

A Multimodel Update on the Detection and Attribution of Global Surface Warming

DÁITHÍ A. STONE

Department of Physics, and Department of Zoology, University of Oxford, Oxford, United Kingdom

MYLES R. ALLEN

Department of Physics, University of Oxford, Oxford, United Kingdom

PETER A. STOTT

Hadley Centre for Climate Prediction and Research, Reading, United Kingdom

(Manuscript received 10 June 2005, in final form 21 April 2006)

ABSTRACT

This paper presents an update on the detection and attribution of global annual mean surface air temperature changes, using recently developed climate models. In particular, it applies a new methodology that permits the inclusion of many more general circulation models (GCMs) into the analysis, and it also includes more recent observations. This methodology involves fitting a series of energy balance models (EBMs) to the GCM output in order to estimate the temporal response patterns to the various forcings.

Despite considerable spread in estimated EBM parameters, characteristics of model performance, such as the transient climate response, appear to be more constrained for each of the forcings. The resulting estimated response patterns are provided as input to the standard fingerprinting method used in previous studies. The estimated GCM responses to changes in greenhouse gases are detected in the observed record for all of the GCMs, and are generally found to be consistent with the observed changes; the same is generally true for the responses to changes in stratospheric aerosols from volcanic eruptions. GCM responses to changes in tropospheric sulfate aerosols and solar irradiance also appear consistent with the observed record, although the uncertainty is larger. Greenhouse gas and solar irradiance changes are found to have contributed to a best guess of ~ 0.8 and ~ 0.3 K warming over the 1901–2005 period, respectively, while sulfate aerosols have contributed a ~ 0.4 K cooling. This analysis provides an observationally constrained estimate of future warming, which is found to be fairly robust across GCMs. By 2100, a warming of between about 1.5 and 4.5 K can be expected according to the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES) A1B emissions scenario.

These results indicate an emerging constraint for global mean surface temperature responses to external forcings across GCMs, which is corroborated in the observed record. This implies that observationally constrained estimates of past warming and predictions of future warming are indeed becoming robust.

1. Introduction

A major feature of the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) is the chapter on the detection and attribution of climate change (Mitchell et al. 2001). This chapter contains a synthesis of the first quantitative and probabilistic methods dealing with the subject and of their application to a number of dynamical climate models

available at the time. Since the publication of the report several more years of observations have been accumulated, the methodologies have been further developed, and new, more advanced general circulation models (GCMs) have been developed and used to produce more expansive ensembles of simulations of past climate. Combined, these factors provide a base for a useful update on the detection and attribution issue (International Ad Hoc Detection and Attribution Group 2005).

In the past several years, multisignal detection and attribution studies have been performed on surface air temperature (SAT) (Tett et al. 1999; Stott et al. 2000, 2001; Tett et al. 2002), ocean heat content (Barnett et al. 2001, 2005), tropopause height (Santer et al. 2003),

Corresponding author address: Dáithí A. Stone, AOPP, Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, United Kingdom.
E-mail: stoned@atm.ox.ac.uk

land precipitation (Lambert et al. 2004; Gillett et al. 2004), and Northern Hemisphere sea level pressure (Gillett et al. 2003). These studies find that observed changes are inconsistent with natural internally generated variability, but are at least partially consistent with externally forced changes simulated by the GCMs, although the degree of agreement varies (International Ad Hoc Detection and Attribution Group 2005). A general weakness of these studies is their dependence on a small number of GCMs, because only a few have been used to generate the multiple ensembles of simulations necessary for a multiple-detection study. Nevertheless, in theory the results of the optimal detection methodology used in these studies should be relatively independent of the GCM used because the response patterns tend to be similar across GCMs. There are now several GCMs worldwide that have been used to generate these necessary simulations, and Stott et al. (2006) find that they produce consistent detection and attribution results for large-scale SAT changes on decadal time scales.

Nevertheless, most GCMs have not been used to generate all of the simulations necessary for a multisignal attribution study. However, a couple of recently developed techniques get around this limitation for SAT (S. A. Crooks et al. 2006, unpublished manuscript, hereafter CR06; Stone et al. 2007), thus permitting a large additional number of GCMs to be included in the study of the attribution of observed SAT changes. In this paper we present an update on the attribution of global mean SAT using the new technique of Stone et al. (2007) applied to the recent suite of GCM simulations provided by the various modeling groups to the IPCC for analysis in the upcoming Fourth Assessment Report (AR4), along with simulations from older GCMs. We also include more recent observations, include higher-frequency data (annual rather than decadal), and separate the forcings.

2. Model and data

We analyze model SAT output from the historical simulations provided by the various modeling centers around the world to the IPCC Fourth Assessment Model Output database (<https://esg.llnl.gov:8443/index.jsp>). We also analyze SAT output from simulations of older GCMs examined extensively in the IPCC Third Assessment Report (McAvaney et al. 2001). The GCMs used in this study are listed in Table 1. Historical simulations are included only if the corresponding sections of the preindustrial control simulations have reached equilibrium, judged by lack of exceedance of a $0.2 \text{ K century}^{-1}$ trend. All models include the changing anthropogenic forcings (ANT) of greenhouse gases

(GHG) and tropospheric sulfate aerosols (SUL), while some also include the changing natural forcings (NAT) of stratospheric volcanic aerosols (VOL) and solar irradiance (SOL). We analyze simulations including ANT and NAT (ALL) separately to those omitting the natural forcings. Some GCMs do not include indirect effects of aerosols, so we may expect the response to SUL forcing to be smaller in their simulations (Table 1). Some GCMs also include other forcings, such as land surface changes and stratospheric ozone depletion, but we ignore these factors because they have a smaller net effect on the global mean surface forcing (Ramaswamy et al. 2001).

Figure 1a shows the global annual mean evolution of the four main radiative forcings from 1891 to the present provided by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) to the various modeling groups for the IPCC AR4 simulations. The SUL forcing is derived from the burden estimates of Boucher and Pham (2002), the VOL forcing follows C. M. Ammann et al. (2006, personal communication), and the SOL forcing follows Lean et al. (1995). Some modeling groups use different estimates, particularly for the natural forcings, reflecting some of the uncertainty in our knowledge. In addition, the radiative forcing estimates used here may not be identical to those seen by the GCM, depending, for instance, on how it converts emissions data into radiative forcings. The SUL forcing has been interpreted to include both the direct and indirect effects, even though some GCMs only include the direct effect. Forcing information has been converted to radiative forcing estimates as in Stone et al. (2007).

We analyze data over the 1901–2005 period. Some historical simulations finish before 2005, so in order to allow for comparison with observations from the past several years, we also include the first several years of the IPCC Special Report on Emission Scenarios (SRES) A1B scenario simulations of future climate with these GCMs where possible. The A1B scenario is selected because it is the most commonly simulated scenario in the IPCC AR4 database. ANT emissions have closely followed this scenario and no major volcanic eruptions have occurred; however, the constant SOL forcing usually applied for these few years may lead to a slight underestimate of the effect of solar forcing. This inclusion of the SRES A1B scenario cannot be applied to all GCMs and so some finish earlier than 2005 (Table 1). Deseasonalized monthly SAT anomalies from the GCM simulations are interpolated to the monthly $5^\circ \times 5^\circ$ gridded observational dataset of deseasonalized monthly SAT anomalies of Jones and Moberg (2003) and Rayner et al. (2003) according to

TABLE 1. GCMs included in the analysis. ANT refers to the inclusion of changing GHG and SUL forcing, while NAT refers to the inclusion of changing VOL and SOL forcing. Some GCMs with SUL forcing include the indirect effect as noted as ISU under the “Other ANT” heading, and some GCMs with ANT forcing include other anthropogenic forcings listed under “Other ANT,” which are ignored in this study. BCA, OZO, and LAN denote the inclusion of changes in black carbon emissions, stratospheric ozone, and land cover, respectively; brackets imply only some of the simulations include these. The last year of the simulations varies between 1999 and 2005, depending on the GCM.

GCM	Group	Sims	Period	Forcings	Other ANT
CCSM3	National Center for Atmospheric Research (NCAR) (United States)	6	1901–2005	ANT, NAT	ISU, BCA, OZO
Coupled General Circulation Model version 3.1 (CGCM3.1) (T47)	Canadian Centre for Climate Modelling and Analysis (CCCma) (Canada)	5	1901–2005	ANT	
CGCM3.1 (T63)	CCCma (Canada)	1	1901–2005	ANT	
CNRM-Coupled Global Climate Model version 3 (CM3)	Centre National de Recherches Météorologiques (CNRM) (France)	1	1901–2005	ANT	
CSIRO-Mark version 2 (Mk2)	Commonwealth Scientific and Industrial Research Organisation (CSIRO) (Australia)	1	1901–2005	ANT	
ECHAM4-Ocean Isopycnal Model (OPYC3)	Max Planck Institut (MPI), Deutsches Klimarechenzentrum (DKRZ) (Germany)	1	1901–2000	ANT, NAT	ISU, LAN
ECHAM5/MPI-Ocean Model (OM)	MPI (Germany)	2 3	1901–2005 1901–2005	ANT ANT	ISU, OZO
ECHAM-Hamburg Ocean Primitive Equation-Global (ECHO-G)	University of Bonn (Ubonn) (Germany), Korea Meteorological Administration (KMA) (South Korea)	3	1901–2005	ANT, NAT	ISU
GFDL-CM2.0	GFDL, National Oceanic and Atmospheric Administration (NOAA) (United States)	3	1901–2000	ANT, NAT	BCA, OZO, LAN
GFDL-CM2.1	GFDL, NOAA (United States)	3	1901–2000	ANT, NAT	BCA, OZO, LAN
GFDL-R15a	GFDL, NOAA (United States)	1	1901–2005	ANT	
GFDL-R30	GFDL, NOAA (United States)	3 3	1901–98 1901–2005	ANT, NAT ANT	
GISS-Atmosphere–Ocean Model (AOM)	Goddard Institute for Space Studies (GISS), National Aeronautics and Space Administration (NASA) (United States)	2	1901–2005	ANT	
GISS-EH	GISS, NASA (United States)	5	1901–99	ANT, NAT	ISU, BCA, OZO
GISS-Model E–R (ER)	GISS, NASA (United States)	9	1901–2003	ANT, NAT	ISU, BCA, OZO
INM-Coupled Model version 3.0 (CM3.0)	Institute of Numerical Mathematics (INM) (Russia)	1	1901–2005	ANT, NAT	
IPSL-Coupled Model version 4 (CM4)	L’Institut Pierre-Simon Laplace (IPSL) (France)	1	1901–2000	ANT	ISU
Model for Interdisciplinary Research on Climate 3.2, medium-resolution version [MIROC3.2 (medres)]	Japan Agency for Marine-Earth Science and Technology (JAMSTEC) (Japan)	3	1901–2005	ANT, NAT	ISU, BCA, OZO, LAN
MRI-CGCM2.3.2	MRI (Japan)	5	1901–2005	ANT, NAT	
Parallel Climate Model (PCM)	NCAR (United States)	4 12	1901–99 1901–99	ANT, NAT ANT	OZO (OZO)
Met Office Second Hadley Centre Coupled Ocean–Atmosphere GCM (UKMO-HadCM2)	Hadley Centre, Met Office (United Kingdom)	4	1901–2005	ANT	
Met Office Third Hadley Centre Coupled Ocean–Atmosphere GCM (UKMO-HadCM3)	Hadley Centre, Met Office (United Kingdom)	4	1901–2002	ANT, NAT	ISU, OZO
		4	1901–99	ANT	ISU, OZO

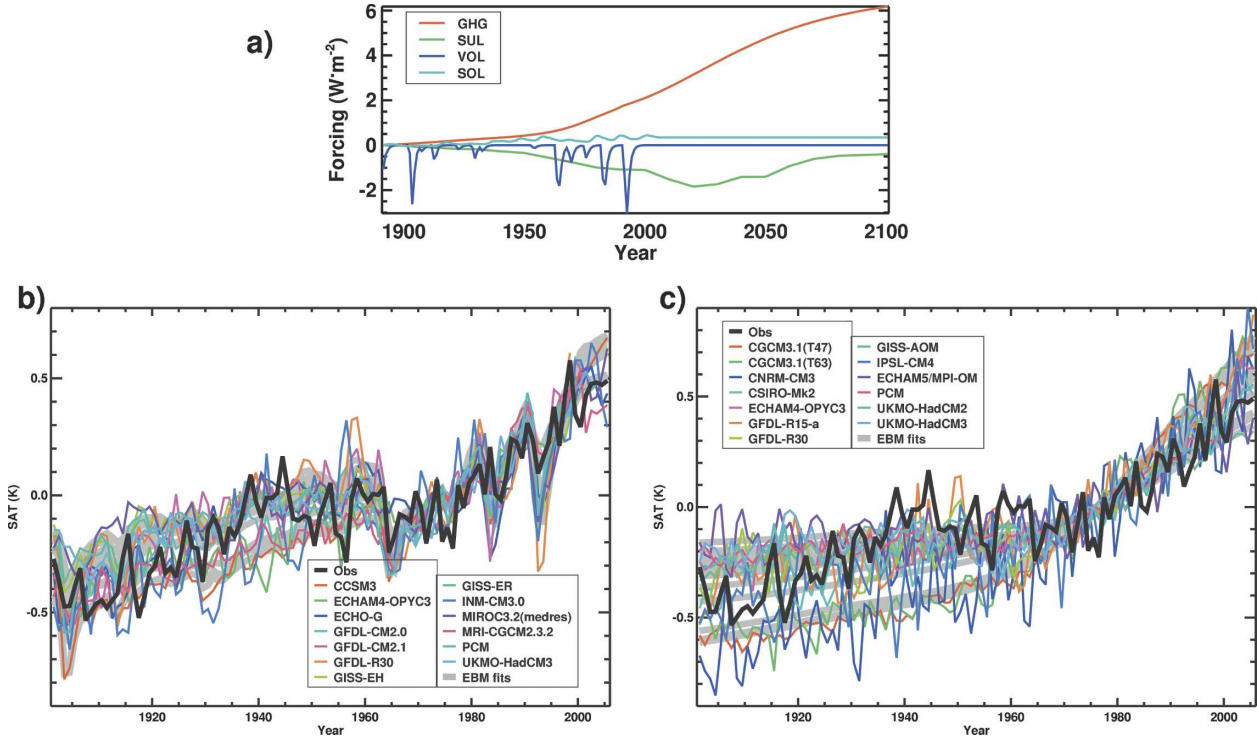


FIG. 1. (a) Global radiative forcing time series estimated from data compiled by PCMDI for use by modeling groups for the IPCC “20C3M” (historical) and “SRES A1B” (future) climate simulations covering 1891 through 2100. (b) Plot of global annual mean SAT time series from 1901 to 2005 for the 13 GCMs with ALL simulations. The ensemble mean is plotted for each GCM. The total EBM fits are shown in thick gray. The observed variations are in black. (c) As in (b), but for the 13 GCMs with ANT simulations.

the availability of observations. Annual global means are calculated for both datasets and are used in the subsequent analysis. Preindustrial and present-day control simulations are also used in this study for estimating the internally generated variability of the climate system. Altogether, these simulations are divided into 96 segments of 105-yr length and are masked onto the observations in the same manner as the historical simulations.

3. Method

Because we lack ensembles of simulations, including subsets of the forcings used in the historical simulations with each GCM, we apply the method of Stone et al. (2007) to extract the GCM responses to the individual forcings. This involves fitting a sum of n different energy balance models (EBMs) corresponding to the n forcings. These EBMs are of the form

$$c_i \frac{\partial T_i(t, z)}{\partial t} = F_i(t) - \lambda_i T_i(t, z) - k_i \frac{\partial^2 T_i(t, z)}{\partial z^2}. \quad (1)$$

Here $T_i(t, z)$ is the global annual mean temperature response as a function of depth in the ocean mixed layer to the evolving forcing $F_i(t)$ shown in Fig. 1a; c_i is

the heat capacity of the mixed layer, $(1/\lambda_i)$ is the climate sensitivity, and k_i is the vertical diffusion parameter in the mixed layer for the forcing i . These parameters are tuned to reproduce the ensemble mean SAT response of the GCM, $\mathbf{T}_i = T_i(t, 0)$. Supposing the responses add linearly, that is, $\mathbf{T} = \sum_{i=1}^n \mathbf{T}_i$, the parameters are tuned to minimize the squared difference between the total EBM time series \mathbf{T} and the mean response \mathbf{T}_{GCM} from the GCM ensemble. Uncertainty in this fit arises from the finite GCM ensemble size and the accuracy of the parameter-fitting algorithm. The EBMs are spun up with 10 yr of varying forcings before the start of the comparison in 1901.

The estimated individual EBM responses (\mathbf{T}_i) are then used in the standard multiple regression methodology (Allen and Tett 1999). Under this, we express observed temperature changes \mathbf{T}_{obs} as a linear sum of the simulated responses determined for each forcing (\mathbf{T}_i) plus a residual (\mathbf{v}_0),

$$\mathbf{T}_{\text{obs}} = \sum_{i=1}^n \mathbf{T}_i \beta_i + \mathbf{v}_0. \quad (2)$$

The β_i s are scaling factors for the response to forcing i estimated in the regression that minimize the variance

of the residual term \mathbf{v}_0 . The regression analysis assumes that the GCMs correctly represent the temporal response patterns to the forcings, and that the EBMs adequately reproduce these patterns. However, these model responses may be incorrect in their estimate of the amplitude of that pattern. The regression is performed on the full masked global annual time series. While it would be nice to include spatial information as well, it is not clear how appropriate the EBM approximation would be at nonglobal scales. This is a limitation of this study because spatial information can be important for detecting climate response signals (CR06; Stott et al. 2006). No optimization or data reduction is used because this is of limited applicability to the temporal data used here.

A control simulation is needed both to estimate the covariance of the residual term \mathbf{v}_0 and to estimate the uncertainty of the β_i scaling parameters. Because of the limited size of the control simulations of each GCM, we aggregate all of the control simulations from all of the GCMs to produce a single large ensemble of control simulations used throughout the analysis. All possible control simulation data are used, and while this implies that GCMs with longer control simulations are given higher weighting, Gillett et al. (2002) find this preferable to using a more limited equal portion from each GCM. Each control is divided into 105-yr-long segments, with each segment being treated identically to the ALL and ANT simulations (e.g., being masked to the observations). Fifty-two of these control segments are used to estimate the covariance of \mathbf{v}_0 , while 44 others are used for an independent estimate of the covariance that is used in estimating the distributions of the β_i s (Allen and Tett 1999). Each of these two sets includes at least one member from each GCM. The regression formulation used here, often referred to as ordinary least squares, does not include any error in the estimate of the simulated responses arising from the finite ensemble size or model deficiencies. Because the EBM does not simulate internally generated variability, the uncertainty from the finite sample size cannot be readily included in an analytical regression framework. However, both of these extra sources of uncertainty can be accounted for through a Monte Carlo approach, which is conducted in section 7.

4. The EBM fits

The time series of ensemble and global mean SAT for each of the GCMs is shown in Figs. 1b,c. The observed SAT is well within the range covered by the GCMs with ALL simulations. The observed record shows a warming to 1940, followed by a stable period,

and then further warming after 1970. The GCM simulations show a similar evolution, although the initial warming event continues past 1940 until the early 1960s, with an abrupt cooling occurring with the eruption of Agung, and then the general warming continuing after that eruption. Responses to the eruptions of Santa María (1902), Agung (1963), El Chichón (1982), and Pinatubo (1991) (see Fig. 1a) are clearly visible in the simulations. The especially good agreement during the 1961–90 interval arises from the use of this period in defining the anomalies. Altogether, the spread of the GCMs with ANT simulations tends to cover the observed SAT record, but individual models clearly do not get the combination of the early century warming and midcentury cooling, as noted previously (Stott et al. 2000).

The total EBM fit for each GCM is also shown in Figs. 1b,c. Despite the lack of labeling on the EBM fits in the figure, the correspondence between fit and GCM ensemble is generally clear in areas where the GCMs diverge. The fits to the GCMs with ALL simulations show distinct responses to several large volcanic eruptions, which are absent in the fits to the ANT simulations.

The EBM parameter values estimated for each model are shown in Fig. 2. In general, the spread of parameter values is rather large (with some minuscule values outside the plotting range). Such a spread is also found in Stone et al. (2007) using a single GCM, implying that the spread arises more from problems with the fitting than with differences in GCM properties. The large range of heat capacities may partly reflect model differences but also clearly indicates some uncertainty in the EBM fit procedure, because it seems improbable that such a large range would actually exist across GCMs. The estimated values of the vertical diffusion parameter k_i show a similarly large spread as that for c_i . There are some consistent differences in climate sensitivity estimates, with SUL and VOL values always smaller than corresponding GHG values. The estimated values of the climate sensitivity (estimated from unmasked data, unlike in Fig. 2) are compared against diagnosed values for each of the GCMs in Table 2. Diagnosed values tend to center around $0.9 \text{ K W}^{-1} \text{ m}^2$, while estimated EBM values tend to center around $0.7 \text{ K W}^{-1} \text{ m}^2$ for ALL simulations and $1.0 \text{ K W}^{-1} \text{ m}^2$ for ANT simulations. The correspondence for individual GCMs is not as good, indicating that results for parameter values from this study should be interpreted in a general sense rather than on an individual basis.

Ultimately the quantity we most want to estimate accurately in the context of modeling transient climates is the transient climate response (TCR; Allen et al.

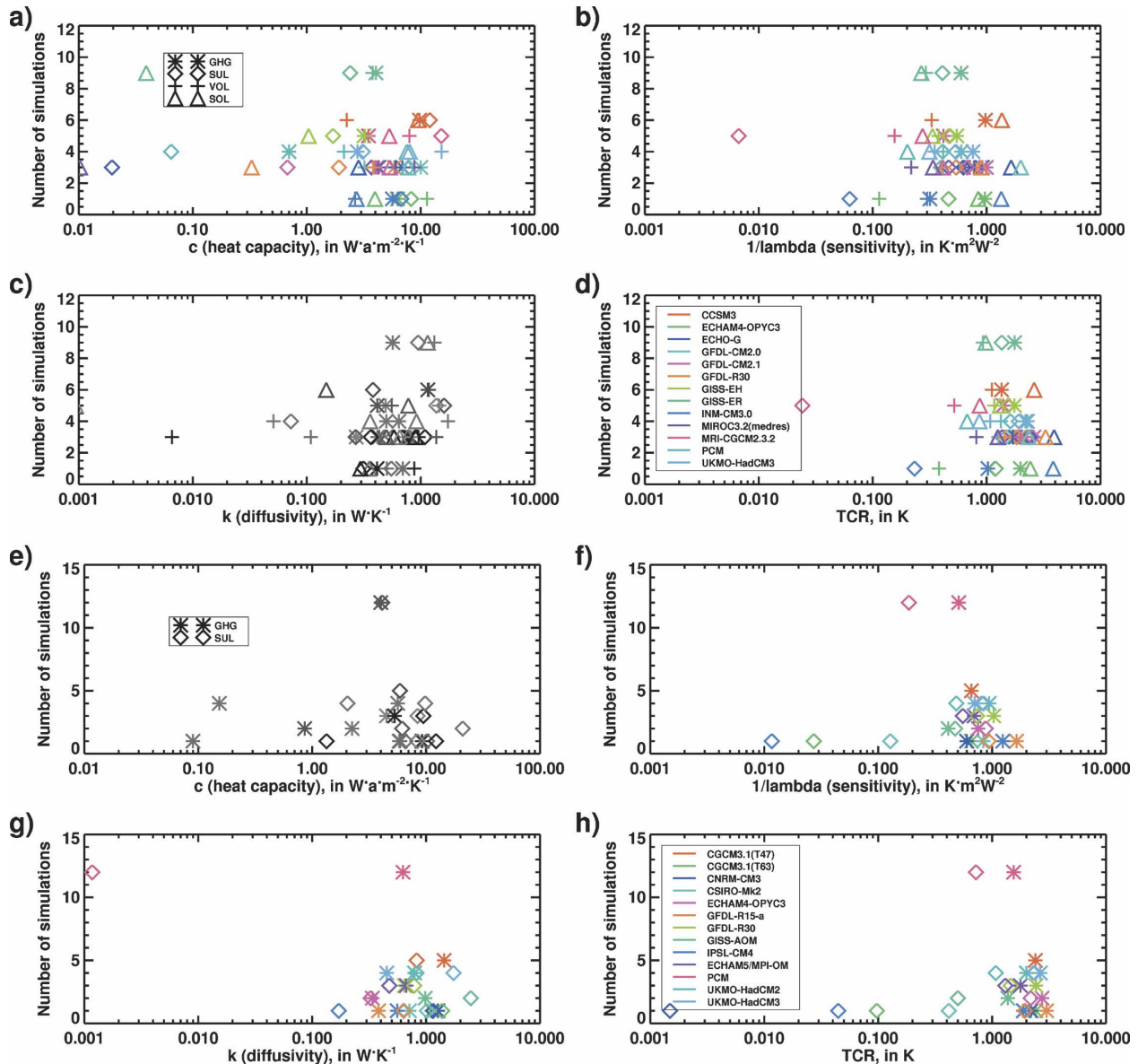


FIG. 2. (a)–(d) The EBM parameter values for each of the forcings and GCMs with ALL ensembles, which provide the best fit to the observed record: (a) c_i , the heat capacity of the ocean mixed layer for forcing i ; (b) $(1/\lambda_i)$, the climate sensitivity; (c) k_i , the parameter for vertical diffusion in the mixed layer; and (d) the TCRs resulting from these parameter sets. (e),(f) As in (a)–(d) but for GCMs with ANT ensembles.

2005). This is the temperature change after an increase in forcing equal to a doubling of CO_2 at a rate of increase in emissions of $1\% \text{ yr}^{-1}$ (over 70 yr). An estimate of the TCR is also shown in Figs. 2d and 2h. This quantity is more tightly constrained for some individual forcings than the other parameters, as expected (Allen et al. 2005; Stone et al. 2007). The GHG and VOL forcings have been strongest historically, and so we may expect these to have the most accurately determined parameters. While estimated parameters for these two

forcings vary widely, the estimates of the TCR vary by only about a factor of 4 and 2, respectively, between GCMs. The results of Stone et al. (2007) suggest that, unlike the EBM parameters, much of the spread in the TCR values may be due to actual differences in GCM properties. The ALL GCMs all have lower TCR estimates for VOL forcing than for GHG forcing, with SUL estimates also being lower than GHG for all but one GCM. The GHG parameter estimates from the ANT simulations are quite similar to those from the

TABLE 2. Diagnosed values for various GCMs of the climate sensitivity and TCR to GHG forcing and corresponding estimates from EBM surrogates for ensembles of simulations including ALL or ANT forcings. All diagnosed values of the climate sensitivity are from information supplied by the modeling groups to the IPCC Fourth Assessment Model Output database, except those marked with a *, which are from Cubasch et al. (2001). Diagnosed values of the climate sensitivity ($1/\lambda_{\text{GHG}}$) marked with a † are supplied by the modeling groups for a doubling of CO_2 and are converted to unit forcing assuming a CO_2 doubling forcing of 3.97 W m^{-2} (Ramaswamy et al. 2001). The diagnosed effective climate sensitivity is shown for ECHAM4 – OPYC3. All diagnosed values of the TCR are estimated from simulations in the IPCC AR4 database, which include a $1\% \text{ yr}^{-1}$ increase in GHG forcing from preindustrial levels to a doubling, except those marked with a *, which are from Cubasch et al. (2001). Estimates from EBM surrogates and diagnosed TCR estimates apply to unmasked data (unlike in the rest of the paper).

GCM	$\frac{1}{\lambda_{\text{GHG}}} (\text{K W m}^{-2})$			TCR _{GHG} (K)		
	Diagnosed	ALL	ANT	Diagnosed	ALL	ANT
CCSM3	0.8	1.1	—	1.4	1.6	—
CGCM3.1 (T47)	—	—	0.7	2.0	—	2.4
CGCM3.1 (T63)	—	—	0.8	—	—	3.0
CNRM-CM3	—	—	0.6	1.6	—	2.3
CSIRO-Mk2	*†1.1	—	0.8	*2.0	—	1.8
ECHAM4-OPYC3	*†0.7	0.6	1.2	*1.4	1.9	2.3
ECHAM5/MPI-OM	1.2	—	0.9	2.2	—	1.8
ECHO-G	†0.8	0.7	—	2.1	1.3	—
GFDL-CM2.0	†0.7	1.1	—	1.5	2.4	—
GFDL-CM2.1	†0.7	1.0	—	1.5	2.4	—
GFDL-R15a	*†0.9	—	1.3	2.3	—	2.7
GFDL-R30	*†0.9	0.7	1.5	*2.0	1.9	2.3
GISS-AOM	—	—	0.7	—	—	1.3
GISS-EH	†0.7	0.5	—	1.6	1.8	—
GISS-ER	†0.7	0.6	—	0.4	1.9	—
INM-CM3.0	0.5	0.3	—	1.7	0.9	—
IPSL-CM4	1.3	—	1.3	2.1	—	2.0
MIROC3.2 (medres)	1.3	0.4	—	2.1	1.4	—
MRI-CGCM2.3.2	0.9	0.4	—	2.3	1.4	—
PCM	*†0.5	0.7	1.0	1.4	2.9	1.7
UKMO-HadCM2	*†1.0	—	0.9	*1.7	—	1.8
UKMO-HadCM3	*†0.8	0.8	1.0	2.2	2.2	3.9

ALL simulations. The comparison of the SUL parameters is messier though, suggesting that this method works best with simulations including more forcings. A comparison of diagnosed TCR values for GHG forcing and values estimated from the EBMs is shown in Table 2 (both diagnosed estimates and estimates from the EBMs in the table are from unmasked data). Despite the fact that both estimates of TCR_{GHG} have similar and relatively small ranges, there is still little correspondence between individual estimated EBM values and diagnosed estimates from GCM simulations. Such discrepancies may result from differences in the radiative forcings used for the EBMs and GCMs or in the difficulty in finding the 12 (or 6) -dimensional best-fit parameter set with only 105 temporal data points. In any case, it appears that results from this application of this methodology are better interpreted in a general sense rather than on an individual basis.

The EBM surrogates estimated here exhibit a wide range of properties that do not always individually match with diagnosed values. The primary aim of this

paper is an overall attribution statement encompassing many plausible responses as encapsulated by many GCMs rather than attribution statements for individual GCMs. Thus, because properties of the surrogate models and the diagnosed properties of the GCMs agree in a general sense, these surrogate models appear to be adequate for this study.

5. The detection

With these estimates of the response components to each of the forcings, we can now proceed with the standard optimal detection methodology (Allen and Tett 1999). The estimates of the 90% confidence intervals on the scalings (β_i) required to fit the observations are shown in Figs. 3a–b. These confidence intervals account for uncertainty in the observations but not in the estimated GCM responses.

The scaling factor for GHG is inconsistent with zero for all of the ALL GCMs, and is consistent with one for all but one of them. This indicates that this forcing is

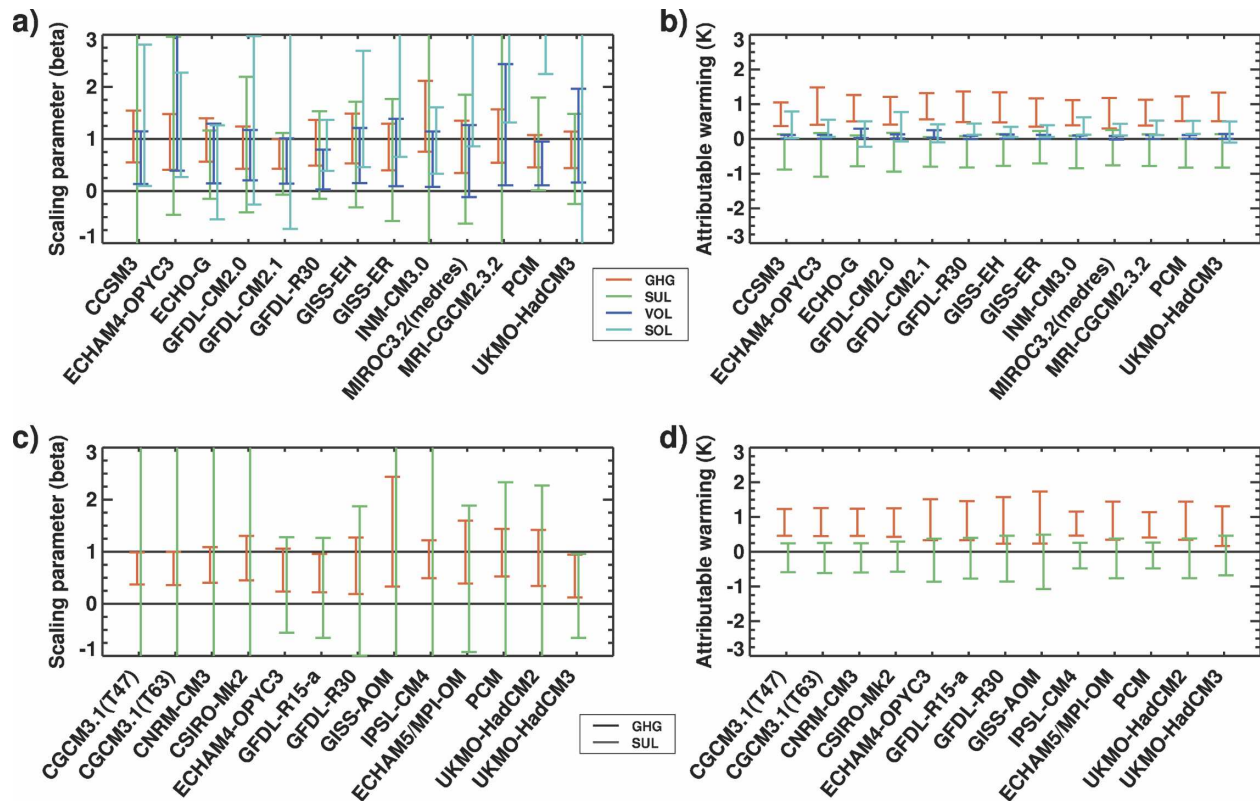


FIG. 3. (a) The 90% confidence intervals of the amplitude scalings (β_i) derived from the regression of observed SAT changes onto the EBM estimates of the GCM response patterns to each of the forcings. Results are shown for the GCMs with ALL simulations. (b) The resulting 90% confidence intervals of the estimated SAT difference over the 1996–2005 period relative to the 1901–10 period that is attributable to each of the forcings. (c)–(d) As in (a)–(b), but for the GCMs with ANT simulations.

necessary to reproduce observed changes and that the amplitude of the response in the GCM simulations is generally consistent with what has been observed. The scaling factor for VOL is also generally inconsistent with zero and generally consistent with one, but there does appear to be a general tendency of the GCMs to overestimate the response amplitude. The uncertainty is much larger on the SUL scalings, and so most are consistent both with zero and one, meaning that no detection is possible. This larger uncertainty is not surprising because the SUL response is partly degenerate with the GHG response in the temporal output examined here. The scaling factors for SOL are usually more uncertain but nevertheless two-thirds are inconsistent with zero. It is difficult to draw any conclusions about the SOL response, a situation that probably arises in part because it is the smallest-amplitude forcing examined here and in part because it is the forcing most likely to have been included in the GCM simulations with a different historical scenario than that which is assumed here (Fig. 1a). Nevertheless, two of the four GCMs that use the Lean et al. (1995) forcing scenario [Geophysical Fluid Dynamics Laboratory (GFDL) Cli-

mate Model version 2.0 (CM2.0), GFDL CM version 2.1 (CM2.1), Meteorological Research Institute (MRI) Coupled Atmosphere–Ocean General Circulation Model version 2.3.2 (CGCM2.3.2), Community Climate System Model version 3 (CCSM3)] have no detectable SOL response.

All of the ANT GCMs have a detectable GHG response, that is, they have scalings inconsistent with zero. While many of the GHG scalings are consistent with one, many ANT GCMs appear to be overestimating the GHG response, and some significantly so. The amplitudes of the SUL scalings, on the other hand, are very poorly constrained. As noted above when examining parameters, the performance of this methodology appears more robust when more forcings are included in the simulations.

With these scaling factors, we can now estimate the amount of SAT change occurring over the historical record that is attributable to the various forcings (Figs. 3c–d). Because the responses generally do not follow a linear trend through time, we estimate these attributable warmings as the difference between the last 10 yr of the EBM responses (1996–2005) and the first 10 yr

(1901–10), scaled by the appropriate scaling factors. All of the ALL and ANT GCMs attribute some warming to GHG forcing, with an average best guess of about 0.8 K. The ALL GCMs also attribute somewhat less warming to SOL, around 0.3 K. Because of the 1902 eruption of Santa María, a small warming of about 0.1 K is attributable to the VOL forcing. Both the ALL and ANT GCMs tend to attribute some cooling (~ 0.4 K) to SUL forcing, but they are generally also consistent with no SUL response.

The residual variability in the observed SAT (\mathbf{v}_0) is consistent at the 10% level with the internally generated variability in the control simulations for all of the ALL GCMs. With the ANT GCMs the residuals are also all consistent with the internal variability but are all only just so, apparently because of their failure to remove NAT responses. We can also examine the variance of the simulation output and of the residuals at various frequencies, as shown in Fig. 4. At interdecadal time scales, the variability in the ALL GCM simulations is generally consistent with the observed variability. This is the case with fewer of the ANT GCMs, however. This total variability includes both externally forced and internally generated components. We have estimated the externally forced component above, so by subtracting the best guess of this externally forced response (i.e., EBM fits for the simulations, and EBM fits scaled according to the best guess of β_i for the observations) from the total variability we can get estimates of the internally generated variability corresponding to each GCM and in the real world. This interdecadal residual is usually consistent between the observations and ALL GCMs, with exceptions related to an overestimate or underestimate of the centennial-scale trend. Exceptions with the ANT GCMs appear to be more due to the absence of the NAT forcings, because the major discrepancies at the 10- and 30-yr time scales correspond to the general time scale between volcanic eruptions during the twentieth century.

6. Constrained predictions of future climate

Allen et al. (2000) note that the scalings estimated through detection and attribution studies effectively provide observational constraints on GCM predictions of future climate change. A major assumption in this standard attribution methodology is that the response of a GCM to a change in a given forcing is independent of the climate state. Thus, the scalings correct for overestimates or underestimates in the response amplitudes but assume that the GCM has the correct response pattern. If we extend this assumption to future forcing values then effectively the estimated scaling factors

provide an observationally based correction to GCM predictions of future climate change. The 90% confidence interval on the adjusted estimates of past and future climate change is shown in Fig. 5a,b. The especially tight intervals over the 1961–90 period arise from its use as the climatological base period.

Over the historical period the scaled ALL GCM predictions overlap very closely. This indicates that the temporal response pattern is similar across models and so suggests that this pattern is robust and that its amplitude is being effectively constrained by the observations. The ANT GCMs similarly closely agree on past changes. In the future all predictions using both the ALL and ANT GCMs continue to follow each other fairly closely, in comparison with their own uncertainties, under the IPCC A1B scenario, although not surprisingly more divergence occurs between GCMs. We choose to assume this scenario because it tends to be favored by the IPCC AR4 modeling groups. This agreement continues through 2100 at which time the lower bound on the 90% confidence range is about 1.0–2.0 K while the upper bound is about 3.5–5.0 K. The general agreement between GCM predictions is expected and supports the GCM independence of recent predictions of future climate using the observational constraints (Allen et al. 2000; Stott and Kettleborough 2002; Stone et al. 2007; Stott et al. 2006). This arises because the GCMs are only needed for determining the response pattern; as long as this pattern is robust, the amplitude is constrained directly by the observed past response and is thus independent of the GCM used.

7. Results from a superensemble across GCMs

The analysis carried out so far ignores three potentially important sources of uncertainty. First, it does not account for uncertainty arising from the use of a finite number of GCM simulations to characterize the GCM response. It also does not account for uncertainty arising from inaccuracies in the EBM fitting procedure. Finally, by assuming for each GCM that the GCM response pattern is identical to the response pattern in the real world, the results also assume that this pattern is identical across models even though we are in fact using different patterns for different models. In this section we attempt to characterize these additional sources of uncertainty by using a Monte Carlo technique similar to that used by Stone et al. (2007). Stone et al. (2007) use output from a single GCM; thus, the application of the methodology here differs in that it is also taking into account so-called structural uncertainty in the GCMs and the arising differences in their response patterns.

All of the ALL simulations are pooled together to

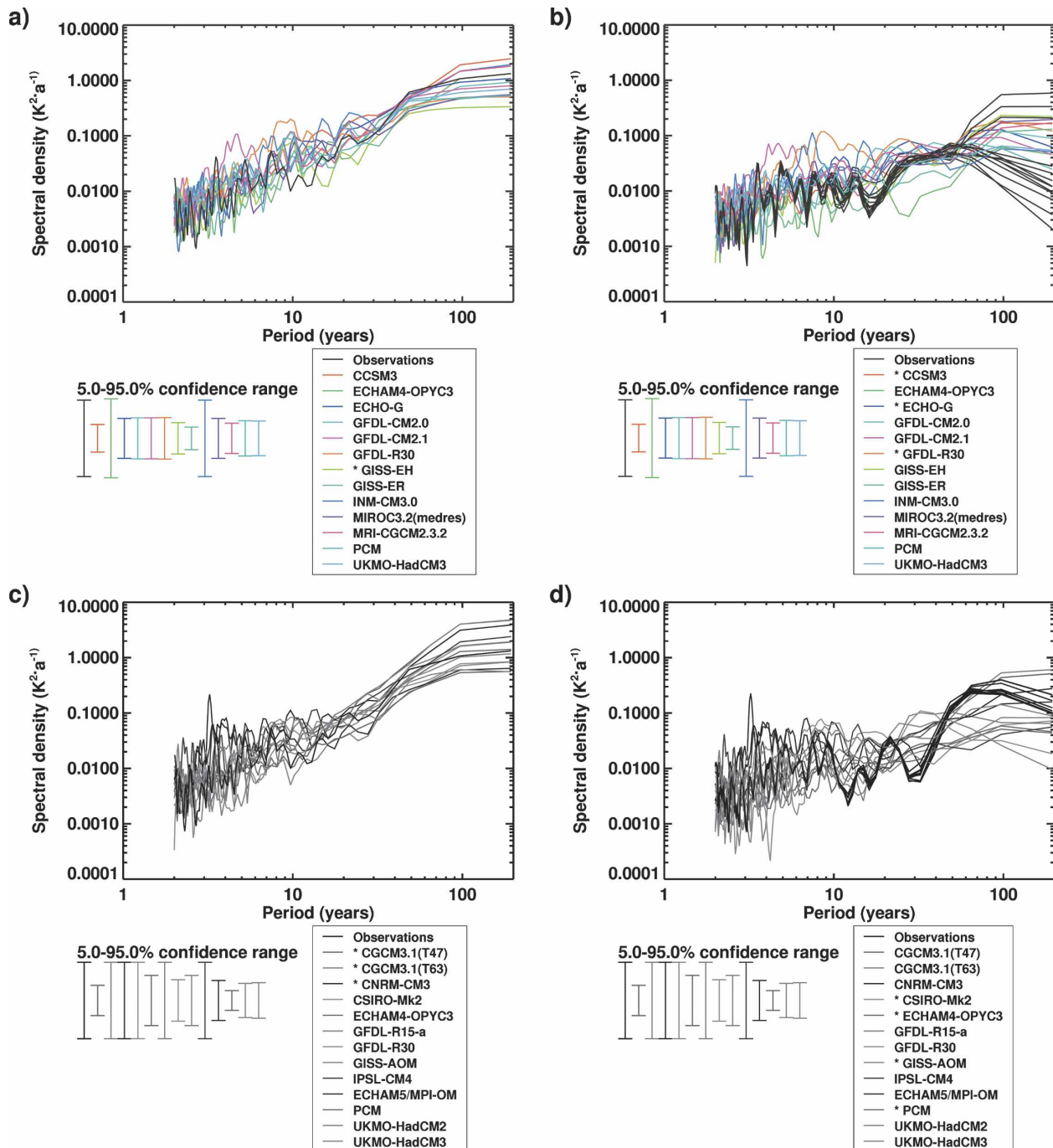


FIG. 4. (a),(b) The spectra of annual global mean SAT from the observed record and the ensemble mean spectra of the individual GCM ALL simulations: (a) without and (b) with removal of the externally forced signal. The signal removed in the estimation of each GCM spectrum is the total EBM response estimated for that GCM, while the signal removed from the observations is the EBM response scaled according to the best-guess scalings estimated in Fig. 3b for that GCM. GCMs marked by asterisks significantly overestimate or underestimate the variability at time scales of 10 yr and longer at the 10% level, as estimated using an F test on the variability integrated over these low frequencies. All spectra are estimated using a Hanning filter of width 97 yr. (c),(d) As in (a),(b), but for the GCMs with ANT simulations.

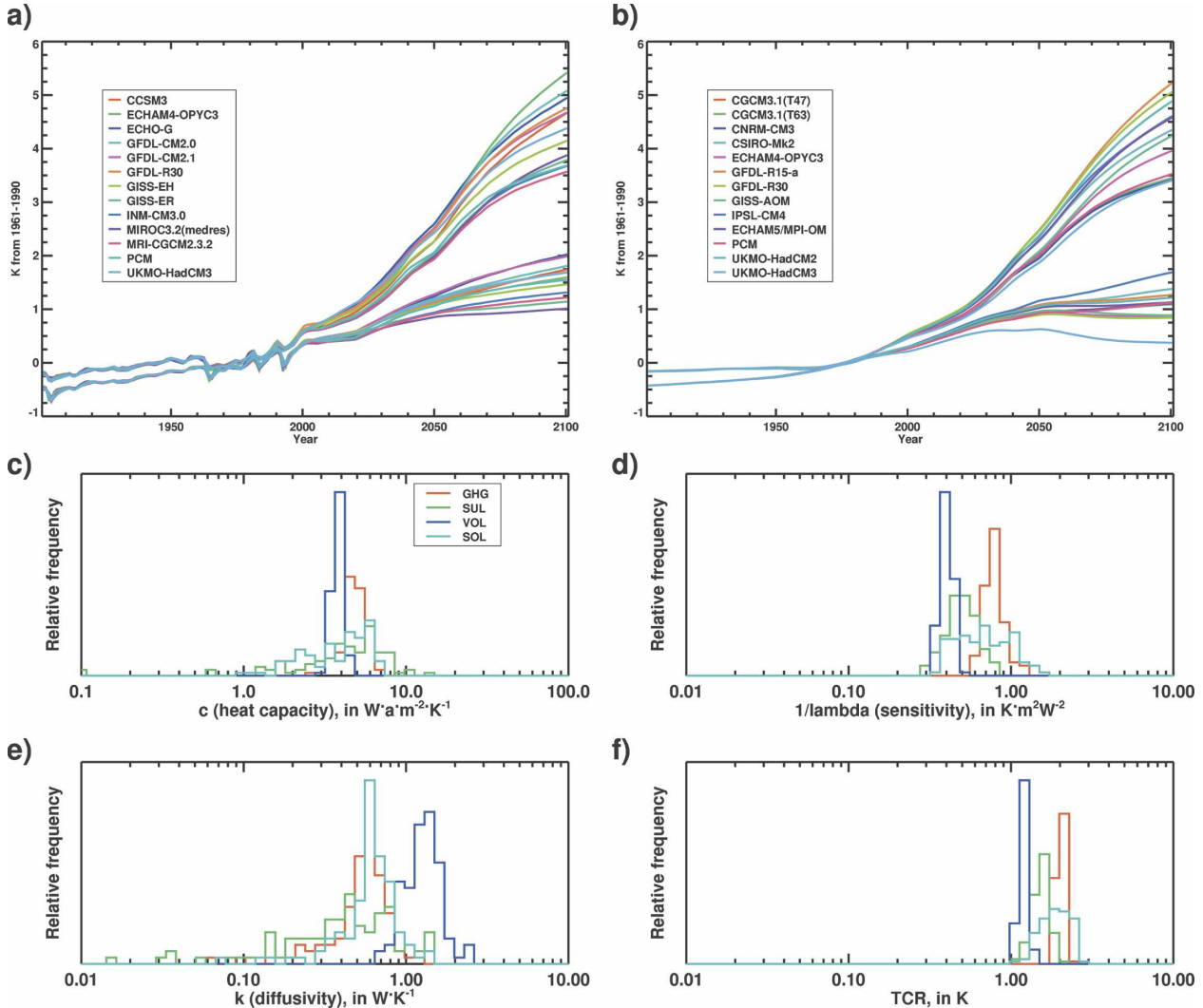


FIG. 5. The estimated 90% confidence interval in the global annual mean SAT climate from 1901 to present and through to 2100 assuming the SRES A1B emissions scenario and constant NAT forcing. These estimates use the EBM fits and scalings derived for the GCMs with (a) ALL ensembles and (b) ANT ensembles. (c)–(f) Similar to Fig. 2, but lumping all of the ALL simulations from all of the GCMs together and showing histograms of the EBM parameters and TCR. The histograms are estimated from 100 Monte Carlo samples where 117 simulations are selected randomly with replacement from a pool of 50 simulations. All 13 GCMs are weighted equally; thus, some of the 50 simulations are more likely to be selected than others.

create a superensemble. We arbitrarily decide to treat all GCMs equally, so some simulations have a higher likelihood of being included. Ensembles of 117 members of these 50 simulations are selected with replacement and are used as input into the analysis as carried out in the previous sections. The ensemble size of 117 is used, because this is the number required to ensure that all simulations could be included once while preserving the even weighting across GCMs. This selection of ensembles is repeated 100 times, allowing the estimation of probability distributions for the various quantities examined that account for the additional sources of uncertainty.

The resulting histograms of EBM parameters and of the TCR are shown in Figs. 5c–f. As in the earlier figure, in general, the VOL and GHG quantities are better constrained than those for the other forcings. The VOL and GHG histograms also tend to differ with very little overlap. The tendency in Fig. 2 for the climate sensitivity and TCR for VOL forcing to be smaller than that for GHG forcing is also visible here.

The distributions of the scaling parameters in the multiple regressions using the superensemble are shown in Fig. 6. Results from a sensitivity analysis in Stone et al. (2007) suggest that the detection and attribution results obtained using the methodology used

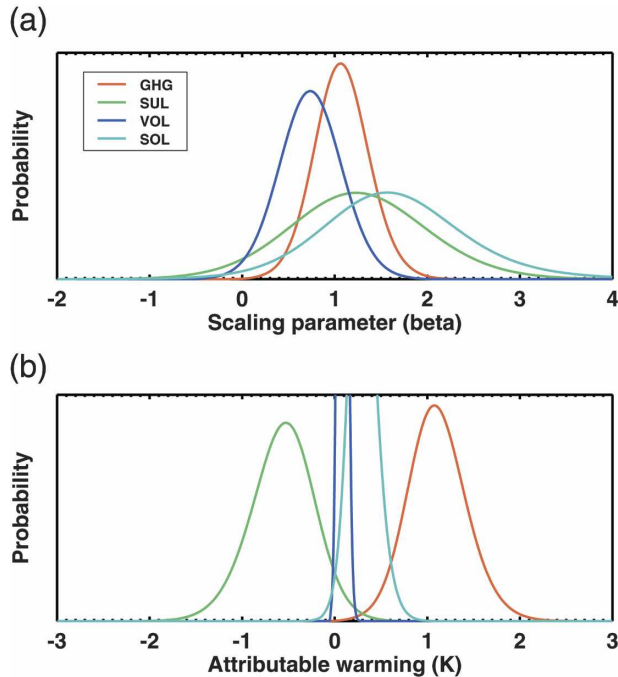


FIG. 6. Similar to Fig. 3, but lumping all of the ALL simulations from all of the GCMs together: (top) The distributions of the amplitude scalings (β_i) derived from the regression of observed SAT changes onto the EBM estimates of the GCM response patterns to each of the four forcings included in the simulations. The distributions are the average distributions across 100 Monte Carlo bootstrap samples. (bottom) The resulting distributions of the estimated change in SAT in the 1996–2005 decade relative to the 1901–10 decade attributable to each of the forcings.

here are fairly insensitive to the ensemble size, and indeed the confidence ranges on these distributions are comparable to those in Fig. 3. Responses to all four forcings are inconsistent with zero amplitude at the 5% significance level. All responses are consistent with a scaling amplitude of one, and the residual variability is consistent with that in the control simulations. As expected, the attributable warmings obtained from using these scalings are similar to those obtained in Fig. 3. The GHG attributable warming estimate tends to be a little larger, and the SUL estimate correspondingly smaller, than those obtained in the earlier section, but the differences are not significant. While this and the earlier estimates are complementary, the earlier estimates have the advantage that it makes more sense fitting an EBM to an ensemble from a single GCM rather than to an ensemble from multiple GCMs with differing representations of physical processes.

As in section 6 we can use the above scaling estimates to provide constraints on past and future climate change. This is shown in Fig. 7, once again with the assumption of the SRES A1B scenario in the future.

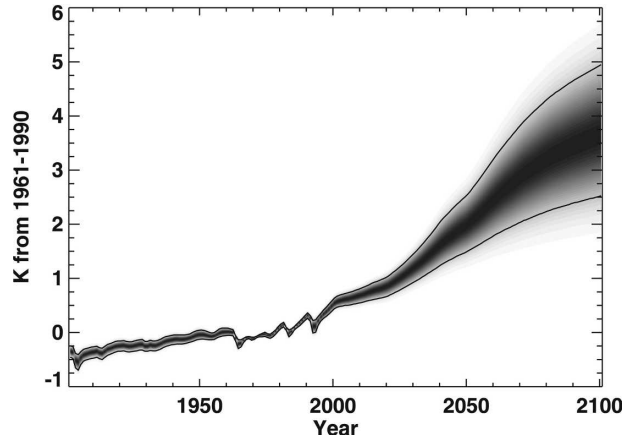


FIG. 7. Similar to Fig. 5a, but lumping all of the ALL simulations from all of the GCMs together. The shading shows the probability density function (PDF) of the global annual mean SAT climate from 1901 to present and through to 2100 assuming the SRES A1B emissions scenario. The lines show the 90% confidence intervals on this evolving PDF. The evolving distributions are the averages across 100 Monte Carlo bootstrap samples and are normalized to a standard peak value for visibility.

The evolution of the SAT changes is similar to but tighter than that in Fig. 5a, with the 90% confidence range for 2100 being 2.5–5.0 K. This is at the higher range of Fig. 5a, consistent with the attributable warming values, which are also on the higher side.

8. Conclusions

This paper provides an update on the multisignal detection and attribution of climate change, using annual data, more recent observations, more advanced GCMs, separated forcings, and a new technique that allows the inclusion of GCMs in the analysis even when multisenario ensembles of simulations have not been performed. While there is some spread in the GCM properties (as expressed in the corresponding EBM parameters), they generally show similar response patterns and amplitudes to the four major forcings examined. The responses to GHG and VOL forcing are consistently detected in the comparison with the observations, while the uncertainty is larger with the SUL and SOL responses and so the comparisons are more ambiguous. The observational constraints provided by this analysis indicate that GHG forcing contributed to about a 0.8-K warming over the 1901–2005 period, while the SOL/SUL forcing contributed a smaller warming/cooling and little warming from VOL forcing. These constraints permit a prediction of a ~ 1.5 – 4.5 K warming by 2100 under the SRES A1B emissions scenario.

The purpose of the EBM fits used here is simply to emulate GCM responses. Therefore, the exact values of the parameters used in the EBMs are not important inasmuch as the output of the EBM is satisfactory. While consistency of results between GCMs suggests that satisfactory output across GCMs has been obtained, the “black box” component of the analysis is highlighted by the large spread in some EBM parameters. Nevertheless, the estimates for the attributable warming and future changes are remarkably consistent across models. Thus, despite the large spread of EBM parameter values, the overall characteristics of the EBM performance are similar across models and this allows the observations to provide similar constraints across models. Still, the implication is that the results of this study should be taken in a broad sense, rather than specifically for each GCM. Thus, while the methodology used here allows the inclusion of many more GCMs and simulations in the study of detection and attribution, it should not be considered a replacement to the traditional method of examining multiple ensembles of simulations with different forcing scenarios.

Two other techniques are being applied to the detection and attribution of global SAT changes with the IPCC AR4 ALL simulations. One uses an approach on the global mean SAT changes in which a Bayesian decision method is used as a tool for classifying observations into given scenarios (Min and Hense 2006b). Min and Hense (2006b) find “decisive” evidence in favor of the ALL simulations versus the preindustrial control simulations in terms of their representation of 1900–99 SAT changes. Min and Hense (2006a) find that all of the ensembles of ALL simulations reproduce historical changes well, while more than half of the ensembles of ANT simulations cannot reproduce the observed changes. The other study uses a space–time separable approach to extract response patterns to the four forcings examined here (CR06). Using their approach, the response to GHG forcing is consistently detected, as is the case here. However, in part because of smaller uncertainties in the scaling parameters, responses to SUL and SOL forcing are also consistently detected in their analysis. On the other hand, VOL forcing is often not detected; CR06 hypothesize that this is due to the use of forcing time series, rather than the time series of responses to the forcings, in their method, which could be expected to have a stronger effect for VOL than for the other forcings. These scalings estimates lead to attributable warming best-guess estimates of about 1.0, –0.5, 0.0, and 0.1 K for the GHG, SUL, VOL, and SOL forcings, respectively, which are comparable to the values found here. We plan to incorporate the EBM fitting technique used here into the method of CR06 in order

to achieve a more comprehensive spatiotemporal detection and attribution analysis.

This paper has demonstrated consistency in the nature of responses to forcings across many GCMs. Furthermore, results are broadly consistent with those from other current studies. This consistency suggests emerging constraints in the past surface temperature changes attributable to the various forcings and to predictions of future temperature changes conditional on a given forcing scenario. While some discrepancies remain and the constraints remain broad, the results appear to be robust enough to make observationally constrained predictions of future warming.

Acknowledgments. The authors wish to thank Caspar Ammann, Curt Covey, Simon Crooks, Mark Hawkins, David Karoly, Hugo Lambert, Gerry Meehl, and Martin Stendel for their help with the data and analysis. We also thank Seung-Ki Min, Francis Zwiers, and two anonymous reviewers for many helpful comments on earlier versions of the manuscript. We especially acknowledge the international modeling groups for providing their data for analysis. For the IPCC AR4 data we acknowledge the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. We acknowledge additional model data from the Geophysical Fluid Dynamics Laboratory (NOAA), the Climate and Global Dynamics Division (UCAR), and the IPCC Data Distribution Centre (Climate Research Unit, UEA). We also acknowledge the HadCRUT2v dataset provided by the Climate Research Unit, UEA. D.A.S. and P.A.S. were supported by the U.K. Department for Environment, Food, and Rural Affairs, and D.A.S. was also partially supported by a Wellcome Trust Showcase Award. M.R.A. received partial support from the U.S. NOAA/DOE International Detection and Attribution Group.

REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- , P. A. Stott, J. F. B. Mitchell, R. Schnur, and T. L. Delworth, 2000: Uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.
- , and Coauthors, 2005: Observational constraints on climate sensitivity. *Avoiding Dangerous Climate Change*, J. S.

- Schellnhuber et al., Eds., Cambridge University Press, 281–289.
- Barnett, T. P., D. W. Pierce, and R. Schnur, 2001: Detection of anthropogenic climate change in the world's oceans. *Science*, **292**, 270–274.
- , and Coauthors, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate*, **18**, 1291–1314.
- Boucher, O., and M. Pham, 2002: History of sulfate aerosol radiative forcings. *Geophys. Res. Lett.*, **29**, 1308, doi:10.1029/2001GL014048.
- Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 525–582.
- Gillett, N. P., F. W. Zwiers, A. J. Weaver, G. C. Hegerl, M. R. Allen, and P. A. Stott, 2002: Detecting anthropogenic influence with a multi-model ensemble. *Geophys. Res. Lett.*, **29**, 1970, doi:10.1029/2002GL015836.
- , —, —, and P. A. Stott, 2003: Detection of human influence on sea level pressure. *Nature*, **422**, 292–294.
- , A. J. Weaver, F. W. Zwiers, and M. F. Wehner, 2004: Detection of volcanic influence on global precipitation. *Geophys. Res. Lett.*, **31**, L12217, doi:10.1029/2004GL020044.
- International Ad Hoc Detection and Attribution Group, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate*, **18**, 1291–1314.
- Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206–223.
- Lambert, F. H., P. A. Stott, M. R. Allen, and M. A. Palmer, 2004: Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.*, **31**, L10203, doi:10.1029/2004GL019545.
- Lean, J., J. Beer, and R. S. Bradley, 1995: Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.*, **22**, 3195–3198.
- McAvaney, B. J., and Coauthors, 2001: Model evaluation. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 471–524.
- Min, S.-K., and A. Hense, 2006a: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophys. Res. Lett.*, **33**, L08708, doi:10.1029/2006GL025779.
- , and —, 2006b: A Bayesian assessment of climate change using multimodel ensembles. Part I: Global mean surface temperature. *J. Climate*, **19**, 3237–3256.
- Mitchell, J. F. B., D. J. Karoly, G. C. Hegerl, F. W. Zwiers, M. R. Allen, and J. Marengo, 2001: Detection of climate change and attribution of causes. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 695–738.
- Ramaswamy, V., and Coauthors, 2001: Radiative forcing of climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 349–416.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Santer, B. D., and Coauthors, 2003: Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science*, **301**, 479–483.
- Stone, D. A., M. R. Allen, F. Selten, M. Kliphuis, and P. A. Stott, 2007: The detection and attribution of climate change using an ensemble of opportunity. *J. Climate*, **20**, 504–516.
- Stott, P. A., and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**, 723–726.
- , S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137.
- , —, —, —, W. J. Ingram, and J. F. B. Mitchell, 2001: Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dyn.*, **17**, 1–21.
- , J. F. B. Mitchell, J. M. Gregory, B. D. Santer, G. A. Meehl, T. L. Delworth, and M. R. Allen, 2006: Observational constraints on past attributable warming and predictions of future global warming. *J. Climate*, **19**, 3055–3069.
- Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth-century temperature change near the Earth's surface. *Nature*, **399**, 569–572.
- , and Coauthors, 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107**, 4306, doi:10.1029/2000JD000028.