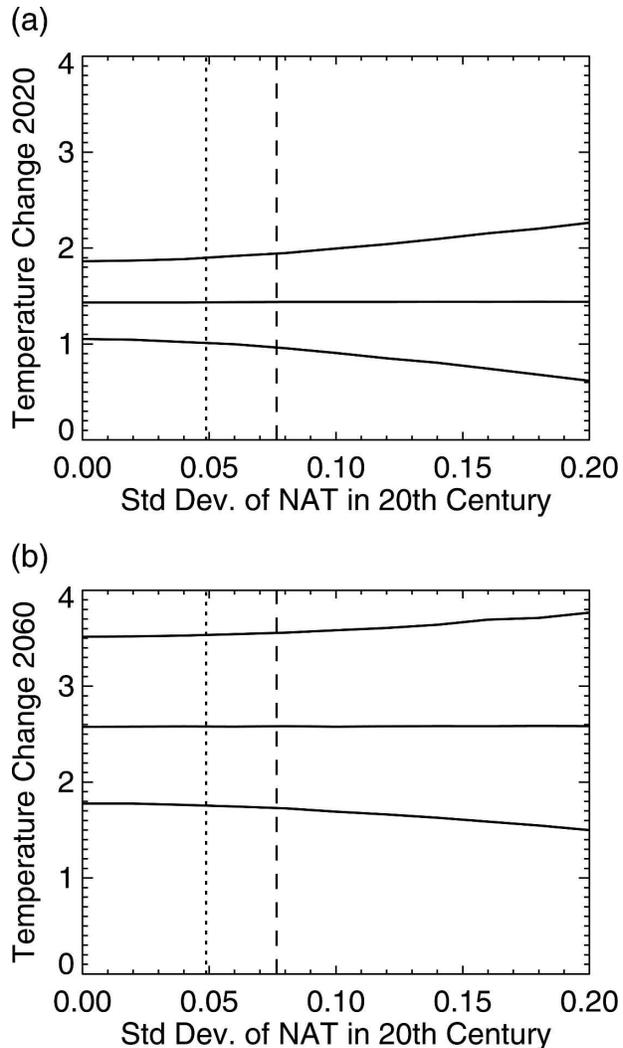


FIG. 7. Sensitivity of the forecast uncertainty to changes in the assumed standard deviation in natural forcing. (a) Forecast temperature in 2020. (b) Forecast temperature in 2060. Shown are the 5%, 50%, and 95% percentiles of the predicted temperature distribution. Vertical dotted line is the standard deviation of the ensemble mean global mean temperature of the naturally forced ensemble of HADCM3 during 1900–2000. The vertical dashed line is the standard deviation of the same ensemble over 1850–2000.

thresholds or where there is an identifiable level of dangerous change (Hulme 2003; Knutti et al. 2005). The probability of the warming since 1995 exceeding 1.5 K is shown in Fig. 8 (approximately equal to a warming of 2 K since the preindustrial era). This has been calculated, for every year of the century, by adding an estimate of the annual subdecadal variability onto each of the possible future temperature time series. Looking at the probability of exceeding 1.5 K very clearly shows the difference between the forcing scenarios. Especially

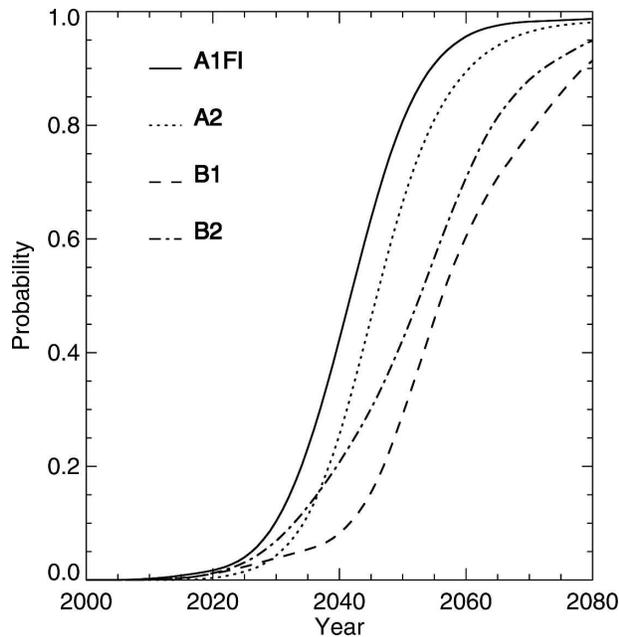


FIG. 8. Probability of global mean warming since 1995 exceeding 1.5 K.

notable is the difference between the A1FI and B1 scenarios. Under the A1FI scenario there is a 20% chance that the global mean temperature will have exceeded 1.5 K by about 2030. In contrast, under the B1 scenario this same probability of warming is delayed for about 15 yr, until about 2045. Under the fossil fuel intensive scenario, A1FI, in 2050 there is over an 80% chance that the global mean temperature will have risen by over 1.5 K. For the less fuel intensive scenarios, B1 and B2, however, the probability of exceeding 1.5 K is less than half of that for A1FI. Figure 8 also demonstrates that, although following a particular scenario may delay warming, there are high probabilities of over 1.5-K global mean temperature increase by the end of this century under all of the scenarios: A1FI, A2, B1, and B2.

6. Discussion and summary

We have given a detailed discussion of the methodology used by ASK in making predictions of uncertainty in global mean temperature change (Allen et al. 2000; Stott and Kettleborough 2002). Estimates of uncertainty based on this method should be relatively robust to changes in model formulation. We have minimized the input of models by 1) assuming a uniform prior in the forecast quantity rather than relying on a model parameter sampling strategy and 2) using a constraint between an observed quantity and a forecast quantity that is robust to known errors. It is worth iden-

tifying the places where the results may be dependent on model errors. Two sets of models are involved (Fig. 1): the EBM ensemble and a set of HadCM3 integrations. The EBM is used for two purposes. The first is to establish the existence of the constraint between twentieth-century and twenty-first-century temperature change. The second is to estimate the error associated with the fact that the constraint may not be exact. This error is the only place that the EBM has a quantitative impact on the results in assessing the uncertainty in the emergent constraint [$P(E_i)$ in section 4].

In section 4 we calculated the uncertainty in using a linear constraint based on the uncertainty in the model parameters of the simple EBM. These are not the only sources of uncertainty that can impact the linearity or compactness of the constraint. Others include model structural uncertainty and historical radiative forcing uncertainty. Both of these could be included using ensembles of more complex models and a range of historical forcing histories. The results of section 4 suggest that the existence of the emergent constraint between twentieth-century and twenty-first-century warming appears to be a consequence of energy conservation, linear feedbacks, and the relative time scales of the forcing and the climate system response. If there are nonlinear feedbacks or different time scales in the real climate system response, then the errors associated with using a linear constraint may become large, in effect making the constraint useless. In the case of structural uncertainty there is some evidence that complex models exhibit nonlinear behavior (Stocker and Schmittner 1997; Cox et al. 2000; Senior and Mitchell 2000), but the robustness of these results to model uncertainties is still to be established. In the case of uncertainty in historical forcing there will be an impact on the constraint if the forcing history introduces new time scales or has a significant impact on the shape of the forcing.

Although we have not accounted for the influence of structural or historical forcing uncertainty on the linear constraint there is some evidence that the additional errors are not large. Stott et al. (2006b) have compared ASK projections of uncertainty in global mean temperatures from three AOGCMs. These projections, just as here, exploit the existence of a linear constraint. The models have structural differences and different radiative forcing histories. The resulting projections are in good agreement, suggesting that use of the linear constraint has not introduced a large error.

The second use of models is in using HadCM3 to make estimates of the uncertainty in recent climate change. In this part of the problem HadCM3 is used both to provide the pattern of recent climate change

and to provide estimates of internal variability. Use of a different model, with different variability, will change the estimates of uncertainty in recent climate change, but as shown in Allen et al. (2000) and Stott et al. (2006b) these changes are relatively small.

The third use of models is the use of HadCM3 to provide the shape of the forecast of future global mean temperature change. For a given future radiative forcing other models should give similar results. This is a consequence of the emergent constraint: model errors tend to scale the same basic warming profile shape, rather than change the shape. One aspect of model predictions that may lead to model dependence in the estimates of uncertainty is when the same emissions scenario leads to different radiative forcing. This could be a particular problem in the prediction of uncertainty in future sulfate cooling. Differences in the chemistry and cloud physics schemes of models can lead to different radiative forcing for the same SO₂ emissions. There may, of course, still be an emergent constraint between twentieth-century and twenty-first-century temperature change even accounting for errors in sulfur chemistry and physics, in which case STAID forecasts would still be possible. In the absence of the constraint we can still make STAID forecasts, but they become explicitly dependent on the radiative forcing time series.

We have shown that for some scenarios constraints may themselves be uncertain. For instance, the emergent constraint for the G component of temperature change is more uncertain for the B1 scenario than the A1FI scenario. Although for predictions made with HadCM3 this is not too big a problem since the emergent constraint on the combined GS response dominates. This suggests, however, that for some scenarios it may be easier to make STAID forecasts of climate change than for others. The uncertainty in the constraint for global mean temperature has little impact on the results.

In this study we have applied the method of ASK to global mean temperatures. The method could be applied to other variables or spatial scales provided there is a compact transfer function between the forecast variables and some quantity or set of quantities for which observationally based uncertainty estimates can be made. For instance Gregory et al. (2002) in effect apply the same method to determine the equilibrium temperature response to a doubling of CO₂. Stott et al. (2006a) apply the method to regional temperature change. Piani et al. (2005) again apply a similar method to the climate feedbacks and climate sensitivity. In the case of Piani et al. (2005) the transfer function emerges from a large perturbed physics ensemble rather than an

EBM. Large perturbed physics ensembles (such as those provided by www.climateprediction.net) will help establish the range of climate variables for which uncertainty forecasts can be made using a method similar to ASK.

We have also calculated the probability of global mean warming exceeding a temperature threshold of 1.5 K for four SRES scenarios. Low fossil fuel scenarios delay the probability of exceeding this threshold at some level by 20–30 yr. Under the fossil fuel intensive scenario (A1FI) the twenty-first-century warming as high as 5.5 K cannot be ruled out at the 95% level. The estimates of twenty-first-century global mean temperature change are relatively insensitive to the future natural forcing because of the dominance of the GHG signal.

Acknowledgments. Thanks to the reviewers: their comments have helped clarify the content of this paper considerably. PAS was funded by the U.K. Department for Environment, Food and Rural Affairs under Contract PECD 7/12/37. MRA acknowledges additional funding from the NOAA/DOE International Detection and Attribution Group. BBBB was funded by a NERC studentship, and JAK by NERC.

REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for internal consistency in optimal fingerprinting. *Climate Dyn.*, **21**, 419–434.
- , and P. A. Stott, 2001: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dyn.*, **15**, 477–491.
- , and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrological cycle. *Nature*, **419**, 224–232.
- , P. A. Stott, J. F. B. Mitchell, R. Schnur, and T. Delworth, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.
- , D. Frame, J. A. Kettleborough, and D. A. Stainforth, 2006: Model error in weather and climate forecasting. *Predictability in Weather and Climate*, T. Palmer and R. Hagedorn Eds., Cambridge University Press, 391–427.
- Andronova, N. G., and M. E. Schlesinger, 2001: Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.*, **106** (D19), 22 605–22 611.
- Cox, P. N., R. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell, 2000: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**, 184–187.
- Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**, 113–117.
- Frame, D. J., B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen, 2005: Constraining climate forecasts: The role of prior assumptions. *Geophys. Res. Lett.*, **32**, L09702, doi:10.1029/2004GL022241.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulations of SST, sea ice extents and ocean heat transports

- in a version of the Hadley centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168.
- Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Climate*, **15**, 3117–3121.
- Hansen, J., G. Russell, A. Lacis, I. Fung, D. Rind, and P. Stone, 1985: Climate response times: Dependence on climate sensitivity and ocean mixing. *Science*, **229**, 857–859.
- Hartmann, D. L., 1994: *Global Physical Climatology*. Academic Press, 411 pp.
- Harvey, L. D., and R. K. Kaufmann, 2002: Simultaneously constraining climate sensitivity and aerosol radiative forcing. *J. Climate*, **15**, 2837–2861.
- Hasselmann, K., 1997: On multi-fingerprint detection and attribution of anthropogenic climate change. *Climate Dyn.*, **13**, 601–611.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 881 pp.
- Hulme, M., 2003: Abrupt climate change: Can society cope? *Philos. Trans. Roy. Soc. London*, **361A**, 2001–2021.
- Huntingford, C., P. A. Stott, M. R. Allen, and F. H. Lambert, 2006: Incorporating model uncertainty into attribution of observed temperature change. *Geophys. Res. Lett.*, **33**, L05710, doi:10.1029/2005GL024831.
- Hyde, W. T., and T. J. Crowley, 2000: Probability of future climatically significant volcanic eruptions. *J. Climate*, **13**, 1445–1450.
- Knutti, R., T. F. Stoker, F. Joos, and G. K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**, 719–723.
- , F. Joos, S. A. Müller, G. K. Plattner, and T. F. Stoker, 2005: Probabilistic climate change projections for CO₂ stabilization profiles. *Geophys. Res. Lett.*, **32**, L20707, doi:10.1029/2005GL023294.
- Lean, J., 2001: Solar irradiance and climate forcing in the near future. *Geophys. Res. Lett.*, **28**, 4119–4122.
- Levitus, S., J. I. Antonov, T. P. Boyer, and C. Stephens, 2000: Warming of the world ocean. *Science*, **287**, 2225.
- Murphy, J. M., D. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.
- Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res.*, **99**, 14 373–14 399.
- Piani, C., D. J. Frame, D. A. Stainforth, and M. R. Allen, 2005: Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Lett.*, **32**, L23825, doi:10.1029/2005GL024452.
- Pope, V. D., M. L. Galani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model—HadAM3. *Climate Dyn.*, **16**, 123–146.
- Reilly, J., P. H. Stone, C. E. Forest, M. D. Webster, H. D. Jacoby, and R. G. Prinn, 2001: Uncertainty and climate change assessments. *Science*, **293**, 430–433.
- Schneider, S. H., 2001: What is dangerous climate change? *Nature*, **411**, 17–19.
- Senior, C. A., and J. F. B. Mitchell, 2000: The time dependence of climate sensitivity. *Geophys. Res. Lett.*, **27**, 2685–2689.
- Stainforth, D. A., and Coauthors, 2005: Evaluating uncertainty in predictions of the climate response to changing levels of greenhouse gases. *Nature*, **433**, 403–406.
- Stocker, T. F., and A. Schmittner, 1997: Influence of CO₂ emission rates on the stability of the thermohaline circulation. *Nature*, **388**, 862–865.
- Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate*, **11**, 32 892–32 904.
- , and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in predictions of twenty-first temperature rise. *Nature*, **416**, 723–725.
- , S. F. B. Tett, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137.
- , M. R. Allen, and G. S. Jones, 2001: Estimating signal amplitudes in optimal fingerprinting, part II: Application to general circulation models. *Climate Dyn.*, **21**, 493–500.
- , J. A. Kettleborough, and M. R. Allen, 2006a: Uncertainty in continental-scale temperature predictions. *Geophys. Res. Lett.*, **33**, L02708, doi:10.1029/2005GL024423.
- , J. F. B. Mitchell, M. A. Allen, T. L. Delworth, J. M. Gregory, G. A. Meehl, and B. D. Santer, 2006b: Observational constraints on past attributable warming and predictions of future global warming. *J. Climate*, **19**, 3055–3069.
- Tett, S. F. B., and Coauthors, 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107**, 4306, doi:10.1029/2000JD000028.
- Webster, M. D., and Coauthors, 2003: Uncertainty analysis in climate change and policy response. *Climate Change*, **61**, 295–320.
- Wigley, T. M. L., and S. C. B. Raper, 2001: Interpretation of high projections for global-mean warming. *Science*, **293**, 451–454.

Copyright of *Journal of Climate* is the property of *American Meteorological Society* and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.