ORIGINAL PAPER

Quantifying anthropogenic influence on recent near-surface temperature change

**M. R. Allen · N. P. Gillett · J. A. Kettleborough ·
G. Hegerl · R. Schnur · P. A. Stott · G. Boer ·
C. Covey · T. L. Delworth · G. S. Jones ·
J. F. B. Mitchell · T. P. Barnett**

Abstract   We assess the extent to which observed large-scale changes in near-surface temperatures over the latter half of the twentieth century can be attributed to anthropogenic climate change as simulated by a range of climate models. The hypothesis that observed changes are entirely due to internal climate variability is rejected at a high confidence level independent of the climate model used to simulate either the anthropogenic signal or the internal variability. Where the relevant simulations are available, we also consider the alternative hypothesis that observed changes are due entirely to natural external influences, including solar variability and explosive volcanic activity. We allow for the possibility that feedback processes, other than those simulated by the models considered, may be amplifying the observed response to these natural influences by an unknown amount. Even allowing for this possibility, the hypothesis of no anthropogenic influence can be rejected at

M. R. Allen (✉)
Atmospheric, Oceanic and Planetary Physics, University of Oxford Clarendon Laboratory,
Parks Road OX1 3PU, UK
e-mail: myles.allen@physics.ox.ac.uk

N. P. Gillett
School of Earth and Ocean Sciences, University of Victoria, Victoria, Canada

Present Address:
N. P. Gillett
Climate Research Unit, School of Environmental Sciences, University of East Anglia,
Norwich, UK

J. A. Kettleborough
Space Science and Technology Department, Rutherford Appleton Laboratory,
Didcot, UK

Present Address:
J. A. Kettleborough
Met. Office, FitzRoy Road, Exeter, UK

G. S. Jones · J. F. B. Mitchell
Met. Office, FitzRoy Road, Exeter, UK

the 5% level in almost all cases. The influence of anthropogenic greenhouse gases emerges as a substantial contributor to recent observed climate change, with the estimated trend attributable to greenhouse forcing similar in magnitude to the total observed warming over the 20th century. Much greater uncertainty remains in the response to other external influences on climate, particularly the response to anthropogenic sulphate aerosols and to solar and volcanic forcing. Our results remain dependent on model-simulated signal patterns and internal variability, and would benefit considerably from a wider range of simulations, particularly of the responses to natural external forcing.

Keywords Climate change · Detection · Attribution

1 Background

The Second Assessment Report (SAR) of the Intergovernmental Panel on Climate Change (IPCC: Houghton et al. 1996) concluded that ''the balance of evidence suggests a discernible human influence on global climate'' but cautiously avoided attempting to quantify the magnitude of this influence. Strictly interpreted, the IPCC statement in the SAR therefore allowed for the possibility of a statistically significant anthropogenic climate change which was too small to have any practical importance, although the SAR authors noted at the time that ''our current inability to estimate reliably the fraction of the observed temperature changes that are due to human effects does not mean that this fraction is negligible''. This proved a remarkably prescient statement: much of the progress in detection and attribution over subsequent years has been in quantifying this anthropogenic contribution to recent observed warming and it has indeed been found to be very far from negligible.

The IPCC Third Assessment Report (TAR: Houghton et al. 2001) was in a position to make a more quantitative assessment of the extent to which recent observed climate changes are attributable to human influence, particularly the impact of anthropogenic greenhouse gas emissions. This is only partly due to an apparently strengthening signal in the observational record, even though the 1990s

G. Hegerl
Nicholas School for the Environment and Earth Sciences, Duke University, Durham, NC, USA

R. Schnur
Max Planck Institute for Meteorology, Hamburg, Germany

P. A. Stott
Met Office, Reading Unit, Dept. of Meteorology, University of Reading, Reading, UK

G. Boer
The Canadian Centre for Climate Modelling and Analysis, Victoria, BC, Canada

C. Covey
PCMDI, Lawrence Livermore National Laboratory, Livermore, USA

T. L. Delworth
NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

T. P. Barnett
Scripps Institution for Oceanography, University of California, San Diego, USA

were significantly warmer than the century as a whole. We use data to September 1996 in this study, which is only 2–3 years longer than Santer et al. (1996b) and Heger et al. (1996). Inclusion of late-1990s data generally strengthens our conclusions (Hegerl and Allen 2002), but we have chosen to exclude these years because the anomalously large El Niño event of 1997–1998, which was associated with record-breaking global temperatures, could introduce a bias into our results which are based on climate models with relatively coarse ocean resolution (Meehl et al. 2001). Since this study was undertaken, progress has been made in the simulation of El Niño in climate models (AchutaRao and Sperber 2006), so studies updating this work (Stone et al. 2006; Stott et al. 2006) have extended the period of interest into the 21st century, which generally strengthens conclusions.

A more important development since the SAR is the availability of a wider range of simulations with coupled atmosphere-ocean general circulation models (A-OG-CMs) of the climate response to different external forcing scenarios, including various combinations of anthropogenic forcing agents and natural external influences such as variations in solar and volcanic activity. A crucial element of any quantitative assessment of the extent to which an observed change can be attributed to a particular external factor is an estimate of the expected change due to that factor, together with estimates of expected changes due to physically plausible alternatives. In simple terms, to assess whether a signal is there, we need to know both what we are looking for and what we need to discriminate against.

The most rigorous approach to comparing multiple explanations of recent climate change using model-based estimates of the signals under investigation is the ''multi-fingerprinting'' algorithm of Hasselmann (1997), as first implemented in Hegerl et al. (1997). This latter study concluded that both anthropogenic greenhouse-gas and sulphate influence were required to account for the observed spatial pattern of boreal summertime warming trends over the past 50 years, while cautiously noting that some influence of solar variability might be required to account for the warming observed in the early decades of the twentieth century. In a number of follow-up papers (Barnett et al. 1999; Hegerl et al. 2000), the authors have examined the sensitivity of this result to the model used to simulate the characteristics of anthropogenic signals and internal climate variability, finding that their original conclusions regarding sulphate influence on climate are particularly sensitive to the model used and assumptions made in the analysis. North and Stevens (1998), taking a formally distinct but functionally equivalent (North et al. 1995; Hegerl and North 1997) approach based on a very different type of climate model, arrived at similar conclusions.

Allen and Tett (1999), proposed an alternative formulation of the Hasselmann (1997), multi-fingerprinting algorithm to allow a more direct interpretation of results. Without any change in the underlying formalism, they re-phrased the problem as an optimal estimation procedure, with the parameters being estimated being the factors by which it is necessary to scale model-simulated signals to reproduce observed climate change. Under their interpretation, a scaling factor consistent with zero implies that a particular model-simulated signal is not detectable in the observed climate record using the diagnostic under consideration, while a scaling factor consistent with unity implies the model-simulated amplitude could be correct. In a very similar vein, Leroy (1998), reformulated the North and Stevens (1998), approach in terms of Bayesian estimation theory: the algorithms of Allen and Tett (1999), and Leroy (1998), are formally identical in the absence of any prior expectation about (or equivalently, given infinite prior variance in) scaling factors to be applied to individual signals.

Tett et al. (1999), applied this approach to observed spatio-temporal patterns of surface temperature change (Hegerl et al. 1996, and subsequent papers focussed on spatial patterns of linear temperature trends) and concluded, like Hegerl et al. (1997), that both greenhouse and sulphate influence are required to account for observed near-surface temperature changes in the latter half of the 20th century, with a possible role for solar variability in the warming that occurred over the period 1910–1940. With ensemble simulations of the response to natural external forcing agents at their disposal, they were able to draw stronger conclusions than previous authors regarding the difficulty of accounting for observed changes in exclusively natural terms. Tett et al. (2002), using a more up-to-date model and a wider range of diagnostics, reached similar conclusions while finding an even clearer role for solar forcing in the early century warming.

Stott et al. (2001), show that the inclusion of seasonal information strengthens the evidence for solar influence on the early part of the century. In contrast, Delworth and Knutson (2000), using a different model to estimate internal variability which displays greater variance on 50–80-year timescales, conclude that the warming that occurred early in the century can be accounted for through a combination of anthropogenic influence and internal variability. There is no inherent inconsistency between these two results: Stott et al. (2001), were using a more powerful analysis, in that they included both seasonal information and a model-based estimate of the spatio-temporal pattern of response to solar forcing that they were looking for (not available to Delworth and Knutson (2000), since the climate model that they used had not yet been run with estimates of changing solar forcing). It is not surprising that a more powerful analysis detects a weak signal when a less powerful one fails, but this sensitivity to the diagnostic and/or model used indicates that the evidence for a solar role in early century warming remains weaker than the evidence for anthropogenic influence in more recent decades.

The earliest fingerprinting studies, such as Santer et al. (1995, 1996b), Hegerl et al. (1996) and (1997), used signals based either on the equilibrium response of a climate model to a particular forcing or on the transient response over some period in the future. In both cases, the forcing would be sufficiently strong that sampling uncertainty in the model-simulated response would be negligible: if the experiment were repeated, essentially the same model-simulated signal would be obtained. In contrast, more recent studies, beginning with Tett et al. (1996), have used signals based on model-simulated responses to forcing changes over the same period as the observations used in the comparison. This approach simplifies interpretation of re-sults, but introduces the complication that climate forcing is not as strong over the 20th century, so model-simulated signals are contaminated by internal variability. This contamination can be reduced through the use of ensemble simulations, as in Tett et al. (1999), or through the use of noise-free climate models, as in North and Stevens (1998). The ensemble approach is expensive, so ensembles are not available for all models, while any climate model that is free of internal variability is also likely to be lacking key non-linear feedback processes that may affect the model-simulated signal.

Allen and Stott (2003), and Stott et al. (2003), describe and test a modification to the standard optimal estimation procedure that accounts explicitly for the presence of sampling noise in model-simulated signals based on single-member or small-size ensembles. They follow the standard ''Total Least Squares'' approach originally proposed by Adcock (1878), and documented in detail by, for example, Ripley and

Thompson ([1987]), and van Huffel and Vanderwaal ([1994]). Stott et al. ([2003]), demonstrate that this revised algorithm is particularly important if small (one- or two-member) ensembles are used in climate model simulations, although it can still make an appreciable difference even if larger ensembles are available, particularly on upper bounds of uncertainty ranges and when considering relatively weak model-simulated signals (such as the response to natural external influences). A range of ensemble-sizes are available to this study, so we use this explicit Total Least Squares approach.

2 Diagnostics, observations and models

In this study, we focus on observed changes in near-surface air temperature and sea-surface temperature as compiled in the Parker et al. ([1994]), dataset, updated. The data are originally available on a $5° \times 5°$ degree monthly spatio-temporal grid and are expressed as non-overlapping decadal mean temperature anomalies about the 1961–1990 climatology, with decades running from September 1906 to August 1996 inclusive. Surface air temperatures from a range of models driven with a variety of scenarios for external forcing together with a number of ''control'' (constant forcing) integrations were likewise expressed as anomalies about corresponding decades and bilinearly interpolated to the observational grid. On the very long (10–90 year) timescales we are interested in, fluctuations in surface air temperature are closely correlated with fluctuations in underlying sea-surface temperatures such that any error resulting from merging the two is likely to be small relative to residual systematic errors in the observations themselves. The impact of random errors and errors due to sampling in the observational data was explored by Jones and Hegerl ([1998]) and found to be small on the timescales of interest here. Systematic errors in the observations would have a much more substantial impact on results. Quantitative estimates of the impact of known sources of potential systematic error, such as the so-called ''urban heat island'' effect, indicate they are likely to have only a minor effect on results Parker ([2004]). The possibility of a completely unknown source of bias contaminating the early century instrumental temperature record will always remain a caveat, but more recent studies using multiple strands of climate proxy data, not all of which need to be calibrated against the instrumental record, provide some independent support (e.g., Hegerl et al. [2003], [2006]).

We will study these changes using essentially the diagnostic proposed in Stott and Tett ([1998]), projecting the data onto spherical harmonics to focus exclusively on large spatial scales. Noting the controversy over the origin of early century temperature changes (Tett et al. [1999]; Delworth and Knuston, [2000]), we focus on the period September 1946 to August 1996, but we also express data as anomalies about the mean of the 90-year period 1906–1996. In locations where no data is available prior to 1946, this naturally makes no difference. In locations with data early in the century, it allows us to exploit the information that the most recent 50 years have been generally warmer than the century as a whole without attempting to fit the details of decadal changes earlier in the century. Our main reason for using this longer climatology base period was to avoid focussing exclusively on trends over the 1946–1996 period (as, for example, in the analysis of Hegerl et al. [1997]) which, it has been argued, may happen to be a particularly effective period for the detection of

climate change not simply because the signal is strongest over this period, but because the noise in the observed record may, by chance, happen to have inflated the apparent strength of the anthropogenic signal.

Six Atmosphere-Ocean General Circulation Models (A-OGCMs) were originally included in this study. These were the HadCM2 model from the Met. Office (Johns et al. 1997); the ECHAM3 and ECHAM4 models from the Max Planck Institute for Meteorology in Hamburg (Cubasch et al. 1994; Roeckner et al. 1999); the R-30 resolution version of the GFDL climate model (Knutson et al. 1999) and two versions of the Canadian Centre for Climate Modelling and Analysis model (Boer et al. 2000). In this paper, we will refer to these as HCM2, ECH3, ECH4, GFDL, CCC1 and CCC2, respectively. A range of external climate forcings were considered, including anthropogenic greenhouse gases (G); the direct radiative forcing due to sulphate aerosols (S); the combination of indirect sulphate forcing with tropospheric ozone changes (I); variations in total incoming solar irradiance from the Hoyt and Schatten (1993), reconstruction (H), extended with satellite data (W. Ingram, *pers. comm.*); solar variations from the Lean et al. (1995) reconstruction (L); and volcanic aerosol forcing from the Sato et al. (1993), reconstruction (V).

In several cases, not all these forcings were imposed in combination, and not all models were run with all forcing agents. Moreover, sulphate aerosol forcing in particular varies significantly between models. Only the direct effect of sulphate aerosols on albedo was included in the HCM2, ECH3, GFDL, CCC1 and CCC2 models, but there are still significant differences between these runs in the spatial pattern and particularly the time-history of the sulphate forcing. The indirect effect of sulphate aerosols on cloud optical properties was represented in the HCM3 and ECH4 models, introducing further differences. In this study, we do not distinguish between differences in forcing and differences in response: each model run, with its associated forcing datasets, is considered a self-contained representation of reality. Identifying the origins of model-data discrepancies requires us to isolate whence those discrepancies arise. Since this study was undertaken, many more integrations have been performed to "fill in the gaps" in Table 1, which is a very welcome

Table 1 Summary of simulations used in this study

Model	Full name	Resolution	Size of ensembles						Length control
			G	GS	GSI	H	L	V	
HCM2	HadCM2	3.75° × 2.5° L19	4	4		4	4	4	1,700 year
ECH3	ECHAM3/LSG	T21L19	4	2		2	2		1,900 year
ECH4	ECHAM4/OPYC3	T42L19	1	1	1				1,000 year*
GFDL	GFDL/R30	R30L14		5					1,000 year
CCC1	CGCM1	T32L10		3					850 year
CCC2	CGCM2	T32L10		3					850 year

All models consist of a dynamical atmosphere coupled to a dynamical (but non-eddy-resolving) ocean. Lxx refers to the number of vertical levels in the model atmospheres, while Txx and Rxx refers to the triangular or rhombic spherical harmonic truncation used in spectral models, with higher numbers corresponding to higher horizontal resolution: for comparison, the grid-point model HadCM2 has a similar effective horizontal resolution as a T42 spectral model

*In the case of the ECHAM4 atmospheric model, we were obliged to rely on the 1,000-year ECHAM4/HOPE control integration, since the 200-year control integration of the ECHAM4/OPYC3 model was too short for either the definition of the detection space or for uncertainty analysis

development. Overall conclusions for the detectability of human influence on climate are essentially unchanged: see, for example, Stott et al. (2006); Stone et al. (2006) for updated conclusions with more recent runs.

A summary of integrations considered in this study (including the available control integrations) is given in the Table 1: further details of the models and forcing timeseries used are available in the literature.

Results from three additional models, HadCM3, CSM and PCM, became available in the course of this study. These are documented in detail elsewhere (in Stott et al. 2000; Tett et al. 2002; Blackmon et al. 2001; Washington et al. 2000) and so are not included in the majority of figures and discussion herein, but have been included (referred to as HCM3, CSM and PCM, respectively) in the summary figures for the sake of completeness.

The gridded observational data were expressed as decadal anomalies about the 1906–1996 time-mean of the relevant grid-box by averaging annual means in each decade, with grid-boxes set to missing if 5 years or more had less than 8 months available data or if fewer than two decades were available for the calculation of the time-mean (other thresholds were explored and found to have little impact on results: the crucial point is that observations and models must be masked similarly before computing time-means and anomalies). Corresponding segments of model output were likewise expressed as decadal averages, bilinearly interpolated onto the observational grid, masked with the pattern of missing data in the observations and expressed as anomalies about their respective 1906–1996 time-mean fields. These were used to define the signals under investigation. As is standard practice in detection and attribution studies, sequences of nine-decade segments of "pseudo-observations" were extracted from the available unforced control integrations, overlapping each other by 8 decades to maximise the number of segments extracted, and interpolated and masked in the same manner. These were used to define the "detection space" (coordinate system) used for model-data comparison and for uncertainty analysis.

Treating the observations, scenario runs and each nine-decade segment of control model output identically, the five decades corresponding to the 1946–1996 period were extracted and projected onto spherical harmonics, truncating at T4 to retain only scales greater than 5,000 km (Stott and Tett 1998). The resulting observation vector, $\mathbf{y}$, thus contains 125 elements, being the 25 spherical harmonic coefficients in a T4 truncation corresponding to each of the 5 decades in the 50-year period. In projecting incomplete fields onto spherical harmonics, missing data were set to zero before the projection and the resulting harmonic coefficients uniformly scaled such that the lowest-order coefficient corresponded to an unbiased estimate of the global mean temperature (other methods of estimating spherical harmonics from incomplete data are discussed in Stott and Tett 1998). The impact of this is to give each decade equal weight in the observation vector even though they may contain different amounts of missing data. Since the fractional coverage of decadal-mean temperature anomalies changes relatively little over the latter half of the twentieth century, this weighting is found to have little impact on results (Stott et al. 2001). Extending the analysis to include earlier decades would make results more sensitive to the relatively arbitrary decisions that need to be made concerning the treatment of missing data.

In addition to the issue of changing coverage, we also choose to focus on the latter half of the 20th century because the origin of the observed trend early in the century

remains controversial (Tett et al. 1999; Delworth and Knutson 2000; Stott et al. 2001). Solar variability may have played a role in that early century warming, and simulations of the response to solar variability are not yet available for all the models considered. In a follow-up studies (Stott et al. 2006) this analysis is extended to the century as a whole, but our priority here is to focus on anthropogenic signals, applying the analysis to as wide a range of models as possible.

3 Qualitative model-data comparison

Before discussing the methodology and results of the quantitative model-data comparison we outline the qualitative response of the different models to the different forcings. This provides a background to help interpret the quantitative results, and also demonstrates some of the effects of the data masking procedure.

Figure 1 shows the global mean decadal mean evolution of near-surface temperatures in observations and various model simulations. The global mean anomalies are taken from the spherical harmonics used in the detection study and so have been masked using the observational data mask. The data points are marked with symbols, the lines are cubic splines between the data points. The splines are not intended to imply there is more information there than the five data points for each integration, rather they have been included to make the figures easier to visualise. Each model is represented by the linestyle given in the legend on panel (a).

The observations show the marked increase in global mean surface temperatures after the 1970's, with the 1990s being about 0.25 K warmer than the period 1906–1996 as a whole. The observed temperature anomalies from the 1950s to the 1970s are about 0.05 K colder than the mean of the 1906–1996 period.

All model integrations that include greenhouse gas forcing show increases in temperatures through the 1980s and 1990s that are comparable to the observations. Inclusion of G alone yields more warming than observed in all cases. For ECH4 and HCM2 the inclusion of the effects of sulphate aerosol brings the modeled global mean temperature changes into closer agreement with observations (especially so for the GSI integration of ECH4). For ECH3 the inclusion of the effects of sulphate aerosol appears to cool the model too much during the 1970s and 1980s. In particular there is very little warming in the period 1966–1986; the warming appears to be delayed by the sulphate aerosol. This anomalous (and, by comparison with the observations, unrealistic) time-evolution of the ECH3 model under GS forcing will prove important in the quantitative analysis based on spatio-temporal response-patterns. Previous work focussing on ECH3 (Hegerl et al. 1997, 2000) used spatial patterns of 50-year linear trends extracted from simulations of the 21st century, suppressing information on the evolution of the signal within any currently observable 50-year period.

The modelled responses to natural forcings do not capture the observed temperature changes during the last two decades of the century. Although the response of HCM2 and ECH3 to solar forcing suggests temperatures in the 1990s that are warmer than the period 1906–1996 as a whole, the warming is not as large as that observed. In particular, solar forcing, even if amplified by some unknown feedback mechanism, seems unlikely to account for the observed acceleration in warming after the 1970s.
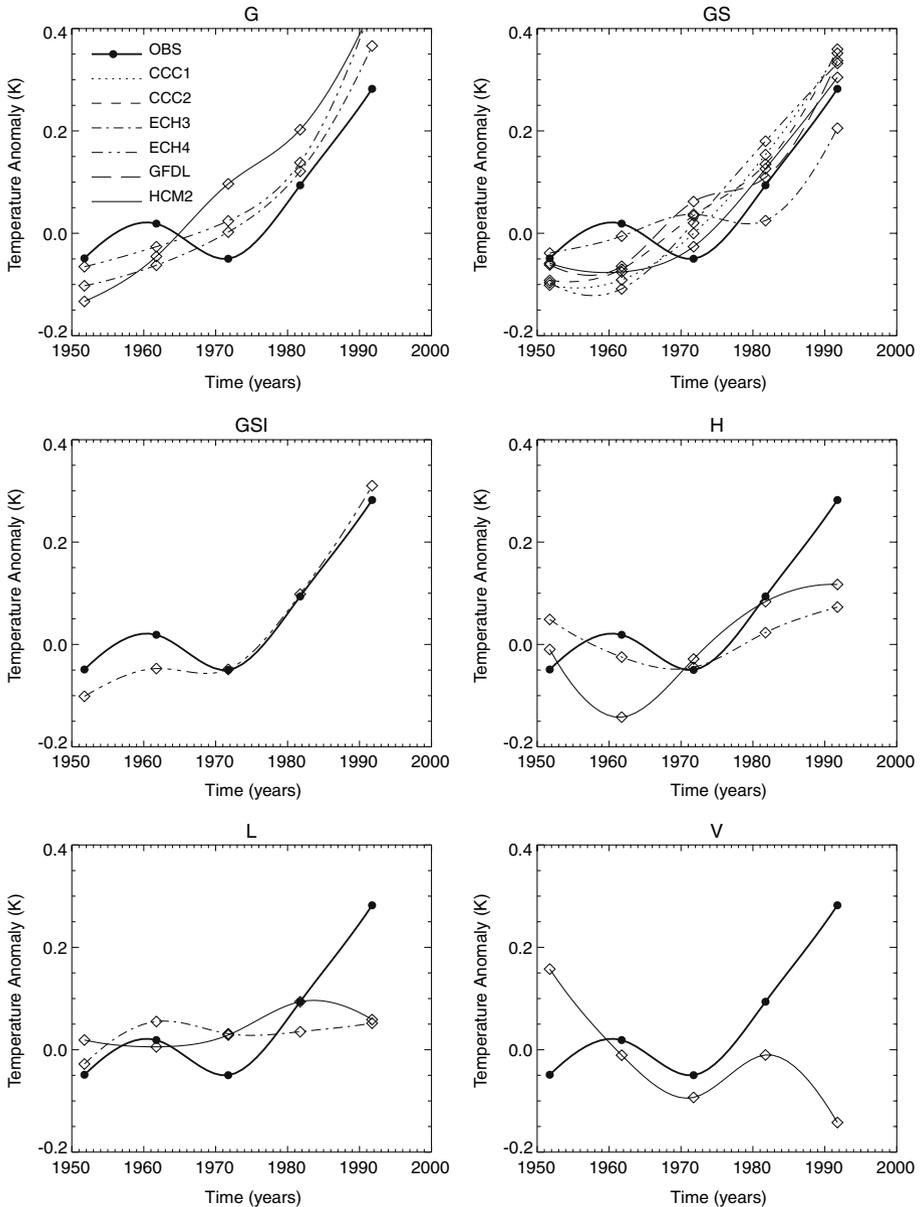
Fig. 1 Globally and decadally averaged temperature anomalies over 1946–1996 relative to the 1906–1996 climatology. Points show decadal averages centred on the mean of the corresponding decade, and lines show cubic spline fits between them. Results shown are for observed (heavy lines on all plots) and in response to (**a**) greenhouse gas forcing alone; (**b**) including direct sulphate aerosol forcing; (**c**) including both direct and indirect sulphate aerosol forcing, (**d**) the Hoyt and Schatten (1993), reconstruction of solar forcing alone. (**e**) the Lean et al. (1995), reconstruction of solar forcing. (**f**) the Sato et al. (1993), reconstruction of volcanic forcing

For both ECH3 and HCM2 the response to L is smaller than the response to H, reflecting the smaller magnitude of estimated interdecadal solar irradiance changes in the Lean et al. (1995), reconstruction. The response of HCM2 to volcanic forcing is in the opposite sense to the observed temperature change, with a general cooling through the 1946–1996 period. This volcanic signal reflects the relatively low volcanic activity during the 1920–1960 period, followed by a period of relatively high volcanic activity during the last four decades of the century (Sato et al. 1993).

The dashed line in Fig. 2 shows the observed trend in global mean temperature over the 1906–1996 period obtained by a simple least-squares fit between decadal mean temperatures 1946–1996, expressed as anomalies about the 1906–1996 climatology, and a straight line anchored to zero in the 1946–1956 decade. This provides an unbiased estimate of the century time-scale trend from temperature anomalies computed in this way, taking into account the fact that the trend may have been accelerating (so simply computing a trend over the most recent 50 years would give a misleadingly high estimate of the century-time-scale rate of warming). The diamonds and vertical bars show mean and estimated 5–95% ranges of the corresponding diagnostic computed from 90-year segments drawn from the model control integrations that are used subsequently for uncertainty analysis. In all cases, we see that the overall drift in the models is small relative to their internal variability. In contrast, the observed rate of warming is well in excess of the highest trends found in unforced model integrations, suggesting that internal variability alone as simulated by these models is very unlikely to account for the observed temperature changes. There are marked differences between the levels of variability observed in the different models, which will prove important for the quantitative analysis.

An indication of the spatial patterns of the responses to the various forcings is given by maps of surface temperature anomaly for the decade 1986–1996 (Figs. 3, 4, and 5). The observed temperature changes (Fig. 3a) show a greater than 0.6 K
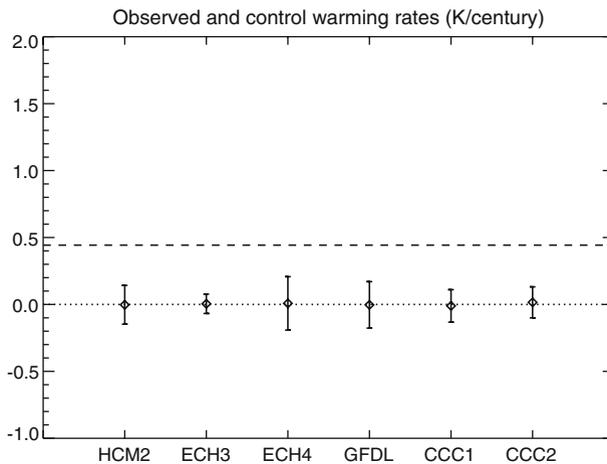

Fig. 2 Dashed line: trend in global mean temperature over the 20th century computed by a least-squares fit over the 1946–1996 period to globally and decadally averaged temperature anomalies about the 1906–1996 climatology. Diamonds and vertical bars: mean and 5–95% range of uncertainty in trends computed similarly from the available 9-decade segments of model control integrations (overlapping successive segments by 8 decades to maximise the sample size)
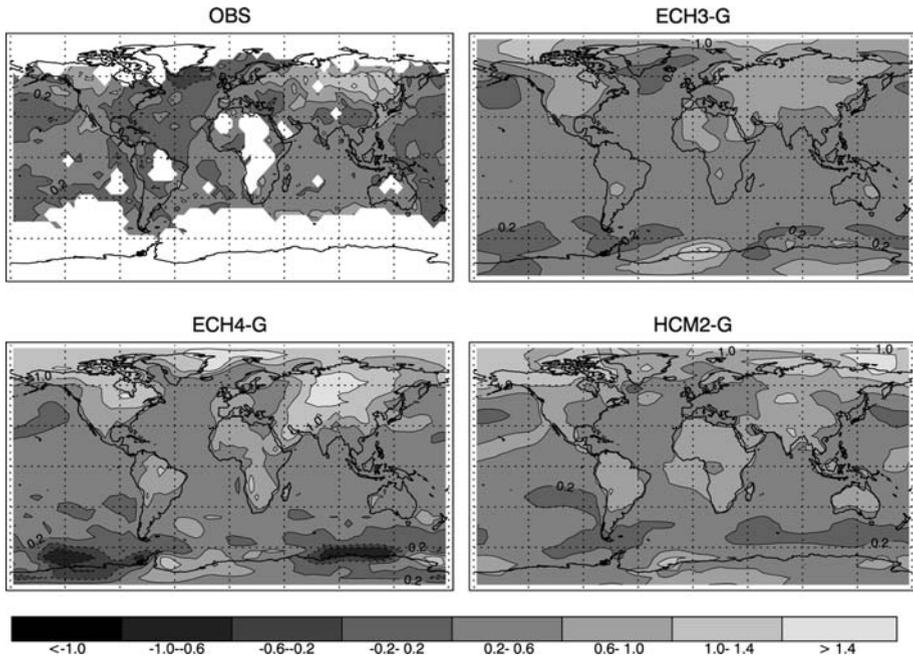
Fig. 3 Modelled pattern of surface temperature response to greenhouse gas forcing, expressed as the 1986–1996 decadal mean anomaly from the 1906–1996 climatology, compared with observed temperatures, processed similarly. (**a**) Observations, (**b**) ECH3, (**c**) ECH4, (**d**) HCM2. Units are K on all panels. Note that the observations and the ECH4 simulation both comprise single realisations and so should be expected to contain more noise than the other two simulations, with the observations also being subject to sampling errors and data gaps in addition to observational error

warming over the northern continental land masses, most notably over Asia, as well as warming over much of the rest of the globe with the exception of the North Atlantic. Data is missing in this decade over Antarctica, the Arctic, much of central Africa and the Amazon basin. Before discussing the temperature changes in the model integrations it is worth discussing the effect of this data mask on the temperature anomalies. Figure 4g shows the corresponding temperature anomaly map for the GS integration of HCM2 interpolated to the observational grid but retaining all data, while Fig. 4h shows the same temperature anomaly field, but taking into account the observational data mask. It is worth underlining at this point that the quantitative analysis described in the following sections uses the model data with the observational data mask: hence care is taken to compare like with like, sampling both externally forced model integrations and segments of model controls used for uncertainty analysis in the same way as the observations have been sampled. Not suprisingly, the introduction of the mask does not have a great deal of impact in regions where data is consistently available throughout the century. There is some impact on the anomaly field in the Pacific ocean where data coverage changes substantially over time.

Overall, the main effect of the data masking is to decrease the modelled temperature response in the grid boxes where the data record does not cover the entire 1906–1996 period. This is a consequence of the fact that the climatological mean temperature for these grid boxes is biased towards the temperatures in the later part
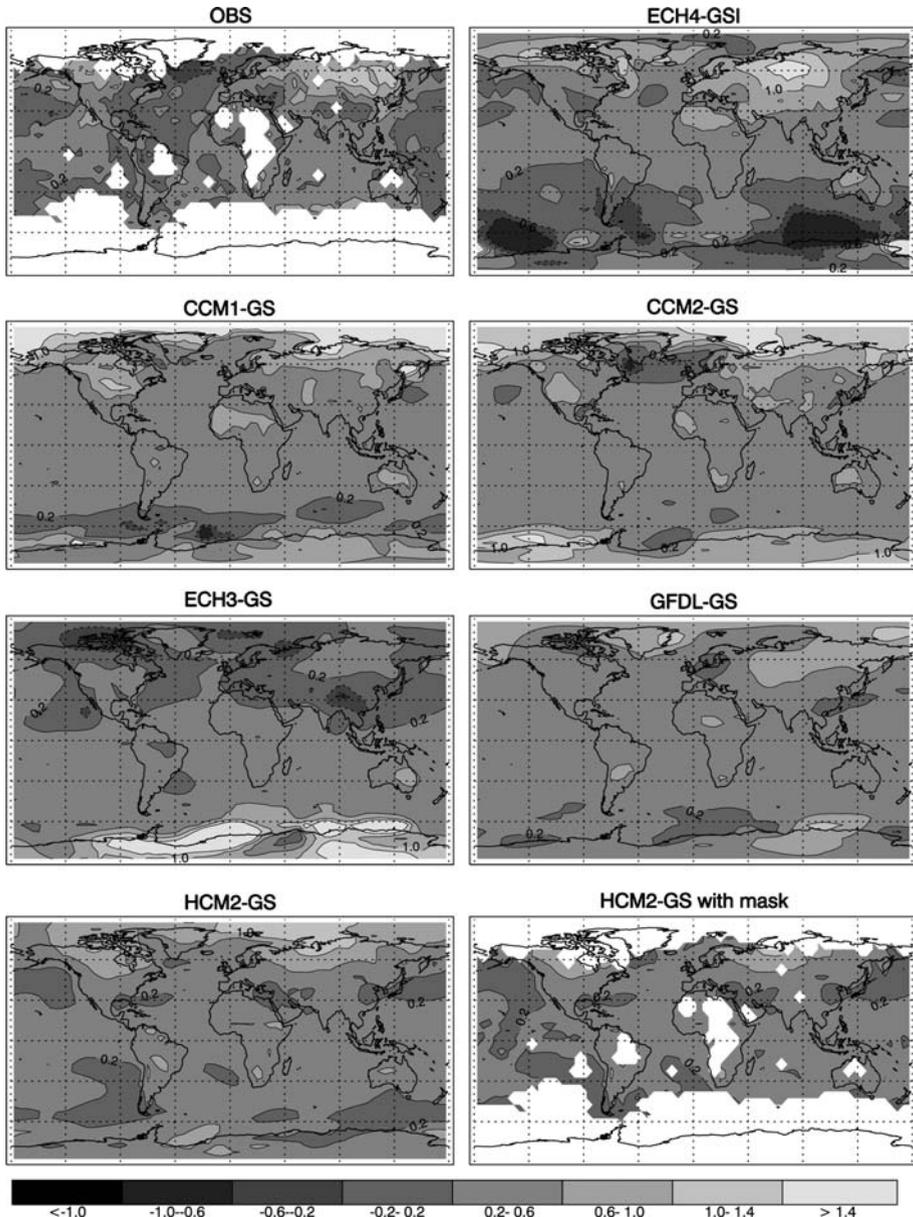
Fig. 4 Same as Fig. 3, but for greenhouse gas and sulphate aerosol forcing. (**a**) Observations, (**b**) ECH4 (from the GSI integration: ECH4 GS integration appears very similar to the ECH4 G integration, since direct sulphate forcing in this model was relatively low) (**c**) CCC1, (**d**) CCC2, (**e**) ECH3, (**f**) GFDL, (**g**) HCM2, (**h**) HCM2, but with the observational data mask applied

of the century and so the temperature anomalies in these grid boxes are biased downward. Obviously, the data mask has most impact where there is no data, notably in polar regions. From the perspective of detection and attribution, this may help to give consistent results by masking out regions where the models' behaviour
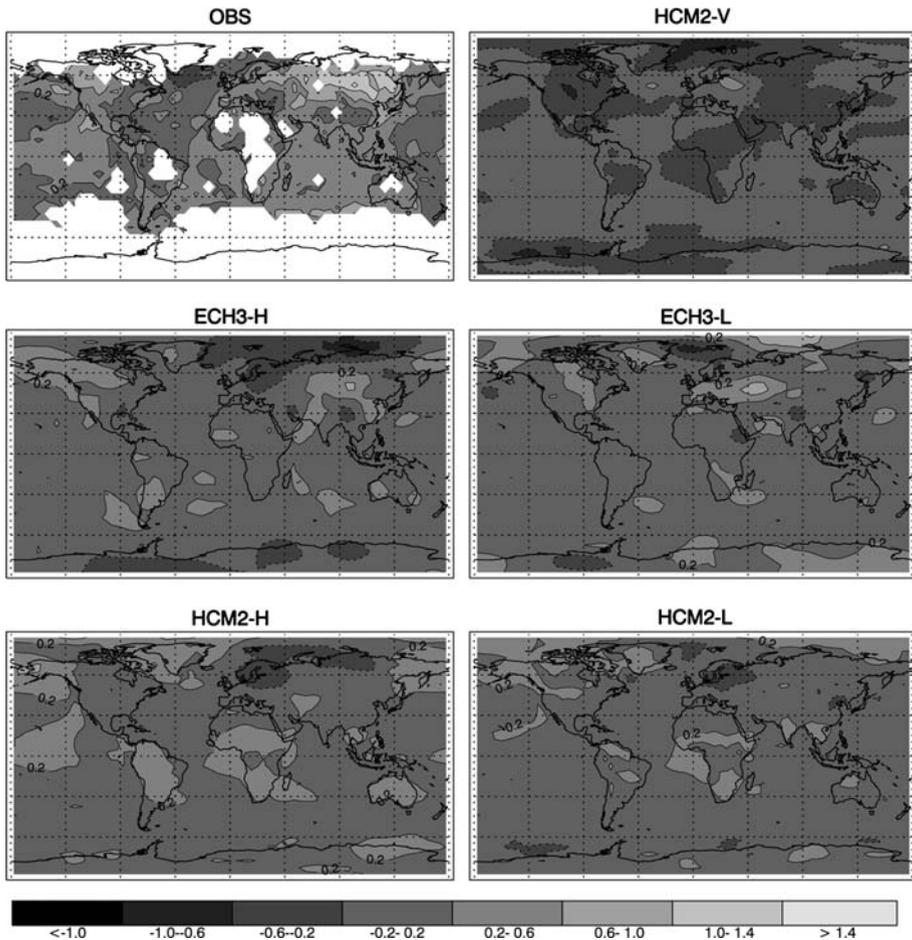
Fig. 5 Same as Fig. 3, but for natural forcings. (**a**) Observations, (**b**) HCM2, volcanic forcing, (**b**) ECH3, Hoyt and Schatten (1993), solar forcing, (**d**) ECH3, Lean et al. (1995), solar forcing, (**c**) HCM2, Hoyt and Shattern solar forcing, (**e**) HCM2, Lean solar forcing

are least likely to match observed changes, either due to model deficiencies or because of high internal variability. For example, maps of the ensemble mean and standard deviation of the control integrations (not shown) indicate that the drift in the models and the internal variability of the models are largest at the ice edge: the use of the data mask means these regions are not included in our analysis. On the other hand, Duffy et al. (2000) show that the exclusion of polar regions significantly reduces the expected global mean anthropogenic warming trend, so the overall impact of excluding polar regions on signal-to-noise is unclear.

The patterns of response of the models to G and GS are shown in Figs. 3 and 4. All of these integrations show the general picture of the largest warming occuring over the northern continental land masses although there is less contrast between the warming over the land and ocean than in the observations. CCC2 GS shows a cooling over the North Atlantic, but this is more spatially confined compared to the observed temperature changes. For the G integrations it has already been noted that

the modelled warming is larger than the observed. This appears to be due to more uniform warming over the oceans and land rather than an intensification of the warming over land. ECH4 is a possible exception to this where the modelled warming over Asia is larger than that observed: it must be stressed, however, that this is a single integration, not an ensemble mean, so results need to be interpreted with caution. Note that the large temperature changes seen in models at the edge of the ice sheet do not affect our analysis because these regions are excluded by the data mask.

The patterns of response of the models to natural forcings are shown in Fig. 5. These are relatively spatially uniform compared both to the observed changes and the responses to anthropogenic forcing. The HCM2 spatial pattern of response to solar forcing appears to be relatively insensitive to the solar reconstruction, with most warming over the tropics and high northern continental land masses, and cooling over the Baltic. ECH3 does not respond with as much warming over the tropics, but more warming over Eurasia. The response of HCM2 to volcanic forcing indicates cooling almost everywhere, maximum cooling over the land masses and only a small region of warming close to the Baltic states. Not suprisingly, having already discussed the global mean temperature response, the modelled responses to natural forcing in the decade 1986–1996 are smaller than that observed, and in the case of volcanic forcing of the wrong sign.

4 Spectra of modelled and observed variability

Since we are relying on model-simulated internal variability for uncertainty analysis in this study, the question naturally arises as to how realistic these model-based variability estimates are. Validation of model-simulated variability on the 30–100-year timescales used for detection is complicated by the likelihood that the observations contain an externally forced component of variability on these timescales (Santer et al. 1996b; Stott et al. 2000). We use a standard check on residuals in the optimal estimation procedure, as proposed by Allen and Tett (1999), to ensure that the statistical model used is formally adequate: that is, that remaining model-data discrepancies can be accounted for by internal variability as simulated by the model in question. An additional check is provided by the power spectra of models and data, shown in Fig. 6. The heavy black line shows the power spectrum of observed global mean temperatures over the period 1861–1999 after removing a linear trend. The various thin lines show the spectra of variability in global mean temperatures in the model control integrations, again after removal of a linear trend from each integration. In each case, a Tukey–Hanning window with a width equal to one-fifth the length of the series was used, giving all spectral estimates approximately equal variance (5–95% range on this log scale shown by the vertical error bar).

There are several ways of estimating power spectra: we adopted the approach used in Fig. 6 for consistency with figure 8.2 of Santer et al. (1996a). There are various remaining sources of bias between the modelled and observed spectra which complicate the comparison. Most straightforwardly, the observations are subject to sources of uncertainty (such as measurement errors and sampling variations from year to year) which mean that the observed spectra are slightly higher than they would be if we had the same error-free observations of the real world as we have of these climate models. Since our principal concern is the possibility of the models

underestimating variance, this is not a particularly important problem. More seriously, as an estimate of real-world internally generated variability, the observed spectrum is, on average, likely to be biased downwards at low frequencies by the detrending procedure (since the trend is estimated from the data, some genuine low-frequency variability is likely to have been thrown out along with it). At the same time, the observed spectrum will be biased upwards at all frequencies by the fact that actual externally forced variability is will not be completely removed by a simple linear trend.

Both these sources of bias can to some extent be accounted for by removing an independent estimate of the externally forced response, provided by the HCM3 model driven with the combination of anthropogenic, solar and volcanic forcing (Stott et al. 2000). The resulting spectrum, OBS2, is shown by the heavy dotted line in Fig. 6. Santer et al. (1996a), noted that removing a model-based estimate of externally driven variability did not necessarily guarantee an accurate spectral
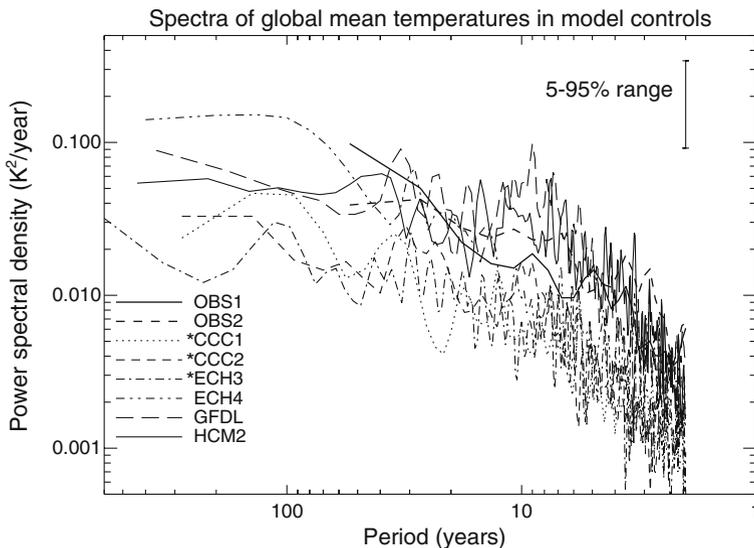

Fig. 6 Thin lines: Power spectra of global mean temperatures in the unforced control integrations that are used to provide estimates of internal climate variability in this paper. All series were linearly detrended prior to analysis, and spectra computed using a standard Tukey window with the window width (maximum lag used in the estimate) set to one-fifth of the series length, giving each spectral estimate the same uncertainty range, shown (see, e.g., Priestley 1981). The first 300 years were omitted from ECH3, CCC1 and CCC2 models as potentially drift-contaminated. Thick solid line: spectrum of observed global mean temperatures (Jones 1994) over the period 1861–1998 after removing a best-fit linear trend. This estimate is unreliable on interdecadal timescales because of the difficulty of unambiguously partitioning observed variability into externally forced and internally generated components. Thick dashed line: spectrum of observed global mean temperatures after removing an independent estimate of the externally forced response to both anthropogenic and natural forcing as provided by the ensemble mean of a coupled model simulation (Stott et al. 2000). This estimate will be contaminated by uncertainty in the model-simulated forced response, together with observation noise and sampling error. However, unlike the detrending procedure, all of these sources of contamination introduce a positive (upward) bias in the resulting estimate of the observed spectrum. The thick dashed line is therefore, if anything, likely to overestimate the level of internal variability in the real world. Power spectral density is defined such that unit-variance uncorrelated noise would have an expected PSD of unity

estimate, but crucially, all principal remaining sources of bias, including sampling error due to the small size of the ensemble and systematic errors in the model's sensitivity or prescribed forcings, would tend to inflate the OBS2 spectral estimate above the spectrum of internal variability in the real world.

We could, in principle, correct for some of these biases: for example, given that the HCM3 model is itself subject to an apparently realistic level of internal variability, this spectrum is likely to over-estimate the true variability by at least 25% since it is based on the difference between the observations and the mean of a four-member ensemble (notice that the heavy dashed line is above the heavy solid line at the highest frequencies at which the impact of external forcing in likely to be least). Measurement and sampling errors in the observations would increase this over-estimate, although probably not by a significant margin. We choose to leave in these sources of bias, and simply show the spectrum of the difference between the observations and the HCM3 "all forcings" ensemble mean, because in this way we know that, if anything, the result is likely to overestimate internal variability in the real world.

Comparing these spectra provides a test of model-simulated variability, which we can formalise by applying an F-test to the ratio observed: modelled power spectral density integrated over the spectral interval 10–60 years. Under the null-hypothesis that the model-simulated variability is an accurate representation of observed variability and that both model and observations behave like linear stochastic processes, this ratio is, to a good approximation, distributed like $F(v_{\mathrm{obs}}, v_{\mathrm{model}})$ where $v = n_i$ $(n_p–2) / (n_s–1)$ (we are grateful to Francis Zwiers for discussions on this point). In this expression, n_i refers to the number of spectral estimates in the 10–60-year interval, n_p the total number of points in the series and n_s the total number of spectral estimates. The correction factors of –2 and –1 arise from the removal of the mean and trend from the original data and our ignoring the constant term in the power spectrum, respectively. If either estimate of the observed spectrum (heavy solid or dotted line) is inconsistent with the estimated spectrum of any of these models, that model is flagged with an asterisk in the legend of Fig. 6.

Under this test, the observed spectrum appears consistent with three of the models used in this study: ECH4, HCM2 and GFDL. This suggests that simulated variability from these models can be used consistently for uncertainty analyses but does not, of course, guarantee their validity. In particular, variance may be accurately simulated in the global mean, but not on smaller (even continental or hemispheric) scales. The spectra of the ECH3, CCC1 and CCC2 models are demonstrably inconsistent with the observed spectrum. While this may be due to the remaining biases in the estimate of the observed spectrum noted above, it is more likely to indicate a deficiency in these models' simulated variability. For completeness, we continue to show results based on these models, but it should be borne in mind that these uncertainty estimates are likely to be underestimated. Hence detection results based on these models are more likely to give a false-positive result (or "type-1 error").

5 Quantitative model-data comparison

5.1 Optimal fingerprinting

The overall approach we adopt to model-data comparison is based on the optimal fingerprinting algorithm of Hasselmann (1979, 1993, 1997), interpreted as an optimal

estimation problem following Northk et al. (1995); Leroy (1998); Allen and Tett (1999); Allen and Stott (2003); Stott et al. (2003). In physical terms, we assume the observations consist of a linear superposition of m model-simulated responses to various external forcings, with the unknown quantities in the estimation problem being the factors (β_i^{true}) by which we have to scale the ith model-simulated response to reproduce observed changes. That is, we assume the "detection model":

$$\mathbf{y}^{\text{true}} = \sum_{i=1}^{m} \mathbf{x}_i^{\text{true}} \beta_i^{\text{true}}, \tag{1}$$

where $\mathbf{y}^{\text{true}}$ and $\mathbf{x}_i^{\text{true}}$ are the real-world and model-simulated responses (if any) to external forcing free of sampling noise due to internal climate variability. These are not directly observable: we never observe climate change in the real world in the absence of internal variability, and although $\mathbf{x}_i^{\text{true}}$ could be approximated by taking the mean of a very large ensemble simulation, at the time this study was undertaken, ensemble sizes typically available to climate change detection studies are of order 1–4. Since then, somewhat larger ensembles have been performed, but sampling noise in model simulations remains an issue, particularly on smaller spatial scales where the signal-to-noise ratio is weaker.

Written in terms of actual observations, $\mathbf{y}$, and model-simulated responses based on averaging finite ensembles, $\mathbf{x}_i$, the detection model becomes

$$\mathbf{y} - \upsilon_0 = \sum_{i=1}^{m} (\mathbf{x}_i - \upsilon_i) \beta_i^{\text{true}}. \tag{2}$$

Because of the presence of sampling noise, υ_0 and υ_i, the coefficients β_i^{true} must be estimated, and the estimator we use, β_i, will always have some level of uncertainty associated with it.

It is important to stress that this detection model neglects so-called "structural uncertainty": that is, the possibility that the model may systematically simulate the wrong shape of response, as well as the wrong amplitude, to a given external forcing. Since this study was undertaken, initial efforts have been made to account for structural uncertainty using the "errors in variables" approach, a straightforward generalisation of the detection model used here incorporating some estimate of "model error" in the error covariance matrix (Huntingford et al. 2006). Estimating model error remains a highly contentious issue, however, so this is a topic of ongoing research.

If β_i is consistent with zero at a given confidence level, then we can conclude that the ith model-simulated signal is not required to account for observed changes under this detection model or, more specifically, that the null-hypothesis of zero amplitude of this signal cannot be rejected. If β_i is consistent with unity, then we can conclude (at the same level of confidence) that the amplitude of this model-simulated signal could be correct. More generally, if $\boldsymbol{\beta}$ is consistent with $\mathbf{0}$, then no climate change of any form can be detected in these observations under this detection model, whereas if all elements of $\boldsymbol{\beta}$ are consistent with unity and the regression residuals are consistent with internal variability then this particular detection model (combination of model-simulated signals) represents an adequate account of observed climate change. A detection model that is impossible to reject when considered in isolation

may, however, still be rejected when more information is brought to bear on the problem, either in the form of additional observational data or additional model-simulated signals.

In contrast to Hasselmann (1997), this interpretation makes clear that ''detection'' (rejecting $\mathcal{H}(\boldsymbol{\beta} = \mathbf{0})$) and ''attribution'' (rejecting $\mathcal{H}(\boldsymbol{\beta} = \mathbf{0})$ and failing to reject $\mathcal{H}(\boldsymbol{\beta} = \mathbf{1})$)) are essentially two aspects of the same procedure: the only distinction is that they address different null-hypotheses. One of the criticisms levelled at Santer et al. (1996a) in the SAR was that ''detection'', considered in isolation, is not particularly informative (it implies only that some climate change has occurred, of unknown origin) while the ''attribution'' problem, considered in isolation, is statistically ill-posed (successful attribution appears to involve failing to reject a null-hypothesis, which is an ambiguous result: either the null-hypothesis is true, or the test simply was not powerful enough to reject it): see Berliner et al. (2000) for a discussion of this point. With a greater range of model simulations than were available at the time of the SAR, it is possible to express both detection and attribution in terms of a single estimation procedure, making the overall problem much simpler: the conclusion that $\boldsymbol{\beta}$, or some element thereof, is consistent with unity is only of interest if the corresponding range of uncertainty is relatively small.

The question of whether a particular element, β_i, includes zero (e.g., ''is greenhouse influence detectable?''), while attracting considerable political attention, may be relatively uninteresting from a physical point of view. We already know, on physical grounds, that increasing greenhouse gas levels, for example, must have some influence on climate. Rejection of the null-hypothesis of no influence provides some reassurance that the spatial pattern of change as simulated by our models bears some relationship to what is going on in the real world, but even this may represent a limited advance in understanding: on basic physical grounds we should expect, for example, any externally driven warming to occur faster over land than over the oceans. What is of interest is the range of values of β_i consistent with recent climate observations, since this indicates the extent to which the climate model may be over-estimating ($\beta_i < 1$) or under-estimating ($\beta_i > 1$) the response to a particular forcing agent. This, in turn, gives some indication of what may occur in the future as these forcings continue (Allen et al. 2000).

Rephrasing detection and attribution as a single estimation procedure does not avoid all these ''philosophical'' problems. Most importantly, Hasselmann (1997) noted that, for any formal detection and attribution procedure to get started, it is necessary to confine attention *a priori* to a relatively small number of competing explanations for observed climate change. If the number of allowed model-simulated signals, m, is too large, then it becomes increasingly likely that at least one signal will closely resemble a linear combination of the others, leading to a so-called ''degenerate'' estimation problem in which the data are insufficient to constrain the β_i. We will consider cases up to $m = 4$ in this paper, which allows us to cover the main known drivers of recent near-surface temperature change: greenhouse gases, anthropogenic aerosols, solar variability and volcanic activity. We will find that, even with $m = 3$ or $m = 4$, many results become ill-constrained by the kind of large-scale data considered in this paper. Extending this approach to much larger values of m, for example to distinguish between different models' simulations of the response to anthropogenic forcing in the hope of identifying the ''best'' model, would require much more detailed input data.

The assumption that responses to forcings of this magnitude may be superimposed linearly on each other is central to this procedure. The evidence available suggests that linear superposition does hold (e.g., Haywood et al. 1997; Penner et al. 1997; Santer et al. 2003; Gillett et al. 20004b; Meehl et al. 2004) for the main anthropogenic drivers of climate change, although Meehl et al. (2003) report evidence for non-linearity in the response to natural (solar) forcing. The possibility of non-linear interactions between the responses to different forcings and between forced and internal variability remains an important caveat, particularly as these approaches are extended to smaller scales and other variables than temperature.

The procedure by which we estimate the β_i and assign confidence intervals is essentially based on a least-squares fit, weighted by the inverse square root of the expected noise variance in each "statistically independent data point" (Hasselmann 1993; North et al. 1995). Complications arise because the model-simulated responses are not known exactly, but only estimated from a small ensemble simulation (in some cases with only a single member). Unbiased estimates of β_i and confidence intervals can be derived using the "Total Least Squares" approach detailed in van Huffel and Vanderwaal (1994); Allen and Stott (2003), and the appendix.

5.2 Choice of detection model

The discussion in the previous subsection took the detection model, Eq. (1), as given. In practice, specification of this model requires a strong element of expert judgement. Climate models cannot be expected to simulate observed climate change in every detail, even accounting for uncertainty due to climate noise. Certain variables, such as ice or soil moisture, may be simulated poorly, and small scale changes near or below the model resolution are not represented at all. Hence y^{true} cannot contain every aspect of observed climate change, but only those aspects that a reasonably accurate model can be expected to reproduce: in the case considered here, continental scale decadal changes in near-surface temperatures. A more practical, and often more restrictive, constraint on the range of scales and variables that can be considered is the availability of sufficiently long control integrations to estimate their covariance structure.

Since this study was undertaken, a number of studies have reported detection and attribution of human influence on climate on smaller spatial scales, including Karoly et al. (2003), Stott (2003), Zwiers and Zhang (2003), Braganza et al. (2004), Gillett et al. (2004a), Stott et al. (2004), Karoly and Bragnza (2005), Zhang et al. (2006). The most robust results remain, however, at the continental scale.

In our approach, the rank of y^{true} is determined by projecting both observations and model-simulated signals onto a small number (10–15) of "extended-EOFs" (E-EOFs, or eigenmodes of the spatio-temporal lag-covariance matrix—see Weare and Nastrom (1982) of a climate model control integration. The number of E-EOFs is based on a compromise: it should be small enough to exclude lower-variance spatio-temporal modes of variability that are likely to be under-represented in the model, but large enough to contain sufficient information to distinguish between the signals included in the detection model. Various studies (e.g., Allen and Tett 1999; Stott et al. 2001; Tett et al. 2002) have considered this "truncation" issue in some detail: we do not explore it in detail here, but key results (particularly detection of a

substantial greenhouse influence on recent near-surface temperatures) are insensitive to varying the truncation level between 10 and 15.

The advantage of projection on to E-EOFs is that they are a relatively objective function of the control integrations available, although there is still an element of subjectivity in the choice of norm defining the E-EOFs. The disadvantage is that the regional origins of detection and attribution results may be obscured by projection onto global spatio-temporal patterns. A valuable alternative approach is to focus on large-scale indices of change, as used by Braganza et al. (2004).

A more obvious issue is the number and choice of signals to include in the detection model: that is, m and the $\mathbf{x}_i^{\text{true}}$ in Eq. (1). The important point to note here is that comparing the relative merits of two different detection models (or candidate explanations of recent climate change—for example, a detection model that includes the signal of anthropogenic influence with a model that does not) is a much better posed problem than evaluating the credibility of a detection model in isolation (Smith et al. 2003). The reason is, as noted above, there will always be aspects of observed climate change that models fail to simulate, so failing a simple "goodness-of-fit" statistic like an F-test on regression residuals (see Allen and Tett 1999) is ambiguous: it could imply an important signal is missing from the detection model, or it could simply mean that the underlying climate model is incapable of simulating this particular set of observations. Failing some test of goodness-of-fit with a particular detection model (signal combination) is only of interest if this test is passed in some other, physically reasonable, detection model, since this implies that the underlying climate model is capable of simulating those observations, and the problem with the first detection model can be attributed to one or more missing signals.

Given that we are primarily interested in relative statements, of the form "detection model A provides a more accurate account of these observations than detection model B, by a significantly greater margin than we would expect if detection model B is actually valid", it is clear that there can be no universally correct detection model: different models allow us to compare different candidate explanations of recent climate change. In the "single-signal" detection models considered first, for which $m = 1$, forcings are combined to give only a single simulation of the climate response to, for example, greenhouse + sulphate (GS) influence on climate. The only alternative to this GS signal allowed by a single-signal detection model is internal climate variability. This, we shall see, can be rejected at a very high confidence level. This provides us with information about the strength of the anthropogenic signal relative to unforced variability, but it does not provide a complete estimate of uncertainty in the size of the anthropogenic signal because other factors are also believed to have played a role in recent climate change. These are considered in the multi-signal detection models, which involve estimating a larger number of free parameters from the same limited set of observations and so tend to yield much greater uncertainties. Nevertheless, as we shall see, a substantial contribution from anthropogenic greenhouse gases consistently emerges as a necessary ingredient in any adequate account of recent climate change.

The more separate signals are included in the analysis (the larger we make m), the more likely it is that one of them will closely resemble another (or a linear combination of others). This so-called "degeneracy" inevitably leads to problems, since increasing the amplitude of the first signal while reducing the amplitude of the second will make no difference to the overall goodness-of-fit. If, however, other

information can be brought to bear limiting the range of amplitudes on one member of a degenerate signal pair then useful information can still be obtained on the other: for example, we might argue on physical grounds that anthropogenic aerosols must cause a net cooling, and hence not allow a change of sign of the aerosol signal. Even in the absence of prior information, we can make assumptions and explore their implications. For example, without positive evidence to the contrary, it is reasonable to assume (while acknowledging that this assumption may turn out to be incorrect) that the net impact of natural forcings on surface temperatures over the past 50 years is negligible, and hence that their amplitude can be set to zero (this is, in effect, what is done in the single-signal GS results quoted below).

It is, however, only legitimate to leave a signal out of the analysis if including it does not significantly improve the overall goodness-of-fit (i.e., if its presence is not detectable). Results are likely to be misleading if based on a detection model which is demonstrably false, either on the basis of residual consistency checks or because an additional, detectable, signal has been omitted. For example, some informal studies have estimated the magnitude of solar influence on climate assuming solar forcing is the only driver of recent near-surface temperature change, even though the influence of greenhouse gases can also be detected on the datasets used (e.g., Friis-Christensen and Lassen 1991). Omitting an undetectable signal is legitimate parsimony, provided the omission is acknowledged, but arbitrarily setting the amplitude of a detectable signal to zero is simply misleading. We have included some results from such "rejectable" detection models for the sake of completeness, but we should therefore avoid interpreting these results physically: this primarily applies to the "all natural" and "greenhouse-gas-only" models considered below.

6 Results

6.1 Single-pattern analyses

We first consider a number of cases in which it is assumed that the observed climate change consists of a single model-simulated response-pattern ($m = 1$). Given the number of scenarios and models listed in Table 1, the range of permutations of scenarios and control simulations is potentially large. To keep the analysis manageable, we will treat each A-OGCM as self-contained, using only those scenarios available with that A-OGCM, using the first half of the available control integration to define the pre-whitening operator **P** (see appendix) and the second half for uncertainty analysis. In all cases, nine-decade segments of control variability were overlapped by 8 decades to maximise the sample size, with an allowance made for this overlap in the computation of degrees of freedom (see appendix). The rank of **P**, or the number of E-EOFs of the control used in the definition of the detection space, is set to 10, as in Tett et al. (1999), in all one- and two-pattern analyses which included models with relatively short control integrations. The truncation was increased to 15 for the three- and four-pattern analyses which involved models with longer control runs (the longer the control, the more E-EOFs we can expect to estimate accurately). Key results (particularly detection of greenhouse influence) hold at both truncations considered.

Results will be sensitive to the truncation level, with uncertainty ranges tending to fall as κ is increased and more information is included in the analysis. The problem is

that some of this information may be misleading if it depends on small-scale spatio-temporal patterns of variability in which variance is likely to be underestimated. Checks for residual consistency can alleviate this problem (Allen and Tett 1999), but if we allow κ to vary we introduce an additional degree of freedom into the analysis raising the question of which detection results to present (and the danger of focussing misleadingly on the positive ones). Results from each model are considered separately, although multi-model approaches, such as Gillett et al. (2002a), might increase signal-to-noise.

We will also consider cases in which the model-simulated response consists of the sum of a number of A-OGCM simulations where different forcings have been prescribed separately. The difference between this and the multi-pattern analyses is that, in generating a single response-pattern by adding up a number of simulations or prescribing a number of forcings in a single simulation, we are assuming the relative amplitude of the response to these different forcings is accurately simulated by the model, and only the total amplitude of the response is unknown. In the multi-pattern analyses, we estimate the response to individual forcings separately. When a number of ensemble simulations are added together to generate a response-pattern, the noise variance in that pattern is simply the sum of the noise variance in the constituent ensembles.

6.1.1 Greenhouse gases alone

Simulations of the 20th century forced with greenhouse gases alone (G) were available to this study for three A-OGCMs: HCM2, ECH3 and ECH4 (many more are now available: see Stott et al. (2006); Stone et al. (2006) for an update). Best-fit scaling parameters are shown in Fig. 7, with estimated 5–95% uncertainty ranges.
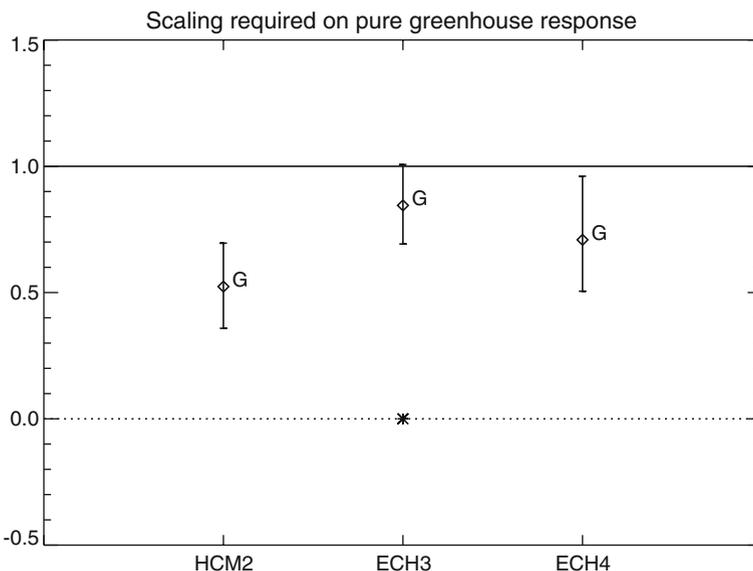

Fig. 7 Best-fit scaling factors and 5–95% uncertainty ranges assuming the observed record consists only of a response to greenhouse gas increases as simulated by three A-OGCMs

In the HCM2 and ECH4 cases, the range of acceptable values for β is greater than zero but less than unity. This implies that, if we assume that greenhouse gases are the only external forcing to have affected climate over this century, the greenhouse signal is easily detectable but the models appear to be over-estimating the amplitude of the response: that is, they need to be scaled down by roughly a factor of 1.5–2 to reproduce the observed record, consistent with the qualitative discussion above.

The asterisk on the horizontal axis in the ECH3 case indicates that the weighted regression residuals (distance of the points from the best-fit line in a scatter plot of model vs. observed climate change in the various components of **y**) are sufficiently large that we can reject the hypothesis that they are attributable only to internal climate variability as simulated by the ECH3 model at the 10% level. The choice of 10% is arbitrary, but we believe it is appropriate to use a higher P-value for the residual check than the level used for confidence intervals because this is not a very powerful test. Any evidence of deficiency in model-simulated variability gives cause for concern, even if only at a relatively low confidence level.

The asterisks could imply either that some other agent is affecting climate that cannot be represented as a simple scaling on the greenhouse response, or that variability in the ECH3 control simulation is unrealistically low. Alternatively, since we are using a relatively high threshold (10%) for this test, they could simply be the result of chance: we would expect in approximately one in ten cases for the test to fail without anything being wrong with either the forcings considered or the model-simulated variability. This ambiguity in interpretation of results from tests of residual consistency makes them difficult to use as a central component of the analysis: unlike Tett et al. (1999), and Stott et al. (2001), we prefer to focus here on which signals are detectable, using the checks for residual consistency simply to flag those cases in which uncertainty estimates are likely to be misleadingly low.

Figure 8 shows the 20th-century warming trend ''attributable'' to greenhouse gas influence under this detection model: see Appendix for the details of how this is computed. Even after these model-simulated responses to greenhouse gas forcing have been scaled down to be consistent with observed climate change, the warming trend attributable to greenhouse gas influence over the 20th century remains substantial: that is, comparable in magnitude to the total observed warming. This is important because the size of sulphate aerosol influence on climate remains uncertain (Ramaswamy et al. 2001). If, despite current evidence to the contrary, sulphate cooling eventually proves to have a negligible impact on climate, then we might have to conclude that current models overestimate the response to greenhouse gas forcing. Even in this hypothetical situation, however, the magnitude of the estimated response to greenhouse gases remains large enough to imply a significant warming in the future (Allen et al. 2000).

Attributable warming trends in Fig. 8 should not be over-interpreted, since in the case of the HCM2 model at least, the hypothesis that recent near-surface temperature changes can be accounted for purely as a scaled model-simulated response to greenhouse gas forcing can be rejected. Even though the residuals of regression are consistent with internal variability as simulated by the HCM2 model (indicated by the absence of an asterisk), other signals omitted from this analysis (specifically sulphate influence) do have a detectable influence. The conditional statement, ''if greenhouse gases had been the only contributor to late twentieth century
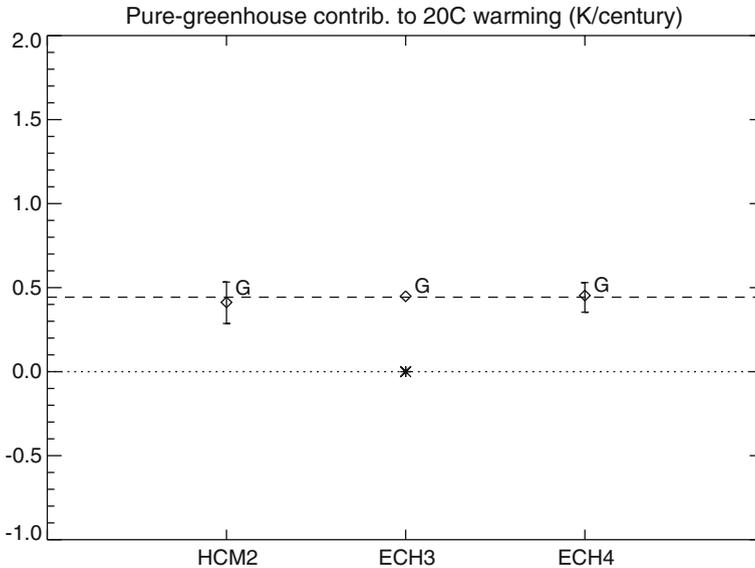
Fig. 8 Implied contribution to warming trends over the 20th century, in degrees per century, assuming anthropogenic greenhouse gases are the sole external contributor to climate change. Confidence intervals are not shown where the regression residuals implied an unreliable noise model—indicated by the asterisks. Dashed line indicates total observed rate of warming in this diagnostic

temperature trends, then they would have accounted for a trend of 0.3–0.5 K/century" is only of limited interest if we have evidence that the condition is not satisfied.

6.1.2 Greenhouse gases and anthropogenic sulphate aerosols

Simulations of the response to the combined influence of greenhouse gases and the direct albedo influence of anthropogenic sulphate aerosols were available for all six A-OGCMs. In one case, ECH4, an additional simulation was available including indirect sulphate forcing and changes in tropospheric ozone. Figure 9 shows that the inclusion of sulphate forcing largely eliminates the need to scale down the model-simulated responses to reproduce observed climate change, with all β ranges with the exception of CCC1 now consistent with unity.

Given that the climate sensitivities of these models vary by almost a factor of two, it may seem surprising that almost all give simulated climate change under GS forcing that are consistent with the observed changes. Sensitivity differences do not translate directly into differences in transient response: more sensitive models take, on average, longer to re-equilibrate after a change in forcing. Hence differences in transient response tend to be smaller than differences in sensitivity for well-understood thermodynamic reasons (Hansen et al. 1985). The transient response is also affected by the details of oceanic heat uptake and there has been some speculation that changes in ocean circulation may act to further surpress differences between transient responses beyond the simple thermodynamic effect: these issues are reviewed by Cubasch et al. (2001).
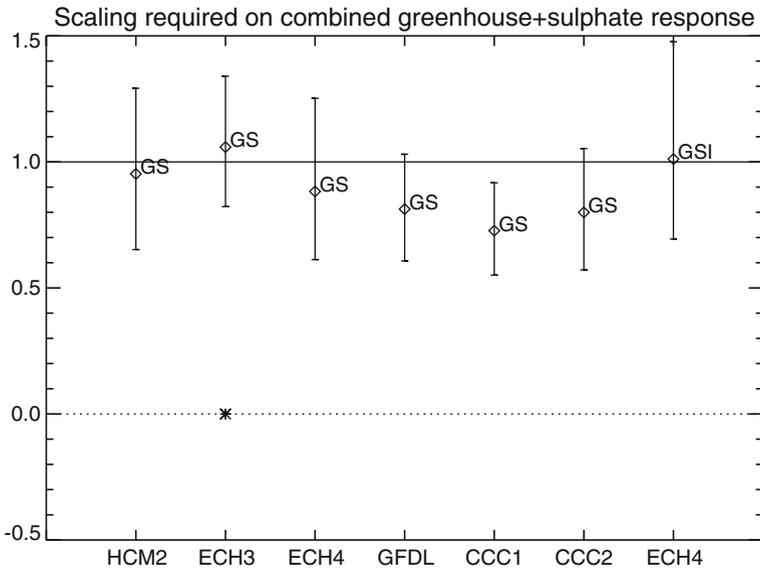
Fig. 9 Signal amplitudes estimated as above, but based on model simulations forced with the combined influence of greenhouse gases and sulphate aerosols, including in one case (GSI) the effect of indirect aerosol and tropospheric ozone

Expressed in terms of contributions to 20th century warming (Fig. 10), we find very similar ranges to those in the previous subsection. To some extent this can be attributed to the fact we are using single-signal detection models in this section. The observed record contains a substantial warming trend so, given only a single signal to explain it, the trend attributable to that signal is very similar to the total observed trend in all cases in which the signals contain a pronounced trend. We are fitting all model-simulated signals to the same data so it is not surprising that, after the fitting is done, the scaled model-simulated responses are brought into agreement with each other. We shall see, however, in section 6.6 below that this similarity of results across single-signal detection models is not simply an artifact of the analysis: even if information as to the size of the global mean warming trend in the observations is intentionally suppressed, consistent results are still obtained across models.

Anticipating results from the multi-pattern analyses, we shall find that in all cases except ECH3 no further signals are detectable beyond the GS signal in this particular diagnostic. This means, in the jargon of detection and attribution, that recent climate change is attributable to GS influence, in that there is no need either to modify the amplitude of the GS response nor to invoke responses to other forcing agents to account for recent observed changes. This does not, of course, mean that other external factors have *not* affected climate over this period, simply that their influence is not strong enough to be detectable in this particular (large-spatial-scale, decadal-time-scale) diagnostic.

Not only are the model-simulated response amplitudes consistent with recent observed changes in Fig. 9, but the range of uncertainty in the scaling factors β is quite small: in most cases the distance between the lower end of the uncertainty range and the zero line is greater than the 5–95% range. We should be cautious
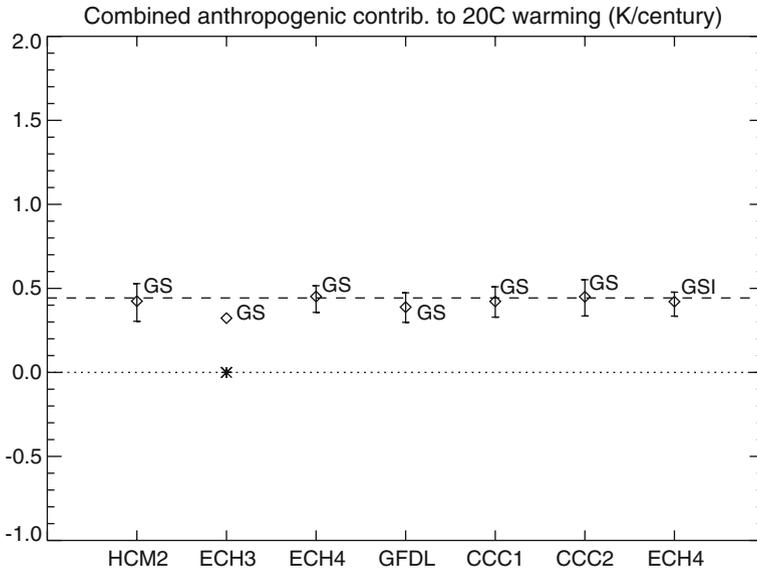
Fig. 10 Contributions to 20th century warming based on model simulations forced with the combined influence of greenhouse gases and sulphate aerosols. Confidence intervals are not shown where the regression residuals implied an unreliable noise model as in Fig. 8

about interpreting this as a ''4-σ result'', because the distribution of estimators under Total Least Squares is non-Gaussian and with the relatively small lengths of control available, we cannot make quantitative statements about very low-probability events (Gillett et al. 2000). Nevertheless, the distance of these uncertainty ranges from zero indicates we are obtaining much higher signal-to-noise levels in this analysis than previous studies such as Hegerl et al. (1997), Tett et al. (1999), and Hegerl et al. (2000). Since this study was undertaken, further progress in signal-to-noise has been made through the use of larger ensembles (Stott et al. 2006) and averaging signals across models (Gillett et al. 2002a; Huntingford et al. 2006).

Our sensitivity studies suggest that the reason for this higher signal-to-noise is primarily our use of a diagnostic which is based on surface temperatures expressed as anomalies about the 1906–1996 climatology, rather than anomalies about the past 50 years as in these previous studies. The fact that temperatures have not only been rising over the past 50 years but have also been generally warm relative to the century as a whole is clearly very powerful information in discriminating against climate noise as simulated by these A-OGCMs. A diagnostic based solely on 50-year trends or anomalies about the climatology of the 50-year period used in the analysis would make no distinction between a 50-year 0.4 K warming that started from anomalously cold conditions during the control integration and one that started from the long-term control climatology. Since the control climate is stationary (and can be relatively accurately modelled as an AR(1) process—Tett et al. 1997), the former event is much more likely to occur by chance than the latter. This enhanced signal-to-noise comes, of course, at a price: our results are more sensitive than those based purely on the past 50 years of data to possible systematic errors in early 20th century temperature observations, which are difficult to quantify. This highlights the

importance of using multiple lines of evidence to support conclusions from the instrumental record (Hegerl et al. 2006).

Unlike the pure-greenhouse case, we generally do not have positive grounds to reject the hypothesis that recent climate change is entirely due to the combination of greenhouse and sulphate forcing and internal variability as simulated by these models (with the exception of ECH3, highlighted, and noting the inclusion of tropospheric ozone in the forcing of ECH4-GSI). The following sections will conclude that the influence of natural forcings on these timescales is small and generally indistinguishable from zero. Thus the statement that the combination of these anthropogenic forcings have been responsible for 0.3–0.5 K/century warming over the 1906–1996 period represents a reasonable summary of our results, and has been supported by subsequent work (e.g., Stott et al. 2006). It is, of course, subject to caveats regarding the simulation of internal variability and the requirement that the net response to natural forcings proves negligible on these timescales, consistent with current evidence.

6.1.3 Solar and volcanic forcing

In the case of two models, HCM2 and ECH3, simulations were also available of the response to the main known natural external influences on climate, solar variability and explosive volcanism. Considered alone, the influence of volcanic activity as represented in the Sato et al. (1993), reconstruction is not a promising candidate to explain all recent observed climate changes, since it is of the wrong sign (corresponding β values are negative). The increase in solar irradiance observed in the reconstructions of Hoyt and Schatten (1993), and Lean et al. (1995), produce a change in the correct sense, but some (in most cases a significant) amplification is required for consistency with the observed record. This in itself would not be sufficient to rule out a solar explanation of recent observed climate change, since mechanisms amplifying solar influence on climate have been proposed in the literature (Haigh 1996; Svensmark and Friis Christensen 1997) that are not represented in these simulations: in both models, only total incoming solar irradiance was altered and there was no explicit representation of any response of either stratospheric ozone or cloud-cover (Figure 11).

In all cases except the HCM2-V simulation (in which the volcanic signal was estimated with an unphysical sign) the regression residuals were found to be inconsistent with the hypothesis that the observed change was attributable to these natural forcings plus internal variability as simulated by these A-OGCMs. This figure should not be interpreted as indicating there is anything necessarily wrong with these models' simulation of the response to natural forcing (although see Robock 2000 for a discussion of the response to volcanic forcing), but that some other signal that is (by chance) anticorrelated with volcanic influence is likely to be confusing the picture.

Single-signal results do provide some indication that purely-natural explanations of recent observed climate change are less credible than explanations involving anthropogenic forcing, but the ambiguity of these residual checks has been noted above (failing a check on residuals could indicate either a missing signal or a deficiency in model-simulated variability). More conclusive evidence against a purely-natural account of recent temperature changes is provided by the multi-pattern detection results. We do not present an "attributed trend" figure for these natural
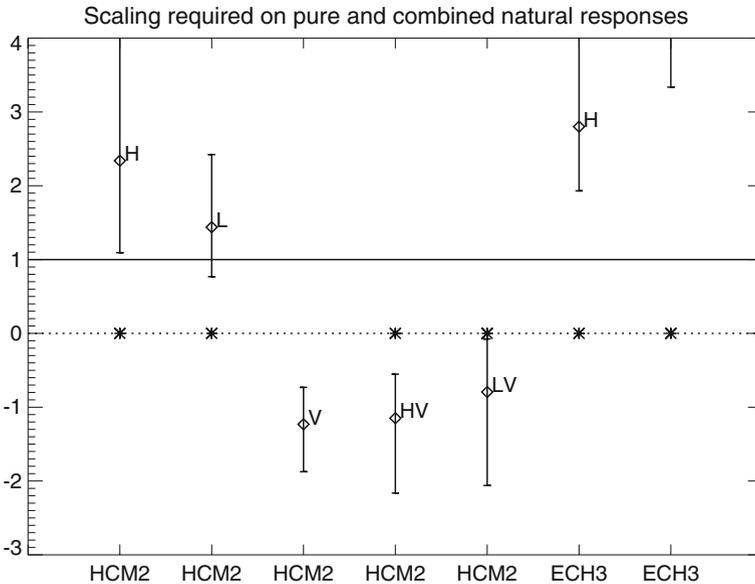
Fig. 11 As previous figure but based on model simulations forced with the Hoyt and Schatten (1993), (H) and Lean et al. (1995), (L) reconstructions of past solar irradiance changes and the Sato et al. (1993), (V) reconstruction of aerosol forcing due to explosive volcanism. Double letters indicate the simulated responses were added up to give a single composite response-pattern in cases where forcings were prescribed separately

forcings: since these can be rejected as adequate accounts of late-20th century temperature change, the size of trend nominally attributable to them is misleading.

6.1.4 Combinations of natural and anthropogenic forcing

On physical grounds, we expect both natural and anthropogenic influences to have played a role in recent observed climate changes. The simplest test of a model's performance in simulating the response is therefore to add up the simulated responses to these various forcings to produce a single "total climate change" signal to compare with the observed record. Alternatively, all forcings could be prescribed jointly in a single ensemble simulation (Stott et al. 2000). This would have the advantage of lower noise in the model-simulated signal (adding up signals from different simulations increases the noise in the total) but provided the linearity assumptions inherent in the detection model (6) are valid, would otherwise be exactly equivalent.

Figure 12 shows results from various combinations of natural and anthropogenic forcing. Model-simulated signals based on the combination of anthropogenic and solar forcing need to be scaled down to give the most accurate reproduction of the observed record, in some cases by a significant margin (the uncertainty range excludes unity). This should be expected, since both of these forcings act in the same sense and the response to anthropogenic forcing alone was found to be consistent with the observed record. The exception here is the ECH3-GSL simulation, which gives a scaling factor consistent with unity but fails the residual check, suggesting
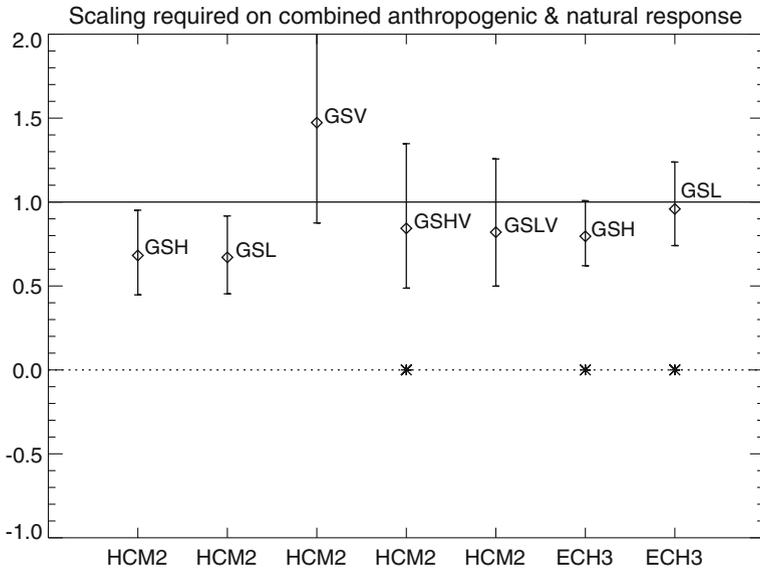
Fig. 12 As previous figure but based on single "total climate change" signals obtained by adding up the responses to anthropogenic and natural forcing where the appropriate simulations are available

something is still unaccounted for or that ECH3 internal variability is incorrect (as was suggested by the spectral analysis). The combination of anthropogenic and volcanic forcing needs to be scaled up, although the discrepancy is not significant: again, this is consistent with the fact that volcanic forcing has opposed anthropogenic over recent decades.

Finally, the combination of anthropogenic, solar and volcanic forcing appears to give a response of approximately the correct magnitude, although only one of the two available solar forcing indices (L) gives a composite response entirely consistent with the observed record. The residual consistency check suggests something is still unaccounted for in the GSHV composite although, as noted above, results from this test should be treated as indicative since it is not particularly powerful and we are using a relatively high confidence threshold. This could be either a gross error in the observations or model response, or it could indicate something amiss in this particular set of forcings: perhaps the Lean et al. (1995) solar reconstruction really is more realistic than the Hoyt and Schatten (1993) reconstruction, or perhaps the difference is simply compensating for another error. The "attributable trend" figure corresponding to Fig. 12 appears similar to Fig. 10, unsurprisingly since the anthropogenic signal is the dominant factor over this period.

Figure 12 provides an overall indication of model performance if we assume both that natural and anthropogenic forcings have contributed to recent observed climate change and that the relative amplitude of the responses to these different forcings are correctly simulated by the climate models. It does not tell us which of these various forcings are necessary to account for recent changes or, more quantitatively, what fraction of recent observed climate change is attributable to the different forcings. This can only be achieved by a multi-pattern analysis, to which we now proceed.

6.2 Multi-pattern analyses

6.2.1 Greenhouse and sulphate forcing

On the assumption that anthropogenic greenhouse gases and sulphate aerosols, as the largest single factors expressed in terms of total radiative forcing, have dominated climate change over the past few decades, the simplest multi-pattern analysis is to consider these two signals together. In all cases, both forcings were prescribed together in a GS ensemble and greenhouse gas influence was prescribed alone in a separate G ensemble. If we use these greenhouse and greenhouse + sulphate response-patterns as the signals in our detection model, the interpretation of the results is as follows. The amplitude of the greenhouse signal, $\beta_{G'}$, represents the amount of *additional* greenhouse influence we need to add to the greenhouse + sulphate response in order to reproduce the observations (the prime indicating that it is a correction). Thus the total scaling required on the amplitude of the greenhouse response is given by the sum, $\beta_G = \beta_{GS} + \beta_{G'}$, while the scaling required on the sulphate response is given simply by $\beta_S = \beta_{GS}$. These issues are discussed in detail in Allen and Tett (1999), and Stott et al. (2001) (Fig. 13).

Allowing for uncertainty in model-simulated signals has a much more substantial impact on results in multi-pattern analyses, particularly in the case of the ECH4 model where only single-member ensembles are available. For reasons discussed in Allen and Stott (2003), this uncertainty primarily affects estimated upper bounds on confidence regions, which in some cases recede to infinity. Averaging patterns across models is likely to be helpful here (Gillett et al. 2002a; Huntingford et al. 2006).
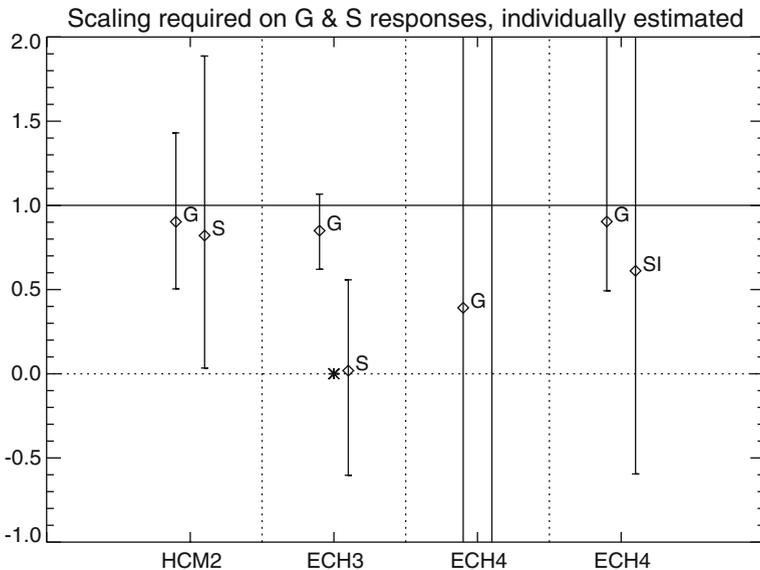

Fig. 13 Amplitude of greenhouse (G) and sulphate (S) signals estimated separately from the observed record for HCM2, ECH3 and ECH4 models (left three boxes). Right box shows corresponding result for the ECH4 sulphate signal including indirect sulphate forcing and tropospheric ozone

In all but one case (ECH4 G & S analysis) we find that greenhouse gas influence is both detectable and consistent in amplitude between model and observations (β_G consistent with unity and inconsistent with zero). We stress that, in this analysis, the sulphate signal is allowed to take whatever amplitude provides the best fit to the observations, so this detection of greenhouse influence is not contingent on the simulated sulphate response being of the correct amplitude. Uncertainties in the sulphate signal amplitude are somewhat larger, but it is detectable (β_S significantly greater than zero) in the HCM2 case. Uncertainties are significantly larger in the case of the ECH4 model. This is likely to be because only single-member ensembles are available, so the noise in the model-simulated response-patterns is as large as the noise in the observations, and possible degeneracy between the G and S signals simulated by this model. As a result, in the case when only direct sulphate influence was included in the ECH4 simulation, signals were sufficiently weak and degenerate that no useful bounds could be placed on the individual G and S amplitudes. We believe this is simply a consequence of poor signal-to-noise in this single-member ensemble, and hence does not appear to be a problem for other models in which ensembles are larger or the responses to forcing is stronger.

Figure 14 shows the contributions to 20th century warming from both greenhouse and sulphate influence, estimated from the observed climate record in this way. While the GS combination, considered as a single signal, was found to contribute 0.3–0.5 K/century to the warming observed over the century, when the individual contributions from greenhouse gases and sulphate aerosols are estimated individually, the greenhouse induced warming is estimated to lie in the range 0.3–1.2 K/century, partially compensated for by a sulphate cooling of up to –0.7 K/century. This separation of attributable warming into different components is highly sensitive to the details of the model simulations, with the ECH3 model indicating much lower
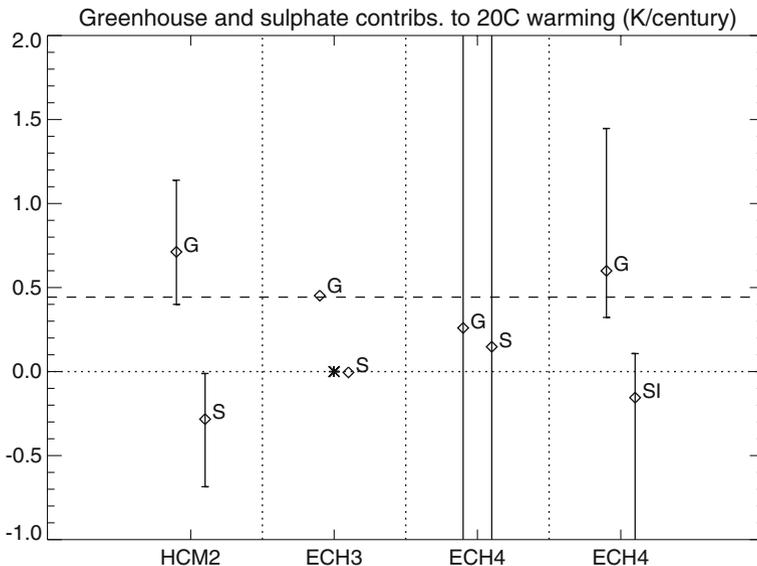

Fig. 14 Estimated contributions to global mean temperature change over the 1906–1996 period due to greenhouse gases (G) and sulphate aerosols (S)—including indirect sulphate and tropospheric ozone (SI) in the case of model ECH4

levels of both greenhouse- and sulphate-induced temperature changes. Again, in the ECH4 case with only direct aerosol forcing, attributable trends are highly uncertain because of the degeneracy between the signals.

6.2.2 Implications of greenhouse and sulphate signals

On the assumption that the net effect of natural forcing on the diagnostic considered here is relatively small, we can examine the implications of the joint distribution of the estimated magnitude of the greenhouse and sulphate signals. These joint distributions are shown in Figs. 15 and 16 for the HCM2, ECH3 and ECH4 models, also including results from the HCM3 model for completeness. The crosses show the best-guess scaling factors on the model-simulated greenhouse and sulphate responses, while the curves enclose the estimated 90% confidence region on the estimate. These regions are strongly tilted, indicating that the uncertainties in these two signals are correlated: that is, if we are underestimating the amplitude of the greenhouse signal, we must also be underestimating the amplitude of the sulphate signal and *vice versa* (Mitchell et al. 2001; Hegerl and Allen 2002). This to be expected, since greenhouse warming is opposed by sulphate cooling. In the case of the HCM2 and HCM3 models, the greenhouse and sulphate signals are sufficiently distinct from each other for both to be detectable: the confidence region does not intersect either axis. In the case of ECH3 and ECH4, only the greenhouse signal is detectable.

Figure 15 shows the implications of these estimated scaling factors for the net anthropogenic warming by the decade 2036–2046 under the IS92a forcing scenario, following the procedure of Allen et al. (2000). This is based on the assumption that the same scaling factors can be applied to the model-simulated future warming as over the past few decades. This assumption is supported by simple model simulations of the sensitivity of global mean temperature to key parameters, but would clearly be invalidated by a sudden non-linear climate change over the forecast period.

Forecast 50-year warming appears to be relatively well constrained by the observed signal, since the isolines of future warming (shown by the dotted contours) are aligned with the orientation of the confidence regions. Under this forcing scenario, net anthropogenic warming relative to pre-industrial by the decade 2036–2046 is estimated to be 1.0–2.1 (HCM2), 1.0–2.3 (HCM3), 1.4–2.3 (ECH3) and 0.6–2.0 K (ECH4-GSI). We should not expect these ranges to be identical, since they are the product of a statistical esimation procedure, but they serve to re-illustrate the point made in Allen et al. (2000): reconciling A-OGCM simulations with the observed record not only serves to provide an uncertainty estimate on individual forecasts, but it also, in principle, draws different model simulations together. The raw ECH3 simulation, for example, predicts a rather lower warming by the mid-21st century than the other three, but this is ''corrected'' by the model-data fitting exercise, largely as a result of down-weighting the amplitude of sulphate cooling.

Less well constrained by the observed signal is the estimated magnitude of sulphate forcing itself, shown in Fig. 16. Recall that the estimated scaling factor, β_i, represents the amount we have to scale the ith model-simulated response to a given forcing to reproduce the observed signal. If both forcing and response are uncertain,
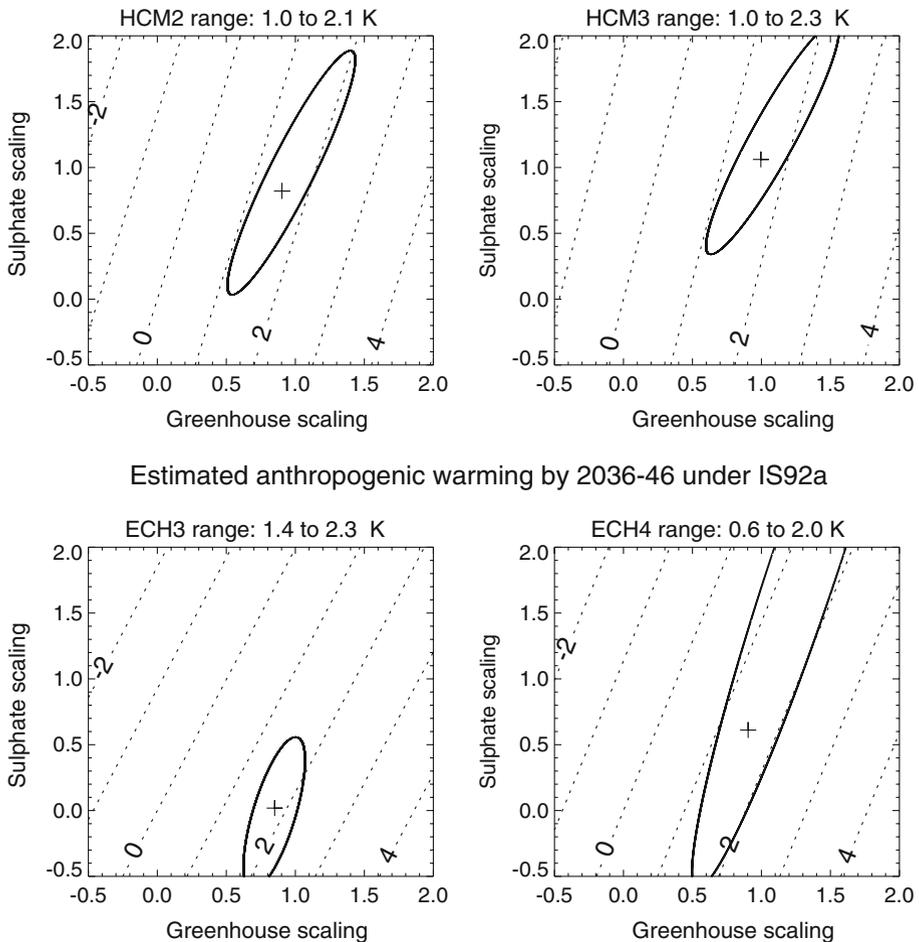
Fig. 15 Joint uncertainty intervals on scaling required on model-simulated greenhouse and sulphate signals to reproduce the observations over the 1946–1996 period: crosses indicate the best fit, curves enclose the 90% confidence region. Dotted contours show the model-simulated total anthropogenic warming by the decade 2036–2046 as a function of the scaling factors applied to the greenhouse and sulphate signals: to the extent that the confidence regions are aligned with these isolines of future warming, the forecast is well constrained by the observed signal. Numbers in titles on individual panels show 5th and 95th percentiles of the estimated distribution of warming by 2036–2046 consistent with the observed signal

then β_i may be written as the product of a scaling on the prescribed forcing, β_i^{F}, and a scaling on the response per unit forcing, β_i^{R}. If we assume that errors in the sulphate response per unit forcing scale with errors in the corresponding greenhouse response (also not strictly justified, since the mechanisms, time-history and spatial patterns of the two forcings differ), then $\beta_{\mathrm{G}}^{\mathrm{R}} = \beta_{\mathrm{S}}^{\mathrm{R}}$. Note that the β_i^{R} correspond to scaling factors applied to model-simulated responses normalised by the forcing, not the normalised response itself in terms of K/(W/m^2), so we are only requiring that errors in the two normalised responses are correlated with each other, not that the responses themselves are the same. Hence,

$$\frac{\beta_G}{\beta_G^F} = \beta_G^R = \beta_S^R = \frac{\beta_S}{\beta_S^F}. \tag{3}$$

If we further assume that greenhouse forcing is correctly specified in these models (not entirely consistent, since 1990 values for greenhouse forcing range from 1.7 to 2.2 W/m^2 in the simulations considered), then $\beta_G^F = 1$, giving

$$\beta_S^F = \frac{\beta_S}{\beta_G}. \tag{4}$$

For example, if an allowed scaling on the greenhouse signal is 0.8 and the greenhouse forcing is assumed correct, this implies that the model is over-estimating the greenhouse response by 25% (that is, $\beta_G = \beta_G^R = \beta_S^R = 0.8$). If, consistent with this scaling on the greenhouse signal, an allowed scaling on the sulphate signal, β_S, is 0.4, then allowing for the same 25% over-estimate in the normalised response, this implies the imposed sulphate forcing is overestimated by 50% (0.4/0.8 = 0.5).

The dotted contours in Fig. 16 show isolines of sulphate:greenhouse scaling factor ratios corresponding to different values of the sulphate forcing in 1990, as estimated from the relevant simulations. The implied range in sulphate forcing amplitude is relatively well constrained only in the case of HCM3 (–0.5 to –1.5 W/m^2 in 1990), and in the case of ECH4-GSI it is impossible to place a lower (most negative) bound on sulphate forcing amplitude, again reflecting the large uncertainties resulting from single-member ensembles. With ECH4-GS, no bounds can be placed on the sulphate response because the signal is too ill-defined. Where multi-member ensembles are available, the only conclusion that may perhaps be drawn from this figure is that very high values of net sulphate forcing are excluded as inconsistent with the observed signal. This is physically reasonable, since we are assuming in this analysis that the net effect of natural forcing on this diagnostic is small: hence if errors in the response are the same for both, the magnitude of the sulphate forcing must be less than that of the greenhouse forcing to account for the observed warming.

In summary, uncertainty in the sulphate forcing and response does not eliminate the need for a substantial response to greenhouse gases to account for the recent observed warming. Even if the net sulphate signal turns out to be very small, and models are therefore overestimating the rate of greenhouse warming, a substantial greenhouse signal remains after this overestimate is corrected. If the sulphate signal turns out to be very large, then the greenhouse warming will need to be larger still (and underestimated by current models) to account for the observed change.

6.2.3 Greenhouse, sulphate and solar forcing

If we include the model-simulated responses to natural as well as anthropogenic forcing into a multi-pattern analysis, a bewildering array of signal combinations present themselves, and it would be impossible to attempt an exhaustive analysis here. The specific combination of signals appropriate to a particular study will depend on the physical questions under investigation. It is misleading and ultimately fruitless to suggest there can be some kind of "global" detection and attribution analysis capable of summarising the climate-related information-content of a dataset
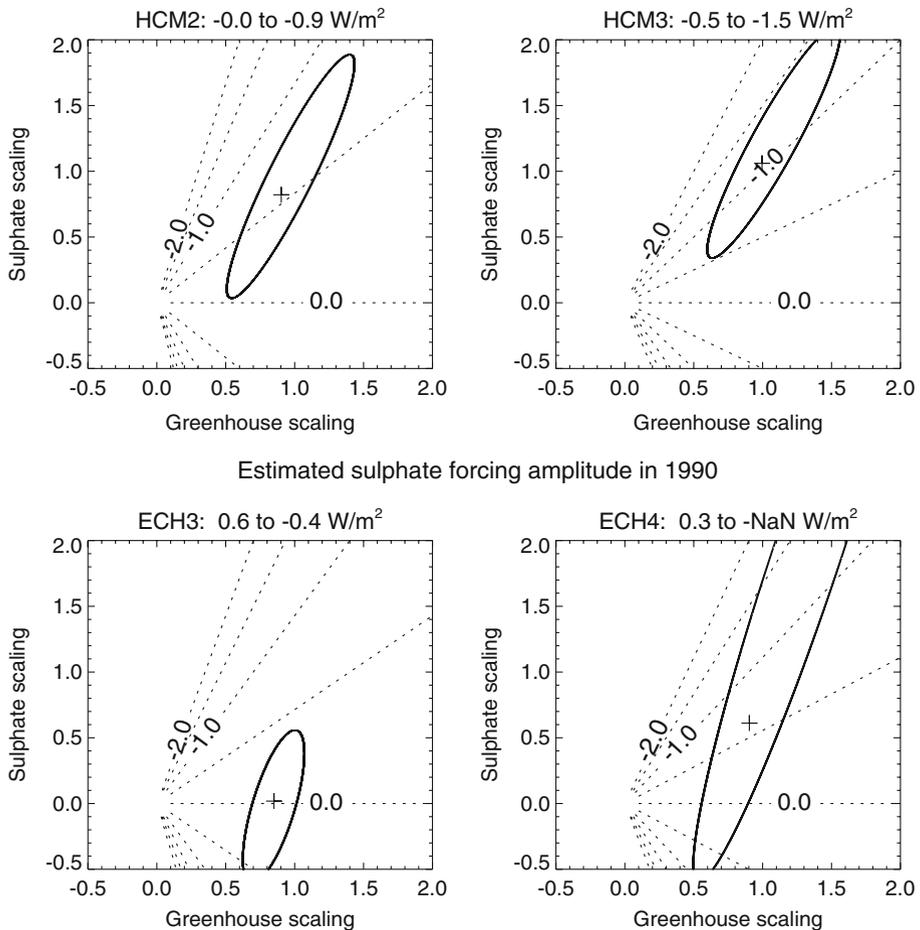
Fig. 16 As previous figure by plotted against isolines of sulphate forcing amplitude in 1990, shown as a function of sulphate:greenhouse scaling factor ratios on the assumption that the greenhouse forcing is correct in each of the model simulations

in a single estimation procedure. For this reason, we are distributing the software used in this study and selected input datasets to allow investigators to apply these techniques themselves to the specific questions that interest them. Nevertheless, to conclude this study, we present results from multi-pattern analyses from the two models for which response-patterns corresponding to natural forcing agents were available. We begin by focussing on the combination of anthropogenic and solar forcing, since two models are available with relevant runs (ECH3 and HCM2). In both cases, the Hoyt and Schatten (1993), and Lean et al. (1995), reconstructions of solar forcing are considered separately.

Including the impact of solar forcing while neglecting the impact of volcanic forcing is a somewhat artificial exercise, since even on interdecadal timescales, volcanic forcing is thought to be as large or larger than, and of the opposite sign to, solar forcing over the latter half of the 20th century on which we focus. Nevertheless,

it is instructive to present these results as a sensitivity study. We use a truncation of 15 for these multi-pattern results: relatively long control runs are available for both of these models mean higher truncations are possible. We note in the text if results are altered by reducing to truncation 10.

Results are shown in Fig. 17. In the ECH3 case, greenhouse influence remains detectable at the 5% level regardless of which reconstruction is used to provide the solar signal, although it is only detectable at the 10% level if the Lean et al. (1995), reconstruction is used and the truncation is reduced to 10. On the other hand, the uncertainty ranges on β_G are less than unity and with the Lean et al. (1995), forcing β_L appears to be larger than β_G. In physical terms, this implies that the model is over-estimating the response to greenhouse forcing slightly while getting the amplitude of the response to solar forcing either approximately correct or moderately underestimated. The discrepancy, however, is not very significant, as indicated by the overlap between the two ranges, and sensitive both to the solar reconstruction used and to the truncation, so it would be incorrect to conclude on the basis of this evidence alone that any process was amplifying the response to solar forcing that is not represented in this model.

It is noteworthy, however, that greenhouse influence remains detectable at at least the 10% level even at the lower truncation and allowing both sulphate and solar signals to fit the data as well as possible. Despite the fact that we are estimating three independent quantities from essentially only 10–15 pieces of information (the number of E-EOFs of the control used in the definition of the detection space), the greenhouse signal in the observed climate record is sufficiently strong for it still to be detectable.
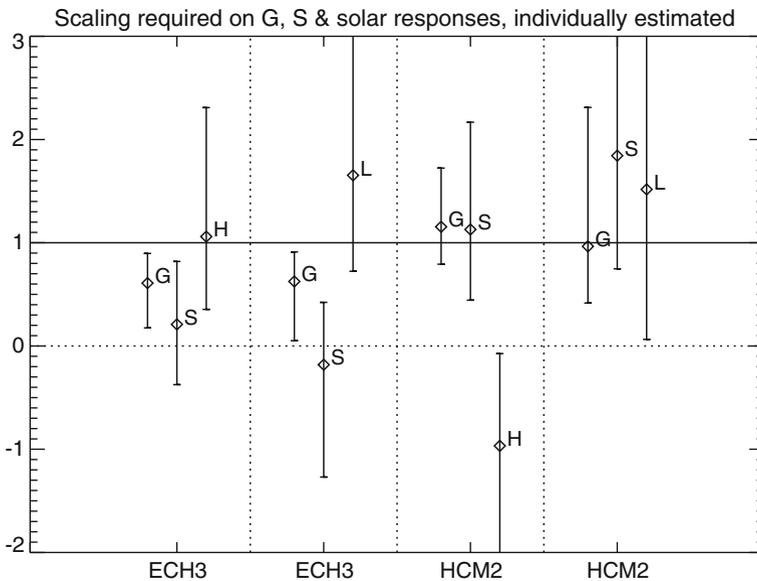

Fig. 17 Left two boxes: Three-way estimation results based on greenhouse, sulphate and solar forcing using the ECH3 model, considering the Hoyt and Schatten (1993), (H) and Lean et al. (1995), (L) solar reconstructions separately. Right two boxes: Three-way estimation results based on greenhouse, sulphate and solar signals simulated by the HCM2 model

With the HCM2 model (right two boxes), both greenhouse and sulphate signals are detectable with both solar reconstructions and both truncations, while the solar signal based on the Lean et al. (1995), forcing timeseries is detectable at truncation 15 but not at truncation 10. We are inclined to view this result with caution because it does not emerge in the lowest-ranked, best-sampled, E-EOFs. Moreover, it is sensitive to the solar reconstuction used: in contrast to the Lean et al. (1995), result, the response to the Hoyt and Schatten (1993), forcing timeseries emerges with the wrong sign: the range of uncertainty on H in the third box is entirely negative. If we consider other periods in the 20th century (Tett et al. 1999) or other diagnostics (Gillett et al. 2000; Stott et al. 2001), the influence of solar forcing on the observed climate record emerges more clearly.

Figure 18 shows trends attributable to anthropogenic and solar signals in these three-way analyses. In the case of HCM2, trends attributable to greenhouse gases and sulphate aerosols are similar to the all-anthropogenic case considered in Fig. 14. Because the scaling factor on H is entirely negative, this three-way analysis suggests a global mean cooling due to the Hoyt and Schatten (1993) reconstruction of solar irradiance, which is unphysical: the most likely explanation is that some other signal or combination thereof is confounding results.

6.2.4 Greenhouse, sulphate, solar and volcanic forcing

Finally, for the HCM2 model, we show results from analyses including the effect of volcanic forcing, this being the only model available to this study for which the relevant ensemble simulations have been performed: more are now available, and see Stott et al. (2006); Stone et al. (2006) for an update. The left two boxes in Fig. 19 show scaling factors assuming solar and volcanic influence were the only
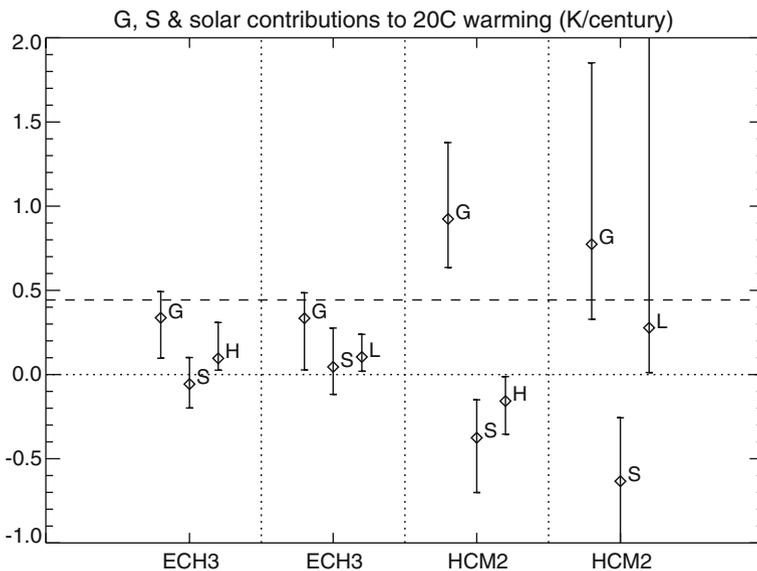

Fig. 18 Global mean temperature trends attributable to greenhouse, sulphate and solar signals presented in the previous figure
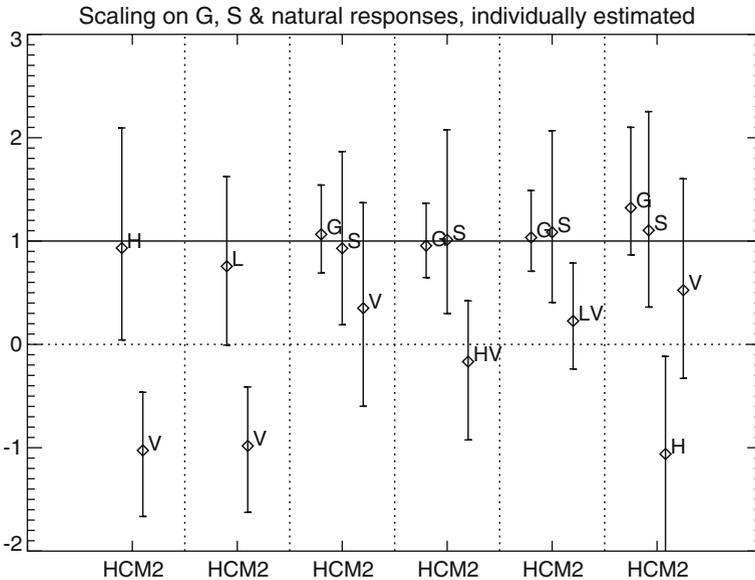
Fig. 19 Left two boxes: Two-way estimation results based on solar and volcanic signals using the HCM2 model: note that the volcanic signal is required to have an unphysical sign. Next three boxes: Three-way estimation results based on greenhouse, sulphate and either volcanic or combined solar + volcanic signals. Right box: Four-way estimation results based on greenhouse, sulphate, solar and volcanic forcing using the HCM2 model, estimating both solar and volcanic signal amplitudes separately

contributors to late 20th century temperature change. In both cases, as observed by Tett et al. (1999), and Stott et al. (2001), the only account consistent with the observations requires the volcanic signal to have an unphysical sign (i.e., enhanced volcanic aerosol loadings causing warming), and so can be rejected. Combinations of anthropogenic and volcanic forcing, or anthropogenic, volcanic and solar forcing, with the natural signals either combined or treated separately, indicate both anthropogenic signals are detectable in this diagnostic. The estimated amplitude of the anthropogenic signals is consistently close to unity, indicating the model-simulated amplitude of the anthropogenic response is approximately correct.

In a multi-pattern detection problem, we are asking the data to pick up on what may be relatively subtle features of the various signals that distinguish them from each other. As we increase the number of candidate signals using the same input dataset, the probability increases that the signals will begin to resemble scaled versions of each other, making estimated amplitudes very unstable: this is the degeneracy problem discussed by Tett et al. (1999). For this reason, results are generally much more sensitive to the precise specification of the problem than is the case in a single-pattern analysis. To illustrate this point, Fig. 19 shows results from the 4-way analysis including the solar signal based on the Hoyt and Schatten (1993), reconstruction: if the Lean et al. (1995), reconstruction is used instead, the level of degeneracy between the signals is such that none of the four signals can be distinguished from either zero or unit amplitude. Given that the main feature distinguishing solar from greenhouse forcing on surface temperatures is the 11-year

cycle in the former, the response to which is intentionally suppressed in this diagnostic, it is likely that other diagnostics will need to be used to resolve this degeneracy: see, for example, North and Stevens (1998), and North and Wu (2001). Our purpose in this paper is to explore the full implications of a single observation-vector.

Trends attributable to the various signals shown in Fig. 19 are displayed in Fig. 20. In all cases, trends attributable to natural agents are relatively small, while between 0.5 and 1.2–1.7 K/century (depending on the analysis) is attributed to greenhouse influence, and up to – 0.7 K/century cooling due to sulphate aerosols. The larger warming rates attributable to greenhouse influence in the final box should be treated with caution, since the best-guess scaling on the solar signal in this diagnostic is unphysically negative. The ''prior ignorance'' assumption may be particularly inappropriate here, in that despite considerable uncertainty regarding recent solar influence on climate, there seems to be a general consensus that solar irradiance has increased over the past century and that this would have a positive impact on surface temperatures. Nevertheless, the figure serves to illustrate how the uncertainty in the trends attributable to anthopogenic factors increase when natural agents are included in the estimation procedure.

6.3 Summary of results and sensitivity to details of the analysis

Figure 21 shows a (necessarily incomplete) summary of results discussed in this study expressed in terms of scaling factors by which various model-simulated signals need to be multiplied to reproduce the observations. For completeness, the HCM3 model has been included although details of this model are discussed elsewhere (Stott et al. 2000; Tett et al. 2002). To re-iterate, models forced with greenhouse
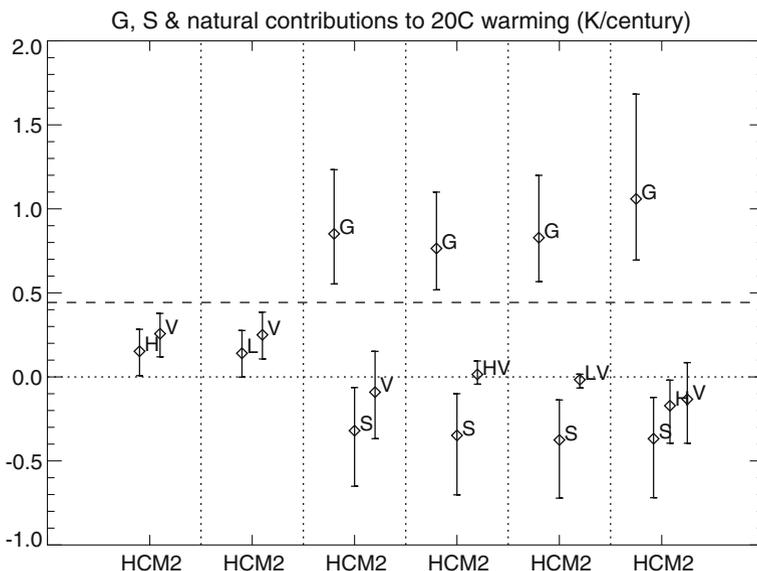

Fig. 20 Global mean temperature trends attributable to the various signals presented in the previous figure
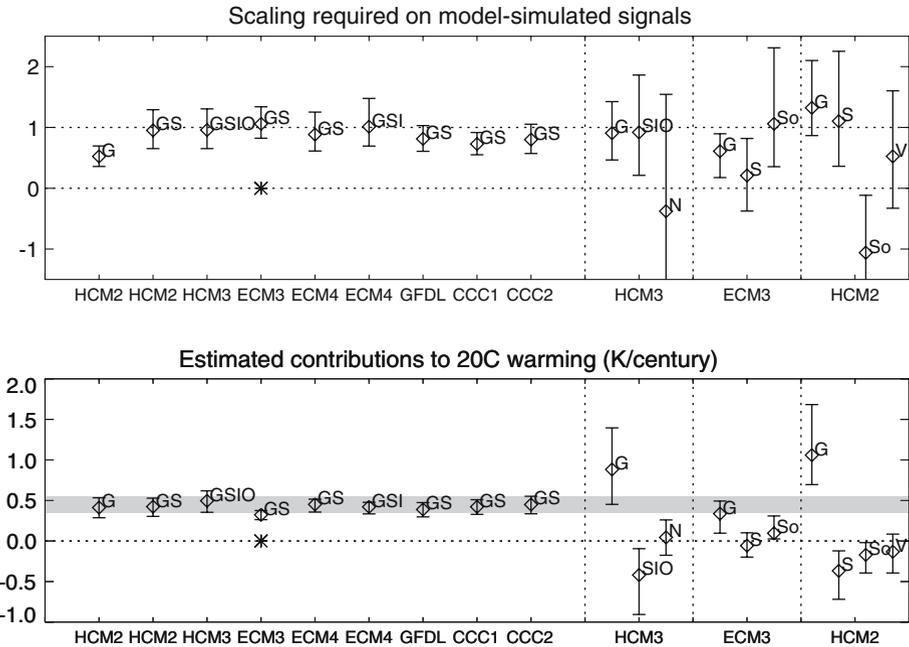
Fig. 21 Summary of results presented in this paper—see text for details

gases alone appear to overestimate the observed response, and hence their response needs to be scaled down. Models forced with the combinations of greenhouse gases and sulphate aerosols appear to simulate the amplitude of the observed response approximately correct, at least at this level of confidence, although the GFDL, CCC1 and CCC2 appear to be overestimating the response (estimated scaling factors less than unity, in the case of CCC1 significantly so).

Selected multi-signal results are shown in the right three panels: in the case of HCM3, these comprise greenhouse; the combination of direct and indirect sulphate, tropospheric and stratospheric ozone forcing; and the combination of solar and volcanic forcing. In the case of ECH3, these comprise greenhouse; direct sulphate; and solar forcing, shown here using the Hoyt and Schatten (1993) reconstruction. In the case of HCM2, these comprise greenhouse; direct sulphate; solar (again using Hoyt and Schatten 1993); and volcanic forcing. These estimation results suggest that the ECH3 model may be overestimating the greenhouse response (although the signal remains detectable), in that the uncertainty range for β_G is entirely less than unity. The lack of a simulation of the volcanic response with this model complicates the interpretation of this result: volcanic cooling has probably masked some greenhouse warming in the observed record (e.g., Christy and McNider 1994), and without a volcanic signal in the analysis, the estimation procedure attributes this to the model overestimating the magnitude of the greenhouse response.

The lower panel shows the trends attributable to different signals under these various accounts of 20th century temperature change, with a grey band indicating the estimated total warming and the corresponding uncertainty range from the most variable model control segment used for uncertainty analysis.

6.4 Results based on ordinary least squares regression

Throughout this paper, we have used a variant on the standard weighted-least-squares approach to detection and attribution (Hasselmann 1997) that takes into account the presence of sampling uncertainty in model-simulated signals based on small ensembles. This algorithm has been shown (Stott et al. 2003) to correct for a low bias in estimated signal amplitudes based on the standard approach: hence we believe the estimated signal amplitudes and attributed trends in Fig. 21 to be as close to unbiased as we can achieve. Nevertheless, as a sensitivity study, Fig. 22 shows the corresponding results based on the standard (ordinary least squares, or OLS) approach, as used by Hegerl et al. (1997), Tett et al. (1999) and Stott et al. (2001). Overall, the message of the figure is little changed except that models now appear to be consistently overestimating the response in the single-pattern results: as stated above, this is an artefact of the bias in the OLS algorithm. Detection of non-greenhouse anthropogenic influence as simulated by the HCM3 model is now marginal, and the solar signal is scaled down in ECH3, being replaced by the greenhouse signal. This latter case illustrates how the biases in OLS are particularly acute for low signal-to-noise responses to natural forcing. Results for HCM2 appear relatively unchanged.

6.5 Sensitivity to control segments used for optimisation and testing

An important issue in detection and attribution studies, discussed at length in Allen an Tett (1999), is the choice of the "detection space" used for model-data
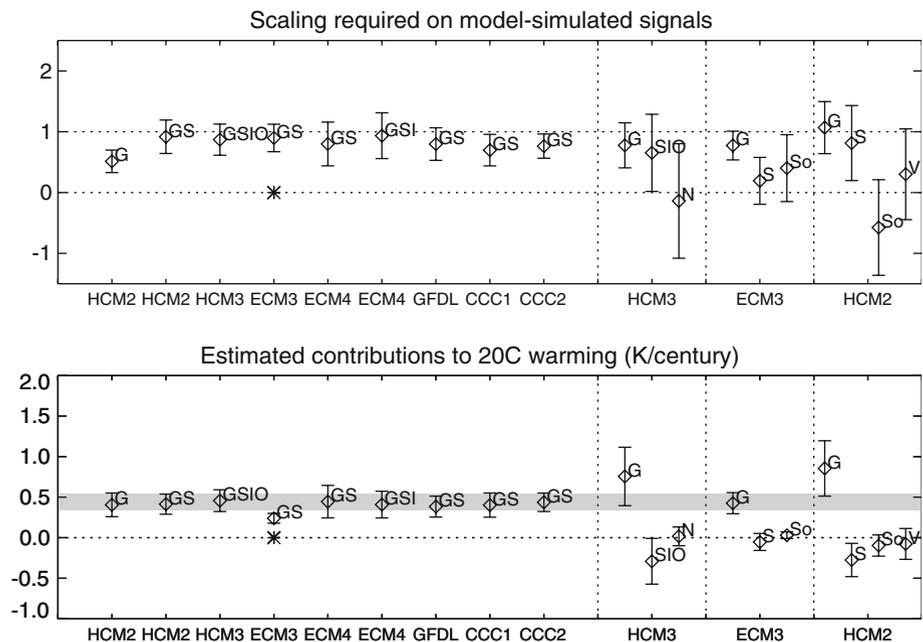

Fig. 22 As previous figure, but based on a weighted ordinary least squares regression rather than the total least squares approach used throughout this paper

comparison. This is conventionally determined by an E-EOF analysis of a segment of a control integration, truncating at a small number of E-EOFs. Since the SVD algorithm used to compute E-EOFs is non-linear, there is scope for unpredictable changes to results emerging from a different choice of control to compute the E-EOF basis. As a simple check on the magnitude of this effect, we exchange the control segments used for determining the detection space and hypothesis testing respectively. Results of this are shown in Fig. 23: the general picture as regards detection is largely unchanged. Among the single-pattern results, ECH3-GS now appears to underestimate the response by a significant margin, but is still flagged as suspect by the residual check. Among the multi-pattern results, the upper bounds on HCM3 signal amplitudes are now virtually undefined while the HCM2 signals are reasonably unaffected—with a longer control available for HCM2, less sensitivity to this change is to be expected. With ECH3 all three signals, greenhouse, sulphate and solar, are now detected and consistent with unity.

The ECH3 result illustrates a problem arising from the large number of relatively arbitrary choices that need to be made in a detection and attribution study to define the diagnostic considered and fix the details of the analysis. If one set of choices makes the model look more consistent with reality than another, what are we to conclude? On the one hand, if a signal is detectable under one approach and not under another, this could simply indicate the second approach was looking in the wrong direction. On the other hand, it is essential that the details of the analysis procedure are not "tuned" to produce a "desired" result, whatever that may be, since the overall algorithm is sufficiently flexible for the outcome of such tuning to be highly misleading. To avoid this problem, we defined our approach at the outset
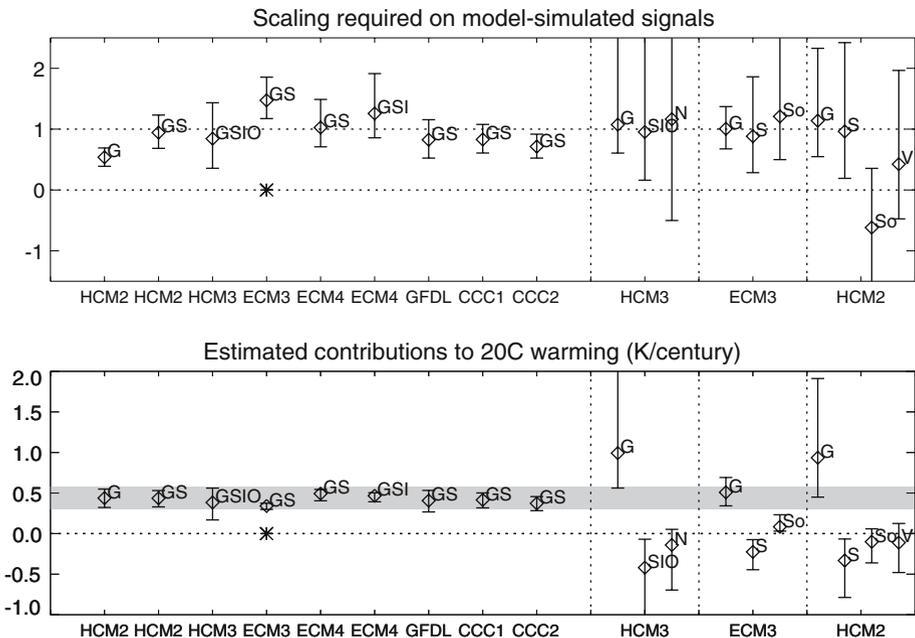

Fig. 23 As summary figure, but exchanging the segments of control used to define the detection space and for optimisation and for hypothesis testing respectively

based simply on the methodology of Allen et al. (2000), and present the results as they first appeared, even in those cases (as here) where a subsequent modification of the algorithm presented a model in a better light. In principle, the impact of these arbitrary choices could be incorporated into the overall procedure, by repeating the analysis many times exploring the full range of options and reporting some kind of weighted sythesis of the results. The problem then becomes quantifying the error rate for such a ''composite'' procedure, although the literature on bootstrapping might help (see, for example, Wilks 1997). For the sake of simplicity, we have simply noted sensitivities where they arise.

6.6 Sensitivity to global mean trend

One reason results based on single signals should be expected to give consistent estimated contributions to 20th century warming trends is that the warming trend is included in the data used for the analysis: if all the information in these diagnostics were contained in the global mean temperature trend, then these ''attributable trends'' would be equal by construction: all we would be doing was fitting the various models to the same data and arriving at the same result. To ascertain whether this is the case, Fig. 24 shows the summary figure based on data from which the global mean temperature trend has been removed from both models and data prior to the detection and attribution analysis. Clearly, important information has been lost, in that without the global mean trend information we
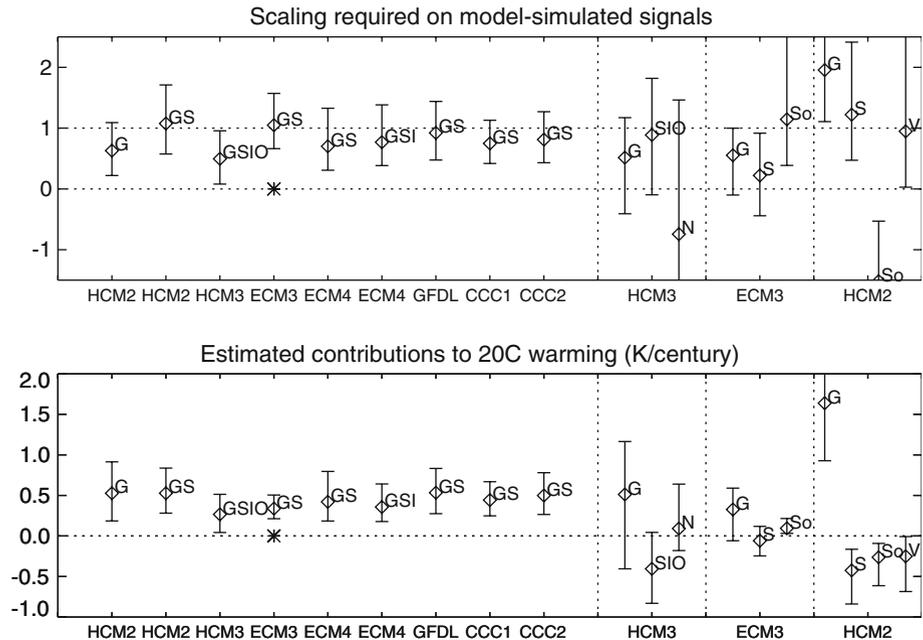

Fig. 24 As summary figure, but preprocessing both observations and model simulations to remove the linear trend in global mean temperatures before estimating scaling factors. Estimated contributions to 20th century warming in such pre-processed data are, of course, identically zero, so the trends removed are added back in after the analysis to generate the lower panel

can no longer detect greenhouse or sulphate signals independently of each other in the HCM3 and ECH3 3-way analyses. Both signals remain detectable in the HCM2 4-way case, but the uncertainty range on β_G no longer includes unity. Nevertheless, the overall estimated magnitude of the signals remains consistent with the base case, indicating this is not simply dictated by the global mean trend. Moreover, the attributable trends shown in the bottom panel, which are now completely independent of the data used in the estimation, are also broadly consistent with the base case.

6.7 Sensitivity to diagnostic used for analysis

Finally, Fig. 25 presents results from a completely different diagnostic: the spatial pattern of 50-year linear trends used by Hegerl et al. (1996), and Hegerl et al. (1997). A more detailed comparison of these two approaches is presented by Gillett et al. (2002b): the only points to note here are that uncertainties are clearly much larger without the information about the time-history of the signals included. In the four-way HCM2 regression, in particular, the confidence intervals are now unbounded: there simply is not enough information in the spatial trend-pattern alone to distinguish between these signals. In the single-pattern cases, the GS signal amplitudes remain generally detectable and consistent with unity and both greenhouse and sulphate influence are detectable in the three-way HCM3 and ECH3 results.
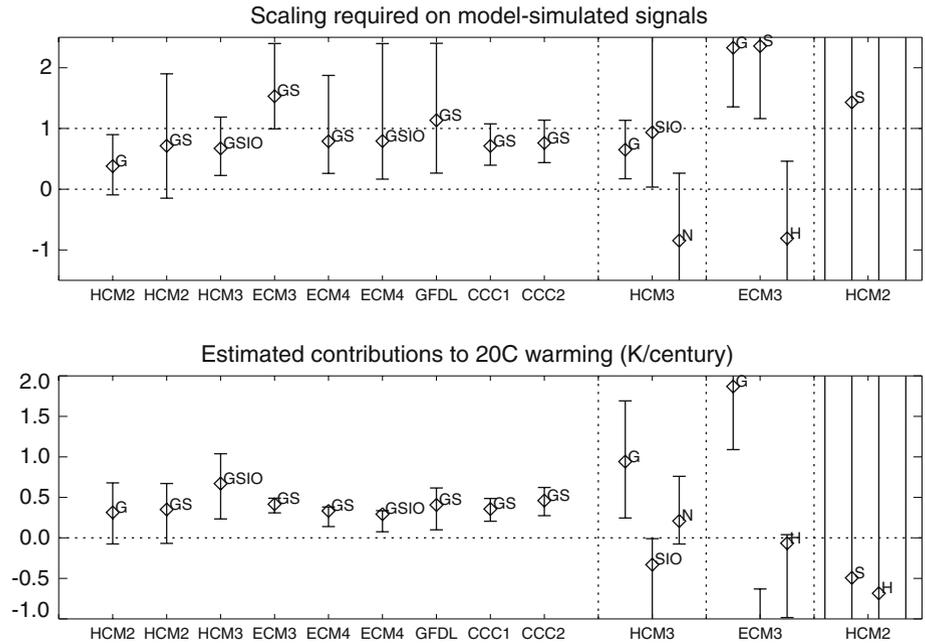

Fig. 25 As summary figure but based on the spatial pattern of 50-year temperature trends rather than the spatio-temporal pattern of temperature change used in previous figures. Lower panel now refers to trends over the 1946–1996 period, which are generally larger than those over the century as a whole

Consistent with the findings of Gillett et al. (2002b), the greenhouse and sulphate signals predicted by the ECH3 model are detected at a much higher signal-to-noise level with this diagnostic than with the spatio-temporal patterns used earlier: indeed, if this were the only information at our disposal, we might conclude that the ECH3 model was significantly underestimating the response to these two forcing agents. This illustrates the sensitivity of quantitative conclusions to the specification of the signal under investigation, particularly when multiple signals are included in the estimation problem. The reason for this is simple: model errors are not global, and a model may overestimate one feature of the observed response while underestimating another. Hence if practical conclusions are to be drawn from detection results, it is essential the the detection diagnostic used is tailored to be as closely linked as possible to the question of interest. Successful detection and attribution of hemispheric-scale temperature changes does not necessarily imply any greater confidence in model-based predictions of regional precipitation changes.

7 Remaining uncertainties and methodological issues

With all these figures, the question naturally arises as to which is the ''best'' estimate of anthropogenic and natural influence on climate over the 20th century. Two conflicting demands mean that there can be no definitive answer to this question. On the one hand, we wish to estimate as much as possible from the observed climate record and to rely on model-simulated signal amplitudes as little as possible. On the other hand, we have only a finite amount of data available and model-based estimates of internal climate variability are only reliable, if at all, on the largest spatio-temporal scales. The more quantities we attempt to estimate, the larger the model-data discrepancy that is required for a parameter-set to be rejected and the larger the confidence intervals become. Greater uncertainty is not inherently undesirable, of course, although it makes it more difficult to draw useful conclusions from the results. A more dangerous problem is that, as we increase the number of signals with the same input data, it becomes more likely that a chance feature of a particular signal combination will account for a significant fraction of the variance in the observed data, giving an apparently good fit for entirely spurious reasons.

The HCM2 three-way and four-way detection results with the solar signal based on the Lean et al. (1995), reconstruction provide an excellent illustration of this point. The observed record cannot be accounted for by the Lean et al. (1995), signal alone (Fig. 11), nor in combination only with volcanic forcing. If this solar signal is combined with the simulated volcanic signal, results from a three-way regression with anthropogenic forcings remain physically reasonable, with β_G and β_S both greater than zero and consistent with unity and β_{LV} relatively small. In the four-way case (Fig. 19) the estimated amplitude of the solar signal is of an unphysical sign: the problem breaks down because we are trying to estimate 4 pieces of information from only 10–15 independent pieces of data (the coefficients on the E-EOFs).

There are two solutions to this problem, either introducing prior constraints based on physically reasonable ranges for these estimated quantities or introducing more information in the data used to constrain them. It is inevitable that, for a given amount of data, the estimation problem will eventually break down as we try to estimate the amplitudes of more and more candidate signals. Hasselmann (1997), remarked on this point, but concluded it would not be too serious a problem since

the number of candidate explanations for recent climate change is relatively small. Unfortunately, the imagination of the scientific community is such that the number of candidate explanations for recent climate change may be unbounded, so we cannot simply estimate all candidate signal amplitudes mechanically from the available data. As in every other problem, progress can only be made by combining prior knowledge with the additional information provided by a particular dataset.

Given that the combined forcing due to solar and volcanic activity over the latter part of the 20th century is relatively small compared to anthropogenic forcing, and the response appears, on the basis of Figs. 17–20, to be indistingishable from zero in this diagnostic, it is not necessarily true that inclusion of these natural signals in the estimation procedure will "improve" (move towards the truth) the estimated amplitude of the anthropogenic signals. In any case, in almost all cases considered, it makes relatively little difference, with greenhouse warming in the range 0.3 to around 1.2 K/century, with the upper bound particularly uncertain, and sulphate cooling up to –0.7 K/century.

8 Conclusions

We have considered a range of possible scenarios to account for recent large-scale near-surface temperature changes, using a combination of model-simulated responses to external forcing and model-simulated internal variability. In addition to the fundamental constraints on any detection and attribution study mentioned at the end of the previous section, we are also subject to a practical constraint, which is to extend the analysis to as wide a range of models as possible when only a very small number of models have been run with the full range of external forcing scenarios.

Our results indicate that the combination of greenhouse and sulphate forcing, as simulated by these climate models, is generally adequate to account for large-scale decadal temperature changes over the period 1946–1996, expressed as anomalies about the 1906–1996 climatology: that is, with one exception, remaining model-data discrepancies can be explained as internal variability as simulated by these models. Under this account, anthropogenic influence is responsible for a 0.3–0.5 K/century warming (5–95% range) over the 1906–1996 period. This is consistent with the range of warming rates simulated by these climate models but, crucially, was estimated from the observed climate record without making any assumption about the correctness or otherwise of the amplitude of the model-simulated responses.

The detection and attribution procedure involves scaling the amplitude of model-simulated signals up or down to fit the observed data: hence, if only a single signal is considered for each model, we should expect, by construction, for the trend attributable to this signal to be similar to the total observed trend and to those estimated from other models. A more direct indication of model performance is given by the scaling factors, β_i, required to bring the model-simulated signals into line with the observations. For the combination of greenhouse and sulphate forcing, these are all found to be close to unity, implying all the models are giving approximately the correct magnitude of response. This may seem a rather surprising result given the factor of two range of climate sensitivities (equilibrium warming on doubling of carbon dioxide) in the models considered in this paper. Part of the explanation is

likely to be that ocean heat uptake is compensating for the stronger atmospheric feedbacks in the more sensitive models, delaying the response. Hence the intermodel range of transient climate responses is much less than the range in equilibrium sensitivities (Hansen et al. 1985; Cubasch et al. 2001).

A more worrying possibility is that the forcing, particularly the very uncertain sulphate component, has been "tuned" (perhaps unconsciously) to the sensitivity of each model to reproduce the overall observed trend. Our results in which we intentionally suppress information on the global mean trend (Fig. 24) provide some reassurance on this point. Without a systematic and formal optimisation procedure, which has certainly not been used by any of the modelling groups concerned, it would be effectively impossible to tune the forcings to anything other than global mean changes. We find that the removal of global mean information, although it naturally increases the uncertainties, does not alter the overall picture of broad model-data consistency.

When greenhouse and sulphate signals are considered separately, in those cases where the necessary runs were available, we find the greenhouse contribution to this warming to lie in the range 0.3–1.2 K/century, with the upper bound particularly sensitive to the details of the analysis, and the magnitude of sulphate-induced cooling being up to –0.7 K/century. Thus the maximum warming rate attributable to greenhouse gas influence over the past century considerably exceeds the net observed warming, and this result is found to be relatively insensitive to the inclusion of model-simulated responses to natural external forcing in the analysis.

Attempts to account for recent near-surface temperature changes in exclusively natural terms were consistently unsuccessful, but a significant caveat in this and all other results is our reliance on only three models (ECH3, HCM2 and HCM3) for our results for anything other than purely anthropogenic response-patterns. One of these (ECH3) has not been run with volcanic forcing and may display an unrealistically low level of internal variability. It is clearly important for more models to be run with a wider range of external forcing scenarios to establish the robustness of the results reported here. To this end, we are providing the software used in this study (written in the IDL data-processing language) and key input datasets to any groups wishing to examine the sensitivity of our results and to apply this basic algorithm to other model simulations.

When both natural and anthropogenic signals are considered, anthropogenic changes appear to account for most of the warming over the past 50 years with only a small (in many cases undetectable) contribution from the natural sources considered (solar variability and volcanic activity). This should not be interpreted as indicating that natural factors have had no influence on climate over this period: simply that they have had no detectable influence on the specific diagnostic used in this study. This, being based on decadal mean data, would tend to suppress any response to the 11-year solar cycle and short-term response to volcanic eruptions. Uncertainty in the magnitude of the response of the system to natural, particularly solar, forcing remains one of the most important outstanding issues for detection and attribution. As the title implies, our focus in this study was on quantifying the anthropogenic contribution to recent near-surface temperature change. In so doing, we considered the response to natural forcing, but only as a potential confounding signal, not as an end in itself. It is to be hoped that more targeted studies will provide tighter constrains on natural contributions to recent climate change (Santer et al. 2001).

Appendix total least squares estimation

The estimation procedure used in this study is based on the assumption that the observed climate record, as characterised by the n-element vector $\mathbf{y}$ ($n = 125$ in our example), can be represented as a linear superposition of m model-simulated response-patterns, $\mathbf{x}_i$, denoted as the columns of the $m \times n$ matrix $\mathbf{X}$. Ideally, large ensemble simulations should be used to minimise the influence of sampling noise on the model-simulated signals (i.e., to pin down exactly what we are looking for in the observations). In practice, to compare results from models given a wide range of available ensemble-sizes, we need to take sampling variability explicitly into account, giving the detection model:

$$\mathbf{y} - \upsilon_0 = \sum_{i=1}^{m}(\mathbf{x}_i - \upsilon_i)\beta_i^{\text{true}} \tag{5}$$

$$= (\mathbf{X} - \mathbf{\Upsilon}_X)\boldsymbol{\beta}, \tag{6}$$

in which υ_0 is the sampling noise in the observations and υ_i (or the ith column of $\mathbf{\Upsilon}_X$) is the noise in the ith model-simulated responses to external forcing.

We assume a linear relationship between the underlying deterministic (noise-free) observed and modelled responses, $\mathbf{y} - \upsilon_0$ and $\mathbf{x}_i - \upsilon_i$, given by the unknown scaling factors, β_i^{true}. The physical interpretation of these scaling factors is as follows: β_i^{true} represents the amount by which we have to scale the ith model-simulated response to reproduce the observations, assuming the shape of the response is accurately simulated (apart from sampling error due to the finite size of the ensemble used to generate it) and only the amplitude is unknown.

Assuming both υ_0 and the υ_i are independent and Gaussian distributed, the probability of observing this particular combination of climate observations $\mathbf{y}$ and model simulations $\mathbf{X}$ for a given set of scaling factors $\boldsymbol{\beta}$—known as the "likelihood" of $(\mathbf{y}, \mathbf{X})$ given $\boldsymbol{\beta}$ and the detection model (6)—is given by:

$$P(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta}) = constant \times e^{-r^2}, \tag{7}$$

where r^2 is the sum of the noise in both $\mathbf{y}$ and $\mathbf{X}$ implied by that value of $\boldsymbol{\beta}$, weighted by their respective inverse covariance matrices, assuming these are known. That is,

$$r^2 = \upsilon_0^T \mathbf{C}_0^{-1} \upsilon_0 + \sum_{i=1}^{m} \upsilon_i^T \mathbf{C}_i^{-1} \upsilon_i, \tag{8}$$

where $\mathbf{C}_0$ and $\mathbf{C}_i$ are the expected $n \times n$ covariance matrix of the noise in the climate observations $\mathbf{y}$ and the ith model-simulated response-pattern, $\mathbf{x}_i$, respectively.

What we ultimately want is the distribution of $\boldsymbol{\beta}$ given this set of climate observations and model simulations, or $P(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$. This also depends on our prior knowledge, requiring an application of Bayes theorem:

$$P(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \frac{P(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta})P(\boldsymbol{\beta})}{P(\mathbf{y}, \mathbf{X})}. \tag{9}$$

In this equation, $P(\boldsymbol{\beta})$ represents our prior knowledge or assumptions about the range of possible values for the β_i^{true}: that is, how probable we consider each value of β_i^{true} before these observations are taken into account. $P(\mathbf{y}, \mathbf{X})$ is the probability of obtaining this particular set of observations and model simulations, given those constraints, irrespective of the true value of $\boldsymbol{\beta}$: it acts as a normalisation, ensuring that all probability density functions integrate to unity. In the cases considered here, we will assume both $P(\boldsymbol{\beta})$ and $P(\mathbf{y}, \mathbf{X})$ are uniform distributions, giving the "standard" or "classical" detection model used, for example, by Hegerl et al. (1997), and Tett et al. (1999). It is, however, important to recognise that there is no fundamental distinction between the "classical" and "Bayesian" approaches. Classical optimal fingerprinting is simply a special case of a Bayesian analysis in which uniform priors are assumed: that is, in which we assume no prior knowledge whatsoever of the parameters we are trying to estimate and the observations we are likely to obtain.

If we assume that the noise in both $\mathbf{y}$ and $\mathbf{X}$ is Gaussian and dominated by internal climate variability as simulated by the A-OGCM, then we can make $\mathbf{C}_i = \mathbf{C}_0 = \mathbf{C}_N$, the covariance between elements of $\mathbf{y}$-like vectors drawn from an unforced control simulation, simply by scaling the ith response-pattern by $\sqrt{\ell_i}$, the number of simulations in the ith ensemble. We also introduce a so-called "pre-whitening operator", $\mathbf{P}$, defined such that

$$\mathbf{P}\mathbf{C}_N\mathbf{P}^T = \mathbf{I}^{(\kappa)}, \tag{10}$$

where $\mathbf{I}^{(\kappa)}$ is a unit matrix of rank κ. The pre-whitening operator, $\mathbf{P}$, will be model-dependent: there is no universal operator to remove correlations in variability simulated by any model. The so-called "rank of the detection space", κ, need not be the same as the number of observations n—if it is less than n, $\mathbf{P}$ not only acts to make the noise equal-variance and uncorrelated in all the elements of the pre-whitened observations and model-simulated response-patterns, but it can also be used to filter out any component of the signal that correlates with modes of variability that are poorly simulated by the A-OGCM. This dual role of the pre-whitening step in climate change detection is discussed at length in Allen and Tett (1999).

The simplest, although by no means unique, way of defining $\mathbf{P}$ is to decompose a sequence of "climate noise" realisations from a control integration of the A-OGCM into a set of spatio-temporal patterns, or extended-EOFs (Weare and Nastrom 1982) and associated principal components, using a singlar value decomposition. The κ rows of $\mathbf{P}$ are then defined as the κ highest-ranked E-EOFs weighted by their

respective inverse singular values. With this choice of weighting, if we were to take another realisation of climate noise, from a different control integration for example, and pre-multiply it by $\mathbf{P}$, we would expect approximately the same noise variance in each of the κ rows of the resulting pre-whitened noise vector, and for the noise to be uncorrelated between rows.

Representing the pre-whitened model-simulated response-patterns, $\mathbf{P\,X}$ as the first m columns of the $\kappa \times (m + 1)$ matrix $\mathbf{Z}$, and the pre-whitened observations, $\mathbf{P\,y}$, as the $m + 1$ column, the detection model (6) can be re-written (Allen and Stott 2003)

$$(\mathbf{Z} - \mathbf{\Upsilon}_Z)\mathbf{v} = \mathbf{0}, \tag{11}$$

where $\boldsymbol{\beta}$ has been encoded into the vector $\mathbf{v}$ (see below). Equation (11) states that a linear relationship exists between the noise-free observations and the noise-free model-simulated response-patterns or, equivalently, that a vector exists which is orthogonal to every row of $\mathbf{Z} - \mathbf{\Upsilon}_Z$. The pre-whitened noise contamination, $\mathbf{\Upsilon}_Z$ is unknown, but an unbiased estimate of the orientation of the vector $\mathbf{v}$ can be obtained from the eigenvector of $\mathbf{Z}^T\mathbf{Z}$ with the smallest eigenvalue, as was first observed by Adcock (1878). This estimate, $\tilde{\mathbf{v}}$, also maximises the likelihood function in Eq. (7) by minimising the *total* (perpendicular) squared distance of the observation-model points (defined by the κ rows of $\mathbf{Z}$) from the m-dimensional plane, orthogonal to $\tilde{\mathbf{v}}$, which represents the best-fit solution: hence the name Total Least Squares solution (van Huffel and Vanderwaal 1994).

Best-fit scaling factors on the individual model-simulated response-patterns, as required to reproduce the observations in model (6) may be obtained from the ratios,

$$\beta_i = -\sqrt{\ell_i}\,\frac{\tilde{v}_i}{\tilde{v}_{m+1}}, \tag{12}$$

recalling that the ith column of $\mathbf{Z}$ has been scaled by $\sqrt{\ell_i}$ to equalise the noise variance in all observables. Best-fit reconstructions of the noise-reduced observations and model-simulated response-patterns are obtained by extracting the component of $\mathbf{Z}$ orthogonal to $\tilde{\mathbf{v}}$, thus:

$$\tilde{\mathbf{Z}} = \mathbf{Z} - \mathbf{Z}\tilde{\mathbf{v}}\tilde{\mathbf{v}}^{\mathbf{T}}. \tag{13}$$

Confidence intervals on the elements of $\tilde{\mathbf{v}}$ are obtained using the formulae for uncertainties in an eigen-decomposition given in North et al. (1982), and Allen and Smith (1996), using an independent estimate of $\mathbf{C}_N$ to re-estimate the eigenvalues of $\mathbf{Z}^T\mathbf{Z}$ to avoid the problem of "artificial skill" noted by Bell (1986)—see Allen and Stott (2003), for details.

Twentieth-century warming trends "attributable" to specific external influences under the detection model (6) are obtained as follows: the columns of $\tilde{\mathbf{Z}}$ in Eq. (13) represent our best estimate of the noise-free (and hence unobservable) modelled and observed climate change respectively. A range of other reconstructions are also consistent with these observations and this model simulation at a given confidence level, found in a similar manner to the uncertainty range in $\tilde{\mathbf{v}}$. We compute linear trends in global mean temperature in these "possible" reconstructions by extracting

the first spherical harmonic for each of the five decades in the observation vector and regressing them onto a vector which is zero in the 1946–1956 decade, increasing linearly thereafter: this provides the best estimate of the linear trend over the 1906–1996 period in data that has been expressed as anomalies about the 1906–1996 climatology. For the sake of brevity, we refer to this as the "attributable trend over the 20th century".

References

AchutaRao K, Sperber KR (2006) ENSO simulation in coupled ocean-atmosphere models: are the current models better? Clim Dynam 27, doi:10.1007/s00382-006-0119-7

Adcock RJ (1878) A problem in Least Squares. Analyst (Des Moines, Iowa) 5:53

Allen MR, Smith LA (1996) Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. J Clim 9:3373–3404

Allen MR, Stott PA (2003) Estimating signal amplitudes in optimal fingerprinting I: theory. Clim Dynam 21:477–491

Allen MR, Tett SFB (1999) Checking internal consistency in optimal fingerprinting. Clim Dynam 15:419–434

Allen MR, Stott PA, Mitchell JFB, Schnur R, Delworth T (2000) Quantifying the uncertainty in forecasts of anthropogenic climate change. Nature 407:617–620

Barnett TP, Hasselmann K, Chelliah M, Delworth T, Hegerl G, Jones P, Rasmusson E, Roeckner E, Ropelewski C, Santer B, Tett S (1999) Detection and attribution of recent climate change—a status report. Bull Am Meteorolog Soc 80:2631–2659

Bell TL (1986) Theory of optimal weighting to detect climate change. J Atmos Sci 43:1694–1710

Berliner LM, Levine RA, Shea DJ (2000) Bayesian climate change assessment. J Clim 13:3805–3820

Blackmon M, Boville B, Bryan F, Dickinson R, Gent P, Kiehl J, Moritz R, Randall D, Shukla J, Solomon S, Bonan G, Doney S, Fung I, Hack J, Hunke E, Hurrell J, Kutzbach J, Meehl J, Otto-Bliesner B, Saravanan R, Schneider EK, Sloan L, Spall M, Taylor K, Tribbia J, Washington W (2001) The community climate system model. Bull Am Meteorolog Soc 82:2357–2376

Boer GJ, Flato G, Reader MC, Ramsden DA (2000) transient climate change simulation with greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the 20th century. Clim Dynam 16:405–425

Braganza K, Karoly DJ, Hirst AC, Stott P, Stouffer RJ, Tett SFB (2004) Simple indices of global climate variability and change: Part II—Attribution of climate change during the 20th century. Clim Dynam 22:823–838

Christy JR, McNider RT (1994) Detecting global warming using a precise but short (15-year) satellite record. Nature 367:325

Cubasch U, Santer BD, Hellbach A, Hegerl G, Höck H, Maier-Reimer E, Mikolajewicz U, Stössel A, Voss R (1994) Monte Carlo climate change forecasts with a global coupled ocean-atmosphere model. Clim Dynam 10:1–19

Cubasch U, Meehl GA, Boer GJ, Stouffer RJ, Dix M, Noda A, Senior CA, Raper S, Yap KS (2001) Projections of future climate change. In: Houghton JT et al (eds) Climate change 2001, The Science of Climate Change. Cambridge University Press, Chap. 9, pp 527–582

Delworth TL, Knutson TR (2000) Simulation of early 20th century global warming. Science 287:2246–2250

Duffy PB, Doutriaux C, Fodor IK, Santer BD (2000) Effects of missing data on estimates of near-surface temperature change since 1900. J Clim 14:2809–2814

Friis-Christensen E, Lassen K (1991) Length of the solar cycle: an indicator of solar activity closely associated with climate. Science 254:698–700

Gillett NP, Allen MR, Tett SFB (2000) Modelled and observed variability in atmospheric vertical temperature structure. Clim Dynam 16:49–61

Gillett NP, Zwiers FW, Weaver AJ, Hegerl GC, Allen MR, Stott PA (2002a) Detecting anthropogenic influence with a multi-model ensemble. Geophys Res Lett 29, doi:10.1029/2002GL015836

Gillett NP, Hegerl GC, Allen MR, Stott PA, Schnur R (2002b) Reconciling two approaches to the detection of anthropogenic influence on climate. J Clim 15:326–329

Gillett NP, Weaver AJ, Zwiers FW, Flannigan MD (2004a) Detecting the effect of climate change on Canadian forest fires. Geophys Res Lett 29, doi:10.1029/2002GL015836

Gillett NP, Wehner MF, Tett SFB, Weaver AJ (2004b) Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. Geophys Res Lett 31, doi:10.1029/2004GL020111

Haigh JD (1996) The impact of solar variability on climate. Science 272:981–984

Hansen J, Russell G, Lacis A, Fung I, Rind D, Stone PH (1985) Climate response times: dependence on climate sensitivity and ocean mixing. Science 229:857–859

Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In: Shawn T (ed) Meteorology of tropical oceans. Royal Meteorological Society, pp 251–259

Hasselmann K (1993) Optimal fingerprints for the detection of time dependent climate change. J Clim 6:1957–1971

Hasselmann K (1997) On multifingerprint detection and attribution of anthropogenic climate change. Clim Dynam 13:601–611

Haywood JM, Stouffer R, Wetherald R, Manabe S, Ramaswamy V (1997) Transient response of a coupled model to estimated changes in greenhouse gas and sulphate concentrations. Geophys Res Lett 24:1335–1338

Hegerl G, Hasselmann K, Cubasch U, Mitchell JFB, Roeckner E, Voss R, Waszkewitz J (1997) On multi-fingerprint detection and attribution of greenhouse gas and aerosol forced climate change. Clim Dynam 13:613–634

Hegerl GC, Allen MR (2002) Origins of model-data discrepancies in optimal fingerprinting. J Clim 15:1348–1356

Hegerl GC, North GR (1997) Comparison of statistically optimal approaches to detecting anthropogenic climate change. J Clim 10:1125–1133

Hegerl GC, von Storch H, Hasselmann K, Santer BD, Cubasch U, Jones PD (1996) Detecting greenhouse gas-induced climate change with an optimal fingerprint method. J Clim 9:2281–2306

Hegerl GC, Stott PA, Allen MR, Mitchell JFB, Tett SFB, Cubasch U (2000) Optimal detection and attribution of climate change: sensitivity of results to climate model differences. Clim Dynam 16:737–754

Hegerl GC, Crowley TJ, Baum SK, Kim KY, Hyde WT (2003) Detection of volcanic, solar and greenhouse gas signals in paleo-reconstructions of Northern Hemispheric temperature. Geophys Res Lett 30:1242–1245

Hegerl GC, Crowley TJ, Allen MR, Hyde WT, Pollack HN, Smerdon J, Zorita E (2006) Detection of human influence on a new, validated 1,500 year temperature reconstruction. J Clim to appear

Houghton JT, Filho LG Meiro, Griggs DJ (eds) (1996) K Maskell. Climate Change 1995: The Science of Climate Change. Cambridge University Press

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, (eds) (2001) CA Johnson. Climate Change 2001: The Science of Climate Change. Cambridge University Press

Hoyt DV, Schatten KHA (1993) discussion of plausible solar irradiance variations. J Geophys Res 98:18895–18906

Huntingford C, Stott PA, Allen MR, Lambert FH (2006) Incoporating model uncertainty into attribution of observed temperature change. Geophys Res Lett 33, doi:10.1029/2005GL024831

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The Second Hadley Centre coupled ocean-atmosphere GCM: model description, spin-up and validation. Clim Dynam 13:103–134

Jones PD (1994) Hemispheric surface air temperature variations: a reanalysis and update to 1993. J Clim 7:1794–1802. updated on http://www.cru.uea.ac.uk/cru/data/temperat.htm.

Jones PD, Hegerl GC (1998) Comparisons of two methods of removing anthropogenically related variability from the near-surface observational temperature field. J Geophys Res 103:13777–13786

Karoly DJ, Braganza K (2005) Attribution of recent temperature changes in the Australian region. J Clim 18:457–464

Karoly DJ, Braganza K, Stott PA, Arblaster JM, Meehl GA, Broccoli AJ, Dixon KW (2003) Detection of a human influence on North American climate. Science 302:1200–1203

Knutson TR, Delworth TL, Dixon KW, Stouffer RJ (1999) Model assessment of regional surface temperature trends (1949–1997). J Geophys Res 104:30981–30996

Lean J, Beer J, Bradley R (1995) Reconstruction of solar irradiance since 1600: implications for climate change. Geophys Res Lett 22:2319–2322

Leroy S (1998) Detecting climate signals, some Bayesian aspects. J Clim 11:640–651

Meehl GA, Gent PR, Arblaster JM, Otto-Bliesner BL, Brady EC, Craig A (2001) Factors that affect the amplitude of El Ninoin global coupled climate models. Clim Dynam 17:515–526

Meehl GA, Washington WM, Wigley TML, Arblaster JM, Dai A (2003) Solar and greenhouse gas forcing and climate response in the 20th century. J Climate 16:426–444

Meehl GA, Washington WM, Ammann CM, Arblaster JM, Wigley TML, Tebaldi C (2004) Combinations of natural and anthropogenic forcings in 20th century climate. J Climate 17:3721–3727

Mitchell JFB, Karoly DJ, Hegerl GC, Zwiers FW, Allen MR, Marengo J (2001) Detection of climate change and attribution of causes. In: Houghton JT et al (eds) Climate Change 2001, The science of climate change. Cambridge University Press, Chap. 12, pp 697–738

North GR, Stevens MJ (1998) Detecting climate signals in the surface temperature record. J Clim 11:563–577

North GR, Wu Q (2001) Detecting climate signals using space-time EOFs. J Clim 14:1839–1863

North GR, Bell TL, Cahalan RF, Moeng FJ (1982) Sampling errors in the estimation of Empirical Orthogonal Functions. Mon Weather Rev 110:699–706

North GR, Kim KY, Shen SSP, Hardin JW (1995) Detection of forced climate signals, 1: filter theory. J Clim 8:401–408

Parker DE (2004) Climate: large-scale warming is not urban. Nature 432:doi:10.1038/432290a

Parker DE, Jones PD, Folland CK, Bevan A (1994) Interdecadal changes of surface temperature since the late nineteenth century. J Geophys Res 99:14373–14399

Penner JE, Wigley TML, Jaumann P, Santer BD, Taylor KE (1997) Anthropogenic aerosols and climate change: a method for calibrating climate forcing. In: Howe W, Henderson-Sellers A (eds) Communicating about climate: the story of the model evaluation consortium for climate assessment. Amsterdam The Netherlands: Gordon and Breach Science Publishers, pp 91–111

Priestley MB (1981) Spectral analysis and time series. vol 1. Academic Press

Ramaswamy V, Boucher O, Haigh J, Hauglaustaine D, Haywood J, Myhre G, Nakajima T, Shi GY, Solomon S (2001) Radiative forcing of climate change. In: Houghton JT, Ding Y, Griggs D J, Noguer M, van der Linden P, Dai X, Maskell K, Johnson CI (eds) Climate change (2001): The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press

Ripley BD, Thompson M (1987) Regression techniques for the detection of analytical bias. Analyst 112:377–383

Robock A (2000) Volcanic eruptions and climate. Rev Geophys 38:191–219

Roeckner E, Bengtsson L, Feichter J, Lelieveld J, Rodhe H (1999) Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. J Clim 12:3004–3032

Santer BD, Mikolajewicz U, Bröggemann W, Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Wigley TML (1995) Ocean variability and its influence on the detectability of gr eenhouse warming signals. J Geophys Res 100:10693–10725

Santer BD, Wigley TML, Barnett TP, Anyamba E (1996a) Detection of climate change and attribution of causes. In: Houghton JT et al (eds) Climate Change 1995, The science of climate change. Cambridge University Press, Chap. 8, pp 411–443

Santer BD, Wigley TML, Doutriaux C, Boyle JS, Hansen JE, Jones PD, Meehl GA, Roeckner E, Sengupta S, Taylor KE (2001) Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. J Geophys Res 106(D22):28033–28059

Santer BD, Wehner MF, Wigley TML, Sause R, Meehl GA, Taylor KE, Ammann C, Arblaster J, Washington WM, Boyle JS, Brüggemann W (2003) Contributions of anthropogenic and natural forcing to recent tropopause height changes. Science 301:479–482

Santer BD, Taylor KE, Wigley TML, Johns TC, Jones PD, Karoly DJ, Mitchell JFB, Oort AH, Penner JE, Ramaswamy V, Schwarzkopf MD, Stouffer RJ, Tett SA (1996b) Search for human influences on the thermal structure of the atmosphere. Nature 382:39–46

Sato M, Hansen JE, McCormick MP, Pollack JB (1993) Stratospheric aerosol depths (1850–1990). J Geophys Res 98:22987–22994

Smith RL, Wigley TML, Santer BDA (2003) bivariate time series approach to anthropogenic trend detection in hemispheric mean temperatures. J Clim 16:1228–1240

Stone DA, Allen MR, Stott PAA (2006) multi-model update on the detection and attribution of global surface warming. J Climate (in press)

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

Stott PA, Tett SFB (1998) Scale-dependent detection of climate change. J Clim 11:3282–3294

Stott PA, Tett SFB, Jones GS, Allen MR, Mitchell JFB, Jenkins GJ (2000) External control of twentieth century temperature by natural and anthropogenic forcings. Science 290:2133–2137

Stott PA, Tett SFB, Jones GS, Allen MR, Ingram WJ, Mitchell JFB (2001) Attribution of twentieth century climate change to natural and anthropogenic causes. Clim Dynam 17:1–21

Stott PA, Allen MR, Jones GS (2003) Estimating signal amplitudes in optimal fingerprinting II: application to general circulation models. Clim Dynam 21:493–500

Stott PA, Stone DA, Allen MR (2004) Human contribution to the European heatwave of 2003. Nature 432:610–614

Stott PA, Mitchell JFB, Allen MR, Delworth TL, Gregory JM, Meehl GA, Santer BD (2006) Observational constraints on past attributable warming and predictions of future global warming. J Clim 19:3055–3069

Svensmark H, FriisChristensen E (1997) Variation of cosmic ray flux and global cloud coverage—a missing link in solar-climate relationships. J Atmos Sol-Terr Phys 20:1225–1232

Tett SFB, Mitchell JFB, Parker DE, Allen MR (1996) Human influence on the atmospheric vertical temperature structure: detection and observations. Science 247:1170–1173

Tett SFB, Johns TC, Mitchell J (1997) Global and regional variability in a coupled AOGCM. Clim Dynam 13:303–323

Tett SFB, Stott PA, Allen MR, Ingram WJ, Mitchell JFB (1999) Causes of twentieth century temperature change near the Earth's surface. Nature 399:569–572

Tett SFB, Jones GS, Stott PA, Hill DC, Mitchell JFB, Allen MR, Ingram WJ, Johns TC, Johnson CE, Jones A, Roberts DL, Sexton DMH, Woodage MJ (2002) Estimation of natural and anthropogenic contributions to twentieth century temperature change. J Geophys Res 107 (D16), art. no. 4306

van Huffel S, Vanderwaal J (1994) The total least squares problem: computational aspects and analysis. SIAM

Washington WM, Weatherly JW, Meehl GA, Semtner AJ, Bettge TW, Craig AP, Strand WG, Arblaster J, Wayland VB, James R, Zhang Y (2000) Parallel climate model (PCM) control and transient simulations. Clim Dynam 16:755–774

Weare BC, Nastrom JN (1982) Examples of extended empirical orthogonal function analyses. Mon Weather Rev 110:481–485

Wilks DS (1997) Resampling hypothesis tests for autocorrelated fields. J Clim 10:65–82

Zhang X, Zwiers FW, Stott PA (2006) Multi-model multi-signal climate change detection at regional scale. J Clim 19:4294–4307

Zwiers FW, Zhang X (2003) Toward regional scale climate change detection. J Clim 16:793–797