

Constraining climate model parameters from observed 20th century changes

By CHRIS E. FOREST*, PETER H. STONE and ANDREI P. SOKOLOV, *Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA 02139, USA*

(Manuscript received 19 December 2007; in final form 4 June 2008)

ABSTRACT

We present revised probability density functions for climate model parameters (effective climate sensitivity, the rate of deep-ocean heat uptake, and the strength of the net aerosol forcing) that are based on climate change observations from the 20th century. First, we compare observed changes in surface, upper-air, and deep-ocean temperature changes against simulations of 20th century climate in which the climate model parameters were systematically varied. The estimated 90% range of effective climate sensitivity is 2–5 K but no corresponding upper bound can be placed on the equilibrium climate sensitivity. The net aerosol forcing strength for the 1980s has 90% bounds of -0.70 to -0.27 W m^{-2} . The rate of deep-ocean heat uptake corresponds to an effective diffusivity, K_v , with a 90% range of 0.04 – $4.1 \text{ cm}^2 \text{ s}^{-1}$. Second, we estimate the effective climate sensitivity and rate of deep-ocean heat uptake for 11 of the IPCC AR4 AOGCMs. By comparing against the acceptable combinations inferred from the observations, we conclude that the rates of deep-ocean heat uptake for the majority of AOGCMs lie above the observationally based median value. This implies a bias in the predictions inferred from the IPCC models alone.

1. Introduction

The recognition that anthropogenic activity is causing global warming (Houghton et al., 2001; Solomon et al., 2007) has emphasized the importance of developing climate models with predictive capability. In recent decades considerable effort has been devoted to evaluating state-of-the-art climate models from this point of view. A good summary of this work is given in Chapter 8 of the latest IPCC report (Randall et al., 2007). Much of the work has focused on evaluating the models' ability to simulate the annual mean state, the seasonal cycle, and the interannual variability of the climate system, since good data is available for evaluating these aspects of the climate system. However good simulations of these aspects do not guarantee a good prediction. For example, Stainforth et al. (2005) have shown that many different combinations of uncertain model subgrid scale parameters can lead to good simulations of current surface temperatures, but do not lead to a robust result for the model's climate sensitivity.

A different test of a climate model's capabilities that comes closer to actually testing its predictive capability on the century time scale is to compare its simulation of changes in the 20th century with observed changes. A particularly common test has been to compare observed changes in global mean sur-

face temperature with model simulations using estimates of the changes in the 20th century forcings (e.g. fig. 9.5 in Hegerl et al., 2007). The comparison often looks good, and this has led to statements such as: "...the global temperature trend over the past century . . . can be modelled with high skill when both human and natural factors that influence climate are included" (Randall et al., 2007). However this transient climate response is not a good measure of the models' predictive capabilities, owing to the great uncertainties that affect the simulated trend (e.g. climate sensitivity, rate of heat uptake by the deep-ocean, and aerosol forcing strength). For example, a model with a relatively high climate sensitivity can simulate the 20th century climate changes reasonably well if it also has a strong aerosol cooling and/or too much ocean heat uptake. Depending on the forcing scenario in the future, such models would generally give very different projections from one that had all those factors correct.

There have in recent years been a number of studies using the observed 20th century temperature to calculate probability density functions (PDFs) for the above mentioned uncertain parameters (Andronova and Schlesinger, 2001; Forest et al., 2002, 2006; Knutti et al., 2003). A meaningful test of a model's capabilities can be provided by comparing properties of different state-of-the-art models with their values, as implied by 20th century changes. Forest et al. (2006) have presented such a comparison for the models used in the IPCC TAR but not for those models used in the IPCC AR4. Here, we present an update of the Forest et al. (2006) results, in which we use the 20th

*Corresponding author.

e-mail: ceforest@mit.edu

DOI: 10.1111/j.1600-0870.2008.00346.x

century observations to constrain the effective climate sensitivity rather than the equilibrium climate sensitivity, while simultaneously constraining the ocean heat uptake and aerosol forcing; and we also now analyse 11 of the IPCC 4AR models for which the necessary data is available. Recent improvements made in the MIT climate model have caused the model's effective and equilibrium sensitivities to differ significantly from each other when the climate sensitivity is large. The concept of effective climate sensitivity, first introduced by Murphy (1995), will be discussed in more detail in Section 2. Here, we note that the effective sensitivity corresponds to the net climate feedback that is active during the transient response to external forcings at a particular time while the equilibrium sensitivity corresponds to the net climate feedback once the climate system has reached equilibrium. The effective sensitivity can be time-dependent. As shown by Raper et al. (2001) and Sokolov et al. (2003), the effective sensitivity is required, rather than the equilibrium sensitivity, to reproduce the transient response of a climate on century time scales, and thus, the 20th century climate changes are more suitable for estimating a constraint on effective sensitivity. Section 2 describes the version of the MIT climate model used in this study, Section 3 describes the method for constraining climate model parameters, Section 4 gives the results, and Section 5 summarizes and discusses the results.

2. MIT 2-D climate model

The model used in this study is the climate component of the MIT Integrated Global System Model, Version 2 (Sokolov et al., 2005). This model is an updated version of the model described in Sokolov and Stone (1998). Here we give a brief summary of the model and of the changes made since Forest et al. (2006)

The model consists of a zonally averaged atmospheric model coupled to a mixed layer Q-flux ocean model, with heat anomalies diffused below the mixed layer. The atmospheric model is derived from the Goddard Institute for Space Studies (GISS) Model II general circulation model (GCM) (Hansen et al., 1983) and uses parametrizations of the eddy transports of momentum, heat and moisture by baroclinic eddies (Stone and Yao, 1987, 1990). The model uses the GISS radiative transfer code which contains all radiatively important trace gases as well as aerosols. The surface area of each latitude band is divided into fractions of land, ocean, land-ice and sea-ice, with the surface fluxes computed separately for each surface type. The version used here has 4° latitudinal resolution and 11 layers in the vertical. The zonal averaging and the relatively low meridional and vertical resolution are necessary to make the model computationally efficient enough so that we can carry out simulations totalling hundreds of thousands of years, as required by our methodology (see next section). The ocean mixed layer model and the thermodynamic sea-ice model have 4° × 5° latitude–longitude resolution and are described by Hansen et al. (1984).

The climate sensitivity of the MIT model can be varied by changing the strength of the cloud feedback (Sokolov, 2006), differences in which have been shown to be the main reason for the differences in model climate sensitivity between different AOGCMs (e.g. Cess et al., 1990; Colman, 2003). The rate of mixing thermal heat anomalies into the deep ocean is controlled by the global mean value of the vertical diffusivity coefficient for mixing anomalies (K_v). Sokolov and Stone (1998) and Sokolov et al. (2003) have shown that the large-scale response of a given coupled atmosphere–ocean GCM (AOGCM) to forcings typical of the 20th and 21st century can be duplicated by the MIT 2-D model with an appropriate choice of these two parameters for any scenario. This ability to mimic the AOGCMs is what allows us to use the MIT 2-D model to explore how consistent different AOGCMs are with observed 20th century temperature changes.

The method for changing cloud feedback in the model has been changed from the method used previously. In the earlier versions of the model the cloud cover at all levels was changed by a fixed fraction, which depended on the changes in global mean surface temperature (Sokolov and Stone, 1998). In this version high cloud covers and low cloud covers are changed in opposite directions by a constant factor, which is again dependent on changes in the global mean surface temperature. The new method is described by Sokolov (2006), who shows that this method is in better agreement with changes simulated by AOGCMs, and does not change the 2-D model's ability to mimic global scale temperature changes simulated by AOGCMs.

The most significant change that has been made in the current version of the MIT 2-D climate model is the replacement of the old GISS land surface scheme by the Community Land Model (CLM2.1) described by Bonan et al. (2002). (See Schlosser et al., 2007 for the description of the coupling to the 2-D model.) This improved the simulation of evaporation and removed the tendency of the land model to be too hot in summer, due to excessive evaporation in spring causing the land to dry out. This was also a problem in the parent GCM. However, replacement of the GISS land–surface model with the CLM2.1 slowed down the transient response of the MIT 2-D climate model to an external forcing. Figure 1 shows changes in surface air temperature in simulations with an instantaneous doubling of CO₂ concentration. While evaporation from land is too small to directly affect the global surface energy budget in a significant way, a small rate of land evaporation response to surface warming leads to a delay in the increase of atmospheric water vapour. This, in turn, causes slower warming by reducing the incoming long-wave radiation at the surface.

The differences in the response to an external forcing between the two versions of the 2-D model result in different relations between equilibrium (S_{eq}) and effective (S_{eff}) climate sensitivities. Effective sensitivity is defined as $S_{eff} = \frac{F_{2x}}{\lambda_{eff}}$, where F_{2x} is the forcing due to CO₂ doubling and λ_{eff} is the climate feedback parameter estimated, in the conventional way, at the time of CO₂ doubling in a scenario where CO₂ increases by 1% per year

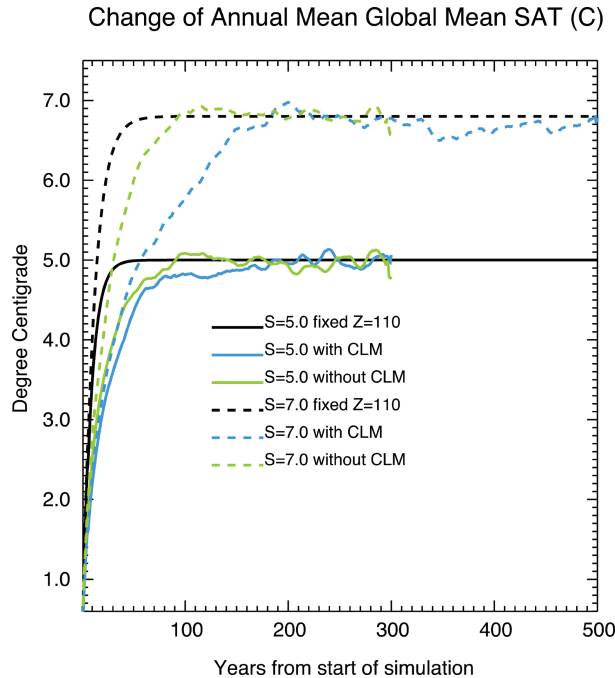


Fig. 1. Global-mean surface air temperature in simulations by MIT 2-D climate model with (blue) and without (green) the CLM land-surface model for an instantaneous doubling of CO_2 concentration. The response by an energy balance model with a 110 m deep ocean mixed-layer is shown by thin black line.

(Murphy, 1995). An increase in the evaporation in response to the surface warming simulated by the CLM2.1 is significantly smaller than an increase simulated by the GISS land surface model. The sensible heat flux increases in simulations with CLM2.1 and decreases in simulations with the GISS land model. These differences in the response of surface fluxes to an increase in surface temperature affect the strengths of the water vapour and cloud feedbacks. The ratios of feedbacks at the time of CO_2 doubling to the feedbacks at equilibrium are larger in the simulation with the GISS land model. Figure 2 shows the relationship between S_{eff} and S_{eq} for both versions of the model. The two sensitivities are virtually equal for $S_{\text{eq}} < 3^\circ$ but in the new model, S_{eff} is considerably less than S_{eq} for large values of S_{eq} . Since the 20th century changes are transient, it is clearly preferable to use the observed changes to constrain S_{eff} rather than S_{eq} . Because the two sensitivities diverge for large values in our model, the transient changes in the 20th century are not effective in constraining large values of S_{eq} .

3. Methods

3.1. Estimation of probability distributions

The methodology for quantifying uncertainty in climate system properties follows the basic method in Forest et al. (2001, 2002,

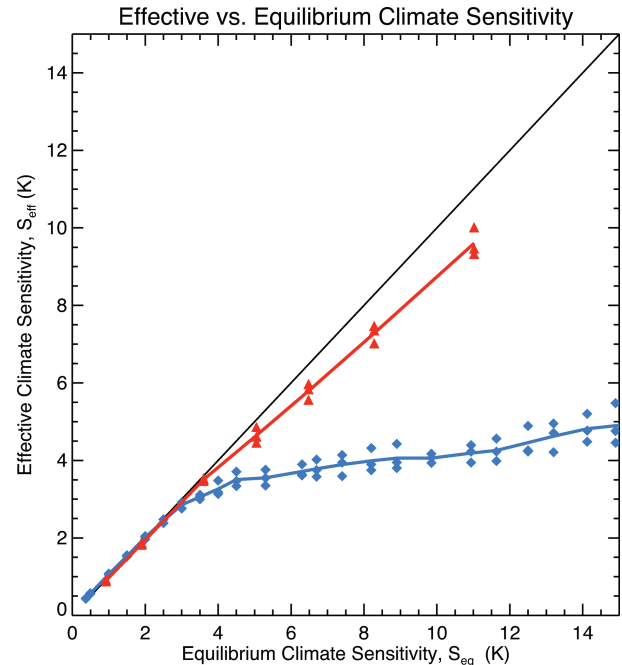


Fig. 2. Comparison of effective and equilibrium climate sensitivities for the MIT 2-D climate model with (blue diamonds) and without (red triangles) the CLM land-surface model. The three values for each equilibrium sensitivity correspond to K_v equal to 0.25, 2.25 and $6.25 \text{ cm}^2 \text{ s}^{-1}$. The red and blue lines are the averages of these three values. The black line indicates where the two sensitivities are equal. We note effective sensitivity can be a time-dependent quantity and as noted in the text, we use the conventional definition estimated at the time of CO_2 doubling in a scenario where CO_2 increases at 1% per year.

2006) with the modifications required to use the climate model. This can be summarized as consisting of two parts: simulations of the 20th century climate record and the comparison of the simulations with observations using optimal fingerprint diagnostics. First, we require a large sample of simulated records of climate change in which climate parameters have been systematically varied. Second, we employ a method of comparing model data to observations that appropriately filters ‘noise’ from the pattern of climate change. The variant of optimal fingerprinting proposed by Allen and Tett (1999) provides this tool and yields detection diagnostics that are objective estimates of model-data goodness-of-fit. In the use of the temperature change diagnostics and the estimation of the posterior probability distribution, the methodology is identical to that in Forest et al. (2006). The three temperature change diagnostics that we use are: (i) the decadal mean surface temperature changes over four equal-area latitude bands for the period 1946–1995 referenced to the 1905–1995 climatology; (ii) the trend in the global mean ocean temperature (down to 3 km depth) during 1948–1995 and (iii) the latitude-height pattern of the zonal mean upper air temperature difference between the 1961–1980 and 1986–1995 periods. The likelihood

functions based on each diagnostic are combined using Bayes' Theorem.

The description of the climate model experiments, the ensemble design, and the algorithm for estimating the joint PDFs are in Forest et al. (2001, 2002, 2006). There are two major differences from Forest et al. (2006) that were required when using the new model. First, a land-use change data set for the twentieth century was not included in these simulations because none was available in the new model's format. However, the contribution of this forcing to the total 20th century forcing is very small (Forster et al., 2007). Thus, the set of applied climate forcings was reduced to: greenhouse gas concentrations, sulphate aerosol loadings, tropospheric and stratospheric ozone concentrations, solar irradiance changes, and stratospheric aerosols from volcanic eruptions. We refer to this set of forcings as GSOSV. (Details on these forcings are in the auxiliary material in Forest et al. (2006)).

The second change was required to accommodate the change from equilibrium to effective climate sensitivity. Because S_{eff} has an upper bound at about 8 K in the new climate model, we truncate the distribution at 8 K rather than 10 K as was done in our previous studies. Thus, for the uniform prior cases, the cumulative probability above 8 K will differ from the results in Forest et al. (2006). In the case where an expert prior is used for S_{eff} , the prior has near zero probability above 8 K and the results are basically unaffected.

When conducting 20th century simulations, we use different values of the strength of the cloud feedback which changes both S_{eff} and S_{eq} . While S_{eff} is a more appropriate measure of transient climate response, results from our previous studies were presented in terms of S_{eq} because, first, S_{eq} and S_{eff} were virtually the same for older versions of the model and second, there is a one to one correspondence between S_{eq} and the strength of the cloud feedback. For the new model, S_{eff} is very different from S_{eq} (for high values), so results of this study are presented in terms of S_{eff} . The value of S_{eff} for a given strength of cloud feedback depends slightly on K_v . On figures below, we use values of S_{eff} averaged over K_v .

3.2. Matching procedure for AOGCMs

As discussed earlier, the large-scale response of the MIT model is controlled by the parameters, S_{eff} (or S_{eq}) and K_v . This flexibility provides the ability to match the large-scale response of AOGCMs by choosing appropriate combinations of these two parameters. Fits for the models were obtained based on the data for surface air temperature (SAT) and thermosteric sea level rise from the simulations with 1% per year increase in CO_2 concentration. Unfortunately the required data are available for only nine (9) AR4 models as part of the CMIP3 data set (Meehl et al., 2007b). Fits for the HadCM3 and five TAR models are based on the results from CMIP2 simulations. In Fig. 3, the values of S_{eff} required to match models' responses are compared with values of S_{eff} published for the corresponding models. Effective

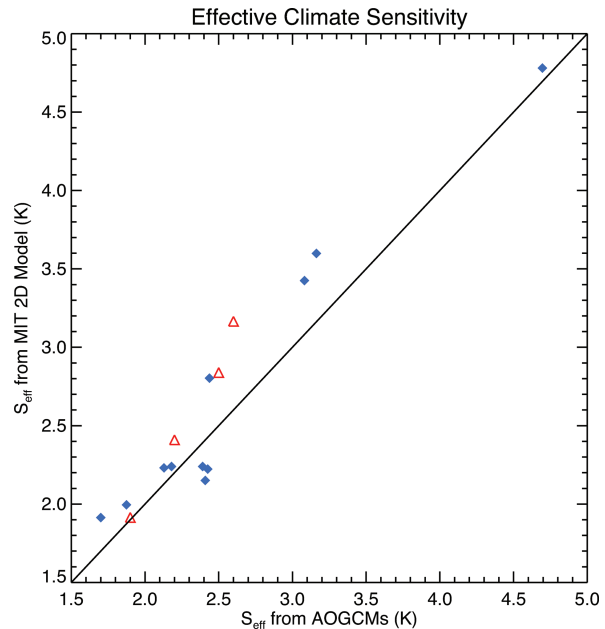


Fig. 3. Comparison of effective climate sensitivity estimated from AOGCM simulations versus effective climate sensitivity required to fit the AOGCM transient response. Blue diamonds refer to AR4 models and red triangles refer to TAR models.

sensitivities for the AR4 models were estimated from the data on 'top of the atmosphere fluxes' from the archived CMIP3 data set (Meehl et al., 2007b) using values of the adjusted radiative forcing due to CO_2 doubling (F_{2X}) given in Table 8S.1 from the IPCC AR4. Values of S_{eff} for CMIP2 models were taken from the literature. It should be noted that for some models F_{2X} is not available and in these cases, a forcing of 3.71 Wm^{-2} was used. Table 1 gives the 2-D model's values of S_{eff} and K_v that match the performance of the listed models.

Heat uptake in the oceans is sometimes measured by a coupled model's heat uptake efficiency, E (Gregory and Mitchell, 1997). We have compared our measure, K_v , with E estimated for the nine coupled AOGCMs using the CMIP3 data sets. They are correlated, with a correlation of 0.837.

4. Results

4.1. Posterior distributions using the new model, IGSM2

The 1-D marginal distributions from the current analysis (Fig. 4) and the Forest et al. (2006) estimates (their Fig. 2) are very similar and indicate the model responses are nearly identical. The fractiles for S_{eff} , K_v and F_{aer} are in Table 2. The aerosol forcing remains well constrained and is similar to other inverse estimates (Hegerl et al., 2007). The distribution for climate sensitivity with the expert prior, as before, has a well-defined mode at 2.8 K while the upper tail remains significant. The expert prior on climate sensitivity remains an important feature of the results with a reduction in the likelihood above 4.5 from 42 to 8% in

Table 1. Values of S_{eff} and K_v for AOGCMs used in the IPCC AR4 (top) and TAR (bottom)

Index	Model	K_v	S_{eff}
1	CGCM3.1(T47)	2.9	3.4
2	ECHO-G	1.3	2.8
3	GFDL-CM2.0	4.	2.2
4	GFDL-CM2.1	4.	2.2
5	INM-CM3.0	0.7	2.0
6	MIROC3.2(medres)	4.0	4.8
7	GISS-EH	1.7	2.2
8	CCSM3	3.4	2.2
9	GISS-ER	3.1	2.2
10	HadCM3	1.9	3.6
11	PCM	2.1	1.9
12	HadCM2	3.0	2.8
13	ECHAM3	1.6	2.4
14	MRI (old)	7.5	3.2
15	CSM	3.8	1.9

the new results. As before, K_v is well constrained by the three diagnostics with the surface temperature providing a strong constraint on the upper bound. Note that the 5 K 95% upper bound on S_{eff} translates into a 95% upper bound on S_{eq} of about 15 K (see Fig. 2).

Our PDF for climate sensitivity can be compared with other PDFs derived from 20th century observations. These have been conveniently summarized in section 9.6.2.1 of the IPCC AR4 (Hegerl et al., 2007). The discussion in Hegerl et al. (2007) describes the climate sensitivities deduced as equilibrium climate sensitivities, but in fact since they are based on transient 20th century changes, they must be regarded as effective sensitivities. As shown in Fig. 2, the two sensitivities can be quite different for large sensitivities. Determining upper bounds on the climate sensitivity is notoriously difficult, in part because of the inverse relation between climate sensitivities and net climate feedback (Hansen et al., 1984) and in part because the uncertainty in the net aerosol cooling allows large climate sensitivities to be consistent with the observed warming. Consequently, the 95% upper bound on effective climate sensitivity found in the different studies is not robust, and ranges from about 5 K to infinity (Hegerl et al., 2007). For low sensitivities, the results appear to be more robust. The 5% lower bound ranges from 1.0 to 2.1 K, and in fact our analysis gives the largest 5% lower bound, 2.1 K. As we showed in our earlier study (Forest et al., 2006) this lower bound comes from our surface temperature diagnostic, which consists of a 2-D (time-latitude) pattern of temperature changes. Our analysis is the only one which uses such a pattern to constrain the uncertainties. None of the other studies have estimated PDFs of ocean heat uptake.

The 2-D marginal distributions are shown in Fig. 5 for the $S_{\text{eff}} - \sqrt{K_v}$ parameter space. The positions of the climate mod-

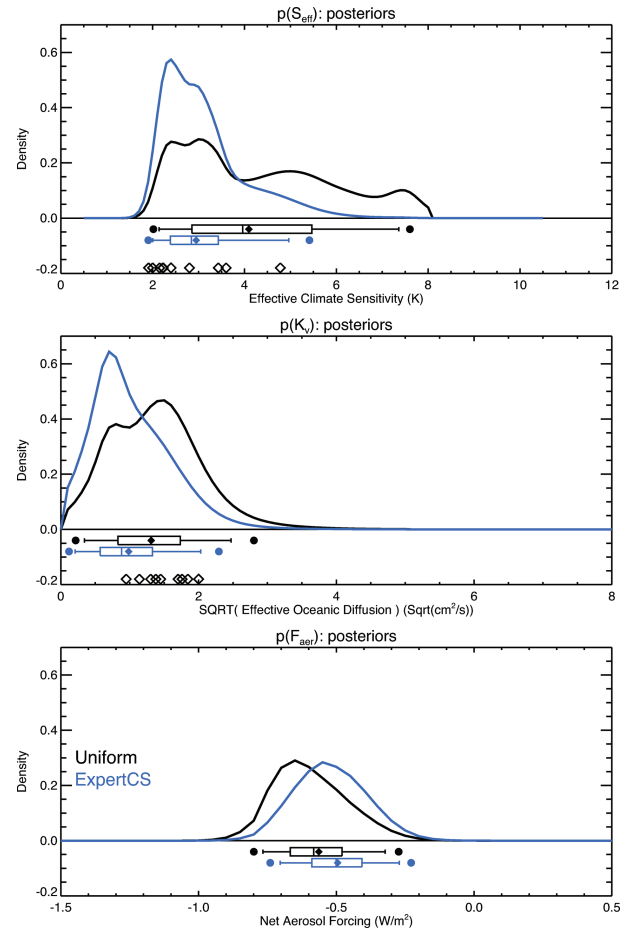


Fig. 4. The marginal *posterior* probability density function for the three climate system properties for two cases with the new model. In each panel, the marginal PDFs are shown for the GSOSV forcings. In one case (black), uniform priors are used on all parameters and in the second case (blue), an expert prior on climate sensitivity (Webster and Sokolov, 2000) is used with uniform priors elsewhere. Marginal distributions are estimated by integrating the density function over the remaining two parameters and renormalizing. The whisker plots indicate boundaries for the percentiles 2.5–97.5 (dots), 5–95 (vertical bar at ends), 25–75 (box ends), and 50 (vertical bar in box). The mean is indicated with the diamond and the mode is the peak in the distribution. The values for S_{eff} and K_v for the AR4 AOGCMs are shown as diamonds below the whisker plots.

els' heat uptake generally remain significantly to the right of the median and mode for the distribution. Given that the mode is an estimate of the most likely value, the AR4 models appear to have a positive bias in their ocean heat uptake, although we have not been able to obtain the data necessary to calibrate 10 of the AR4 models.

We also show the distributions for TCR and SLR (respectively, changes in SAT and thermosteric sea level rise averaged over years 61–80 in simulations with 1% per year increases in CO_2 concentration) as estimated from our new distribution and also

Table 2. Fractiles for marginal PDFs with and without expert prior on S_{eff}

	0.05	0.50	0.95	Mean	Mode
Expert prior					
S_{eff}	2.0	2.8	5.0	2.9	2.4
K_v	0.04	0.78	4.1	0.97	0.49
F_{aer}	-0.27	-0.50	-0.70	-0.50	-0.55
Uniform Prior					
S_{eff}	2.1	4.0	7.4	4.1	3.0
K_v	0.12	1.7	6.1	1.7	2.2
F_{aer}	-0.32	-0.58	-0.77	-0.56	-0.65

as estimated for the AOGCMs (Fig. 6). We note that our PDF is very similar to those of Stott et al. (2006), as replotted in fig. 9.21 of Hegerl et al. (2007), and Stott and Forest (2007), which were also based on 20th century observations. Values of TCR are available for eight of the AR4 AOGCMs for which SLR data was not available (M. Meinshausen, 2007, personal communication) and we have included those values in Fig. 6a as well as those for the models in Table 1. The 19 AR4 models do show a low bias, with a mean of 1.76 K compared to the mean of our PDF of 1.90 K. The bias is necessarily smaller than the bias in K_v , because TCR depends on S_{eff} as well as K_v , and the AR4 models show no apparent bias in S_{eff} . The two parameter comparison shown in Fig. 5 is a much better test of a model's capabilities. We have in fact carried out probabilistic projections of 21st century warming based on our PDFs using the same SRES scenarios as the AR4 (Meehl et al., 2007a). These will be reported in a

separate paper, but suffice it to say here that our PDFs imply a significantly stronger warming in the 21st century than the AR4 model ensemble. Finally we note that there is a high bias in the AOGCMs for SLR, but this is partly compensated by their low warming bias.

4.2. Robustness of the ocean heat-uptake results

Since there is a significant discrepancy between the AOGCMs' simulations of ocean heat uptake and the uptake we estimate from observations, we have explored the sensitivity of our posterior distribution for K_v to various diagnostics. First we show in Fig. 7 how our 1D marginal distributions change when we remove information associated with the three different diagnostics. In particular, we compare our standard results based on all three diagnostics with what happens if: (i) we leave out the upper-air diagnostic, (ii) we leave out the deep-ocean temperature change diagnostic and (iii) we replace the surface temperature change diagnostic using 4 latitude bands, z4, by one using only hemispheric averages, z2, but still retaining the decadal time series. In the last case we retain the contrast in hemispheric temperature averages that reflects the aerosol forcing being larger in the Northern Hemisphere, but remove the polar amplification component in the z4 diagnostic. In all cases we see that the PDFs for S_{eff} and F_{aer} are not much affected and we conclude that these PDFs are relatively robust. However in the case of the K_v distribution we see that removing any of the diagnostics weakens the constraint on K_v , with the removal of the deep-ocean temperature diagnostic showing the most effect. Nevertheless the mode for K_v is relatively robust, and indeed it is smallest when the

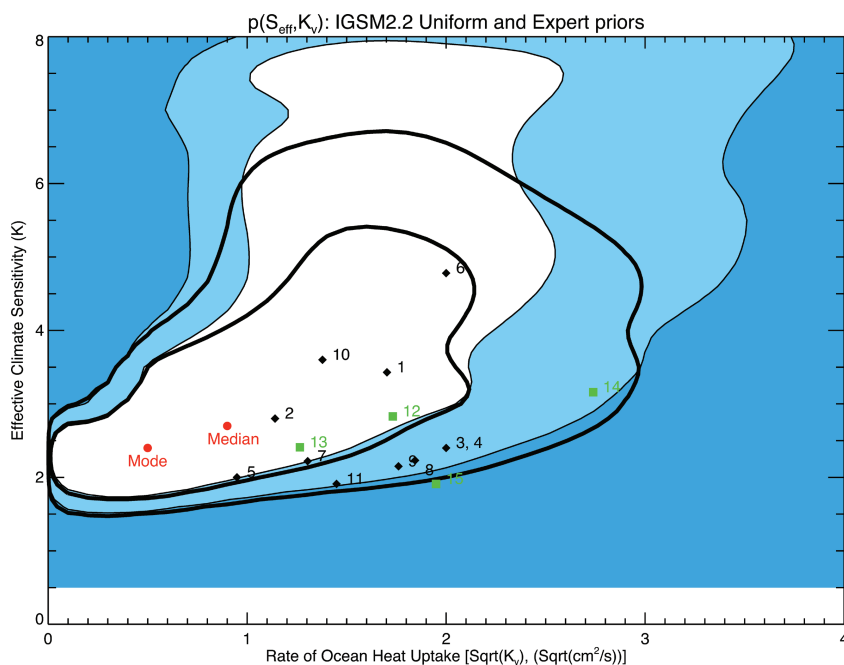


Fig. 5. The marginal posterior probability density function for GSOSV results with uniform (shading) and expert prior on S (thick contours) for the $S_{\text{eff}} - K_v$ parameter space. The shading denotes rejection regions for a given significance level—10%, and 1%, light to dark, respectively. The positions of AOGCMs (diamonds and squares) represent the parameters in the MIT IGSM2 model that match the transient response in surface temperature and thermal expansion component of sea level rise under a common forcing scenario. Model names and parameter values are listed in Table 1. Lower K_v values imply less deep-ocean heat uptake and hence, a smaller effective heat capacity of the ocean. Eleven AOGCMs used in the IPCC AR4 report (black diamonds) and four used only in the IPCC TAR (from Sokolov et al., 2003) (green squares) represent the models with sufficient information available. The median and mode (red circles) are shown for the case with the expert prior.

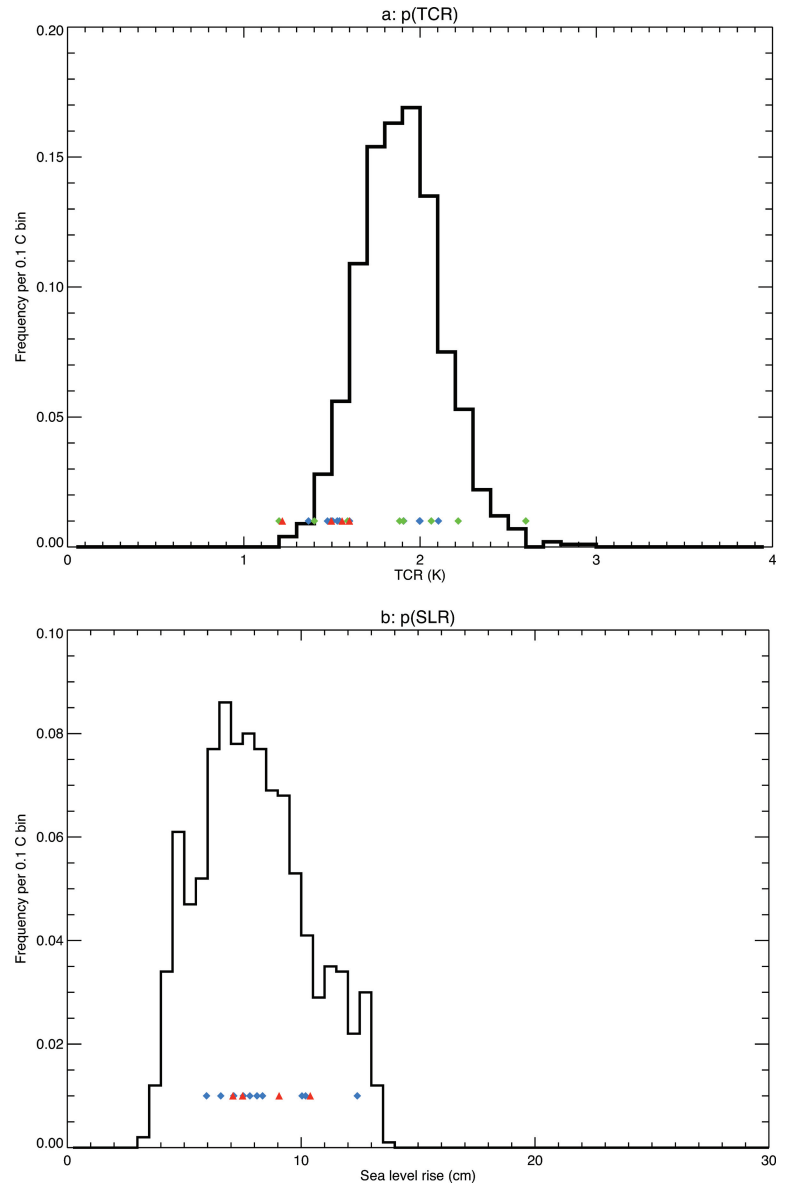


Fig. 6. (a) $p(\text{TCR}|\text{data})$ and (b) $p(\text{SLR}|\text{data})$ from 2-D model results based on 1000 member Latin Hypercube sample. The values of TCR and SLR for the AOGCMs in Table 1 are shown with blue diamonds for the AR4 models and red triangles for the TAR models. Additional AR4 models with only TCR values are shown with green diamonds.

deep-ocean temperature diagnostic is removed. This is because of the small thermal inertia of the atmosphere compared to the oceans. The decadal changes in the atmosphere are controlled by the thermal inertia of the oceans, and thus all the atmospheric diagnostics contribute to the discrepancy between our estimate of the deep-ocean heat uptake and the uptake simulated by the AOGCMs. However, the discrepancy is most significant when the deep-ocean temperature diagnostic is included.

In our standard analysis we used the ocean trend and error estimates given by Levitus et al. (2005). It has recently come to light that the XBT data that they used in their analysis contained systematic errors (Gouretski and Koltermann, 2007). The Gouretski and Koltermann analysis indicates that the Levitus

et al. trend should be reduced by 37%. On the other hand another new analysis reported at the April 2008 meeting of the EGU (C. Domingues, 2008, personal communication) indicates that the Levitus et al. trend should be increased by 30%. We repeated our analysis using the Gouretski and Koltermann result with the trend reduced by 37% and with their error estimate which is significantly larger than that of Levitus et al. The results (not shown) had the mode for K_e reduced to a value consistent with that when the ocean diagnostic was removed (Fig. 7) and the distribution was somewhat broadened, but not as much as when the ocean diagnostic was removed. The error estimate for the newer analysis has not yet been given, so we cannot repeat our analysis using that result. However if the error estimate is similar

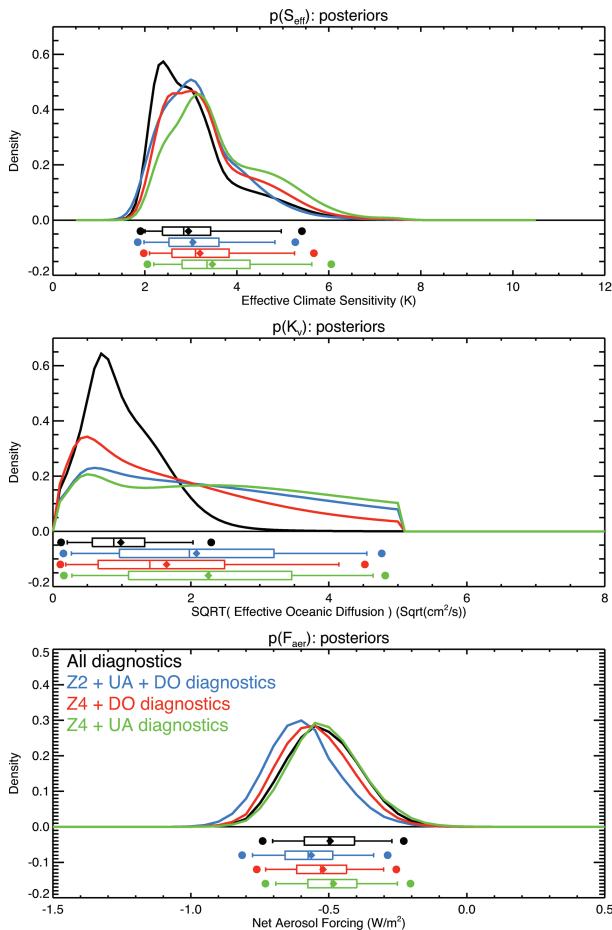


Fig. 7. Posterior distributions using alternative combinations of climate change diagnostics. Standard diagnostics (black) using surface (z4), upper-air (UA) and deep-ocean (DO) temperatures; z2 + UA + DO (blue), hemispheric averages replace four equal-area zonal bands; z4 + DO (red), no upper-air temperatures; and z4 + UA (green), no deep-ocean temperatures.

to Gouretski and Koltermann's, then we would expect a similar result, that is, the ocean data would not have much impact on our result and the discrepancy with the model's would remain.

We also note two recent studies based on the Levitus et al. (2005) analysis of the ocean heat uptake that also indicate that AOGCMs are overestimating the 20th century heat uptake. Pierce et al. (2006) compared 20th century simulations of the heat uptake using the PCM and HadCM3 models with the Levitus et al. (2005) results using the observational data mask. Their fig. 11 shows that both models are overestimating the ocean heat uptake, particularly below the mixed layer. Andrews and Allen (2008) compared the performance of the AR4 AOGCMs with 20th century changes in surface temperature and ocean heat uptake, and found that the AOGCMs were generally overestimating the effective heat capacity of the climate system, which is of course equivalent to mixing heat into the ocean too efficiently.

Finally, we note that there is uncertainty in the forcing aside from that associated with aerosols which our analysis takes into account. In particular, there is uncertainty in the estimates of the volcanic forcing. We addressed this issue in the auxiliary material of Forest et al. (2006). There we showed that a 25% error in our estimate of the volcanic forcing did not significantly affect our PDFs.

5. Discussion and conclusions

We present two new results in this paper. First, we have estimated the S_{eff} and K_v values that correspond to eleven (11) of the AR4 AOGCMs models. This serves to characterize the 'ensemble of opportunity' (EOP) in terms of transient responses. Together, these two properties provide a good metric for comparing the behaviour of different AOGCMs with one another and with respect to the distributions for these properties as estimated from climate change observations. Second, we present the updated probability distribution for the three climate system properties, $\theta = \{S_{\text{eff}}, K_v, F_{\text{aer}}\}$, with S_{eq} replaced by S_{eff} . These distributions are similar to those from Forest et al. (2006), because the forcings are almost identical (no land-use change in this case) and the climate change diagnostics were identical.

From the positions of the AOGCMs within this distribution, we can estimate the AOGCMs' projections under specific forcing scenarios. As noted by many (e.g. Prinn et al., 1999), the total uncertainty in the climate change projections is a combination of the uncertainties in both the forcings and the climate system response. By considering the AOGCM positions within the context of the $p(\theta|\Delta T, C_N)$ distributions, one can infer the range of uncertainty in the climate system response that is represented by their projections. Furthermore, we can track the change in the uncertainty implied by the projections in the various IPCC reports. As shown in Fig. 5, the projections from both the TAR and AR4 indicate a significant shift in the climate model response as estimated by the AOGCMs and their means, medians, and ranges. Although the complete set of models is not available, we still find a clear indication that the AOGCMs, as a whole, overestimate the rate of deep-ocean heat uptake as implied by the observations. The positions of the AOGCMs in the joint distribution of $S_{\text{eff}} - K_v$ (Fig. 5) suggest that the AR4 AOGCM results are biased low for projected surface temperature changes while biased high for sea level rise due to thermal expansion of sea water.

6. Acknowledgments

This work was supported in part by the Office of Science (BER), U.S. Dept. of Energy Grant No. DE-FG02-93ER61677, NSF, and by the MIT Joint Program on the Science and Policy of Global Change. We thank Adam Schlosser for his work on coupling CLM in the climate model and two anonymous reviewers for helpful comments. We acknowledge the modelling groups,

the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multimodel data set. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

References

- Allen, M. R. and Tett, S. F. B. 1999. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419–434.
- Andrews, D. G. and Allen, M. R. 2008. Diagnosis of climate models in terms of transient climate response and feedback response time. *Atm. Sci. Lett.* **9**(1), 7–12.
- Andronova, N. G. and Schlesinger, M. E. 2001. Objective Estimation of the Probability Density Function for Climate Sensitivity. *J. Geophys. Res.* **106**(D19), 22 605–22 612.
- Bonan, G. B., Oleson, K. W., Vertenstein, M., Lewis, S., Zeng, X. and co-authors. 2002. The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. *J. Clim.* **15**, 3123–3149.
- Cess, R. D., Potter, G. L., Blanchet, P., Boer, G. J., Del Genio, A. D. and co-authors. 1990. Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *J. Geophys. Res.* **95**, 16601–16615.
- Colman, R. 2003. A comparison of climate feedbacks in general circulation models. *Clim. Dyn.* **20**, 865–873.
- Forest, C. E., Allen, M. R., Sokolov, A. P. and Stone, P. H. 2001. Constraining Climate Model Properties Using Optimal Fingerprint Detection Methods. *Clim. Dyn.* **18**, 277–295.
- Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. and Webster, M. D. 2002. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117.
- Forest, C. E., Stone, P. H. and Sokolov, A. P. 2006. Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.* **33**, L01705, doi:10.1029/2005GL023977.
- Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R. and co-authors. 2007. Changes in atmospheric constituents and in radiative forcing. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis and co-editors), Cambridge University Press, Cambridge, UK and New York, NY, USA, 129–234.
- Gouretski, V. and Koltermann, K. P. 2007. How much is the ocean really warming? *Geophys. Res. Lett.* **34**, L01610, doi:10.1029/2006GL027834.
- Gregory, J. M. and Mitchell, J. F. B. 1997. The climate response to CO₂ of the Hadley Centre coupled AOGCM with and without flux adjustment. *Geophys. Res. Lett.* **24**, 1943–1946.
- Hansen, J., Russell, G., Rind, D., Stone, P., Lacis, A. and co-authors. 1983. Efficient three-dimensional global models for climate studies: models I and II. *Mon. Weath. Rev.* **111**, 609–662.
- Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P. and co-authors. 1984. Climate sensitivity: analysis of feedback mechanisms. In: *Climate Processes and Climate Sensitivity, Geophysical Monograph* (eds J. E. Hansen and T. Takahashi), Volume 29, American Geophysical Union, Washington, DC, 130–163.
- Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y. and co-authors. 2007. Understanding and attributing climate change. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis and co-editors), Cambridge University Press, Cambridge, UK and New York, NY, USA, 663–746.
- Houghton, J. T., Ding, Y., Griggs, D. K., Noguer, M., van der Linden, P. J., (eds.) 2001. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Knutti, R., Stocker, T. F., Joos, F. and Plattner, G.-K. 2003. Probabilistic climate change projections using neural networks. *Clim. Dyn.* **21**, 257–272.
- Levitus, S., Antonov, J. and Boyer, T. P. 2005. Warming of the World Ocean, 1955–2003. *Geophys. Res. Lett.* **32**, L02604, doi:10.1029/2004GL021592.
- Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T. and co-authors. 2007a. Global climate projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis and co-editors), Cambridge University Press, Cambridge, UK and New York, NY, USA, 747–846.
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B. and co-authors. 2007b. The WCRP CMIP3 multimodel dataset: a new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**(9), 1383–1394.
- Murphy, J. 1995. Transient response of the hadley centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: analysis of global-mean response using simple models. *J. Clim.* **8**, 496–514.
- Pierce, D. W., Barnett, T. P., AchutaRao, K. M., Gleckler, P. J., Gregory, J. M. and co-authors. 2006. Anthropogenic warming of the oceans: observations and model results. *J. Clim.* **19**, 1873–1900.
- Prinn, R., Jacoby, H., Sokolov, A., Wang, C., Xiao, X. and co-authors. 1999. Integrated Global System Model for climate policy assessment: feedbacks and sensitivity studies. *Climatic Change* **41**(3/4), 469–546.
- Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichet, T. and co-authors. 2007. Climate models and their evaluation. In: *Climate Change 2007, The Physical Science Basis* (eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis and co-editors) Cambridge University Press, Cambridge, UK, 589–662.
- Raper, S. C. B., Gregory, J. M. and Stouffer, R. S. 2001. The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *J. Clim.* **15**, 124–130.
- Schlosser, C. A., Kicklighter, D. and Sokolov, A. 2007. *A Global Land System Framework for Integrated Climate-Change Assessments, MIT JP Report 147*. Tech. rept. <http://web.mit.edu/globalchange/www/MITJPSPGC.Rpt147.pdf>. MIT, Joint Program on the Science and Policy of Global Change, Room E40-428, 77 Massachusetts Ave., Cambridge, MA 02139.
- Sokolov, A. P. 2006. Does model sensitivity to changes in CO₂ provide a measure of sensitivity to other forcings? *J. Clim.* **19**, 3204–3306.
- Sokolov, A. P. and Stone, P. H. 1998. A flexible climate model for use in integrated assessments. *Clim. Dyn.* **14**, 291–303.

- Sokolov, A. P., Forest, C. E. and Stone, P. H. 2003. Comparing oceanic heat uptake in AOGCM transient climate change experiments. *J. Clim.* **16**, 1573–1582.
- Sokolov, A. P., Schlosser, C. A., Dutkiewicz, S., Paltsev, S., Kicklighter, D. W. and co-authors. 2005. *The MIT Integrated Global System Model (IGSM) Version 2: Model Description and Baseline Evaluation, MIT JP Report 124*. Tech. rept. http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt124.pdf. MIT, Joint Program on the Science and Policy of Global Change, Room E40-428, 77 Massachusetts Ave., Cambridge, MA 02139.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis and co-editors 2007. *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, Cambridge, UK.
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N. and co-authors. 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403–406.
- Stone, P. H. and Yao, M.-S. 1987. Development of a two-dimensional zonally averaged statistical-dynamical model. Part II: the role of eddy momentum fluxes in the general circulation and their parametrization. *J. Atmos. Sci.* **44**(24), 3769–3786.
- Stone, P. H. and Yao, M.-S. 1990. Development of a two-dimensional zonally averaged statistical-dynamical model. Part III: the parametrization of the eddy fluxes of heat and moisture. *J. Clim.* **3**(7), 726–740.
- Stott, P. A. and Forest, C. E. 2007. Ensemble climate predictions using climate models and observational constraints. *Phil. Trans. R. Soc. A* **365**, 2029–2052. doi:10.1098/rsta.2007.2075.
- Stott, P. A., Allen, J. F. B., Mitchell, M. R., Delworth, T. L., Gregory, J. M. and co-authors. 2006. Observational constraints on past attributable warming and predictions of future global warming. *J. Clim.* **19**, 3055–3069.
- Webster, M. D. and Sokolov, A. P. 2000. A methodology for quantifying uncertainty in climate projections. *Climatic Change* **46**(4), 417–446.