

Observational Constraints on Past Attributable Warming and Predictions of Future Global Warming

PETER A. STOTT,* JOHN F. B. MITCHELL,⁺ MYLES R. ALLEN,[#] THOMAS L. DELWORTH,[@]
 JONATHAN M. GREGORY,[&] GERALD A. MEEHL,** AND BENJAMIN D. SANTER⁺⁺

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

+ Met Office, Exeter, United Kingdom

Physics Department, Oxford University, Oxford, United Kingdom

@ NOAA/GFDL, Princeton University, Princeton, New Jersey

& Centre for Global Atmospheric Modelling, Meteorology Department, University of Reading, Reading, and Hadley Centre for
 Climate Prediction and Research, Met Office, Exeter, United Kingdom

** National Center for Atmospheric Research, ## Boulder, Colorado

++ Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California

(Manuscript received 13 June 2005, in final form 16 November 2005)

ABSTRACT

This paper investigates the impact of aerosol forcing uncertainty on the robustness of estimates of the twentieth-century warming attributable to anthropogenic greenhouse gas emissions. Attribution analyses on three coupled climate models with very different sensitivities and aerosol forcing are carried out. The Third Hadley Centre Coupled Ocean–Atmosphere GCM (HadCM3), Parallel Climate Model (PCM), and GFDL R30 models all provide good simulations of twentieth-century global mean temperature changes when they include both anthropogenic and natural forcings. Such good agreement could result from a fortuitous cancellation of errors, for example, by balancing too much (or too little) greenhouse warming by too much (or too little) aerosol cooling.

Despite a very large uncertainty for estimates of the possible range of sulfate aerosol forcing obtained from measurement campaigns, results show that the spatial and temporal nature of observed twentieth-century temperature change constrains the component of past warming attributable to anthropogenic greenhouse gases to be significantly greater (at the 5% level) than the observed warming over the twentieth century. The cooling effects of aerosols are detected in all three models.

Both spatial and temporal aspects of observed temperature change are responsible for constraining the relative roles of greenhouse warming and sulfate cooling over the twentieth century. This is because there are distinctive temporal structures in differential warming rates between the hemispheres, between land and ocean, and between mid- and low latitudes. As a result, consistent estimates of warming attributable to greenhouse gas emissions are obtained from all three models, and predictions are relatively robust to the use of more or less sensitive models. The transient climate response following a 1% yr⁻¹ increase in CO₂ is estimated to lie between 2.2 and 4 K century⁻¹ (5–95 percentiles).

1. Introduction

Quantitative comparisons between observed and modeled patterns of temperature change yield valuable

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Peter A. Stott, Hadley Centre for Climate Prediction and Research (Reading Unit), Meteorology Building, University of Reading, Reading RG6 6BB, United Kingdom.

E-mail: peter.stott@metoffice.gov.uk

information about the causes of past climate change (e.g., Stott et al. 2000; Tett et al. 2002; Karoly et al. 2003; Santer et al. 2003a). Such comparisons led the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC; Houghton et al. 2001) to conclude that “In the light of new evidence and taking into account the remaining uncertainties, most of the observed warming over the last 50 years is likely to have been due to the increase in greenhouse gas concentrations.” This statement was based on results from optimal detection analyses, which seek to quantify the contributions to past climate change from greenhouse

gases and other external forcing factors (see Mitchell et al. 2001 for more details).

While there is increasing confidence in a significant greenhouse gas contribution to observed warming, there remains a very large uncertainty in aerosol forcing estimated from knowledge of the pertinent aerosol physics and chemistry (so-called forward calculations of aerosol forcing). Estimates range from -4 to 0 W m^{-2} (Anderson et al. 2003). The question remains therefore as to whether attribution studies could be in error by not taking fully into account the actual uncertainty in aerosol forcing and the ranges of possible climatic responses to this highly uncertain forcing.

An important capability of the optimal detection methodology is that it is able to distinguish the greenhouse warming component from other components of anthropogenic climate change. These latter components are likely to be dominated by cooling due to the net effects of aerosols. Given the large uncertainty in forcing from aerosols, if the only information available was a single number representing a trend or change in global-mean near-surface temperature, then a very large or small greenhouse warming could be balanced by a very large or small aerosol cooling and it would be impossible to discriminate between different possible accounts of the observed warming. Optimal detection algorithms that employ information on both space and time scales are able, potentially, to distinguish between different aspects of the greenhouse and aerosol fingerprints. An aim of this paper will be to determine which information provides the most discrimination between the greenhouse gas and sulfate aerosol responses.

Knowledge of the contributors to past temperature change in turn provides information about likely future rates of change. The technique for determining observationally constrained climate predictions was developed by Allen et al. (2000), based on the assumption that a model that under- or overestimates the climate response will tend to under- or overestimate the climate response in the future. The validity of this linear scaling assumption was examined in detail by Kettleborough et al. (2005, manuscript submitted to *J. Climate*, hereafter KET05), who showed that for most scenarios for which there are steadily increasing anthropogenic emissions, this is a robust approach. Stott and Kettleborough (2002), using the Third Hadley Centre Coupled Ocean–Atmosphere GCM (HadCM3), calculated observationally constrained predictions for a range of future emissions scenarios and examined the changing relationship between emissions uncertainty and response uncertainty during the twenty-first century.

The predictions of Stott and Kettleborough (2002) are observationally constrained and therefore should

be relatively model independent, despite being based on results from a single model. Here we additionally analyze two other coupled models, which have also been run with a range of natural and anthropogenic forcings, the Parallel Climate Model (PCM) and Geophysical Fluid Dynamics Laboratory (GFDL) R30 models. All three models give good simulations of twentieth-century global mean temperatures when both anthropogenic and natural forcings are included (Stott et al. 2000; Meehl et al. 2003, 2004a; Broccoli et al. 2003). This is despite the fact that the PCM has a lower sensitivity than the other models (Cubasch et al. 2001) and projections of twenty-first-century temperature change made using the PCM model are smaller than those made with the HadCM3 and GFDL models.

The aim of this study is to investigate whether there is sufficient information in observations of past near-surface temperature change to constrain predictions, whichever model is used, and, if so, to investigate which aspects of the observed temperature record are responsible for discriminating between models. If Stott and Kettleborough (2002) is correct, future warming rates following a particular emissions path are likely to be higher than projected by the PCM model and are more likely to be closer to those of the GFDL R30 and HadCM3 projections.

The structure of this paper is as follows. Section 2 provides a brief description of the models and the simulations considered, and the basic methodology is described briefly in section 3. Section 4 shows the attributed changes for the twentieth century, and section 5 discusses what these imply for the twenty-first century for a scenario including emissions of both greenhouse gases and sulfur and for the transient climate response to a $1\% \text{ yr}^{-1}$ increase in carbon dioxide. Section 6 summarizes the main conclusions of the paper and discusses the limitations of this approach. Further work will use much larger multimodel ensembles to more fully characterize modeling uncertainty, which will be more important for constraining regional predictions.

2. Climate model simulations

Three models are considered: HadCM3, PCM, and GFDL R30. They each have different climate sensitivities, ocean heat uptake efficiencies, and aerosol forcings, and these are summarized in Table 1. The relative roles of climate sensitivity (expressed in terms of the climate feedback parameter, λ) and the efficiency of ocean heat uptake (κ) in determining rates of transient warming in models are discussed by Raper et al. (2002).

TABLE 1. Climate feedback parameter, ocean heat uptake efficiency, aerosol forcing (relative to preindustrial levels), and TCR for the HadCM3, PCM, and GFDL R30 models.

	Climate feedback parameter (λ , $\text{W m}^{-2} \text{K}^{-1}$)	Ocean heat uptake efficiency (κ , $\text{W m}^{-1} \text{K}^{-1}$)	Aerosol forcing (W m^{-2})	TCR (K century^{-1})
HadCM3	1.2	0.6	-1.1	2.86
PCM	2.0	0.6	-0.58	1.89
GFDL	0.9	0.9	-0.64	2.80

a. HadCM3

The atmospheric component of HadCM3 (Pope et al. 2000) has a horizontal grid spacing of 2.5° in latitude by 3.75° in longitude and 19 vertical levels. The oceanic component (Gordon et al. 2000) has 20 vertical levels on a 1.25° by 1.25° grid. Unlike HadCM2 (its predecessor), HadCM3 does not require flux adjustments of heat and water at the air–sea interface to maintain a stable climate for multicentury integrations. Model years start on 1 December and annual mean data from both the model and the observations are calculated from 1 December to 30 November. As a shorthand, dates quoted in the text are given as, for example, 1910 when referring to the period 1 December 1909 to 30 November 1910.

We consider ensembles of climate model simulations, each ensemble consisting of four simulations starting from a different initial condition taken from a long multicentury control run of HadCM3 (CONTROL). Members of the first ensemble we consider are forced with changes in well-mixed greenhouse gases (GHGs). The second ensemble (ANTHRO) is forced with changes in well-mixed greenhouse gases plus anthropogenic sulfur emissions and their implied changes to cloud albedos, and tropospheric and stratospheric ozone changes. The third ensemble is forced with changes in both solar irradiance and stratospheric aerosol following volcanic eruptions (NATURAL). These simulations are identical to those described in more detail by Tett et al. (2002). In addition a fourth ensemble (ALL) contains all of the above forcings. Stott et al. (2000) showed that external forcings in the ALL ensemble are the main contributors controlling near-surface decadal mean temperature changes on global and continental land scales. Whereas the HadCM3 simulations that include both natural and anthropogenic forcings successfully simulate twentieth-century temperatures, simulations that include just natural forcings fail to capture warming observed over recent decades.

b. PCM

The Department of Energy (DOE) PCM model is a coupled ocean–atmosphere general circulation model

that does not use flux correction. It has a resolution of T42 in the atmosphere (roughly 2.8° by 2.8°) with 18 levels in the vertical. The ocean resolution is roughly $2/3^\circ$, reducing to $1/2^\circ$ in the equatorial Tropics, with 32 levels. It is described in detail by Washington et al. (2000) and has been used in many studies (Ammann et al. 2003; Meehl et al. 2003; Santer et al. 2003b,a; Meehl et al. 2004a). A large suite of ensemble simulations of the PCM model has been made, which includes five forcing agents in various combinations. These are changes in solar output, stratospheric volcanic aerosols, well-mixed greenhouse gases, changes in tropospheric and stratospheric ozone, and the direct effect of sulfate aerosols. A four-member ensemble has been made including all the forcings listed above as have four-member ensembles of the combined anthropogenic forcings and four-member ensembles of the combined natural forcings. For each forcing agent there is also a four-member ensemble including that forcing only. Meehl et al. (2004a) show that the late twentieth century can only be reproduced in the PCM model when anthropogenic forcings are included. Their results also indicate that the early twentieth-century warming is mainly caused by natural forcing in the model, mainly from increases in solar output (0.2-K warming in the 1940s compared to the beginning of the century from simulations including changing forcing from increases in solar irradiance only, warming that is larger than that estimated from simulations with changing volcanic forcing only).

c. GFDL R30

The GFDL R30 model is a coupled ocean–atmosphere general circulation model that uses flux adjustments for heat and water that vary seasonally and spatially but are fixed from one year to the next, to maintain a stable multicentury climate simulation. The atmospheric model employs the spectral transform method with rhomboidal truncation at zonal wavenumber 30 corresponding to approximately 2.2° latitude by 3.75° longitude with 14 levels in the vertical. The ocean component has a horizontal resolution of 2.25° latitude by 1.875° longitude with 18 vertical levels. The model is

described in detail by Delworth et al. (2002). Broccoli et al. (2003) describe a series of ensemble integrations of the GFDL R30 coupled model. The first ensemble includes time-varying forcings from greenhouse gases only (G), and in the remaining three ensembles forcings from the direct effects of anthropogenic sulfate aerosols (GS), solar variability (GSS), and stratospheric volcanic aerosols (GSSV) are added progressively, so that the fourth ensemble contains all four of these forcings. Each ensemble consists of three simulations initialized from different states in a multicentury control simulation of the model. They find from comparison with observations that the addition of the natural forcings improves the simulation of global multidecadal trends in temperature, precipitation, and ocean heat content. Solar and volcanic forcings are important contributors to early twentieth-century warming, and cooling from volcanoes reduces the warming simulated for the late twentieth century.

d. Simulations of twentieth-century temperature changes

Global mean warming was not steady during the twentieth century (Fig. 1a, black line). Instead there was a warming of about $0.2 \text{ K decade}^{-1}$ from the early 1970s and a similar rate of warming in the early part of the century, in between which there was a period of more than three decades when temperatures showed no long-term increase. All three models reproduce the main features of the observed evolution of global mean near-surface temperature during the twentieth century when they include both anthropogenic and natural forcings (Fig. 1a). They also agree in showing that the warming observed in the last three decades of the twentieth century is not consistent with simulations that do not include anthropogenic forcings (Fig. 1b). All these analyses agree in finding that recent warming is largely anthropogenic, whereas natural forcings play a larger role in causing the early century warming in the models.

The models differ much more in their predictions of future warming rates following a particular scenario than in their simulations of past temperature change (Fig. 1c). In the past, differences between the models' overall warming response to forcing from greenhouse gases and from aerosols tend to cancel, whereas the model's climate sensitivity is much more important for future warming, when forcing changes are dominated by the increasing forcing from well-mixed greenhouse gases.

There are also greater differences in the models' simulations of the large-scale spatial patterns of temperature change. Figure 2 compares decadal mean ob-

servations and simulations of four basic indices of near-surface temperature change that have been used previously in a number of detection studies (e.g., Karoly and Braganza 2001; Braganza et al. 2003). They are global mean near-surface temperature (Fig. 2a), the inter-hemispheric difference in near-surface temperature expressed as mean Northern Hemisphere temperature minus mean Southern Hemisphere temperature (Fig. 2b), the contrast between land and ocean surface temperature calculated as the difference between mean surface air temperature over land and mean sea surface temperature (Fig. 2c), and the mean meridional temperature gradient in the NH midlatitudes defined as the difference of the zonal band 52.5° – 67.5° N minus the zonal band 22.5° – 37.5° N (Fig. 2d). Whereas the models all have good simulations of global mean temperature changes, they differ in how well they capture changes in the other indices.

3. Methodology

The optimal detection analysis used by Stott and Kettleborough (2002) for HadCM3 is additionally applied to the PCM and GFDL simulations. The methodology is described briefly below; further details are in Tett et al. (2002).

We express observed decadal mean near-surface temperature changes \mathbf{y} as a linear sum of simulated changes from \mathbf{x}_1 (GHG), \mathbf{x}_2 (ANTHRO), and \mathbf{x}_3 (either NATURAL for HadCM3 and PCM or GSSV for GFDL) plus noise, \mathbf{v}_0 :

$$\mathbf{y} = \sum_{i=1}^3 (\mathbf{x}_i - \mathbf{v}_i)\beta_i + \mathbf{v}_0, \quad (1)$$

where β_i is the vector of unknown scaling factors to be estimated in the regression. By including the additional noise term \mathbf{v}_i in the regression equation (Allen and Stott 2003), we take into account the statistical uncertainty introduced by taking the model-simulated responses from a finite ensemble, which differ from the underlying noise-free responses that would be obtained from a hypothetical infinite ensemble.

The regression analysis is carried out in the space spanned by the leading p empirical orthogonal functions of the covariance matrix of internal variability, \mathbf{C}_{N_i} . A consistency test (Allen and Tett 1999) is used to test whether the residuals of regression are consistent with internal variability. This test is indicative of when things are going very wrong, for instance when the truncation is too high and modes of variability that are poorly sampled by the available model representations

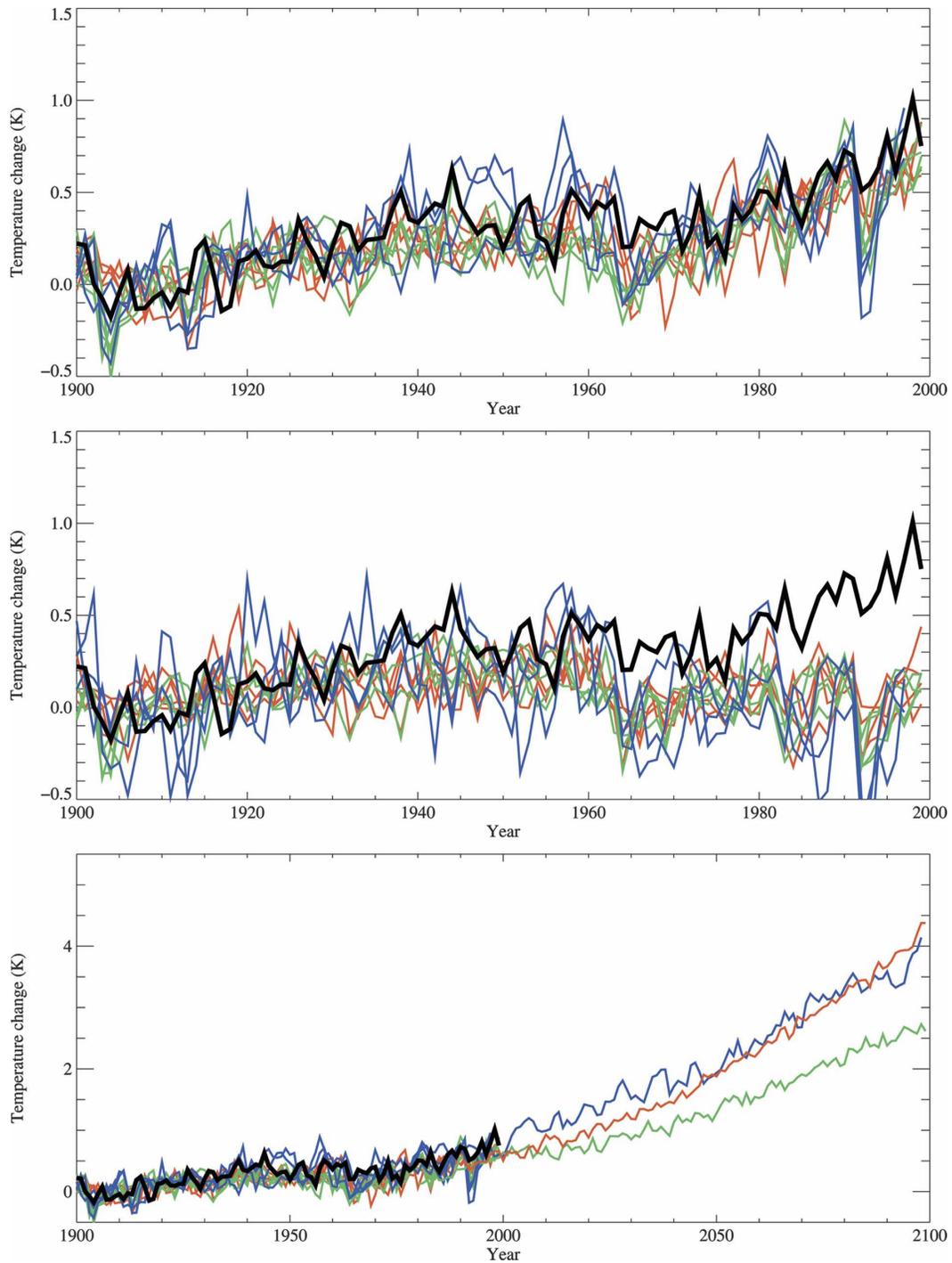


FIG. 1. Global mean temperatures from observations (black lines), and simulations of the HadCM3 (red), PCM (green), and GFDL R30 (blue) models when models are forced with (top) both anthropogenic and natural forcings, (middle) natural forcings only, and (bottom) both anthropogenic and natural forcings to 2000 followed by anthropogenic forcings to 2100 according to the SRES A2 scenario. HadCM3 predictions are the mean of three simulations whereas PCM and GFDL R30 predictions are from a single model simulation.

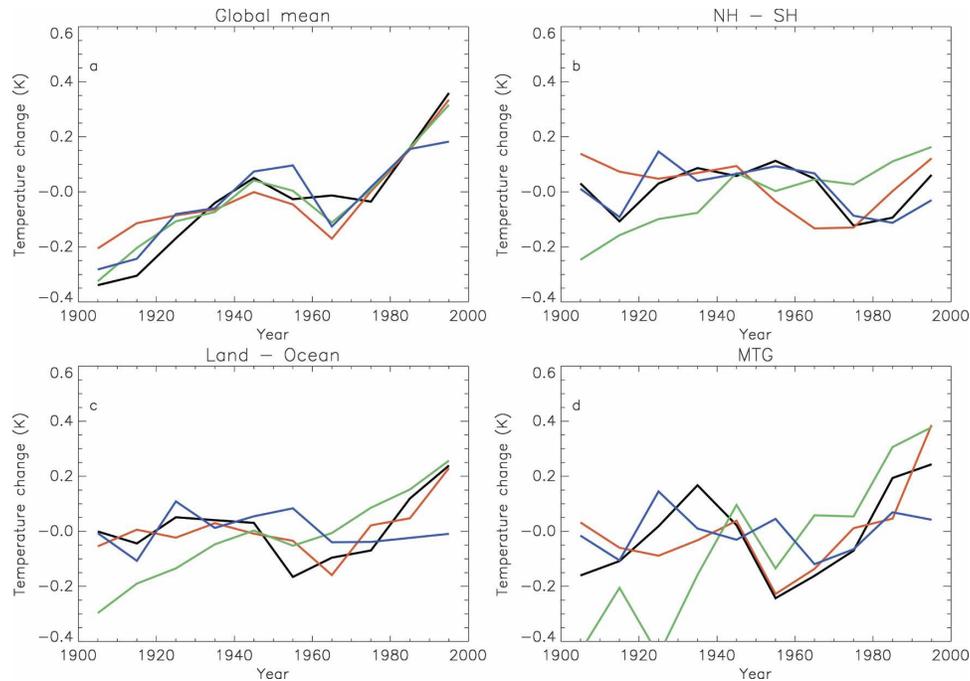


FIG. 2. Temperatures (relative to century mean) from observations (black lines), and simulations of the HadCM3 (red), PCM (green), and GFDL R30 (blue) models for both anthropogenic and natural forcings for (top left) global mean, (top right) difference between Northern Hemisphere and Southern Hemisphere, (bottom left) difference between land and ocean, and (bottom right) difference between temperatures averaged over the 52.5° to 67.5° N latitude band and the average over the 22.5° to 37.5° N latitude band [meridional temperature gradient (MTG)].

of internal variability are being included in the optimization (Allen and Tett 1999). However, nonfailure of the consistency test does not always indicate lack of problems, such as results being very sensitive to truncation. Therefore, the approach taken here is to require that the consistency test is passed and that the results are stable for a wide range of truncations. Where appropriate, results are shown for a range of truncations to illustrate sensitivity to choice of truncation.

To avoid any systematic drift in the simulations affecting the results, for each anomaly simulation, a linear trend calculated from the appropriate control segment is subtracted at each space and time point. For each of the three models we analyze the 100-yr period 1900–99. The observational dataset we use is an updated version of the combined dataset of 1.5-m air temperature over land and sea surface temperatures of Parker et al. (1994).

To replicate the analysis of Stott and Kettleborough (2002), decadal mean patterns of temperature change are projected onto spherical harmonics, using a triangular truncation and truncating at T4 resolution, which filters the data to retain only scales greater than 5000 km (Stott and Tett 1998). Some sensitivity studies are carried out, in which, rather than projecting the data

onto spherical harmonics, the data are expressed in terms of the indices of large-scale patterns of temperature change discussed in section 2d.

For all three models we estimate the scaling factors on the greenhouse gas-only signal (G), the contribution from other anthropogenic factors (mainly aerosols; S) and from natural factors (NAT). For the HadCM3 and PCM models, these are obtained from the GHG, ANTHRO, and NAT simulations by linear transformation, assuming that the climate response to these forcings is linearly additive, an assumption that appears reasonable on large spatial scales (Meehl et al. 2004a; Gillett et al. 2004). The scaling factors for the G, S, and NAT signals are obtained for the GFDL model from a linear transformation of the scaling factors on the G, GS, and GSSV signals.

Differences between temperature changes attributable to aerosols and natural factors arise from differences in the shapes of the models' spatiotemporal patterns of response. The regression methodology used here accounts for uncertainty in the patterns of response arising from having finite ensembles of model simulations, which means that internal variability noise overlies the underlying model pattern of response. In addition, model patterns can be scaled up or

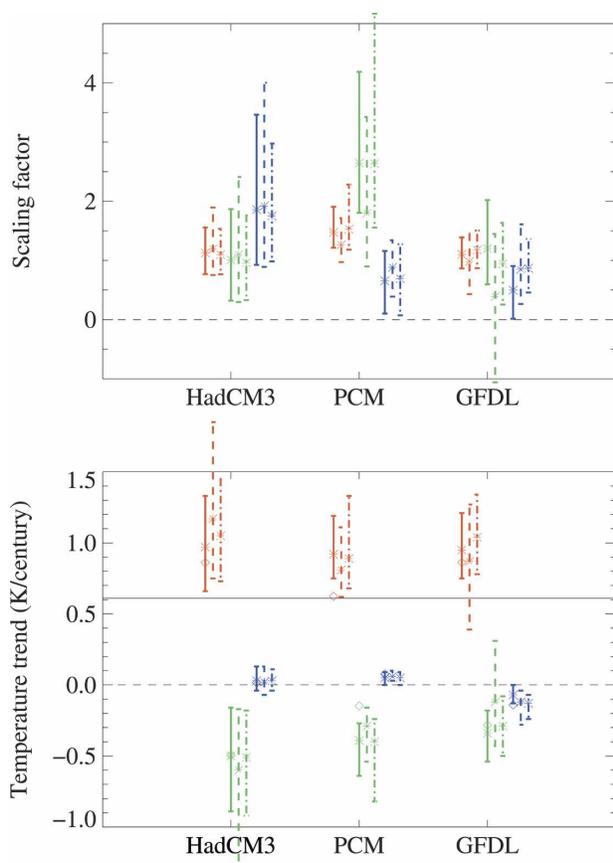


FIG. 3. (top) Scaling factors and (bottom) derived trends over the twentieth century for the G (red), S (green), and NAT (blue) contributions for HadCM3, PCM, and GFDL models. Solid lines denote analyses including both global mean and subglobal information, dash lines denote analyses including just global mean data, and dotted-dashed lines denote analyses including indices. The solid black line in the bottom panel shows observed temperature trend over the century. Diamonds show temperature changes from models unsealed by comparison with observations. Truncations for the analyses including both global mean and subglobal information are 15, 40, and 50 for HadCM3, PCM, and GFDL models, respectively; 10 for global mean analyses for all models; and 20, 20, and 30 for analyses using indices for HadCM3, PCM, and GFDL models, respectively. Truncations were chosen to pass the consistency test and so that results were stable over a range of truncations (see Fig 4).

down [the beta factors in Eq. (1)] to compensate for an over- or underestimate of the observed climatic response to different forcings.

4. Attributable temperature change over the twentieth century

The scaling factors from the three model analyses for the truncations chosen for each model analysis are shown by the solid lines in Fig. 3a and the correspond-

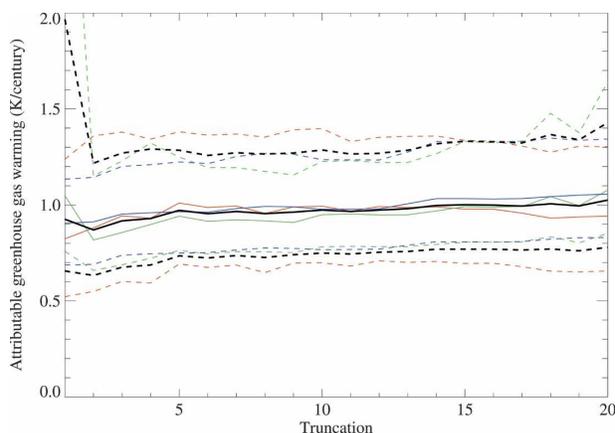


FIG. 4. Trends over the twentieth century (K) attributable to GHGs calculated from the HadCM3 (red), PCM (green), and GFDL (blue) models and for the average of all three models (black) over a range of truncations. Solid lines show 50 percentiles of the distributions and dashed lines 5 and 95 percentiles. Note that truncations are different for each model and are plotted over the range of 10–30 for HadCM3, 31–50 for PCM, and 41–60 for GFDL R30.

ing derived twentieth-century trends are shown by the solid lines in Fig. 3b. For all three models, the 5 percentiles of the warming attributable to greenhouse gases is larger than the observed warming of 0.6 K (solid line in Fig. 3b). The warming attributable to greenhouse gases is between 0.7 and 1.3 K. Good agreement between the three models is seen over a wide range of truncations (Fig. 4). From these results it seems very likely that greenhouse gases have caused more warming than has been observed over the century.

The best estimates of aerosol cooling over the century range from 0.33 to 0.49 K (Fig. 3b, solid lines). There is a larger fractional uncertainty over the amount of cooling attributable to the net effects of aerosols than the warming attributable to greenhouse gases, although for all three analyses the 5 percentiles of the corresponding scaling factors are positive, consistent with detection of aerosols in all three models (Fig. 3a). Aerosols significantly reduce the warming that would otherwise have been observed. Over the century as a whole, there is a relatively small contribution from natural factors. The HadCM3 and PCM analysis indicates a small warming from natural factors (0.05 K), whereas the GFDL analysis indicates a small cooling of -0.07 K, which combines with a smaller aerosol cooling than the other two models (Fig. 3b).

The reconstructed contributions to twentieth-century global mean temperature changes from the three model analyses are shown in Fig. 5. The results from the HadCM3 and PCM analyses are rather similar; most

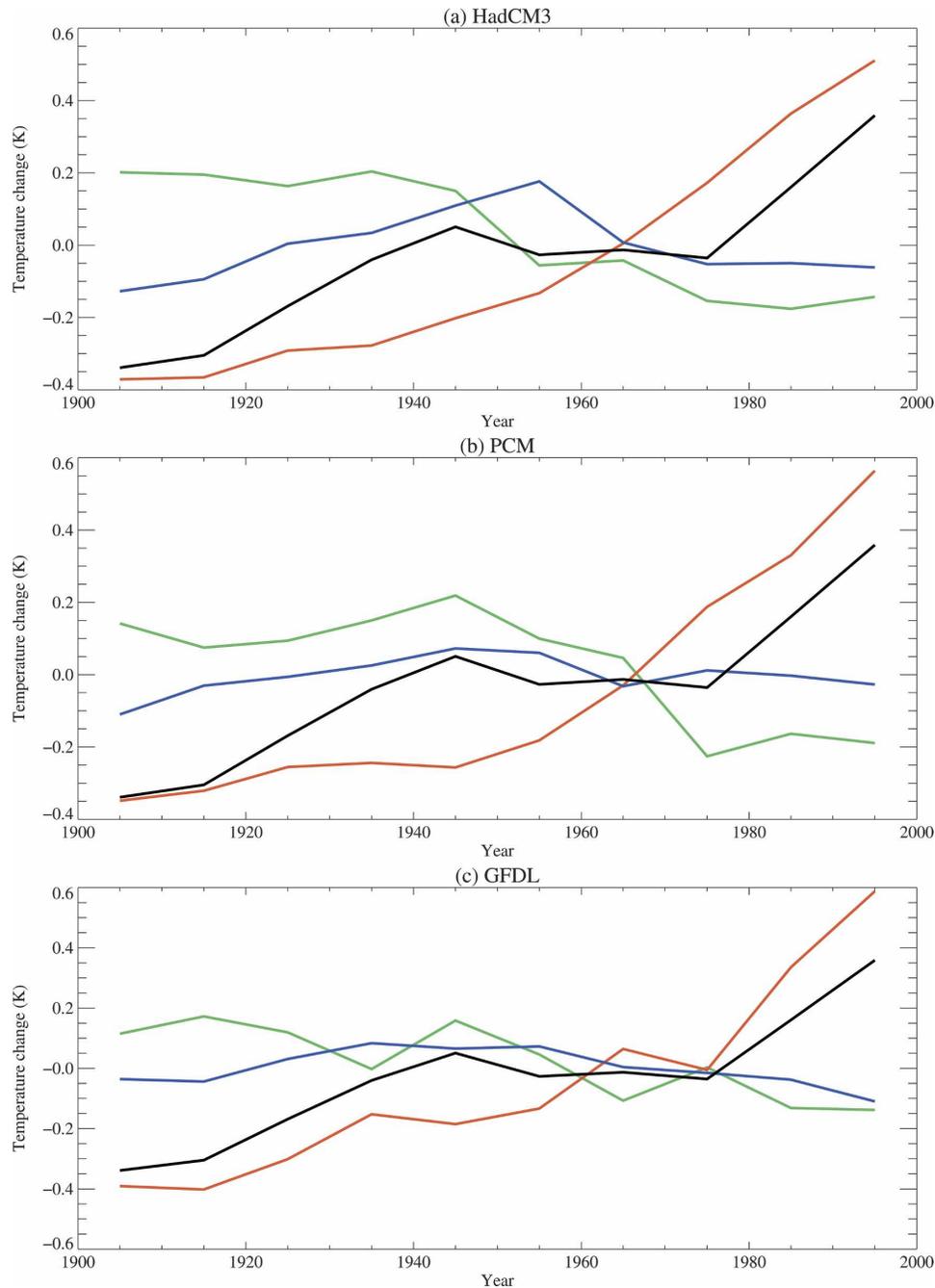


FIG. 5. Observed decadal mean global mean temperature changes over the twentieth century (black lines) and best-fit reconstructions to twentieth-century global mean temperature change for contributions from G (red), S (green), and NAT (blue) for (a) HadCM3, (b) PCM, and (c) GFDL.

aerosol cooling occurs in the middle part of the twentieth century, and there is a small early century warming from natural factors. Aerosol cooling estimated from the GFDL model is smaller, but it is still mostly concentrated in the middle part of the century. The greatest consistency between the shapes of the global

mean temporal response from the three analyses is for the attributable greenhouse warming, which steadily increases during the century and amounts to a net warming of approximately 1 K over the century. The attribution analyses bring greenhouse warming from the three models into better agreement. Scaling factors on

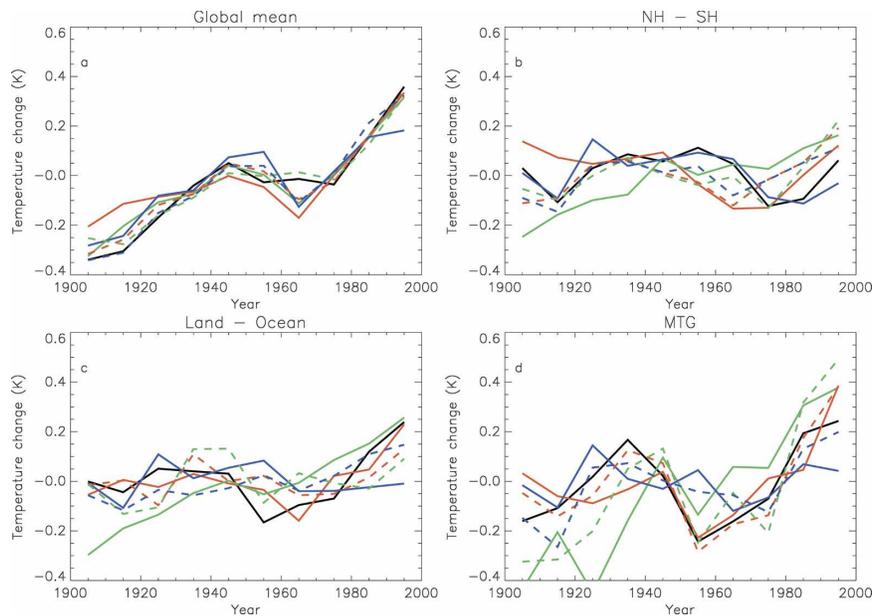


FIG. 6. Same as in Fig. 2, but with solid lines showing unscaled model simulations including both anthropogenic and natural forcings (also shown in Fig. 2) and dashed lines showing best estimates of temperatures reconstructed from the scaled contributions from G, S, and NAT.

both G and S are significantly greater than 1 (at the 5% significance level) for the PCM model (Fig. 3), which brings together the attributable greenhouse warming from the lower-sensitivity PCM model and the warming from the higher-sensitivity HadCM3 and GFDL models.

The importance of subglobal information in constraining the attributable temperature changes is demonstrated by the dashed bars in Fig. 3. These show the results when only global mean temperatures are included in the optimal detection analyses. The uncertainty estimates (5–95 percentiles) from the global mean-only analyses include the best estimates from the full analyses, but uncertainties are generally larger when only global mean data are used and there is a greater spread between models. For the GFDL analysis, aerosols are no longer detected and aerosol cooling is no longer significant. Based on an analysis of just global mean information (including the temporal evolution of global mean temperatures), the overall warming attributable to greenhouse gases could be less than the observed warming.

There is a greater difference between the models' simulations of large-scale spatial patterns of temperature change over the twentieth century than their simulation of the global mean (Fig. 2). Further investigation of the importance of subglobal information in discriminating between the climate response to the different forcings is carried out using the indices of large-scale

temperature introduced in section 2d. These are the global mean near-surface air temperature, the inter-hemispheric temperature contrast, the land-ocean temperature difference and the Northern Hemisphere meridional temperature gradient. The dashed-dotted lines in Fig. 3 (i.e., the third of each set of colored bars) show the results of an optimal detection analysis when decadal mean changes in these four indices are included. Using these four indices alone makes only small differences to results from the full spatiotemporal analysis in which patterns of temperature change are projected onto spherical harmonics, truncated at T4 resolution. This demonstrates that these indices summarize the main features of the patterns of near-surface temperature response.

Figure 6 shows that the fitted reconstructions from the three models are generally closer to the observations. The greatest improvement is for the PCM model, which had the smallest aerosol cooling. The relative lack of cooling in the middle part of the century in the PCM model results in a lack of temporal structure in the indices (solid green lines in Fig. 6). This explains why, in the PCM model, the temperature difference between Northern and Southern Hemispheres and the temperature difference between the higher- and lower-latitude bands continues to increase through the century rather than decrease in the middle part of the century and then increase in recent decades. An increase in both the greenhouse warming and the aerosol cooling

components (scaling factors larger than 1 for PCM in Fig. 3) brings the model simulations into better agreement for the subglobal indices while maintaining a good simulation of the global mean temperature evolution (Fig. 6a).

These results demonstrate that the changes with time of large-scale patterns of temperature change are responsible for constraining the relative roles of greenhouse warming and sulfate cooling over the twentieth century. Santer et al. (1996) showed that a temporal pattern in hemispheric temperature contrast would be expected in the second half of the twentieth century with the Southern Hemisphere warming more than the Northern Hemisphere in the 1950s and 1960s and the Northern Hemisphere warming more than the Southern Hemisphere subsequently. This pattern results from temporal changes in the relative strengths of the greenhouse gas and aerosol forcings. The fact that all three models provide realistic simulations of twentieth-century global mean temperature changes could be taken to mean that each model is equally good at representing the temperature changes from greenhouse gases and other climate forcings and that therefore predictions from each of the three models are equally likely. However, the detection results reported here indicate that subglobal information, in addition to the temporal pattern of global mean data, constrains the possible range of greenhouse warming and aerosol cooling likely to be consistent with the observed record.

Estimates of the likely range of aerosol cooling inferred from these analyses are given in Table 2. These are derived by scaling the raw aerosol forcing in each model (Table 1) 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 greenhouse gas forcing is well known and that errors in the response of the model to different forcings scale equally. These estimates (a combined 5–95 percentile range of 0.4–1.4 W m^{-2}) are broadly consistent with other inverse estimates of aerosol forcing based on observational constraints (Forest et al. 2006; Andronova and Schlesinger 2001; Knutti et al. 2002; Knutti et al. 2003) and appear to exclude larger magnitudes of aerosol forcing derived from forward calculations (Anderson et al. 2003).

A sensitivity study has been carried out by perturbing the HadCM3 aerosol pattern to determine how robust results are to uncertainty in the relative roles of Northern and Southern Hemisphere cooling in response to aerosol forcing. If the Southern Hemisphere aerosol cooling is increased from its model value of 35% of Northern Hemisphere cooling to over 50%, the effects of aerosols are no longer detected. Attribution results

TABLE 2. Estimates of TCR and net aerosol forcing (relative to preindustrial levels) derived from optimal detection analyses carried out on the HadCM3, PCM, and GFDL R30 models and when the PDFs derived from the three models are combined into a single PDF by averaging.

	Net aerosol forcing	TCR
	($\text{W m}^{-2} \text{K}^{-1}$)	(K century^{-1})
	5, 50, 95 percentiles	5, 50, 95 percentiles
HadCM3	0.40, 1.01, 1.38	2.12, 3.22, 4.46
PCM	0.77, 1.06, 1.47	2.20, 2.77, 3.65
GFDL	0.32, 0.68, 1.09	2.28, 3.07, 3.93
Combined	0.42, 0.93, 1.38	2.19, 3.02, 4.01

appear to be most sensitive to perturbations to the time signature of the aerosol fingerprint. If aerosol cooling is shifted earlier by two decades, agreement with the observations can only be reached by combining a much-reduced greenhouse warming with a nonphysical net warming from aerosols.

We have also investigated the sensitivity of attribution results to the length of period of the analysis by comparing an analysis for 1950–99 with the standard 1900–99 analysis and the resultant derived attributable temperature trends calculated over the 1950–99 period (Fig. 7). Anthropogenic factors (both G and S) are detected using signals derived from all three models in both the full-century analyses (Fig. 7a) and in the 1950–99 analyses (Fig. 7b). However, natural factors are only detected in the full-century analyses. Natural external forcings make a larger contribution relative to the total forcing in the first half of the century than the second half of the century, and detection of natural factors in the early part of the twentieth century but not the latter part has been seen in previous detection analyses (Tett et al. 1999, 2002). The largest difference between the two analyses is for the GFDL model, with the 50-yr analysis finding that the observations are consistent with a small greenhouse warming and a much larger warming from natural factors (Fig. 7d), a result of a negative (unphysical) NAT scaling factor (Fig. 7b). Generally, however, the 50-yr analysis supports the findings from the 100-yr analysis for a significant attributable warming from greenhouse gases that is likely to be greater than the observed warming over the last 50 yr of the century.

5. Observationally constrained predictions

Observationally constrained predictions of twenty-first-century temperature rise following the Special Report on Emissions Scenarios (SRES) A2 emissions scenario have been calculated for the three model analyses

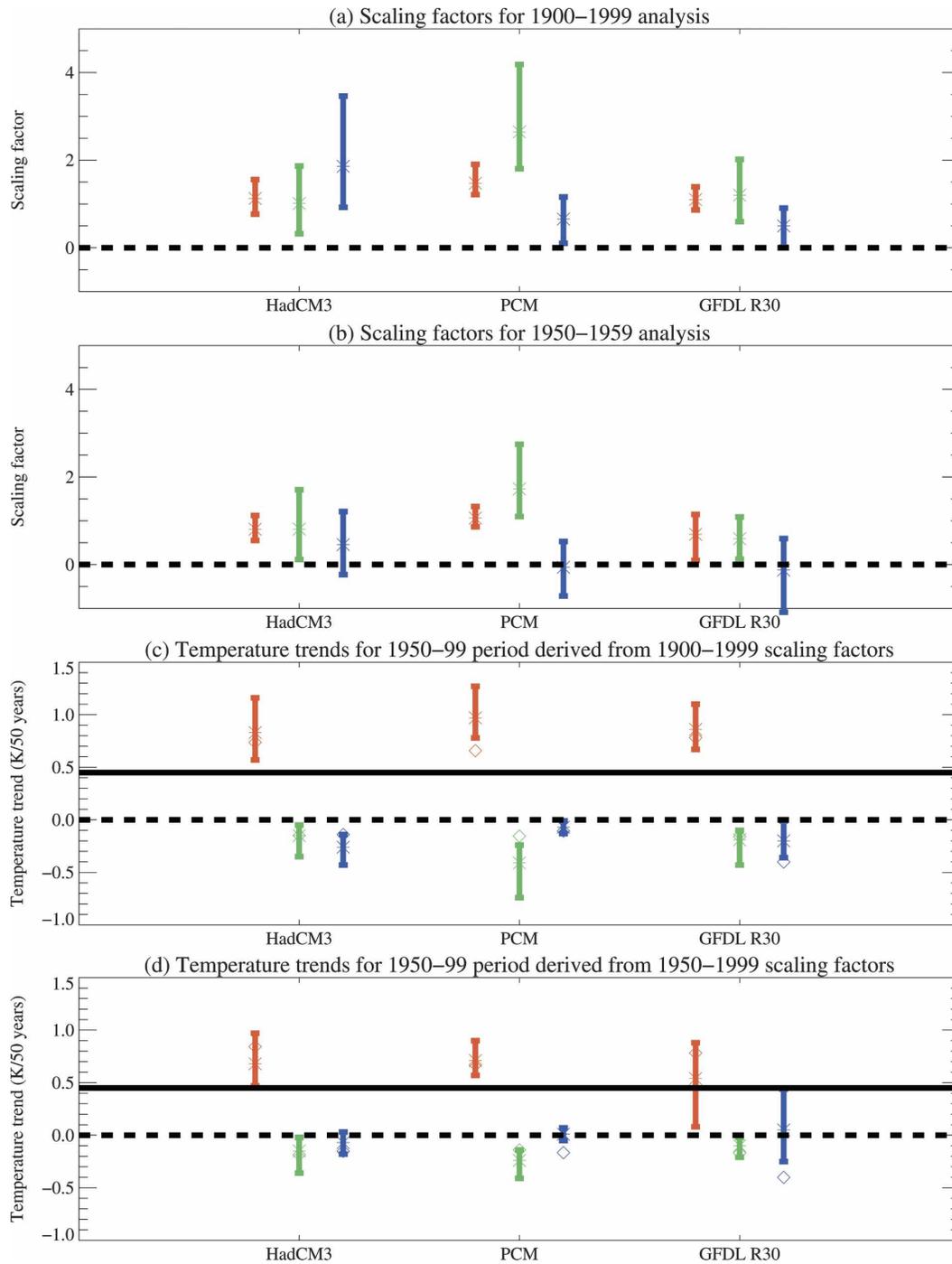


FIG. 7. Scaling factors and derived trends for the G (red), SO (green), and NAT (blue) contributions for HadCM3, PCM, and GFDL models. Scaling factors derived (a) from the 1900–99 analysis and (b) from the 1950–99 analysis; they are used to derive 1950–99 trends (c) from the 1900–99 scaling factors and (d) from the 1950–99 scaling factors. Truncations for 100-yr analysis are given in the caption to Fig. 3 and truncations for 50-yr analysis are 15, 15, and 30 for the HadCM3, PCM, and GFDL models, respectively.

in the same way as described by Stott and Kettleborough (2002) and are shown in Fig. 8a. Probability density functions (PDFs) of the scaling factors on GHG and ANTHRO [derived from Eq. (1)] are used to scale SRES A2 predictions where the GHG scaling factor scales a prediction including increases in well-mixed greenhouse gases only and the ANTHRO scaling factor scales predictions including anthropogenic forcings (where we took the mean of three predictions of HadCM3 and used single predictions of the PCM and GFDL models). We take account of sampling uncertainty due to the small number of predictions from forecast ensembles. We also include uncertainty due to future forced natural variability, which we estimate from the modeled response to past natural forcing. Finally, we also take account of the uncertainty in the actual possible future evolution, which will differ from the forced evolution because of internal variability. Further details are given in Stott and Kettleborough (2002).

The overall uncertainty in predictions of future temperature rise derived from the three models is shown by the gray shading in Fig. 8a. The scaled predictions for the three models (best estimates shown by the solid lines in Fig. 8b) are in much better agreement than the raw model predictions (dashed lines in Fig. 8b). These results support the claim of Stott and Kettleborough (2002) that these predictions are observationally constrained and therefore much less model dependent than raw model predictions. Model-related uncertainty remains, reflecting the spread of patterns of possible model-simulated responses to the different forcings, and can be seen in the spread of the 5 and 95 percentiles calculated from the three models (the colored lines in Fig. 8a). The 5 percentiles of the PDF of warming rates over the century derived from the three models range from 2.5 to 3.1 K century⁻¹ and 95 percentiles range from 4.1 to 5.3 K century⁻¹. Observational data provide information on the climate's transient response to the relatively well-known forcing from well-mixed greenhouse gases (a combination of climate sensitivity and rate of ocean heat uptake) and also on the likely range of aerosol forcing (see Table 2). Taking this information into account provides estimates of the likely range of future warming that are consistent with the observed record. Such forecasts that depend largely on observations and are relatively robust to changes in models are known as Stable Inference from Data (STAID) forecasts (KET05), whose uncertainty is expected to reduce in a predictable way as the signal strengthens (Stott and Kettleborough 2002).

Finally, we calculate PDFs of the transient climate response (TCR) to a 1% yr⁻¹ increase in carbon diox-

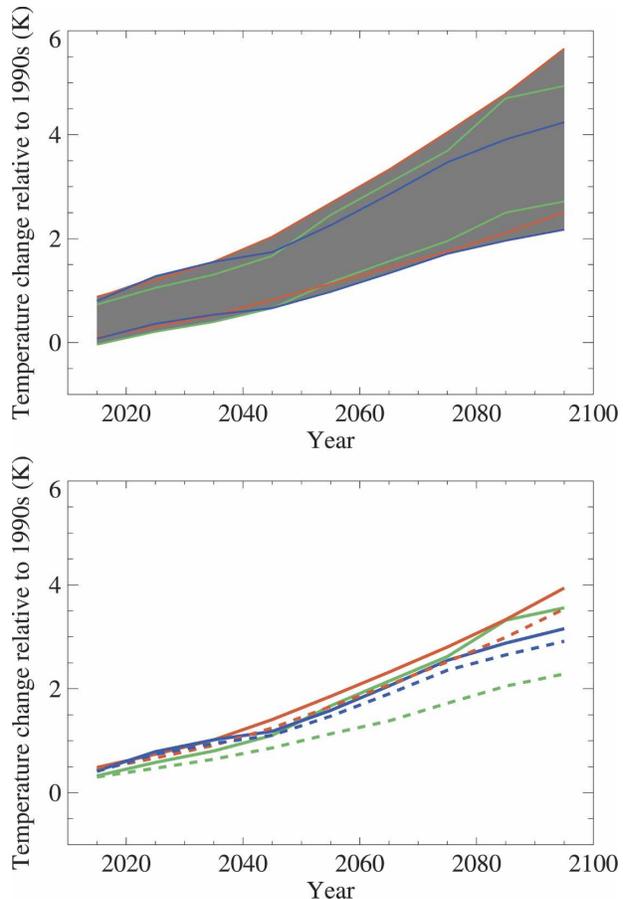


FIG. 8. (top) Uncertainty ranges of observationally constrained predictions for the three models (gray shading) and their 5–95 percentiles uncertainty ranges for HadCM3 (red lines), PCM (green), and GFDL (blue). (bottom) Comparison of observationally constrained predictions (best estimates, solid lines) with raw model predictions (dashed lines).

ide (Cubasch et al. 2001), which is calculated from the globally averaged surface air temperature difference between the average temperature over the 20-yr period around the time of CO₂ doubling and the mean temperature of the control run (which has preindustrial levels of CO₂), expressed as a warming rate over the century. The unscaled values for TCR for the three models are shown in Table 1 (Cubasch et al. 2001; Meehl et al. 2004b; Cubasch et al. 2001). PDFs of TCR are calculated in the same way as for temperature change under the SRES A2 scenario, the main difference being that only the scaling factors on the greenhouse gas component constrains the TCR, it being the response under greenhouse gas forcing only. As for the temperature predictions shown in Fig 8, we take account of uncertainty in the observationally constrained TCR due to internal variability and having only one prediction of TCR from each model. To account for

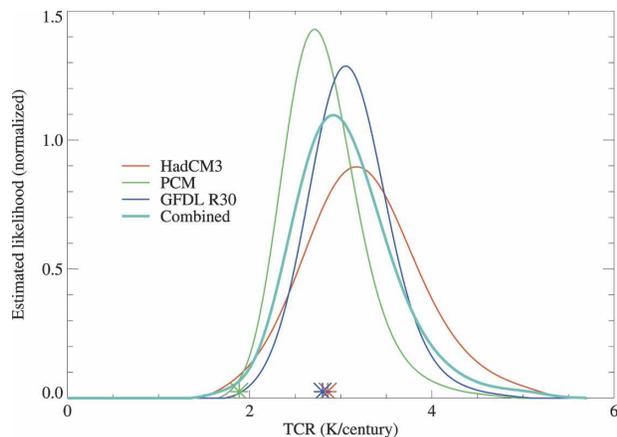


FIG. 9. Probability distributions of TCR (expressed as warming rates over the century), as constrained by observed twentieth-century temperature change, for HadCM3 (red), PCM (green), and GFDL (blue). Colored stars show each model's TCR.

this, the variance of the normally distributed perturbations is calculated from the control run (in this case for 20-yr mean temperature) and added to the uncertainty in TCR. The three PDFs calculated using each model are shown in Fig 9. Past observed temperature change indicates that the TCR is likely to be higher than the PCM model's TCR, and this is the case even when the analysis is carried out solely using PCM data. Also shown in Fig 9 is the PDF calculated by weighting each model equally and combining the three individual PDFs into one single probability distribution by averaging. The mean TCR from this distribution is 3.0 K with 5 and 95 percentiles of 2.2 and 4.0 K, respectively.

6. Summary and discussion

Results from optimal detection analyses, similar to that described by Stott and Kettleborough (2002), when applied to three climate models with different sensitivities and forcings, show that global mean warming over the twentieth century attributable to anthropogenic greenhouse gas emissions is well constrained by the observational record. All three models, when constrained by observations, suggest that warming attributable to greenhouse gases is between 0.7 and 1.3 K (5 and 95 percentiles) and therefore probably greater than the observed warming of 0.6 K over the century.

The features that constrain the likely temperature response to anthropogenic and natural forcings are the temporal structure of the observed global mean temperature changes over the twentieth century and the large-scale features of the spatial patterns of response. Distinctive temporal structures in differential warming

rates between the hemispheres, between land and ocean, and between mid- and low latitudes help to discriminate between models and determine the relative roles of greenhouse warming and sulfate cooling.

These results emphasize that care should be taken not to overinterpret good agreement between climate models and past observed global mean warming. With large uncertainties in climate forcings, especially that due to aerosols, agreement when models include all the most important anthropogenic and natural forcings could be obtained fortuitously as a result of, for example, balancing too much (or too little) greenhouse gas warming by too much (or too little) aerosol cooling. Optimal detection analyses determine which parts of the responses of different models to particular climate forcings are likely to be consistent with the estimated observed response.

As is discussed in detail by KET05 an emergent constraint of the climate system provides a linear relationship between past and future warming because the fractional error in models' simulations of global mean temperature stays approximately constant over time. The results presented here demonstrate that observationally constrained predictions produced in this way are STAID, that is, they are relatively robust to changes in models.

A limitation of our approach is that modeling uncertainty is represented by a single scaling factor on the pattern of a model's response to a particular forcing. To avoid this "perfect model" assumption, it will be necessary to include an estimate of the covariance structure of modeling uncertainty in the analysis. An attempt to combine modeling uncertainty in a single analysis was made by Gillett et al. (2002), who calculated the mean response patterns from five models [HadCM2, HadCM3, the Coupled GCM, versions 1 and 2 (CGCM1 and CGCM2), and ECHAM3], which they used as fingerprints in a detection of greenhouse gas and sulfate aerosol influence. Their estimate of modeling uncertainty was obtained by a simple rescaling of the variability estimated from a long control run, thereby assuming that intramodel uncertainty has the same structure as internal variability. Recently, a more sophisticated methodology for incorporating modeling uncertainty into detection analyses has been developed, by including an estimate of the model error covariance structure, obtained from intermodel differences, in the regression method (Huntingford et al. 2006).

Estimation of modeling uncertainty is dependent on the model simulations available and large multimodel ensembles using different models to estimate structural uncertainty will be required before a more systematic

estimate of uncertainty can be derived (Murphy et al. 2004; Stainforth et al. 2005). The analysis reported here indicates that uncertainty in attribution of global mean temperature is relatively small; there are likely to be much larger uncertainties for other climate variables such as precipitation changes. Murphy et al. (2004) showed that scaling patterns of surface temperature response to doubled CO₂ from one model according to the climate sensitivity reproduces fairly well the patterns from response from 52 other model members of a perturbed physics ensemble. However, this scaling approach works much less well for precipitation changes. Therefore, multimodel ensembles will be required to estimate attributable changes in precipitation on regional scales.

All our results are subject to unquantified uncertainties arising from missing forcings and imperfect representation of climate physics. Nevertheless, the results presented here demonstrate that a forecast determined from observational constraints is relatively robust to changes in models. A major contributor in the future to reducing uncertainty in predictions of global mean temperatures is likely to be the strengthening signal of climate change (Stott and Kettleborough 2002). Our current estimate is that transient climate response under a 1% increase in carbon dioxide has a range of 2.2–4.0 K (5–95 percentiles). Warming rates consistent with observational constraints lie between 2.5 and 5.3 K century⁻¹ under the SRES A2 scenario of anthropogenic emissions.

Acknowledgments. This work was partly funded by the U.K. Department of the Environment, Food and Rural Affairs under Contract PECD 7/12/37 (PAS, JMG) and by the Met Office Corporate Overhead (JFBM) with some additional funding from the NOAA/DOE International Detection and Attribution Group (MRA). A portion of this study (GAM) was supported by the Office of Biological and Environmental Research, U.S. Department of Energy, as part of its Climate Change Prediction Program, and the National Center for Atmospheric Research.

REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- , and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting. Part I: Theory. *Climate Dyn.*, **21**, 477–491.
- , —, J. F. B. Mitchell, R. Schnur, and T. L. Delworth, 2000: Uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.
- Ammann, C. M., G. A. Meehl, W. M. Washington, and Z. Zender, 2003: A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.*, **30**, 1657, doi:10.1029/2003GL016875.
- Anderson, T. L., R. J. Charlson, S. E. Schwartz, R. Knutti, O. Boucher, H. Rodhe, and J. Heintzenberg, 2003: Climate forcings by aerosols—A hazy picture. *Science*, **300**, 1103–1104.
- 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.
- Braganza, K., D. J. Karoly, A. C. Hirst, M. E. Mann, P. A. Stott, R. J. Stouffer, and S. F. B. Tett, 2003: Simple indices of global climate variability and change: Part I—Variability and correlation structure. *Climate Dyn.*, **20**, 491–502.
- Broccoli, A. J., K. W. Dixon, T. L. Delworth, and T. R. Knutson, 2003: Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing. *J. Geophys. Res.*, **108**, 4798, doi:10.1029/2003JD003812.
- Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 525–582.
- Delworth, T. L., R. J. Stouffer, K. W. Dixon, M. J. Spelman, T. R. Knutson, A. J. Broccoli, P. J. Kushner, and R. T. Wetherald, 2002: Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Climate Dyn.*, **19**, 555–574.
- Forest, C. E., P. H. Stone, and A. P. Sokolov, 2006: Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.*, **33**, L01705, doi:10.1029/2005GL023977.
- Gillett, N. P., F. W. Zwiers, A. J. Weaver, G. C. Hegerl, M. R. Allen, and P. A. Stott, 2002: Detecting anthropogenic influence with a multi-model ensemble. *Geophys. Res. Lett.*, **29**, 1970, doi:10.1029/2002GL015836.
- , M. F. Wehner, S. F. B. Tett, and A. J. Weaver, 2004: Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.*, **31**, L14201, doi:10.1029/2004GL020111.
- Gordon, C., C. Cooper, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation 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.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, G. Meehl, 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.
- 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.
- Karoly, D. J., and K. Braganza, 2001: Identifying global climate change using simple indices. *Geophys. Res. Lett.*, **28**, 2205–2208.
- , —, P. A. Stott, J. Arblaster, G. Meehl, A. Broccoli, and K. W. Dixon, 2003: Detection of a human influence on North American climate. *Science*, **302**, 1200–1203.
- Knutti, R., T. F. Stocker, 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.

- , —, —, and —, 2003: Probabilistic climate change projections using neural networks. *Climate Dyn.*, **21**, 257–272.
- Meehl, G. A., W. M. Washington, T. M. L. Wigley, J. M. Arblaster, and A. Dai, 2003: Solar and greenhouse gas forcing and climate response in the twentieth century. *J. Climate*, **16**, 426–444.
- , —, C. M. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi, 2004a: Combinations of natural and anthropogenic forcings in twentieth-century climate. *J. Climate*, **17**, 3721–3727.
- , —, J. M. Arblaster, and A. Hu, 2004b: Factors affecting climate sensitivity in global coupled models. *J. Climate*, **17**, 1584–1596.
- Mitchell, J. F. B., D. J. Karoly, G. C. Hegerl, F. W. Zwiers, M. R. Allen, and J. Marengo, 2001: Detection of climate change and attribution of causes. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 525–582.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **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.
- Pope, V. D., M. L. Gallani, 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.
- Raper, S. C. B., J. M. Gregory, and R. J. Stouffer, 2002: The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *J. Climate*, **15**, 124–130.
- Santer, B. D., and Coauthors, 1996: Reply to “Human effect on global climate?” *Nature*, **384**, 522–525.
- , and Coauthors, 2003a: Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science*, **301**, 479–483.
- , and Coauthors, 2003b: Influence of satellite data uncertainties on the detection of externally-forced climate change. *Science*, **300**, 1280–1284.
- Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.
- Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate*, **11**, 3282–3294.
- , and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in twenty-first century temperature rise. *Nature*, **416**, 723–726.
- , —, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137.
- Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth century temperature change near the earth’s surface. *Nature*, **399**, 569–572.
- , and Coauthors, 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107**, 4306, doi:10.1029/2000JD000028.
- Washington, W. M., and Coauthors, 2000: Parallel climate model (PCM) control and transient simulations. *Climate Dyn.*, **16**, 755–774.