

REVIEW ARTICLE

How reliable are climate models?

By JOUNI RÄISÄNEN*, *Department of Physical Sciences, Division of Atmospheric Sciences, P.O. Box 64, FIN-00014 University of Helsinki, Finland*

(Manuscript received 4 July 2006; in final form 8 September 2006)

ABSTRACT

How much can we trust model-based projections of future anthropogenic climate change? This review attempts to give an overview of this important but difficult topic by using three main lines of evidence: the skill of models in simulating present-day climate, intermodel agreement on future climate changes, and the ability of models to simulate climate changes that have already occurred. A comparison of simulated and observed present-day climates shows good agreement for many basic variables, particularly at large horizontal scales, and a tendency for biases to vary in sign between different models, but there is a risk that these features might be partly a result of tuning. Overall, the connection between model skill in simulating present-day climate and the skill in simulating future climate changes is poorly known. An intercomparison of future climate changes between models shows a better agreement for changes in temperature than that for precipitation and sea level pressure, but some aspects of change in the latter two variables are also quite consistent between models. A comparison of simulations with observed climate changes is, in principle, a good test for the models, but there are several complications. Nonetheless, models have skilfully simulated many large-scale aspects of observed climate changes, including but not limited to the evolution of the global mean surface air temperature in the 20th century. Furthermore, although there is no detailed agreement between the simulated and observed geographical patterns of change, the grid box scale temperature, precipitation and pressure changes observed during the past half-century generally fall within the range of model results. Considering the difficulties associated with other sources of information, the variation of climate changes between different models is probably the most meaningful measure of uncertainty that is presently available. In general, however, this measure is more likely to underestimate than overestimate the actual uncertainty.

1. Introduction

The ongoing anthropogenic changes in the atmospheric composition, particularly the increase in CO₂ and other greenhouse gases, have the potential to cause substantial climate changes during the coming decades and centuries (Houghton et al., 2001). Major research efforts are being put in attempts to forecast these climate changes, or at least to estimate the changes that would follow from a given future development (scenario) of anthropogenic greenhouse gas and aerosol emissions. How much can we trust these projections¹?

The computer models that are used for generating projections of future climate are in many respects similar to the models used for weather prediction on a daily-to-biweekly time-scale. Yet, the reliability of long-term climate change projections is much harder to estimate than that of weather forecasts. The latter can be quickly verified against the weather evolution that actually happened and, although the accuracy of the forecasts varies from time to time, their typical quality can be quantified by collecting verification statistics over a sufficient number of cases. For climate change projections, this approach is not practical, particularly as there are no earlier well-observed analogies of the type of primarily greenhouse-gas-induced climate change that is expected in the future. The reliability of these projections can therefore only be estimated by indirect methods.

In this review, I discuss the question of climate model reliability using three main lines of evidence: the skill of models in simulating present-day climate, intermodel agreement on future climate changes, and the ability of models to simulate climate changes that have occurred during the instrumental period. Other

*Correspondence.

e-mail: jouni.raisanen@helsinki.fi

DOI: 10.1111/j.1600-0870.2006.00211.x

¹The word *climate projection* is used instead of *forecast* or *prediction* to emphasize the dependence of the climate outcome on the assumed future evolution of human activities, the prediction of which is outside the field of natural sciences.

sources of information such as the performance of climate models in initial value prediction on daily-to-seasonal time-scales (e.g. Phillips et al., 2004; Graham et al., 2005) and in the simulation of paleoclimates (e.g. Kohfeld and Harrison, 2000) are also available but are omitted here. Even with this definition of the subject, a wide range of topics would be relevant. In this review, I attempt to keep the discussion at a relatively general level, trying to give an overview that is useful even for readers with a limited experience on climate models.

Several types of climate models have been developed for different purposes, ranging from simple energy balance type climate models that describe the state of the climate system with at most a few tens of numbers (Harvey et al., 1997) via Earth System Models of Intermediate Complexity (EMICs) (Claussen et al., 2002) to three-dimensional global General Circulation Models (GCMs) and Regional Climate Models (RCMs). The focus in this review is on Atmosphere-Ocean GCMs (AOGCMs), which attempt to explicitly simulate the atmospheric and oceanic processes that regulate the direction and magnitude of climate changes on global and at least large regional scales. To illustrate the performance of state-of-the-art AOGCMs, the literature review is complemented with a number of examples based on a multimodel ensemble of recent AOGCM simulations. These examples include three variables that have also been used in many other studies to characterize the time mean surface climate – surface air temperature, precipitation and sea level pressure. Changes in higher-order climate statistics, such as various types of extremes, are also an important issue, but they are not considered explicitly in this review.

The following section gives a brief discussion of some basic issues associated with general circulation models and their application to the simulation of climate changes. This discussion, building on textbooks such as Trenberth (1992), Mote and O'Neill (2000), Houghton (2004) and McGuffie and Henderson-Sellers (2005), is directed especially to readers from outside the climate modelling community. Section 3 discusses first the ability of models to simulate the present-day climate and then the more difficult question of what model skill in the simulation of present-day climate tells about the skill in the simulation of climate changes. The question of how well models agree with each other in the simulation of future climate changes is addressed in Section 4. Section 5 studies the ability of models to simulate the climate changes that occurred in the past 50–100 years. This is followed by a separate section reviewing model- and observation-based estimates of global climate sensitivity (the equilibrium global mean warming resulting from a doubling of atmospheric CO₂ concentration), which is widely regarded as one of the most important numbers in climate change research. Some additional issues associated with uncertainties in external forcing in the future and the limited resolution of global climate models are discussed in Section 7. The key points of the review are summarized and some concluding remarks are given in Section 8.

2. Climate modelling: some basic issues

The atmospheric weather and the ocean circulation are governed by the fundamental laws of physics that describe the conservation of mass, energy and momentum. These basically well-known laws – rather than, for example, empirical correlations between temperature and greenhouse gas concentrations – also form the backbone of climate models. This solid physical basis gives a strong reason to believe that the models are a useful tool for exploring the behaviour of the climate system and its response to changes in external forcing such as increases in greenhouse gas concentrations.

Yet, the full complexity of the real world cannot be presented in any model. Due to limitations in computing power, the atmospheric components in current AOGCMs typically have a grid spacing of 200–300 km in the horizontal direction. In the vertical direction there are typically about 30 levels between the surface and the model top at 30–50 km height, the spacing of the levels increasing from a few hundred metres in the boundary layer to several kilometres in the stratosphere. Processes acting on scales smaller than the grid spacing cannot be resolved explicitly. One consequence of this is that the models cannot simulate local variations in climate. More importantly, many unresolved processes also affect climate on larger horizontal scales. The impact of these processes needs to be *parametrized*, that is, estimated indirectly from the grid scale weather conditions simulated by the model.

Let us take the atmospheric thermodynamic equation as an example. In an isobaric Cartesian coordinate system

$$\frac{\partial T}{\partial t} + u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} + \omega \frac{\partial T}{\partial p} = \frac{\alpha}{c_p} \omega + \frac{Q}{c_p}. \quad (1)$$

The first left-hand side term represents the local rate of change of temperature and the following three terms the advection of temperature in the zonal, the meridional and the vertical directions. The first term on the right-hand side gives the adiabatic warming (or cooling) of air associated with the work done by pressure forces during descending (or ascending) air motion. All these terms are relatively easy to calculate, even though truncation errors resulting from the finite spatial and temporal resolution (climate models typically have a time-step of a few tens of minutes) in the calculations contaminate the solution particularly on the smallest resolved scales (Williamson and Laprise, 2000).

The trouble-maker in eq. (1) is the last term that represents diabatic heating. This term includes the heating or cooling associated with the phase changes of water, as well as the absorption and emission of radiation, plus a very small contribution by molecular diffusion. In computer models, which cannot explicitly simulate the smallest scales of atmospheric motions, this term also includes a contribution from subgrid scale mixing that, in principle, is part of the advection.

The phase changes of water and the transfer of radiation are ultimately microphysical processes, which could only be simulated explicitly with a model able to track the state of individual molecules. However, if model grid boxes were homogeneous in a macrophysical sense, the grid box means of the resulting diabatic heating could be calculated with good accuracy by using variables that describe the grid box mean conditions (of air motion, temperature, concentrations of water vapour and other atmospheric gases, aerosol properties etc.) and are, in principle, predictable by the model equations. Unfortunately, the grid boxes are not homogeneous. For example, a grid box with a mean relative humidity of 90% might be completely clear (if the humidity was evenly distributed) or mostly cloudy (if part of the grid box was very dry and the rest saturated), or anything in between, and these alternatives have different implications on the transfer of radiation and condensation heating.

Computation of grid box average diabatic heating thus requires information on the subgrid scale variation of meteorological conditions. This information must be provided, in one way or another, by parametrization equations based on the grid box mean variables predicted by the model. Similar considerations apply to the forecast equations of momentum and humidity, and also to the model components describing ocean, sea ice and land surface conditions.

The parametrization schemes used in climate models are based on a combination of physical theory, observational evidence and, in some cases, simulation by much higher-resolution models (e.g. Noh et al., 2003). Nevertheless, these sources of information often give only loose guidelines on how the parametrization should be made. Consequently, the parametrization schemes used in different models differ both in their basic structure and in the numerical values used in their equations. Moreover, there tends to be a trade-off between the potential accuracy of the scheme and the computing time that it requires. In many cases, the most sophisticated schemes that have been developed are computationally too expensive to be used in climate models.

Among the several processes that need to be parametrized in climate models, those associated with clouds appear to be the most problematic. Several intercomparison studies have shown that, in model experiments with increased greenhouse gas concentrations, different changes in cloudiness and cloud properties explain a majority of the intermodel differences in global mean surface warming (see Section 6.1). Due to the non-linearity of the climate system, however, not all of the cloud-related uncertainty is associated directly with the parametrization of cloudiness; other parts of the model such as the boundary-layer scheme may also be important (IPCC WG1, 2004).

All together, the simulated global warming appears to be determined primarily by the atmospheric component of the climate model (Meehl et al., 2004a). The model components representing the ocean, sea ice and land surface play secondary roles in this respect, even though they may be important for climate changes on regional scales. Meehl et al. suggest that, to a greater degree

than the other components, the atmospheric model ‘manages’ the feedback processes such as changes in water vapour, clouds and sea ice albedo that modulate the top-of-the-atmosphere radiation balance and thereby regulate the global mean warming.

Another issue that has implications for the interpretation of both model results and observed climate changes is the chaotic nature of weather and climate caused by the non-linearity of the governing equations (Lorenz, 1963). In numerical weather prediction, the skill of the forecasts deteriorates rapidly with time. This is caused partly by errors in the forecast models, but perfect-model studies have shown that a substantial fraction of the forecast errors results from the non-linear growth of small errors in the initial state of the forecast (e.g. Savijärvi, 1995). Even for a perfect model, the detailed daily evolution of weather would be predictable with a useful skill for only about two weeks.

The inherent unpredictability of weather also puts an upper limit on the potential accuracy of climate simulations. When the same climate model is run several times with the same external conditions but with different initial states, the resulting time-series differ so that, for example, the timing of individual warm or cold years is uncorrelated between the simulations (e.g. Stott et al., 2000). However, the magnitude of the differences saturates rapidly, rather than continuing the exponential growth characterizing the first days of weather forecasts. If the increase in greenhouse gas concentrations or other external forcing applied in the simulations is strong enough, the climate changes associated with this forcing will become larger than the internal variability resulting from the chaotic nature of the system. This is analogous to the seasonal cycle of weather: despite inter-annual variability, there is a distinct difference between winter and summer conditions forced by the seasonal cycle in solar elevation.

Internal variability may either add to or subtract from forced climate changes, both in the real world and in model simulations. However, the magnitude of the variability decreases with increasing temporal and spatial averaging. Thus, the associated uncertainty in multidecadal global means is much smaller than that in individual yearly values at a single location.

3. Simulation of present-day climate

An overarching aim in climate model evaluation is to assess the extent to which the behaviour of a model resembles the behaviour of the real climate system. The confidence that can be put on those aspects of model results that cannot be verified directly, such as projections of future anthropogenic climate change, is generally thought to increase with the strength of this resemblance. What has turned out much more difficult is to put this idea in quantitative terms. How well should a model mimic reality to be believed in? Which aspects of model behaviour are most important?

There are two broad aspects of model behaviour that can be evaluated against observations: (i) the ability of a model to

simulate the present-day climate and (ii) its ability to simulate externally forced climate changes and variations. For estimating the reliability of models in simulating future climate changes, the latter form of evaluation provides, in principle, a more direct test. On the other hand, as discussed in Sections 5 and 6, internal climate variability and uncertainties in external forcing and observations put severe limitations on how well the simulation of past changes can be used as a test for the future.

This section focuses on model evaluation based on present-day climate characteristics. First, some general aspects of the evaluation problem are discussed (Section 3.1). After that, a limited set of examples of the performance of state-of-the-art AOGCMs in simulating the present-day time mean surface climate are given (Section 3.2). Finally, in Section 3.3, the most difficult part of the issue is discussed, namely, what model skill in the simulation of present-day climate tells us about the skill in the simulation of anthropogenic climate changes.

3.1. General discussion

The meteorological literature includes countless examples of studies evaluating the simulation of present-day climate in individual models. However, the accumulation of knowledge has been greatly advanced by various model intercomparison projects (MIPs). Major examples include the Atmospheric Model Intercomparison Project, AMIP (Gates, 1992; Gates et al., 1999), the Coupled Model Intercomparison Project, CMIP (Meehl et al., 2000; Covey et al., 2003), and a recently started ambitious exercise known unofficially as the IPCC AR4 intercomparison (http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php), which was motivated by the need to gather material for the upcoming Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC). There have also been several dozen other, more specialized MIPs (see <http://www.ifm.uni-kiel.de/other/clivar/science/mips.htm> for a catalogue). By making output from a large number of models available to a large number of researchers, MIPs have made it much easier to study the common features and differences between models. Although the identification of a common error in several models does not necessarily mean that the cause of the error would be easy to identify, MIPs have also most likely accelerated the improvement of models by pointing out issues that need more attention from modellers.

For several reasons, evaluation of climate models is a challenging task. First, there are many aspects of model behaviour that should be compared with the real world. Boer (2000) divides these to three broad categories. The first is the *morphology of climate* as given by the spatial distribution and structure of means, variances, covariances, and possibly other statistics of basic climate parameters. Secondly, *budgets, balances and cycles* of quantities like energy, momentum and angular momentum should be evaluated. Thirdly, the information from the two previous categories should be complemented by *process studies of*

climate, which investigate particular aspects of the climate system such as the monsoons, blocking, convective processes, etc. In practice, most evaluation studies (including the one in Section 3.2 below) have focused on the morphology of climate, and particularly the time mean climatic conditions that are easiest to compare with observations. However, for estimating the reliability of model-simulated climate changes, detailed process-level comparisons of, for example, simulated and observed cloud behaviour may be even more important, since they give more direct information on the functioning of the subgrid scale parametrizations in the models (e.g. Bony et al., 2004).

Secondly, model evaluation is complicated by lack of and errors in observations. As illustrated by McAvaney et al. (2001) for time mean surface air temperature, precipitation and sea level pressure, differences between alternative observational data sets are in some cases almost as large as the differences between the best models and observations. In addition, both the observed climate and model simulations are affected by internal variability. For these reasons, not even a perfect model would be expected to agree completely with observations.

Thirdly, to a smaller or larger extent, models are tuned to reproduce the observed climate. This is inevitable because the parametrization schemes that are used for the description of unresolved processes include numeric constants that cannot be deduced accurately from theory or process-level observations. In the absence of other information, the choice of these constants tends to be guided by the ability of the model to simulate the observable aspects of present-day climate. However, tuning may introduce compensating errors. An evaluation of the present-day climate of a skillfully tuned model may therefore give a too optimistic impression of the process-level performance of the model, at least if the evaluation focuses on those variables that have been used most extensively in the tuning process.

The most heavily debated form of ‘tuning’ is artificial flux adjustments (Sausen et al., 1988), which are used in many models to keep the present-day distributions of sea surface temperature and salinity close to those observed. However, the need for flux adjustments has been gradually reduced by improvements made to the models. In contrast to the situation that still prevailed at the time of the IPCC Third Assessment Report, most of the models in the IPCC AR4 data set do not use flux adjustments.

Finally, although it is possible to use objective measures like the root-mean-square (rms) error and spatial or temporal correlation to characterize the agreement between models and observations for individual variables, there is no generally accepted figure of merit for measuring model performance as a whole. Together with the fact that different models show varying strengths and weaknesses, this makes attempts to “rank” models difficult. This issue is closely linked to the fact that the connection between model performance in the simulation of present-day climate and in the simulation of climate changes is still poorly understood (Section 3.3).

Table 1. Verification statistics from a global full-field comparison between simulated present-day annual mean climates in the IPCC AR4 models and the corresponding observational estimates. T(2m) = surface air temperature; Prec = precipitation; MSLP = mean sea level pressure. Mean, Min and Max are the average, lowest and highest values of the rms error and the spatial correlation among the 21 models. 21M gives the same statistics for the 21-model mean climate

	rms error				Spatial correlation			
	Mean	Min	Max	21M	Mean	Min	Max	21M
T(2m) (°C)	2.32	1.58	4.56	1.43	0.989	0.968	0.994	0.996
Prec (mm d ⁻¹)	1.35	0.97	1.86	0.95	0.775	0.597	0.867	0.872
MSLP (hPa)	3.96	2.06	8.33	2.65	0.880	0.497	0.984	0.945

3.2. Skill of models in simulating time mean surface climate

Among the several aspects of model results that should be evaluated, the most frequently studied is the time mean climate defined by multidecadal seasonal and annual means of meteorological variables. In particular, many studies have focused on variables such as surface air temperature, precipitation and sea level pressure that are also of major interest when considering the climate changes resulting from enhanced greenhouse gas forcing (e.g. Lambert and Boer, 2001; McAvaney et al., 2001; Covey et al., 2003). Here, I illustrate the performance of the IPCC AR4 models in simulating these three variables.

The model data set includes results from 21 models and is described in Appendix A. The observational data used in the comparison are detailed in Appendix B. Maps illustrating the comparison between the simulated annual mean climates and the corresponding observational estimates are shown in Fig. 1. In Table 1, the similarity between the simulated and observational fields is quantified by the spatial correlation and the rms error. Spatial correlations for seasonal means of climate are similar to those in Table 1, whereas the rms errors are typically 10–20% larger (not shown).

The first two rows of Fig. 1 reveal a high pattern similarity between the multimodel-mean simulated climate and the observed climate, particularly so for temperature and to a slightly lesser extent for sea level pressure and especially precipitation. The global spatial correlation is 0.996 for temperature, 0.95 for sea level pressure and 0.87 for precipitation (Table 1). The rms difference between the multimodel annual mean temperature and the observational estimate is 1.4°C, which is only 10% of the spatial Standard Deviation of the observational field. The values for precipitation and mean sea level pressure are 1.0 mm d⁻¹ and 2.7 hPa, respectively. These correspond to about 50% and 30% of the spatial Standard Deviations of the corresponding observational estimates.

The differences between the multimodel mean simulated and observed climates are shown in the third row of Fig. 1. The simulated temperatures vary on both sides of the observational estimate over the oceans but are slightly too low in most land

areas (Fig. 1g). A relatively large cold bias of about 6°C occurs in north-western Russia and over the Barents Sea. Analogously to the large regional variability of precipitation, the multimodel mean bias in precipitation also shows a rather complicated pattern (Fig. 1h). The multimodel mean sea level pressure is slightly too low in the higher mid-latitudes in both hemispheres (Fig. 1i), except for North America, and too high in the polar regions. This suggests a negative bias in the magnitude and an equatorward bias in the position of the simulated mid-latitude surface westerlies. However, the apparently large sea level pressure bias over Antarctica is difficult to interpret because sea level pressure over the high ice sheet may be sensitive to the method of extrapolating the pressure below the ground.

The multimodel mean biases hide substantial variation between the models. For all three variables, the Standard Deviation between the models exceeds the absolute value of the mean bias in most parts of the world (last two rows of Fig. 1). Thus, the biases are more random than systematic between the models. Areas where all 21 models are ‘wrong’ in the same direction (assuming that the observational estimate is correct) only cover from 2% (for temperature and sea level pressure) to 11% (precipitation) of the world. The unsystematic nature of the present-day biases gives some justification to the hope that climate changes in the future would also generally fall within the range of model projections, but with the important caveat that the tendency of model results to cluster around the observational estimates might result partly from tuning. This issue will be revisited in Section 5.2, where the model simulations are compared with recently observed climate changes.

On the other hand, because the biases in the individual models partly cancel each other in the multimodel mean, the multimodel mean fields give a too optimistic impression of the performance of individual models. In agreement with the findings of Lambert and Boer (2001), the biases in the individual model simulations are almost invariably larger than the biases in the multimodel mean fields. For both temperature and precipitation, although not for sea level pressure, the rms errors are lower and the spatial correlation higher for the multimodel mean than for any of the individual models (Table 1).

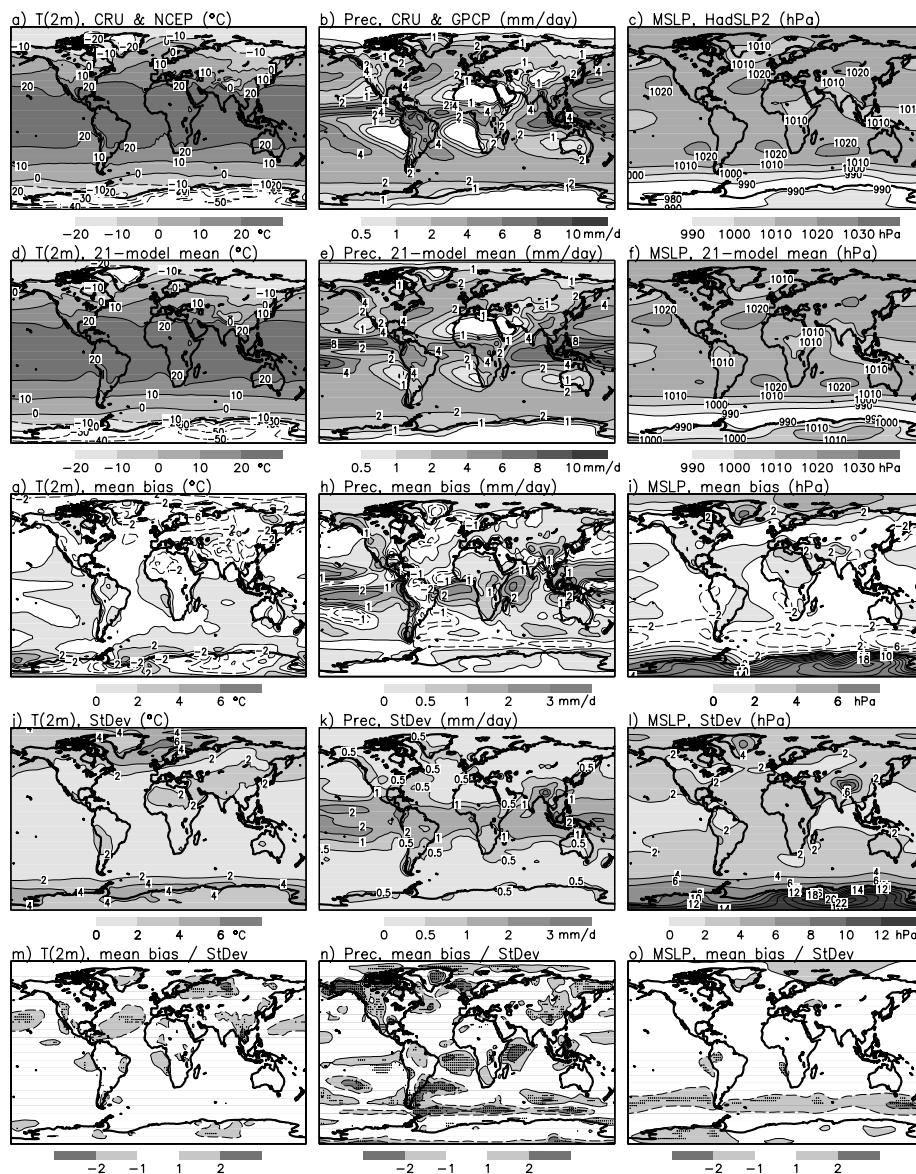


Fig. 1. Comparison between observational estimates of present-day annual mean climate and IPCC AR4 model simulations. Left: surface air temperature; middle: precipitation; right: mean sea level pressure. From top to bottom: observational estimates, multimodel means, average bias (multimodel mean minus observations), Standard Deviation between models, and ratio between the multimodel mean bias and Standard Deviation. The bottom panels also indicate with the dots the areas where the observational estimate is outside the range of the model simulations.

The global rms errors for temperature vary by a factor of 3, those for precipitation by a factor of 2 and those for sea level pressure by a factor of 4 among the 21 models (Table 1). However, the relative performance of the models depends on the variable considered. The intermodel cross-correlation between the rms errors in temperature and precipitation is 0.06, that for temperature and sea level pressure 0.13 and that for precipitation and sea level pressure 0.27. None of these correlations is statistically significant. This simple example illustrates the difficulty of ranking the models in any universal way.

3.3 What does the skill in the simulation of present-day climate tell about model reliability in simulating climate changes?

Climate changes are estimated from model simulations by comparing the simulated future climate with the simulated (rather than observed) present-day climate. This so-called delta change method is based on the assumption that biases in simulated present-day and future climates should tend to cancel each other, making the errors in the simulated climate changes smaller than those in the present-day climate. This assumption is supported by

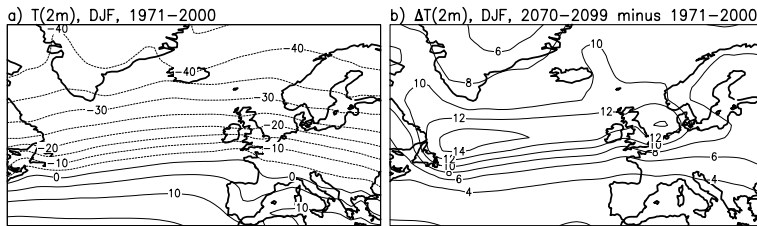


Fig. 2. (a) Winter (DJF = December–January–February) mean temperature ($^{\circ}\text{C}$) in the FGOALS-g1.0 model in the years 1971–2000 and (b) the change in DJF mean temperature in the same model from 1971–2000 to 2070–2099 under the SRES A1B scenario.

intercomparison between climate models (Kittel et al., 1998; see also Section 4): the climate changes simulated for the next hundred years differ in absolute terms substantially less between models than the present-day climate does. A precise simulation of the present-day values of the variable of interest may therefore not be as important for the simulation of climate changes as is sometimes suggested (e.g. Pan et al., 2001). A more crucial issue is the simulation of the feedback processes that regulate the response of the climate system to external forcing.

Nevertheless, because the realism of simulated feedback processes is more difficult to evaluate against observations than the time mean climate is, some studies have assumed a relationship between model skill in the simulation of time mean climate and climate changes (e.g. Giorgi and Mearns, 2002, 2003; Murphy et al., 2004). Such an assumption can be motivated by two lines of thought. First, large biases in the simulated present-day climate demonstrate that at least some processes are represented deficiently by the model. Yet, without an exact knowledge of the nature of the deficiencies, it is difficult to estimate how strongly they affect the simulated climate changes. This issue is further complicated by the risk that a skilfully tuned model might simulate at least some aspects of the present-day climate quite well even when there are large but compensating process-level errors.

Secondly, in specific cases, biases in present-day climate may have a direct impact on some feedback processes. A rather extreme example is given in Fig. 2. The model used for this figure overestimates the area of Northern Hemisphere sea ice cover by a factor of 3, with ice extending South to about 50°N in the North Atlantic Ocean (Zhang and Walsh, 2006). Thus, the simulated present-day winter climate in the North Atlantic and nearby areas is extremely cold. In a simulation forced by an increase in greenhouse gas concentrations following the SRES A1B scenario (Nakićenović and Swart, 2000), the ice edge retreats several hundreds of kilometres northwards by the end of the 21st century. This results in a narrow zone of very large (up to $12\text{--}14^{\circ}\text{C}$) warming at $50\text{--}60^{\circ}\text{N}$. None of the other models in the IPCC AR4 data set simulates a warming larger than $3\text{--}4^{\circ}\text{C}$ in this area. For another demonstration of the sensitivity of simulated climate changes to present-day sea ice conditions, see Hewitt et al. (2001).

A different example was discussed by Mitchell et al. (1987). Greenhouse-gas-induced warming in model simulations is accompanied by an increase in moisture transport by the Hadley

circulation, which tends to increase precipitation within the Intertropical Convergence Zone (ITCZ). Thus, if the ITCZ in the present-day simulation is mislocated, this is also likely to be the case with the simulated greenhouse-gas-induced increases in tropical precipitation. In this case, however, the connection between the present-day climate and climate changes is complicated by the fact that the atmospheric circulation, including the location and width of the ITCZ, may also change along with other changes in climate (Watterson, 1998; Neelin et al., 2006).

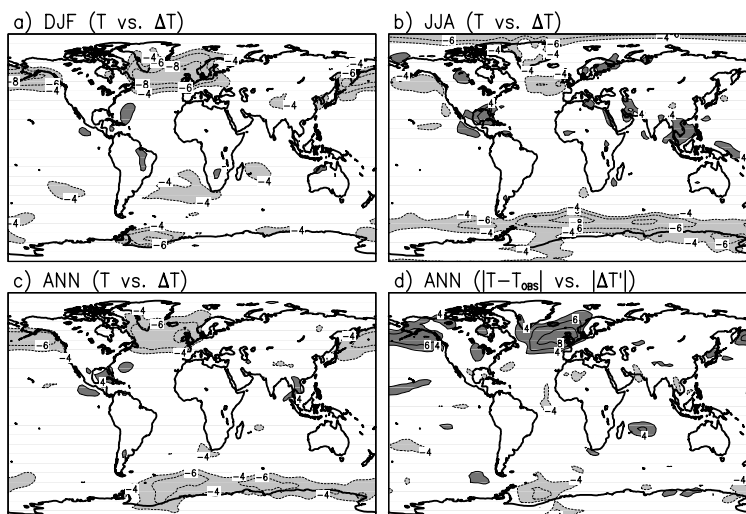
Although there are cases in which model response to anthropogenic forcing is clearly affected by deficiencies in the simulated present-day climate, it is hard to use this information for ranking models. Are, for example, biases in surface air temperature important compared with various sorts of other biases and differences between models?

A possible way to approach this question is to systematically study the relationship between control climate and climate change between different models. An example is given in Fig. 3. The first three panels show the cross-correlation between the late 20th century temperature and the simulated 21st century temperature change for the 21 AR4 models, separately for the Northern Hemisphere winter and summer and the annual mean. The shading begins from an absolute value of 0.4 which approximately represents the lower limit of statistical significance at the 10% level as estimated from a two-sided permutation test that does not require normally distributed data.

A strong negative correlation occurs in the northern North Atlantic, northern North Pacific and over the high-latitude Southern Ocean, particularly during the local winter. These are areas where some models have ice in the present-day climate and others not, and only the former, colder models can simulate a large warming as the ice retreats. In other parts of the world, the correlations are much weaker and their interpretation is complicated by the fact that some areas of apparently significant correlation will necessarily arise from pure chance. In total, the correlation is significant at the 10% risk level in only about 20% of the world (this number does not vary greatly with season).

The absence of linear correlation does not preclude more complicated relationships between two variables. As another test, the absolute value of the bias in the present-day mean temperature was correlated with the absolute difference between the climate change in a given model and the 21-model mean change. Now, strong positive correlations would mean that models that simulate the present-day climate badly tend to be outliers (i.e.

Fig. 3. Intermodel cross-correlation between the mean temperature in 1971–2000 and the change from 1971–2000 to 2070–2099 in (a) DJF, (b) JJA (June–July–August) and (c) annual mean. (d) Correlation between the absolute values of the annual mean temperature bias in 1971–2000 and the deviation from the multimodel mean temperature change. Values are multiplied by 10; contours at ± 4 , ± 6 and $\pm 8 \times 10^{-1}$. Dark (light) shading is used for correlations above 4×10^{-1} (below -4×10^{-1}).



far above or below others) in the simulated climate changes. However, as shown for the annual mean case in Fig. 3d, strong (although mostly positive) correlations were restricted to the same areas where the linear correlation between temperature and temperature change was strong. Thus, with the exception of high-latitude sea areas, there is little evidence of any relationship between simulated present-day mean temperatures and temperature changes. Studies based on earlier model simulations support this conclusion (e.g. Giorgi and Mearns, 2002).

The example in Fig. 3 illustrates one valuable use of multimodel ensembles: they allow an exploration of the relationships between model-simulated future climate changes and those aspects of present-day climate (or simulated past climate changes) that can be compared with observations. If robust relationships were found, they would be potentially useful for reducing the uncertainty in climate changes that might happen in the real world (Allen and Ingram, 2002). Yet, the search for such relationships is a complicated task that requires both statistical expertise and physical insight into the functioning of the climate system, and this task is not made easier by the limited number of quasi-independent climate models that are available for the analysis. Most of the research in this field has focused on the issue of global climate sensitivity, which will be discussed in Section 6.

4. Intercomparison between climate change simulations

4.1 General discussion

When run under similar scenarios of anthropogenic forcing, different climate models agree on many aspects of climate change but disagree on others. An investigation of what the models agree and disagree on has been both a major focus in the IPCC Assessments (Mitchell et al., 1990; Kattenberg et al., 1996; Cubasch et al., 2001) and the topic of a large number of individual journal papers. Examples of the latter include Grotch and MacCracken

(1991), Whetton et al. (1996), Kittel et al. (1998), Giorgi and Francisco (2000), Räisänen (2001), Giorgi and Mearns (2002), Covey et al. (2003) and Harvey (2004) – just to mention a few with a multimodel analysis of changes in surface air temperature and / or precipitation.

In contrast to the evaluation of simulated present-day climates, an intercomparison of climate changes between different models gives a quantitative estimate of uncertainty. However, there are caveats that complicate the interpretation of this estimate. The true uncertainty may be larger than the variation between existing model simulations indicates, but in some specific cases it might also be smaller.

The risk that the uncertainty in the real world exceeds the variation between model results is obvious: even if all models agreed perfectly with each other, this would not prove that they are right. From a more physical perspective, some authors have argued that the differences between the parametrization schemes used in existing models do not cover the actual uncertainty in the representation of subgrid scale processes (Allen and Ingram, 2002; Palmer et al., 2005). As a partial remedy to this, the so-called perturbed-parameter technique has been proposed. The initial results discussed in Section 6.1 indicate that model simulations based on this technique cover a wider uncertainty range than traditional multimodel ensembles, at least as regards the magnitude of the global mean temperature change.

The contrasting possibility that the variation of model results would exaggerate the true uncertainty relates to the fact that some models may be less credible than others. If models that appear as outliers in simulated climate change could be shown to be less credible than others, either because of a poor simulation of present-day climate or because of some other major weakness, then this would imply that these models should be downweighted or excluded when deriving estimates of uncertainty. An obvious case of this situation was shown in Fig. 2 but (as discussed in Section 3.3), in general, the connection between control run biases and simulated climate changes seems much less clear.

4.2. Intercomparison of IPCC AR4 climate change simulations

How well do different climate models agree in their simulated response to anthropogenic forcing? This question is addressed here by using the IPCC AR4 simulations, but findings from earlier studies are also discussed where appropriate. All the results in this section are based on simulations made for the SRES A1B emission scenario, which is, in terms of the greenhouse gas emissions and the magnitude of simulated climate changes, in the midrange of the SRES scenarios (Nakićenović and Swart, 2000). The variation of the model results, as shown in this section, should therefore not be interpreted as a measure of total uncertainty in the real world, where the future evolution of greenhouse gas and aerosol precursor emissions is also an important issue (Cubasch et al., 2001). Furthermore, as discussed in Section 7.1, emissions uncertainty is not the only aspect of forcing

uncertainty, the remaining aspects of which are also likely to be underrepresented by the IPCC AR4 ensemble.

Statistics of annual mean temperature, precipitation and sea level pressure change in the 21 models are shown in Fig. 4. The changes are computed as differences in 30-yr mean climate between the years 2070–2099 and 1971–2000.

The patterns of multimodel mean climate change in the first row of Fig. 4 are remarkably similar to those in studies based on an earlier generation of climate models (Cubasch et al., 2001; Räisänen, 2001; Covey et al., 2003; Harvey, 2004). The simulated warming (Fig. 4a) is at a maximum over the Arctic Ocean, where it is amplified by a decrease in ice cover and thickness. With this exception, the warming is larger over the continents than over the oceans. Over the Southern Ocean and the northern North Atlantic, the surface warming is retarded by the deep vertical mixing in the ocean, which acts to warm the water well below the surface but keeps the temperature increase at the

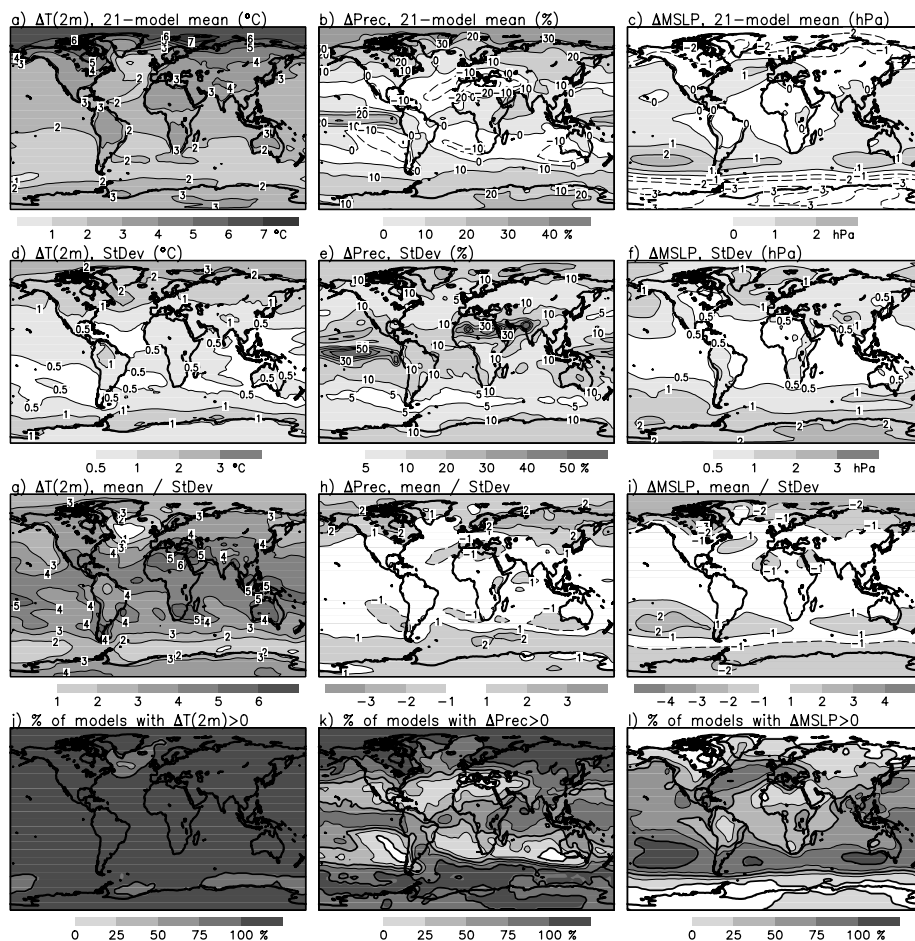


Fig. 4. Simulated changes in annual mean temperature (left; in $^{\circ}\text{C}$), precipitation (middle; in per cent of the mean in 1971–2000) and mean sea level pressure (right; in hPa) from 1971–2000 to 2070–2099 under the SRES A1B scenario. (a)–(c): 21-model mean changes; (d)–(f): Standard Deviation between models; (g)–(i): ratio between the mean change and the Standard Deviation; percentage of models with increase. Slight smoothing (j)–(l) is applied for legibility.

surface relatively modest. Elsewhere over the oceans, the vertical mixing plays a smaller role, but other processes keep the simulated warming smaller than that over the surrounding land areas. First, for a wet surface, increases in temperature tend to lead to an increase in evaporation that counteracts the warming (e.g. Hartmann, 1994). By contrast, the evaporation over most land areas is at least occasionally limited by lack of water, which makes this negative feedback less efficient. In some land areas, this is further exacerbated by a decrease in precipitation. Finally, the positive feedback associated with reduced snow cover acts to enhance the warming over the mid- to high-latitude continents but not over ice-free oceans. Because of these factors, a distinct land–sea contrast in warming also occurs in simulations of the equilibrium climate response to a doubling of CO₂, in which the larger heat capacity of the oceans plays no role (Manabe et al., 1991).

The simulated multimodel mean precipitation increases in high latitudes in both hemispheres and in most areas in the tropics, but decreases in many areas in the subtropics and in the lower mid-latitudes (Fig. 4b). Much of the large-scale pattern in Fig. 4b can be related to the increased moisture transport capacity of a warmer atmosphere, which, in the absence of other changes, tends to make differences between precipitation and evaporation larger in a warmer climate (Manabe and Wetherald, 1987; Mitchell et al., 1987). Where precipitation exceeds evaporation in the present climate, such as in the polar regions and in the ITCZ, the increased moisture transport acts to increase precipitation. The reverse happens where evaporation exceeds precipitation, particularly over the subtropical oceans. However, other mechanisms such as changes in atmospheric circulation and relative humidity modify this pattern, making the changes in precipitation difficult to predict in detail (Watterson, 1998; Rowell and Jones, 2006).

Multimodel mean sea level pressure decreases in polar regions, with a larger decrease in the South than in the North (Fig. 4c). Compensating increases in pressure take place in the lower mid-latitudes, although the belt of increases extends round the latitude circle only in the Southern Hemisphere. There is no simple physical argument to tell how sea level pressure should change with increased greenhouse gas forcing, but Yin (2005) relates the meridional pattern seen in Fig. 4c to an expansion of the Hadley circulation and a poleward shift of the mid-latitude storm tracks. In the Northern Hemisphere, in particular, there is also a hint of a monsoon-type pressure response in the east–west direction; the land–sea contrast in the warming being reflected in a slight relative decrease in pressure over the continents compared to the oceans.

The second row of Fig. 4 shows the Standard Deviations of temperature, precipitation and pressure change between the 21 models. From a comparison of Fig. 4d with Fig. 1j and Fig. 4f with Fig. 1l, the absolute intermodel differences in temperature change and pressure change are smaller (typically by at least a factor of 2) than the corresponding differences in present-day

mean climate. The same is also true for precipitation changes, although the different units in Figs 4e and 1k preclude a visual comparison. This feature, which was already discussed by Kittel et al. (1998), suggests that the delta change method works approximately as expected: models with a cold present-day climate also tend to be cold in terms of the climate simulated for the late 21st century (at least when compared with other models), and so on. However, for both temperature and sea level pressure (and also for precipitation, when the present-day values and the changes are expressed in the same units), there is a pattern similarity between the Standard Deviation of changes and the Standard Deviation of present-day values. In this sense, areas that are difficult for models in the simulation of present-day climate also appear to be difficult in the simulation of climate changes.

The last two rows of Fig. 4 characterize the relative agreement between the 21 models by two measures, the ratio between the all-model mean change and the Standard Deviation and a simple sign-of-change count. According to both measures, the models agree much better on changes in temperature than in the other two variables. In 95% of the global area, temperature increases in all individual models. By contrast, there are only 17% (12%) of areas where all 21 models agree on the sign of precipitation (sea level pressure) change.

For temperature changes, the ratio between the mean and the Standard Deviation is lower in high than in low latitudes. Both the average warming and the differences between the models tend to increase from the tropics to high-latitude areas, but the increase in the differences is larger. For precipitation changes, however, the same measure of relative agreement is at a maximum in high latitudes, particularly the high-latitude Southern Ocean and in northern Eurasia and northern North America. This relatively good agreement presumably reflects the relatively direct thermodynamic link from increased temperature to increased moisture flux convergence and precipitation in high-latitude areas (Manabe and Wetherald, 1987). A poleward shift of mid-latitude storm tracks, which appears to occur quite consistently in the IPCC AR4 models (Yin, 2005), may also contribute to the intermodel agreement on increasing high-latitude precipitation.

As a further illustration, scatter plots of simulated temperature and precipitation change in a single grid box in southern Finland (60°N, 25°E) are shown in Fig. 5, both for the annual mean and for the winter and the summer seasons. In this grid box, all 21 models agree on an increase in temperature in all seasons and on at least a slight increase in winter and annual mean precipitation. Summer precipitation increases in 17 out of the 21 models. Nevertheless, even in those cases in which all models agree on the sign of the change, quantitative differences between the models are considerable. Unsurprisingly, the intermodel variation tends to be larger for the seasonal than for the annual mean changes, since some of the seasonal differences between the models are averaged out in the annual mean.

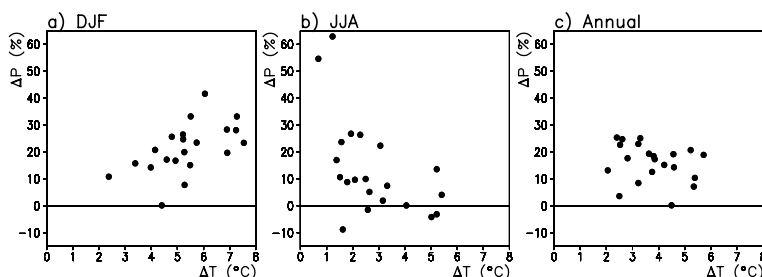


Fig. 5. Changes in temperature (horizontal axis) and precipitation (vertical axis) in southern Finland (60°N, 25°E) from 1971–2000 to 2070–2099 under the SRES A1B scenario, as simulated by 21 IPCC AR4 climate models. (a) DJF, (b) JJA, and (c) annual mean.

4.3. Further remarks

A factor that needs attention in the interpretation of model-simulated climate changes is internal climate variability, which may either add to or subtract from the change that would result directly from the forcing applied in the simulations. Even if all models shared the same noise-free response to anthropogenic forcing, the simulated climate changes would still be somewhat different because of the different realizations of internal variability.

For strong enough forcing and multidecadal averages such as those used in Fig. 4, directly model-related differences in climate change generally dominate over the effects of internal variability (e.g. Giorgi and Francisco, 2000; Räisänen, 2001). However, for weaker forcing comparable with the increase in greenhouse gas concentrations expected within the next few decades the situation is different: the directly model-related differences in climate change are smaller and the noise associated with internal variability is therefore relatively more important (Räisänen, 2001). Similarly, relative agreement between model simulations decreases with decreasing forcing, although absolute intermodel differences in climate change are smaller for weak than for strong forcing.

Both the agreement between climate change simulations and the importance of internal variability depend on the horizontal scale considered. With increasing geographical averaging, internal variability decreases. Partly for this reason and partly because of a decrease in directly model-related differences in climate change, the agreement between model simulations tends to increase with increasing scale. Räisänen (2001) found this effect to be particularly pronounced for changes in precipitation, with much more intermodel agreement on the large-scale features of precipitation change than on the local details. Studies that have focused on climate change on the subcontinental scale ($\sim 10^7$ km²) (e.g. Giorgi and Francisco, 2000; Ruosteenoja et al., 2003) therefore tend to give a somewhat too optimistic idea of how well the models agree at their smallest resolved scales.

The variation of climate changes between different models does not mean that the models would not give potentially useful information, but the uncertainty implied by this variation needs to be taken into account. Accordingly, several authors have proposed that climate change projections should be used

in a probabilistic way. By applying a simple decision-making model, Räisänen and Palmer (2001) and Palmer and Räisänen (2002) showed that probabilistic climate change projections may have considerable value even for variables such as precipitation for which the agreement between different models is relatively low. Because these studies used a cross-verification framework which implicitly assumes that the probability distribution of climate changes will coincide with the distribution of model results, the real value of the forecasts would be lower if models turned out to be more similar to each other than to reality. However, this caveat is not expected to affect the general conclusion that a probabilistic approach is likely to give more valuable information than deterministic approaches based on using the results of a single model or the multimodel mean change.

The best way to convert available model results to probability estimates is, however, still an open question. Several methods have been proposed but their relative strengths and weaknesses are still poorly known. Räisänen and Palmer (2001) and Palmer and Räisänen (2002) derived probabilities by a simple count-of-models method, assuming that all models deserve the same weight in the calculations. By contrast, Giorgi and Mearns (2003) weighted models using a performance criterion based on the simulation of the present-day climate and a (arguably hazardous; see Lopez et al., 2006) convergence criterion based on the proximity of their simulated climate changes to the all-model mean change. More recently, Tebaldi et al. (2005) and Greene et al. (2006) used Bayesian methods to derive continuous probability distributions of climate change. Still another method was developed by Harris et al. (2006), who used a perturbed-parameter ensemble rather than a traditional multimodel ensemble as the basis in their calculations.

5. Skill of models in simulating observed climate changes

The concentrations of CO₂ and several other greenhouse gases have increased throughout the industrial era with most of this increase taking place during the last few decades. At the same time, changes have occurred in the global climate, including an increase of about 0.6°C in the global mean surface air temperature during the last hundred years (Folland et al., 2001a). The ability of models to reproduce the observed climate changes

provides, in principle, a more direct test of model reliability than either the simulation of present-day climates (which has only an indirect connection to the simulation of climate changes) or intercomparison of climate change projections between models (which may understate the uncertainty if models are more similar to each other than to the real world).

In practice, there are important complications. First, in both models and in the real world, externally forced climate changes are accompanied by internal variability. A comparison between simulated and observed climate changes gives therefore no exact answer to whether the forced changes agree.

The second complication is the lack of reliable observations. The time-series of the global mean surface air temperature was analysed in some detail by Folland et al. (2001a,b). These authors accompanied their best-estimate warming trend of 0.6°C during the 20th century with a $\pm 0.2^{\circ}\text{C}$ uncertainty (covering ± 2 Standard Errors around the mean), with most of the uncertainty resulting from changes in observation techniques and the sparseness of the observational network during the early part of the record. For many other variables such as precipitation (e.g. Groisman and Easterling, 1994) and temperature in the free atmosphere (Thorne et al., 2005), difficulties associated with changes in instrumentation and in the interpretation of the measurements make uncertainties in observed changes even larger.

The third complication concerns external forcing. Although the radiative impact of increased greenhouse gas concentrations is known with a good accuracy, there are large uncertainties in other sources of climate forcing, particularly the direct radiative effects of anthropogenic aerosols and the impact of aerosols on clouds (Ramaswamy et al., 2001). Thus, differences between simulated and observed climate changes could result both from a misrepresentation of the external forcing in the model and from a misrepresentation of the feedback processes that regulate the climate response to a given forcing. Alternatively, errors in the representation of the forcing and the feedbacks might compensate each other, resulting in a misleadingly good simulation of observed climate changes. Although these problems are most obvious in the case of individual model runs, they might also affect the interpretation of multimodel ensembles if the simulations collectively misrepresent some component of the forcing such as the impact of anthropogenic aerosols (e.g. Anderson et al., 2003).

Because internal climate variability increases towards smaller horizontal scales, most of the published comparisons between observed and simulated climate changes have focused on changes in the global mean temperature and other large-scale aspects of climate. Some of these large-scale studies are reviewed in Section 5.1. However, the ability of models to simulate observed regional and local climate changes is also of interest. This topic is studied in Section 5.2 using data from the IPCC AR4 simulations.

In addition to the long-term evolution of climate during the 20th century, forced climate variations on shorter time-scales

(such as in the next few years after large volcanic eruptions) and in the pre-instrumental past provide opportunities for testing climate models. Short-term and pre-instrumental climate variations are excluded in this section, but they will be discussed in Section 6 in the context of the global climate sensitivity issue.

5.1. Large-scale studies

Climate models have successfully simulated many global-scale aspects of the climate changes observed during the instrumental period. Stott et al. (2000) showed that their model simulated the 20th century evolution of the global mean surface air temperature remarkably well. They found the warming during the early 20th century to have been mainly caused by changes in solar and volcanic activity, but the warming in the second half of the century was only reproduced when the simulation included the anthropogenic increase in greenhouse gas concentrations. Studies with other models support both these findings (Broccoli et al., 2003; Meehl et al., 2004b; Knutson et al., 2006), even though internal climate variability could also have played a substantial role in producing the warming in the early 20th century (Delworth and Knutson, 2000). However, given the large uncertainties in forcing, particularly the anthropogenic aerosol forcing, there is a risk that the close agreement between many models and observations is partly fortuitous. In principle, the right magnitude of warming could also be obtained from simulations in which too great (or small) negative aerosol forcing balanced too large (or small) model sensitivity (Schwartz, 2004).

In addition to surface air temperature, good agreement between simulations and observations has been found for the changes in several other, more or less directly temperature-related indicators of the global climate. For example, Ramaswamy et al. (2006) found a high degree of similarity between the simulated and observed evolution of global lower stratospheric temperatures during the past 25 yr. The gradual cooling trend at this height appears to have been mainly caused by ozone depletion, with increases in CO_2 and other well-mixed greenhouse gases amplifying the cooling and big volcanic eruptions leading to episodes of intermittent warming. Increases in average tropopause height, which are consistent with a warming of the troposphere and a cooling of the stratosphere, also agree between simulations and observations (Santer et al., 2004). Good agreement between model simulations and observations has likewise been reported for decreases in Arctic Ocean ice cover in the past 30 yr (Gregory et al., 2002a) and changes in water temperature within the top 700 m of the world's oceans since 1960 (Barnett et al., 2005; Pierce et al., 2006).

For precipitation, poor data coverage over the oceans before the satellite era largely limits the comparison between observed and model-simulated changes to land areas. Simulated variations in the global land area mean precipitation on interannual to interdecadal scales agree qualitatively with observations (Gillett et al. 2004; Lambert et al. 2004, 2005), although models appear

to underestimate the magnitude of the variations. However, both the observed and the simulated variations on these time-scales seem to be affected more strongly by volcanic aerosols than anthropogenic forcing. Nevertheless, observation-based estimates of precipitation trends in the 20th century, with increases in most land areas in high latitudes and in the tropics and decreases in the subtropics (Hulme et al., 1998; Folland et al., 2001a), bear some similarity to the response of models to increasing greenhouse gas concentrations. Observed increases in heavy precipitation in many parts of the world (e.g. Groisman et al., 2005) also appear to be consistent with increases that are expected to occur with warming, although it is unclear whether these changes are distinguishable from natural variability (Kiktev et al., 2003).

Climate models predict an increase in the atmospheric water vapour content with increasing temperature, approximately proportional to the increase in saturation humidity which is about 7% for each 1°C warming (e.g. Hartmann, 1994). Because water vapour is a powerful greenhouse gas, this is expected to provide a strong positive feedback effect that amplifies temperature changes. Changes in upper tropospheric water vapour are particularly important for the greenhouse effect but also most difficult to infer from observations, and some authors have questioned the ability of models to simulate the processes that regulate these changes (Lindzen, 1990; Lindzen et al., 2001). However, recent analysis of satellite measurements suggests that upper tropospheric water vapour has increased during the past two decades approximately at the same rate as predicted by models (Soden et al., 2005). Satellite measurements also indicate that the total atmospheric water content (which is dominated by water vapour in the lower troposphere) over the oceans has increased at a rate consistent with model predictions of unchanged relative humidity, whereas radiosonde measurements suggest increases in many but not in all land areas (Trenberth et al., 2005).

Some recently observed climate changes have been more difficult to reconcile with model results. The warming during the last half-century has been accompanied by a decrease in the diurnal temperature range (DTR) in most land areas. Vose et al. (2005) reported, for the years 1950–2004 and a domain covering 71% of the global land area, almost 50% larger an increasing trend in average daily minimum ($0.20^{\circ}\text{C}/10\text{ yr}$) than in maximum ($0.14^{\circ}\text{C}/10\text{ yr}$) temperatures. Although climate models also simulate a decrease in land area mean DTR when forced by anthropogenic changes in greenhouse gases and aerosols, this decrease is much smaller than that observed (e.g. Braganza et al., 2004). Braganza et al. linked the difference to observed increases in cloud cover not captured by the model simulations. Note, however, that the observed decrease in DTR was much faster in 1950–1980 than after 1980, when the minimum and the maximum temperatures have increased at almost the same rate (Vose et al., 2005).

During the last 30–50 yr, there has been a decrease in mean sea level pressure in both polar regions, particularly in the boreal winter, and an associated increase in westerly winds in the higher mid-latitudes. The changes in the Southern Hemisphere

appear to be consistent with model results and attributable to a combination of increased greenhouse gas concentrations and stratospheric ozone depletion, but in the Northern Hemisphere the observed changes are much larger than (although to the same direction as) on the average simulated by models (Gillett et al., 2005). Whether the difference in the Northern Hemisphere represents a real discrepancy (resulting from a misrepresentation of either external forcing or internal processes in climate models) or is explainable by internal climate variability is still an open issue (Selten et al., 2004). However, the idea that models may underestimate the sensitivity of the Northern Hemisphere mid-to-high-latitude circulation to greenhouse gas forcing is favored by the fact that models also tend to underestimate the increase in mid-latitude surface westerlies observed after large volcanic eruptions (Miller et al., 2006). Miller et al. suggested that this may be caused by a too weak coupling between stratospheric and tropospheric processes in the models.

A third, intensely debated case of possible model-observation discrepancy concerns differences in temperature trends between the surface and the free lower-to-mid troposphere. At the time of the IPCC Third Assessment Report (Folland et al., 2001a), it appeared that the globally averaged warming since the beginning of satellite measurements in 1979 had been much smaller in the free troposphere than at the surface, which was very difficult to reconcile with model simulations. Since then, however, new sources of error have been found from the satellite and radiosonde records used to estimate the temperature trends in the free troposphere (CCSP, 2006). It now appears likely that the free troposphere has been warming at approximately the same rate as the surface since 1979, whereas radiosonde records suggest a slightly larger warming in the free troposphere than at the surface since 1958. Thus, the suspected discrepancy between models and observations appears to have been mostly dissolved. However, there is still some discrepancy in the tropics, where most of the available observational data sets suggest a slower warming in the free atmosphere than at the surface since 1979, whereas all model simulations indicate that the warming should have been larger aloft.

5.2. Regional climate changes during the last half-century

To complement the picture obtained from the large-scale studies, a comparison between observed and simulated linear trends in temperature, precipitation and sea level pressure during the years 1955–2005 is presented in Fig. 6. This 50-yr period was chosen for analysis considering both the reliability of observational data (which is expected to be best for the last few decades) and signal-to-noise ratio issues (the shorter the period, the more the trends are affected by internal variability).

The maps in the first row show observational estimates of annual mean climate trends from 1955 to 2005; for the reasons detailed in Appendix B, the temperature and precipitation trends

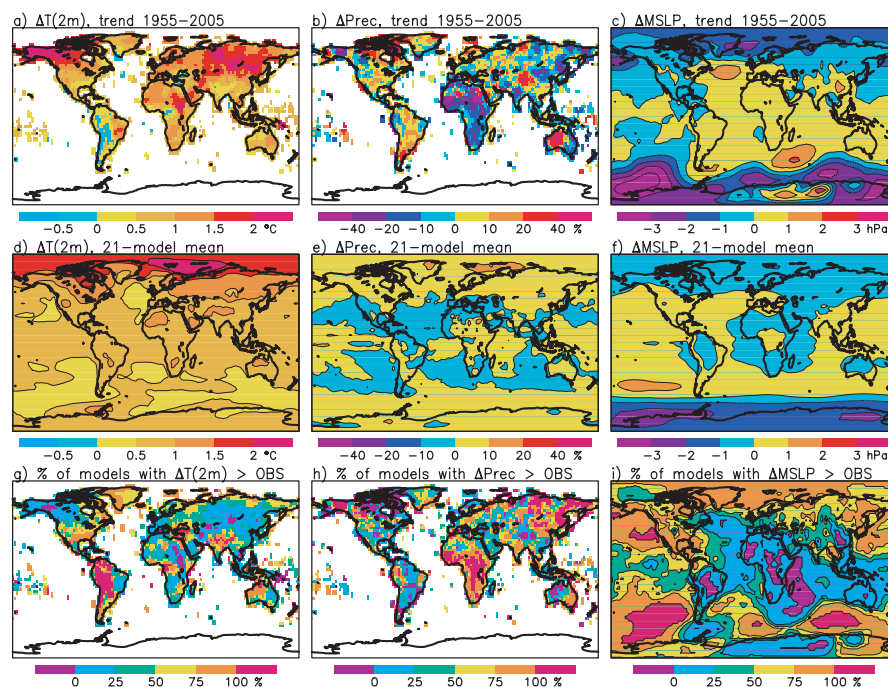


Fig. 6. Comparison between observed and simulated linear trends in annual mean climate from 1955 to 2005. (a)–(c) Observed changes in annual mean temperature (left; in $^{\circ}\text{C} / 50 \text{ yr}$), precipitation (middle; per cent / 50 yr) and mean sea level pressure (right; hPa / 50 yr); (d)–(f): 21-model mean changes; (g)–(i): percentage of models with the change exceeding the observed change.

are only given for land areas including islands but excluding Antarctica. The second row depicts the corresponding 21-model mean changes from the IPCC AR4 ensemble. A visual comparison of the two sets of maps reveals some similarity but also many differences. The fact that the models simulate the observed climate trends imperfectly is to be expected, if not for other reasons, simply because the observed trends are affected by internal climate variability. Internal variability also affects the trends simulated by the models, but most of it is averaged out in the multimodel means. As a result, the multimodel mean trends in Fig. 6 have much smoother geographical patterns and smaller local extremes than that the observed trends have.

The spatial correlation between the observational estimates and the multimodel mean trends is 0.48 for temperature and 0.23 for precipitation (over the area covered by observational data), and 0.56 for mean sea level pressure (over the whole world). Thus, the simulations capture only a relatively small part of the geographical variation in the observed trends. On the other hand, the simulated overall level of warming is similar to that observed. Accordingly, the uncentred correlation (e.g. Barnett and Schlesinger, 1987) between the simulated and observed temperature trends is as high as 0.88 (for trends in the other two variables, the centred and uncentred correlations are almost identical). Thus, the agreement between the simulated and observed trends is better for temperature than that for precipitation and sea level pressure.

Climate changes in the individual simulations are more strongly affected by internal variability than the multimodel

mean changes, have more irregular geographical patterns, and are in most cases less similar to the observed changes than the multimodel means. For example, the spatial correlation between the simulated temperature trends in the individual models and the observed trend varies from 0.10 to 0.52 with a mean of 0.30, exceeding the correlation for the multimodel mean (0.48) in only two out of the 21 models.

Although the observed trends and the multimodel mean trends differ, the observed trends are in most areas within the range covered by the individual model simulations (bottom row of Fig. 6). The observed temperature trend falls outside the range of model results in only 12% of the verification area, whereas the corresponding numbers for precipitation and sea level pressure are 23% and 14%. For comparison, if one assumes the observed trends to be a member of the same statistical population as the model simulations, the expected value of this fraction is $2 / (21 + 1) = 9\%$. These results lend some support (at least for temperature and sea level pressure) to the idea that the variation of climate changes between model simulations may be a reasonable measure of uncertainty.

Nevertheless, the past is not a perfect analogy for the future, and the conclusions made regarding the ability of multimodel ensembles to capture the uncertainty in future climate change must therefore be regarded as tentative. A key issue in this context is the fact that the climate changes projected for the rest of this century are much larger than those during the past 50 yr. As a result, the sources of uncertainty for the past and the future are different: the relative contribution of internal variability to the

uncertainty in the future will be smaller and that of modelling uncertainty larger. The fact that the climate changes observed during the last 50 yr were in most parts of the world within the range of the multimodel ensemble does not prove that this would also be the case for the larger changes expected in the future.

6. Global climate sensitivity and time-dependent global warming

Although the practical impacts of climate change depend primarily on regional climate changes, the change in the global mean surface air temperature is an important parameter for at least two reasons. First, model simulations suggest that the magnitude of regional climate changes increases quasi-linearly with the change in global mean temperature (e.g. Santer et al., 1990; Mitchell et al., 1999; Huntingford and Cox, 2000; Mitchell, 2003; Harvey, 2004), at least for forcing scenarios dominated by increased greenhouse gas concentrations. Secondly, changes in global mean temperature are expected to have a high signal-to-noise ratio, so that the effects of increased greenhouse gas concentrations and other external factors should be relatively easy to discern from internal variability. The latter property makes the change in global mean temperature a useful parameter for testing climate models.

Uncertainties in the projections of global mean warming result both from forcing uncertainty (evolution of greenhouse gas concentrations etc.) and from modelling uncertainty. This section focuses on the latter. The modelling uncertainty is most commonly discussed in terms of the *equilibrium climate sensitivity* (ECS). ECS denotes the change in global mean surface air temperature caused by a doubling of atmospheric CO₂ concentration, with no changes in other external factors and once the climate has had sufficient time (at least several centuries after the increase in CO₂ levelled off) to reach a new statistical equilibrium. ECS estimates from model simulations are discussed in Section 6.1, and efforts to estimate this parameter from observations are reviewed in Sections 6.2 and 6.3. However, while ECS gives information on the long-term response of the climate to increased greenhouse gas forcing, the time-evolution of the global mean temperature is also affected by the rate at which heat is consumed to warming ocean water. Observation-based estimates of the transient response of the global mean temperature to increasing greenhouse gas concentrations are discussed in Section 6.4.

6.1. Equilibrium climate sensitivity: estimates from models

As early as in the late 1970s, it was concluded that ECS would probably be within the range 1.5–4.5°C (Charney, 1979). This range, initially based on the results of just two models, was still quoted in the IPCC Third Assessment Report (Cubasch et al.,

Table 2. Statistics of equilibrium climate sensitivity (°C) in the models used in the IPCC assessments. The numbers for the Fourth Assessment Report (2007) are preliminary and may be subject to changes. N = number of models; StDev = Standard Deviation

Year	N	Mean	StDev	Range
1990	17	3.7	1.0	1.9–5.2
1996	15	3.8	0.8	1.9–5.2
2001	17	3.5	0.9	2.0–5.1
2007	18	3.2	0.7	2.1–4.4

2001) despite being slightly different from the actual range of model results available at that time.

Statistics of ECS for the models used in the IPCC Assessments (Mitchell et al., 1990; Kattenberg et al., 1996; Cubasch et al., 2001) are given in Table 2, including also preliminary numbers for the Fourth Assessment Report to be published in the year 2007 (Gerald Meehl, private communication). Although there are slight hints of a decrease in both the multimodel mean and the intermodel variation of ECS with time, it is remarkable how modest these changes have been, despite the considerable amount of work put in improving the models during this 16-yr period.

The numbers in Table 2 are based on climate models developed at a large number of research institutions. However, because modellers share ideas and in some cases even full model components with each other, the individual models are at best quasi-independent. This together with the relatively limited number of models implies that the variation of climate changes within these multimodel ensembles may not capture the full range of modelling uncertainty. As an alternative, although equally incomplete way of exploring the uncertainty, two recent projects, Quantifying Uncertainty in Model Prediction (QUMP; Murphy et al., 2004) and ClimatePrediction.net (CPDN; Stainforth et al., 2005) have used a so-called perturbed-parameter technique. In this technique, which requires huge computing resources, a large number of versions of the same parent climate model are created by varying the numerical values used in parametrization schemes within their estimated uncertainty ranges.

Although the members of a perturbed-parameter ensemble share the same parent model, and therefore the same basic structure of subgrid scale parametrizations, this technique is able to create model versions with widely varying sensitivity. In particular, both the QUMP and CPDN ensembles include model versions with ECS much above the range found in conventional multimodel ensembles. The highest ECS within a 128-member ensemble of QUMP simulations was 7.1°C (Webb et al., 2006), whereas 4.2% of the over 2000 CPDN simulations documented by Stainforth et al. (2005) had ECS exceeding 8°C, with an absolute maximum of 11.5°C. However, evaluation of the CPDN ensemble against observations of present-day climate suggests that climate sensitivity is less likely to be extremely high than

would be inferred directly from this ensemble (Piani et al., 2005; Knutti et al., 2006; see Section 6.3).

A long-standing and still valid result, based on diagnostic studies of the global radiation budget in climate models, is that a majority of the variation in ECS is due to different changes in cloud amount, optical properties and altitude distribution in different models (Cess et al., 1990, 1996; Colman, 2003; Soden and Held, 2006; Webb et al., 2006). A more detailed analysis by Webb et al. (2006) indicated that, among the models participating in QUMP and the Cloud Feedback Intercomparison Project, changes in low-altitude (i.e. stratus and stratocumulus) clouds were a particularly important contributor to variations in climate sensitivity.

As shown by Colman (2003) and Soden and Held (2006), feedback processes associated with changes in atmospheric water vapour content and vertical temperature distribution also vary in magnitude between climate models. However, their combined contribution to intermodel differences in ECS is substantially smaller than that of cloud feedbacks. Surface albedo feedback due to changes in snow and ice cover, although important for climate change in high latitudes, also appears to be relatively unimportant for the uncertainty in ECS.

6.2 *Estimating equilibrium climate sensitivity from observations: methodological issues*

In recent years, a great deal of research has been conducted to estimate climate sensitivity from observational data. These studies have used several approaches and several types of observations. The first approach uses, in one way or other, the equation

$$F = Q - \lambda \Delta T. \quad (2)$$

Here, F is the globally averaged rate of change in the energy content of the climate system (which is approximately equal to the net heat flux into the ocean), Q is radiative forcing, ΔT is the change in global mean temperature and λ a feedback parameter. ECS is related to λ by the equation

$$ECS = Q_0 / \lambda. \quad (3)$$

where $Q_0 \approx 3.7 \text{ W m}^{-2}$ is the radiative forcing caused by a doubling of the CO_2 concentration. From estimates of F , Q and ΔT , estimates of λ and ECS can be derived. If F can be neglected, which is justified when ΔT represents the temperature difference between two long-term mean climate states thought to be near equilibrium, only Q and ΔT are needed. However, this method is based on the assumption that λ is a universal constant, which is not exactly true. Model simulations indicate that λ depends on the nature of the forcing agent. Although this variation is small among different well-mixed greenhouse gases, it may be of the order of several tens of per cent for more horizontally and vertically inhomogeneous forcings (Hansen et al., 1997, 2005; Joshi et al. 2003). The basic state of the climate also affects many of the feedbacks that together determine climate sensitivity

(Senior and Mitchell, 2000; Boer and Yu, 2003). ECS estimates derived, for example, for glacial conditions may therefore not be fully representative for the current climate.

Another strategy to estimate ECS is to use climate models to find a statistical relationship between some observable quantity and the equilibrium global mean warming simulated by the same model in response to a doubling of CO_2 . In searching for such a relationship, simulations with different models or (more commonly) different versions of the same model are used. A spectrum of models ranging from GCMs through EMICs to simple energy balance models have been used in these studies. Note that GCMs and many EMICs simulate feedback processes explicitly and are therefore free of the assumption that λ should be the same for all forcing mechanisms and in all conditions. By contrast, energy balance models and the simplest EMICs treat λ as an external parameter that can be varied between different simulations but is independent of climate and the forcing mechanism.

Various types of observable quantities can be used. Many studies have used changes in surface air temperature during the instrumental era, in some cases together with the temperature changes observed in the oceans and in the atmosphere. Alternatively, estimates of temperature changes during earlier periods, such as during glacial–interglacial variations, can be used. Some studies have also attempted to infer ECS from the short-term global cooling following large volcanic eruptions. Finally, properties of the present-day annual mean climate and the seasonal cycle have been used in a few studies.

There are several difficulties. Uncertainty in observations needs to be taken into account even in studies based on the 20th century climate evolution but is even more important in studies based on paleoclimatic data. Furthermore, internal climate variability complicates the interpretation of the observed changes. This is a major problem particularly for studies based on the relatively small and short-term temperature changes following volcanic eruptions.

In studies based on temperature evolution during the instrumental period, the largest problem is the uncertainty in external forcing, particularly the poorly known extent to which the positive greenhouse-gas-induced radiative forcing has been compensated by negative aerosol forcing (Boucher and Haywood, 2001; Ramaswamy et al., 2001). To alleviate this problem, several studies (e.g. Andronova and Schlesinger, 2001; Forest et al., 2002, 2006; Gregory et al., 2002b; Harvey and Kaufmann, 2002) have attempted to estimate the magnitude of the aerosol forcing simultaneously with climate sensitivity, using information on the geographical and interdecadal variations of temperature. The forcing issue also affects studies based on volcanic eruptions and pre-industrial climate variations. Note, however, that ice core records of greenhouse gas concentrations (Petit et al., 1999; Siegenthaler et al., 2005) and information on the variation of Earth's orbital parameters (Berger, 1978) allow at least some components of the forcing to be quantified quite well for the past several hundred thousand years.

In studies based on 20th century temperature evolution and on climate response to volcanic eruptions, uncertainty in the magnitude of ocean heat uptake F is also an issue. Although rough observational estimates of F are available for the last half-century (Levitus et al., 2000, 2005; Hansen et al., 2006), these have seldom been used directly (for an exception, see Gregory et al., 2002b). Instead, most studies have relied on model simulations. However, in some models the parameters that regulate ocean heat uptake can be varied and their most probable values and uncertainty ranges can be estimated from observations (e.g. Forest et al., 2002, 2006).

Many studies have attempted to derive a probability distribution for ECS based on some goodness-of-fit statistics between observations and model simulations and taking into account uncertainties associated with forcing, observations and internal variability. However, because few if any studies have included all thinkable sources of uncertainty, the results of all individual studies should be viewed with some caution. Another subtle point is the dependence of the results on prior assumptions (Frame et al., 2005). Many of the probabilistic studies have used a Bayesian framework, in which an assumed prior probability distribution is modified using constraints derived from observations. Beginning from the prior assumption that all values of ECS between 0.17°C and 20°C are equally probable, Frame et al. derived for ECS a 5–95% uncertainty range of 1.2–11.8°C. By contrast, assuming in the beginning equal probability for all values of the feedback parameter λ for which ECS is in the 0.17–20°C range resulted in a 5–95% uncertainty range of 0.6–4.0°C. In studies based on more complex models that do not need λ or ECS as an

input parameter, the most common prior assumption is that all available model versions are equally likely (Murphy et al., 2004; Piani et al., 2005).

6.3. Observation-based estimates of equilibrium climate sensitivity

A number of observation-based studies on ECS are listed in Table 3. It is impractical to discuss all of them in detail but a few remarks are useful. First, most of the studies based on the decadal-to-centennial scale temperature evolution during the instrumental period have found very high (7.7°C or more for the 95th percentile) upper limits of ECS. This is mainly associated with the large uncertainty in aerosol forcing, as discussed above. Exceptions include Frame et al. (2005), but only when using the assumption that λ rather than ECS should be sampled uniformly in the prior distribution, and Harvey and Kaufmann (2002). The latter authors found ECS exceeding 3.0°C inconsistent with the relatively modest decadal-scale cooling that followed the Mount Krakatau eruption in 1881, but they noted that this conclusion may be sensitive to errors in the radiative forcing associated with this eruption. Except for Frame et al. (2005) with uniform sampling on λ , none of these studies indicates a substantial chance of ECS being below 1.0°C.

Of the four studies based on pre-instrumental climate variations and forcing estimates, Hoffert and Covey (1992) used data for the Last Glacial Maximum (LGM; 21 000 yr ago) and the Cretaceous warm period (about 100×10^6 yr ago), both Annan et al. (2005) and Schneider von Deimling et al. (2006) data for

Table 3. Estimates of equilibrium climate sensitivity from observational studies. The second column indicates the type of data used (I = temperature changes during the instrumental period; P = pre-instrumental climate variations; V = climate response to volcanic eruptions; C = present-day climatology)

Reference	Data	Median	Range
Andronova and Schlesinger (2001)	I	2.0°C	1.0–9.3°C (5–95%)
Harvey and Kaufmann (2002)	I	2.0°C	1.0–3.0°C (subjective)
Forest et al. (2002)	I	2.9°C	1.4–7.7°C (5–95%)
Gregory et al. (2002b)	I	6.1°C	1.6°C– ∞ (5–95%)
Frame et al. (2005)	I	Not given	1.2–11.8°C / 0.6–4.0°C (5–95%)
Forest et al. (2006)	I	3.6°C	2.1–8.9°C (5–95%)
Hoffert and Covey (1992)	P	2.3°C	1.4–3.2°C (mean $\pm$ 1 StDev)
Annan et al. (2005)	P	4.5°C	< 7% chance of ECS > 6°C
Schneider von Deimling et al. (2006)	P	Not given	1.2–4.3°C (5–95%)
Hegerl et al. (2006)	P	3.4°C	1.2–8.6°C (5–95%)
Wigley et al. (2005)	V	1.5 to 3.0°C	0.3 to 1.8–5.2 to 7.7 °C (2.5 - 97.5%)
Murphy et al. (2004)	C	3.5°C	2.4–5.4°C (5–95%)
Piani et al. (2005)	C	3.3°C	2.2–6.8°C (5–95%)
Knutti et al. (2006)	C	3–3.5°C	1.5 to 2–5 to 6.5°C (5–95%)
Forster and Gregory (2006)	C	1.6°C	1.0–4.1°C (5–95%)
Hegerl et al. (2006)	IP	2.8°C	1.5–6.2°C (5–95%)
Annan and Hargreaves (2006)	IPV	2.7°C	1.7–4.9°C (2.5–97.5%)

the LGM, and Hegerl et al. (2006) data for the years 1270–1850. The ECS estimates from these studies are by and large consistent with the studies based on instrumental data. Two of these four studies appear to preclude high climate sensitivities exceeding 4.5°C but the others do not.

Wigley et al. (2005) studied the evolution of the global mean surface air temperature after the three largest volcanic eruptions in the second half of the 20th century. Their best estimates for ECS, derived with an energy balance climate model from the cooling following the Agung, El Chichón and Pinatubo eruptions were 2.8°C , 1.5°C and 3.0°C , respectively. The 2.5–97.5% uncertainty ranges, accounting for internal climate variability but not for uncertainties in the model and magnitude of radiative forcing, ranged from 1.8 – 5.2°C for the strongest (Pinatubo) to 0.3 – 7.7°C for the weakest (El Chichón) eruption.

Of the four studies in Table 3 that used observations of present-day climate to constrain climate sensitivity, Murphy et al. (2004) used perturbed-parameter ensemble simulations from the QUMP project. They derived a somewhat arbitrary summary statistics measuring the similarity between simulated and observed climates across a wide range of variables, to weigh the different ensemble members when deriving their probability distribution. Piani et al. (2005) applied the same idea to the much larger CPDN ensemble, but used a more rigorous statistical model to find a relationship between the simulated time mean climate and ECS. This relationship, together with its statistical uncertainty, was used to derive a probability distribution for ECS. Knutti et al. (2006) also used the CPDN ensemble, but in their statistical method only the seasonal cycle of surface air temperature was used for linking present-day climate and ECS. The probabilistic ECS estimates obtained in these three studies were reasonably similar (5–95% ranges of approximately 2 – 6°C and best estimates of 3 – 3.5°C). By contrast, Forster and Gregory (2006) derived a lower range (1.0 – 4.1°C) and best estimate (1.6°C) from the seasonal cycles of temperature and top-of-the-atmosphere radiation balance, assuming that the same feedback processes that regulate the seasonal cycle also determine the long-term climate sensitivity.

Almost all studies in Table 3 appear to preclude a climate sensitivity of less than 1°C , and most of them agree on a best estimate of about 2 – 3.5°C . However, the upper end of the uncertainty range varies widely. If all studies included all relevant uncertainties, then the highest values for the upper limit of ECS (and the lowest values for the lower limit) found in some of the individual studies could be rejected simply because other studies find a narrower uncertainty range. Unfortunately, this condition is most likely not fulfilled.

Nevertheless, the authors of two recent studies (Annan and Hargreaves, 2006; Hegerl et al., 2006) argue that very high values of ECS are much more unlikely than suggested by some of the individual estimates. Annan and Hargreaves (2006) noted that the studies based on the warming during the instrumental period, the cooling following volcanic eruptions, and pre-instrumental cli-

mate variations use largely independent sources of information. Combining the uncertainty estimates from these three lines of research within a Bayesian framework, they concluded that ECS should be with a 95% probability within the range of 1.7 – 4.9°C . Similarly, Hegerl et al. (2006) found that the uncertainty in ECS can be narrowed by combining the information on climate variations in the instrumental and the pre-instrumental period. Yet, although these new studies suggest that ECS is quite likely to be within the uncertainty range derived directly from model results, they do not yet allow a narrowing of the directly model-based range.

6.4. *Observational constraints on time-dependent warming*

The wide uncertainty ranges obtained from most observational studies of ECS may seem discouraging. Fortunately, the relative uncertainty in the warming that would be experienced under a given forcing scenario during, for example, the rest of this century is probably smaller than that in the long-term equilibrium warming. This is a result of ocean heat uptake.

From simple theoretical arguments, ocean heat uptake is expected to retard the warming more for large than for small climate sensitivity (Hansen et al., 1985; Wigley and Schlesinger, 1985), making the initial rate of warming to depend less than linearly on ECS. In addition, the rate of warming in the 20th century has been affected by the combined effects of ECS and ocean heat uptake efficiency, rather than by ECS alone. Because the warming in this century will also be affected by both these factors, it is expected to be constrained more tightly by the 20th century temperature evolution than ECS is (Allen et al., 2000; Frame et al., 2005).

In one of the first studies presenting observationally constrained uncertainty estimates for the warming in the 21st century, Stott and Kettleborough (2002) found that the global mean warming from the decade 1990–2000 to the decade 2020–2030 would have a 90% probability to be between 0.3°C and 1.3°C . These numbers take into account both the uncertainty in anthropogenic warming and the forced and unforced natural variability, and they were found to be insensitive to the choice among the SRES emission scenarios. The warming estimates for the end of this century (2090–2100) depended much more on the emission scenario, the 5–95% uncertainty range for the SRES B1 (A1FI) scenario being 1.2 – 3.3°C (3.0 – 6.9°C). For both the B1 and the A1FI scenarios, the uncertainty ranges of Stott and Kettleborough (2002) extend higher than the corresponding model-based uncertainty ranges in the IPCC Third Assessment Report (Cubasch et al. 2001, fig. 9.14). In a more recent observational study, however, Frame et al. (2005) found only a small chance (a few per cent) that the warming from 1990 to 2100 would exceed 3.0°C under the B1 scenario or 5.0°C under the A1FI scenario.

In another study based on 20th century temperature changes, Stott et al. (2006a) derived a probability distribution for a

parameter known as transient climate response (TCR). TCR denotes the global mean warming that would occur at the time of the doubling of the CO_2 concentration, assuming that CO_2 is increased at a rate of $1\% \text{ yr}^{-1}$ compound, so that the doubling takes 70 yr. Although TCR is clearly an idealized number, it may be more closely linked to the rate of warming in the 21st century than ECS is. Stott et al. (2006a) found that TCR would have a 90% chance to be in the range $1.5\text{--}2.8^\circ\text{C}$. Model-based estimates of TCR are in good agreement with these numbers. The range of TCR among the models used in the IPCC Third Assessment Report was $1.1\text{--}3.1^\circ\text{C}$ (Cubasch et al., 2001), whereas the range for the IPCC AR4 data set is $1.3\text{--}2.6^\circ\text{C}$ (Gerald Meehl, private communication).

All the numbers given above refer to globally averaged temperature changes. Constraining future regional climate changes with available observations is more difficult. Using a technique similar to those used in the global studies discussed above, Stott et al. (2006) derived observationally constrained uncertainty ranges for 21st century temperature changes on the continental scale. For extratropical continents with strong internal variability, the derived uncertainty ranges were very wide: for example, the 5–95% range of warming in Europe from 1990 to 2100 under the SRES A2 emission scenario extended from 0°C to 11°C . For low-latitude continents with smaller variability, narrower but still considerable uncertainty ranges were derived. The authors note, however, that their method probably exaggerates the uncertainty by assuming that temperature changes on different continents are independent of each other.

As the increase in greenhouse gas concentrations continues, more precise observation-based estimates of future global mean warming (and other climate changes) will gradually become possible. A perfect-model test by Stott and Kettleborough (2002) suggests that the uncertainty in the global mean temperature change that will occur by the year 2100 could be potentially halved within the next two decades, as new observational evidence of the response of the climate system to increasing greenhouse gas concentrations accumulates.

7. Some additional issues

This section discusses briefly two additional issues that may make the uncertainty in future climate changes larger than that which is indicated by the variation of AOGCM-based projections. First, in addition to the obvious uncertainty in the magnitude of anthropogenic greenhouse gas and aerosol emissions, there are other uncertainties associated with climate forcing in the future, and these uncertainties are at most partially covered by existing ensembles of AOGCM simulations (Section 7.1). Secondly, in particular in areas of complex geography, some aspects of climate change might vary considerably on scales that are too small to be properly or at all resolved by AOGCMs (Section 7.2).

7.1. Uncertainties in forcing

The implications on future climate of different scenarios of greenhouse gas and aerosol emissions have been studied extensively. For projections of global temperature change during the 21st century, the emissions uncertainty appears to be of similar magnitude as the uncertainty associated with model differences. The ‘best-estimate’ warming from 1990 to 2100 in the IPCC Third Assessment report varied from 2.0°C for the lowest to 4.6°C for the highest SRES scenario (Cubash et al., 2001, fig. 9.14). For shorter-term projections, the emissions uncertainty is less important both because the emissions are better predictable in the near future and because of the long lifetime of many greenhouse gases and the inertia in the response of the climate system to changes in forcing (Cubash et al., 2001; Stott and Kettleborough, 2002). Regional, as well as global, climate changes are expected to increase in magnitude with increasing greenhouse gas emissions but the simulated geographical patterns of climate change appear to be less sensitive to differences between emission scenarios than to differences between climate models (e.g. Harvey, 2004).

Emission uncertainty is not synonymous to concentration uncertainty, which is also affected by the uncertainty in the physical, chemical and biological processes that regulate the sinks and natural sources of greenhouse gases and aerosols. These aspects of uncertainty are not presented well by existing AOGCM simulations, which generally use prescribed rather than interactively calculated greenhouse gas concentrations. The frequently quoted $1.4\text{--}5.8^\circ\text{C}$ uncertainty range for the global warming from 1990 to 2100 (Cubash et al., 2001) also excludes this uncertainty.

First, uncertainty in the modelling of the carbon cycle may be important (Cox et al., 2000; Friedlingstein et al., 2003, 2006; Andrae et al., 2005). In the Coupled Climate-Carbon Cycle Model Intercomparison Project (C⁴MIP) 11 models were used to perform coupled climate–carbon cycle simulations using CO_2 emissions for the SRES A2 scenario (Friedlingstein et al., 2006). The CO_2 concentrations in the year 2100 varied from 730 to 1020 parts per million in volume (ppmv), to be compared with a prescribed value of 836 ppmv in the IPCC AR4 simulations for this relatively high emission scenario. Intermodel differences in the feedback from climate change to carbon cycle explained about 60% (180 ppmv) of the range of CO_2 concentrations obtained in C⁴MIP, but differences in other processes such as in the CO_2 fertilization also contributed to this range. For comparison, the best-estimate CO_2 concentrations (as given in appendix II of Houghton et al., 2001) for the six illustrative SRES scenarios in the year 2100 vary approximately from 540 to 960 ppmv. These numbers suggest that the carbon cycle uncertainty is smaller than emission uncertainty but nevertheless a significant fraction of the latter.

Secondly, simulations of future climate change do not include all potentially significant forcing agents. For example, changes in land use are excluded in almost all the IPCC AR4 models,

although some of the SRES scenarios suggest substantial anthropogenic changes in land cover during this century. These changes are unlikely to be important for the evolution of the global mean temperature, but they might induce regional temperature changes of up to $\pm 2^\circ\text{C}$ in some areas and also affect other aspects of regional climates (e.g. Feddema et al., 2005). Almost all model simulations also exclude future changes in solar and volcanic activity, for the obvious reason that they are unpredictable. The externally forced part of natural variability is therefore lacking from the simulations. However, at least as far as century-scale changes in the global mean temperature are considered, changes in natural forcing are extremely unlikely to offset the anthropogenic increase in greenhouse gas concentrations (Bertrand et al., 2002).

Thirdly, there is substantial uncertainty in the direct and indirect climatic effects of anthropogenic aerosols. The extent to which existing multimodel ensemble simulations of future climate change cover this uncertainty is unclear. Although the treatment of aerosol forcing in model simulations has developed markedly in recent years, some significant components of the direct aerosol forcing (e.g. black and organic carbon and mineral dust) are excluded even in most of the IPCC AR4 simulations. Indirect effects of aerosols on cloud properties and lifetime are also included in only a minority of these models. It is important to note, however, that the SRES scenarios indicate a decrease in anthropogenic sulphate burden towards the end of the 21st century, in contrast with the substantial increase in greenhouse gas concentrations following from the same scenarios. Despite eventual increases in some other aerosol types (e.g. appendix II of Houghton et al., 2001), this suggests that aerosol uncertainty is relatively less important for projections of future climate change than for the interpretation of 20th century climate changes.

7.2. Variation of climate changes on small horizontal scales

The typical horizontal resolution of current AOGCMs is of the order of 250 km. The lack of resolution is widely regarded as a problem in the simulation of extremes, but it may also affect the simulation of time mean climate changes (e.g. Giorgi et al., 2001). A prominent example of the latter is precipitation changes in areas of complex topography (Fig. 7). Figs 7a and b show precipitation changes in northern Europe in a global cli-

mate model (Roeckner et al., 1999) and in an RCM driven by this global model (Räisänen et al., 2004). The RCM simulates a large (up to 70%) local increase in precipitation at the west coast of Norway and a sharp gradient in the change across the Scandinavian mountains. This reflects a large increase in westerly winds in this simulation (see Räisänen et al., 2004) which forces more orographic uplift at the western slopes of the mountain range. The global model shares essentially the same change in circulation but, reflecting the coarser resolution (about 300 km, in contrast to 49 km in the RCM) and consequently smoother topography, the maximum increase in precipitation is smaller and dislocated to the West. Clearly, the higher resolution gives in this case a more physically consistent solution. In fact, the 49-km resolution of this RCM may not be sufficient to fully reveal the combined effects of topography and circulation change on precipitation. Statistical downscaling (e.g. Hellström et al., 2001; Hansen-Bauer et al., 2003) indicates that precipitation changes may be even more variable on small spatial scales than that which is suggested by RCM simulations.

However, higher resolution alone is not a shortcut to smaller uncertainty. When driven by another global model (Hudson and Jones, 2002), which did not simulate an increase in westerly winds, the same RCM only simulated a 0–10% increase in precipitation in western Norway (Fig. 7c; see also Räisänen et al., 2004).

Most of the IPCC AR4 simulations do indicate an increase in westerly winds in northern Europe during this century, although the increase is generally smaller than that in the older simulations presented in Figs 7a and b. Thus, if these AOGCMs had higher resolution but simulated the same changes in circulation, they would on the average simulate a larger increase in precipitation in western Norway than they actually do. However, the variation of precipitation changes between the models would also be larger, because the higher resolution would make the precipitation changes more sensitive to the intermodel differences in circulation change.

A general message from this example is that the true uncertainty on small spatial scales is likely to be larger than that indicated by the variation of global model results, particularly for changes in precipitation but to some extent also for changes in temperature (e.g. Hansen-Bauer et al., 2003). For a precise simulation of climate changes in geographically complex areas, such as near mountain ranges and coastlines, both a high resolution

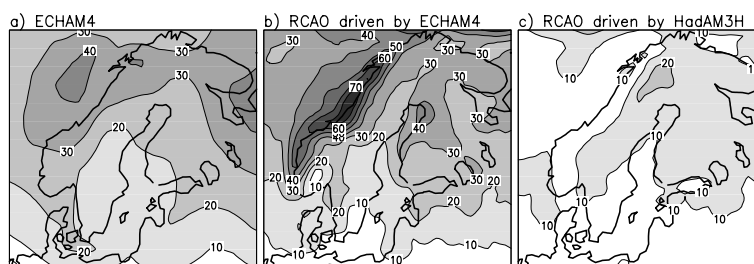


Fig. 7. Per cent changes in annual mean precipitation from 1961–1990 to 2071–2100 under the SRES A2 scenario. (a) the ECHAM4/OPYC3 global model; (b) the regional RCAO model driven by ECHAM4/OPYC3; and (c) RCAO driven by another global model, HadAM3H.

and a realistic simulation of large-scale atmospheric circulation changes are important.

8. Concluding remarks

How much can we rely on model simulations of future climate change? The preceding sections of this review have studied this question based on three main lines of evidence: the ability of models to simulate present-day climate, intermodel agreement on future climate changes, and the agreement between simulated and observed climate changes during the instrumental period. This section gives first a synthesis of the main points that have been addressed, and then, some thoughts on further research needs are presented.

Although there are many reasons to believe that climate models can give useful information on future climate, the question on model reliability has no simple quantitative answer. Below, I first list some key arguments that suggest that models do give reliable projections of climate change or, at least, that the uncertainty is reasonably well captured by the variation between different models:

1. Models are built on well-known physical principles. Despite the approximations needed in the description of some processes, this gives a priori reason to expect that models should be able to provide useful information on climate changes.
2. Many large-scale aspects of present-day climate are simulated quite well by the models. In addition, biases in the simulated climate tend to be unsystematic, so that observational estimates of present-day climate fall within the variation of model results.
3. When compared with each other, different climate models agree qualitatively or semi-quantitatively on several aspects of climate change. Moreover, many large-scale aspects of simulated greenhouse-gas-induced climate change are understood well in physical terms – one example of this is the general increase in high-latitude precipitation allowed by a larger moisture transport capacity of a warmer atmosphere.
4. Models have successfully simulated several large-scale aspects of climate change observed during the instrumental period. Although there is no detailed agreement between observed and simulated changes on smaller horizontal scales, this is largely as expected from the internal variability in the climate system. In most parts of the world, the temperature, precipitation and pressure changes observed during the past half-century fall within the range of model-simulated changes. Exceptions do occur, but not much more frequently than would be expected in the case that the simulated and observed changes belonged to the same statistical population.
5. Observation-based estimates of global climate sensitivity are, although uncertain, consistent with model results.

On the other hand, there are a number of issues that weaken the arguments given above and complicate their interpretation:

1. Many small-scale processes that cannot be simulated explicitly in current climate models are important for the feedback effects that regulate the response of climate to changes in external forcing. Cloud processes are the most important example.
2. The good agreement between simulated and observed present-day climates, and the tendency of the biases to vary in sign between different models, might arise partly because observations of present-day climate are used in tuning the models.
3. Models do not agree on all aspects of future climate change, particularly not on small horizontal scales. Overall, the agreement on changes in precipitation and atmospheric circulation is worse than the agreement on temperature changes.
4. A comparison between simulated and observed climate changes is complicated by uncertainty in the forcing factors (particularly the magnitude of aerosol forcing) that have affected 20th century climate. In addition, the climate changes projected for the rest of the 21st century are much larger than those observed this far. The impact of possible common model errors on the simulated climate changes will therefore also be larger for the future than for the past.
5. Because of uncertainties associated with forcing, observations and internal climate variability, key properties of the climate system such as the equilibrium climate sensitivity are still difficult to estimate from observations with a useful accuracy. Regional aspects of greenhouse-gas-induced climate change are even more difficult to constrain by observations.
6. Although climate models have been run for different emission scenarios, other aspects of forcing uncertainty are not covered well by existing multimodel ensemble simulations of future climate.

Despite these complications, the balance of evidence appears favourable for climate models. It seems likely that the real climate system will respond to increased greenhouse gas concentrations in many respects in a way similar to that models suggest. On the other hand, it is important to realize that the simulated climate changes vary (to a larger or smaller extent, depending on the variable, season, geographical area and horizontal scale considered) between different models.

Considering the difficulty of deriving observation-based estimates of uncertainty even for parameters such as the global climate sensitivity, the intermodel variation of climate changes probably gives the most meaningful estimates of uncertainty that are presently available – at least when going beyond globally averaged numbers. However, this measure of uncertainty should not be viewed uncritically. In general, it may be more likely to underestimate than overestimate the actual uncertainty, but the reverse is also possible in specific cases.

The possibility that climate changes in the real world might fall outside the range of model projections should be considered seriously especially when all models share similar errors in the simulation of present-day climate (which is relatively uncommon) and/or in the simulation of climate changes that have

been observed this far. On the other hand, when exceptionally large or small climate changes in a model have a clear physical connection to a major bias in the simulated present-day climate, such as a large error in the location of the sea ice edge, then it is well justified to discard the result of such a model (e.g. Fig. 2). However, the idea of rejecting or downweighting the climate-change projection from a specific model just because the climate change in this model differs a lot from that of the others involves a risk of circular reasoning. Such a procedure, although used in some studies (e.g. Giorgi and Mearns, 2002, 2003), is not recommended by the author of this review.

The question on the reliability of climate change simulations can be divided into two parts: (i) how well do models agree with each other, and (ii) does the variation between model results give a good estimate of uncertainty? The second question is the more difficult of these two, but more work is also needed on the first one. There are still important aspects of climate change for which the agreement between model simulations has not been properly quantified. For example, although changes in near-surface wind speed are of interest for many climate impact researchers, this variable has generally been excluded from the data bases of multimodel intercomparison projects (including the IPCC AR4 data set, which is in other respects more complete than data bases for earlier intercomparisons such as CMIP). This might be because modellers have concerns about the ability of global climate models to simulate realistic near-surface winds. Nevertheless, even if these concerns are justified, it would be useful to learn how well or badly models agree on greenhouse-gas-induced changes in wind speed. For example, if changes in wind speed would turn out to be mostly regulated by changes in the large-scale atmospheric circulation, then they might not be very sensitive to such biases in the present-day wind climate that are caused by deficiencies in the modelling of boundary layer processes.

Most multimodel studies of intermodel agreement have used simulations with prescribed greenhouse gas concentrations. Although this makes the simulations easier to conduct and the intermodel differences in climate change easier to interpret, more honest uncertainty estimates would be obtained from simulations with greenhouse gas concentrations computed from prescribed emissions. The C⁴MIP project discussed in Section 7.1 presents a valuable step in this direction.

The second question defined above cannot be addressed without evaluating model simulations against observations of the present climate and climate changes during the instrumental period and in the pre-instrumental past. However, in doing this it is also necessary to ask what really matters. Which aspects of present-day climate and previous climate changes are important to simulate correctly for a realistic simulation of future climate changes? If this question can be answered in a sufficiently quantitative manner, the answer would provide not only an objective means for ranking models but also, hopefully, a possibility of reducing the uncertainty in future climate changes.

Because observations of future climate do not exist, the question of what matters cannot be answered by looking at observations alone. However, multimodel and perturbed-parameter ensembles do provide a means to study the connections between past climate changes, present climate and future climate changes – naturally provided that the connections found for models are also applicable to the real world. Some of the climate sensitivity studies reviewed in Section 6 have already attempted to apply this idea (e.g. Piani et al., 2005; Knutti et al., 2006; Schneider von Deimling et al., 2006), and the wide uncertainty ranges obtained in these studies demonstrate that the exercise is far from simple. On the other hand, none of these studies used all available sources of information. Piani et al. (2005) and Knutti et al. (2006) only searched for connections between climate sensitivity and some aspects of the present-day climate, whereas Schneider von Deimling et al. (2006) used the simulated glacial-to-pre-industrial temperature changes (after first eliminating model versions that simulated the present climate badly). It is conceivable that tighter constraints on both the global climate sensitivity and the regional greenhouse-gas-induced climate changes could be derived by simultaneously including in the analysis the present-day climate and climate changes in both the instrumental era and the pre-instrumental past.

Regardless of the as-yet-unknown extent to which a more sophisticated analysis of ensemble simulations will help us to extract more information from the observational data that exist today, the situation will at least slowly improve in the future due to the continuing increase in greenhouse gas concentrations. As more observational evidence on the response of climate to increasing greenhouse gases accumulates, it will also gradually become easier to estimate how climate will change later in the future.

9. Acknowledgments

I acknowledge the international modelling groups for providing their data for analysis, 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, US Department of Energy.

10. Appendix A. Model data

The 21 ‘IPCC AR4’ climate models used in this study are listed in Table 4. For all these models, two simulations are used: a simulation covering the 20th century and forced by a mixture of anthropogenic and (in most models) natural forcing factors, and a 21st century simulation with anthropogenic greenhouse

Table 4. The models used in this study

Model	Institution
CCSM3	National Center for Atmospheric Research, USA
CGCM3.1 (T47)	Canadian Centre for Climate Modelling and Analysis
CGCM3.1 (T63)	Same as previous
CNRM-CM3	Météo-France
CSIRO-MK3.0	CSIRO Atmospheric Research, Australia
ECHAM5/MPI-OM	Max Planck Institute (MPI) for Meteorology, Germany
ECHO-G	University of Bonn and Model & Data Group, Germany; Korean Meteorological Agency
FGOALS-g1.0	Chinese Academy of Sciences
GFDL-CM2.0	Geophysical Fluid Dynamics Laboratory, USA
GFDL-CM2.1	Same as previous
GISS-AOM	Goddard Institute for Space Studies, USA
GISS-EH	Same as previous
GISS-ER	Same as previous
INM-CM3.0	Institute for Numerical Mathematics, Russia
IPSL-CM4	Institut Pierre Simon Laplace, France
MIROC3.2 (hires)	Center for Climate System Research, National Institute for Environmental Studies and Frontier Research Center for Global Change, Japan
MIROC3.2 (medres)	Same as previous
MRI-CGCM2.3.2	Meteorological Research Institute, Japan
PCM	National Center for Atmospheric Research, USA
UKMO-HadCM3	Hadley Centre for Climate Prediction and Research / Met Office, UK
UKMO-HadGEM	Same as previous

gas and aerosol forcing based on the SRES A1B emission scenario (Nakićenović and Swart, 2000). Although parallel runs started from different initial conditions are available for some of these models, only one 20th and 21st century simulation for each model are used in this study.

In terms of the greenhouse gas emissions and the magnitude of simulated climate changes, A1B is in the midrange of the SRES scenarios. The global mean warming between the periods 1971–2000 and 2070–2099 varies from 1.9°C (GISS-AOM and PCM) to 4.0°C (MIROC), with a g1-model mean value of 2.6°C.

The details of the forcing vary with model, but all models include at least the increase in major anthropogenic greenhouse gases and some representation of anthropogenic aerosols in both the 20th and 21st century simulations. However, indirect aerosol effects on clouds are included in only a minority of the models. Most of the 20th century simulations also include variations in solar and volcanic activity, but only two models (GISS-EH and GISS-ER) include a stochastic representation of these natural forcing mechanisms in the 21st century. The differences in climate change between the 21 models result from differences in forcing, internal variability and differences between the models themselves. At least in the late 21st century when the forcing is dominated by increased greenhouse gas concentrations and the simulated climate changes are large compared with internal variability, the last of these three factors is expected to be most important.

For the analysis in this study, all the model results were interpolated to a common $2.5^\circ \times 2.5^\circ$ latitude–longitude grid. The original resolution of the atmospheric model components varies from $1.1^\circ \times 1.1^\circ$ to $4^\circ \times 5^\circ$, the number of levels from 12 to 56, and the model top from 10 hPa (about 30 km) to 0.05 hPa (about 80 km). The horizontal resolution of the ocean components varies from $0.2^\circ \times 0.3^\circ$ to $4^\circ \times 5^\circ$ and the number of levels from 13 to 47. Flux adjustments for heat and freshwater are used in five out of the 21 models. Further details are available at <http://www.pcmdi.llnl.gov>.

11. Appendix B. Observational data sets

The observational estimates of surface air temperature, precipitation and sea level pressure used in this review are specified below.

Surface air temperature. Two data sets are used: the University of East Anglia Climate Research Unit (CRU) TS 2.0 data set (Mitchell et al., 2004) and the National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP-NCAR) reanalysis (Kistler et al., 2001). The CRU data set, which is based on an objective interpolation of station observations, is available over land areas including islands but excluding Antarctica and covers the years 1901–2002. The NCEP-NCAR reanalysis is global and available from the year 1948.

Precipitation. Two data sets are used: CRU and GPCP (Global Precipitation Climatology Project) Version 2 (Adler

et al., 2003). The global GPCP data set is based on a combination of satellite and rain gauge data and is available from the year 1979.

Sea level pressure. The HadSLP2 data set (Allan and Ansell, 2006) is used. This global data set is based on a reduced-space optimal interpolation procedure applied to pressure observations from over 2000 stations around the world. The data are available from the year 1850.

In Figs 1a and b, data from CRU are used over land excluding Antarctica. Elsewhere, the NCEP–NCAR (GPCP) data set is used for temperature (precipitation). The distributions of temperature and sea level pressure in Fig. 1 represent (for both the models and the observational data) the period 1971–2000. For precipitation, the years 1979–2002 are used, as dictated by the common period of the CRU and the GPCP data sets.

Because the GPCP data set is only available from the year 1979 and the NCEP–NCAR reanalysis is unsuitable for trend analysis before the beginning of the satellite era, the 50-year (1955 to 2005) trends in Figs 6a and b are only shown for areas covered by the CRU data set. The temperature (precipitation) values for the years 2003–2005 were taken from the NCEP–NCAR (GPCP) data set, after adjusting these for the average absolute (relative) difference from the CRU data in the years 1979–2002.

Observational data sets may contain significant errors particularly over areas where actual observations are sparse. Although the conclusions regarding the general performance of the models are not likely to be highly sensitive to errors in the data, the detailed results shown in Figs 1 and 6 should be taken with some caution.

References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. and co-authors. 2003. The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *J. Hydrometeorol.* **4**, 1147–1167.
- Allan, R. J. and Ansell, T. J. 2006. A new globally complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2004. *J. Climate*, in press.
- Allen, M. R. and Ingram, W. J. 2002. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232.
- Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. and Delworth, T. L. 2000. Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* **407**, 617–620.
- Anderson, T. L., Charlson, R. J., Schwartz, S. E., Knutti, R., Boucher, O. and co-authors. 2003. Climate forcing by aerosols – a hazy picture. *Science* **300**, 1103–1104.
- Andreae, M. O., Jones, C. D. and Cox, P. M. 2005. Strong present-day aerosol cooling implies a hot future. *Nature* **435**, 1187–1190.
- Andronova, N. G. and Schlesinger, M. E. 2001. Objective estimation of the probability density function of climate sensitivity. *J. Geophys. Res.* **106**, 22605–22611.
- Annan, J. D. and Hargreaves, J. C. 2006. Using multiple observationally-based constraints to estimate climate sensitivity. *Geophys. Res. Lett.* **33**, L06704.
- Annan, J. D., Hargreaves, J. C., Ohgaito, R., Abe-Ouchi, A. and Emori, S. 2005. Efficiently constraining climate sensitivity with paleoclimate simulations. *Scientific Online Letters on the Atmosphere* **1**, 181–184.
- Barnett, T. P. and Schlesinger, M. E. 1987. Detecting changes in global climate induced by greenhouse gases. *J. Geophys. Res.* **92**, 14772–14780.
- Barnett, T. P., Pierce, D. W., AchutaRao, K. M., Gleckler, P. J., Santer, B. D. and co-authors. 2005. Penetration of human-induced warming into the world's oceans. *Science* **309**, 284–287.
- Berger, A. 1978. Long-term variations of caloric solar radiation resulting from the Earth's orbital elements. *Quaternary Res.* **9**, 139–167.
- Bertrand, C., van Ypersele, J.-P. and Berger, A. 2002. Are natural climate forcings able to counteract the projected anthropogenic global warming? *Climatic Change* **55**, 413–427.
- Boer, G. 2000. Analysis and verification of model climate. In: *Numerical Modeling of the Global Atmosphere in the Climate System* (eds P. Mote and A. O'Neill). Kluwer Academic Publishers, Dordrecht, 59–82.
- Boer, G. J. and Yu, B. 2003. Climate sensitivity and climate state. *Climate Dyn.* **21**, 167–176.
- Bony, S., Dufresne, J. L., Le Treut, H., Morcrette, J. J. and Senior, C. A. 2004. On dynamic and thermodynamic components of cloud changes. *Climate Dyn.* **22**, 71–86.
- Boucher, O. and Haywood, J. 2001. On summing the components of radiative forcing of climate change. *Climate Dyn.* **18**, 297–302.
- Braganza, K., Karoly, D. J. and Arblaster, J. M. 2004. Diurnal temperature range as an index of global climate change during the twentieth century. *Geophys. Res. Lett.* **31**, D13217, doi:10.1029/2004GL019998.
- Broccoli, A. J., Dixon, K. W., Delworth, T. L., Knutson, T. R., Stouffer, R. J. and Zeng, F. 2003. Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing. *J. Geophys. Res.* **108**, D4798, doi:10.1029/2003JD003812.
- CCSP (US Climate Change Science Program) 2006. *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*. United States Climate Science Program and the Subcommittee on Global Change Research, Washington DC, USA, 164 pp.
- Cess, R. D., Potter, G. L., Blanchet, J. 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.
- Cess, R. D., Zhang, M. H., Ingram, W. J., Potter, G. L., Alekseev, V. and co-authors. 1996. Cloud feedback in atmospheric general circulation models: an update. *J. Geophys. Res.* **101**, 12791–12794.
- Charney, J. G. 1979. *Carbon Dioxide and Climate: A Scientific Assessment*. National Academy, Washington, DC, USA, 22 pp.
- Claussen, M., Mysak, L. A., Weaver, A. J., Crucifix, M., Fichefet, T. and co-authors. 2002. Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate Dyn.* **18**, 579–586.
- Colman, R. 2003. A comparison of climate feedbacks in general circulation models. *Climate Dyn.* **20**, 865–873.
- Covey, C., Achuta Rao, K. M., Cubasch, U., Jones, P., Lambert, S. J. and co-authors. 2003. An overview of results from the Coupled Model Intercomparison Project. *Global Planet. Change* **37**, 103–133.

- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. and Totterdell, I. J. 2000. Acceleration of global warming due to carbon cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187.
- Cubasch, U., Meehl, G. A., Boer, G. J., Stouffer, R. J., Dix, M. and co-authors. 2001. Projections of future climate change. In: *Climate change 2001. The Scientific Basis* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 525–582.
- Delworth, T. L. and Knutson, T. R. 2000. Simulation of early 20th century global warming. *Science* **287**, 2246–2250.
- Feddema, J. I., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E. and co-authors. 2005. The importance of land-cover change in simulating future climates. *Science* **310**, 1674–1678.
- Folland, C. K., Karl, T. R., Christy, J. R., Clarke, R. A., Gruza, G. V. and co-authors. 2001a. Observed climate variability and change. In: *Climate change 2001. The Scientific Basis* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 99–181.
- Folland, C. K., Rayner, N. A., Brown, S. J., Smith, T. M., Shen, S. S. P. and co-authors. 2001b. Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.* **28**, 2621–2624.
- 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 forcing. *Geophys. Res. Lett.* **33**, L01705, doi:10.1029/2005GL023977.
- Forster, P. M. de F. and Gregory, J. M. 2006. The climate sensitivity and its components diagnosed from Earth Radiation Budget data. *J. Climate* **19**, 39–52.
- Frame, D. J., Booth, B. B., Kettleborough, J. A., Stainforth, D. A., Gregory, J. M. and co-authors. 2005. Constraining climate forecasts: the role of prior assumptions. *Geophys. Res. Lett.* **32**, L09702, doi:10.1029/2004GL022241.
- Friedlingstein, P., Dufresne, J.-L., Cox, P. M. and Rayner, P. 2003. How positive is the feedback between climate change and the carbon cycle? *Tellus* **55B**, 692–700.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W. and co-authors. 2006. Climate – carbon cycle feedback analysis: results from the C⁴MIP model intercomparison. *J. Climate* **19**, 3337–3353.
- Gates, W. 1992. The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.* **73**, 1962–1970.
- Gates, W. L., Boyle, J., Covey, C., Dease, C., Doutriaux, C. and co-authors. 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.* **80**, 29–55.
- Gillett, N. P., Weaver, A. J., Zwiers, F. W. and Wehner, M. F. 2004. Detection of volcanic influence on global precipitation. *Geophys. Res. Lett.* **31**, L12217, doi: 10.1029/2004GL020044.
- Gillett, N. P., Allan, R. J. and Ansell, T. J. 2005. Detection of external influence on sea level pressure with a multimodel ensemble. *Geophys. Res. Lett.* **32**, L19714, doi: 10.1029/2005GL023640.
- Giorgi, F. and Francisco, R. 2000. Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.* **27**, 1295–1298.
- Giorgi, F. and Mearns, L. O. 2002. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *J. Climate* **15**, 1141–1158.
- Giorgi, F. and Mearns, L. O. 2003. Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. *Geophys. Res. Lett.* **30**, L1629, doi:10.1029/2003GL017130.
- Giorgi, F., Hewitson, B., Christensen, J., Hulme, M., Von Storch, H. and co-authors. 2001. Regional climate information – evaluation and projections. In: *Climate change 2001. The Scientific Basis* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 583–638.
- Graham, R. J., Gordon, M., McLean, P. J., Ineson, S., Huddleston, M. R. and co-authors. 2005. A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus* **57A**, 320–339.
- Greene, A. M., Goddard, L. and Lall, U. 2006. Probabilistic multimodel regional temperature change projections. *J. Climate* **19**, 4326–4343.
- Gregory, J. M., Stott, P. A., Cresswell, D. J., Rayner, N. A., Gordon, C. and Sexton, D. M. H. 2002a. Recent and future changes in Arctic sea ice simulated by the HadCM3 AOGCM. *Geophys. Res. Lett.* **29**, L2175, doi:10.1029/2001GL014575.
- Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A. and Rayner, N. A. 2002b. An observationally based estimate of the climate sensitivity. *J. Climate* **15**, 3117–3121.
- Groisman, P. Ya. and Easterling, D. R. 1994. Variability and trends of total precipitation and snowfall over the United States and Canada. *J. Climate* **7**, 184–205.
- Groisman, P. Ya., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C. and Razuvaev, V. N. 2005. Trends in intense precipitation in the climate record. *J. Climate* **18**, 1326–1350.
- Grotch, S. L. and MacCracken, M. C. 1991. The use of general circulation models to predict regional climate change. *J. Climate* **4**, 286–303.
- Hansen, J., Russell, G., Lacis, A., Fung, I. and Rind, D. 1985. Climate response times: dependence on climate sensitivity and ocean mixing. *Science* **229**, 857–859.
- Hansen, J., Sato, M. and Ruedy, R. 1997. Radiative forcing and climate response. *J. Geophys. Res.* **102**, 6831–6864.
- Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A. and co-authors. 2005. Efficacy of climate forcings. *J. Geophys. Res.* **110**, D18104, doi:10.1029/2005JD005776.
- Hansen, J. E., Nazarenko, L., Ruedy, R., Sato, M., Willis, J. and co-authors. 2006. Earth’s energy imbalance: confirmation and implications. *Science* **308**, 1431–1435.
- Hansen-Bauer, I., Førland, E. J., Haugen, J. E. and Tveito, O. E. 2003. Temperature and precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. *Climate Res.* **25**, 15–27.
- Harris, G. R., Sexton, D. M. H., Booth, B. B. B., Collins, M., Murphy, J. M. and Webb, M. J. 2006. Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. *Climate Dyn.* **27**, 357–375.
- Hartmann, D. L. 1994. *Global Physical Climatology*. Academic Press, San Diego, CA, USA, 411 pp.
- Harvey, L. D. D. 2004. Characterizing the annual-mean climatic effect of anthropogenic CO₂ and aerosol emissions in eight coupled atmosphere-ocean GCMs. *Climate Dyn.* **23**, 569–599.

- Harvey, L. D. D. and Kaufmann, R. K. 2002. Simultaneously constraining climate sensitivity and aerosol radiative forcing. *J. Climate* **15**, 2837–2861.
- Harvey, D., Gregory, J., Hoffert, M., Jain, A., Lal, M. and co-authors. 1997. An introduction to simple climate models used in the IPCC Second Assessment Report. IPCC Technical Paper 2 (eds J. T. Houghton, L. G. Meira Filho, D. Griggs and K. Maskell), IPCC, Geneva, Switzerland, 51 pp.
- Hegerl, G. C., Crowley, T., Hyde, W. T. and Frame, D. 2006. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440**, 1029–1032.
- Hellström, C., Chen, D., Achberger, C. and Räisänen, J. 2001. A Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Res.* **19**, 45–55.
- Hewitt, C., Senior, C. and Mitchell, J. 2001. The impact of dynamic sea-ice on the climatology and climate sensitivity of a GCM: a study of past present and future climates. *Climate Dyn.* **17**, 655–668.
- Hoffert, M. I. and Covey, C. 1992. Deriving global climate sensitivity from paleoclimatic reconstructions. *Nature* **360**, 573–576.
- Houghton, J. 2004. *Global Warming, The Complete Briefing*, 3rd Edition. Cambridge University Press, Cambridge, 382 pp.
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden P. J. and Xiaosu, D. (eds) 2001. *Climate Change 2001. The Scientific Basis*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.
- Hudson, D. A. and Jones, R. G. 2002. Simulations of present-day and future climate over southern Africa using HadAM3H. Hadley Centre Technical Note 38, Met Office, Bracknell, United Kingdom, 36 pp.
- Hulme, M., Osborn, T. J. and Johns, T. C. 1998. Precipitation sensitivity to global warming: comparison of observations with HadCM2 simulations. *Geophys. Res. Lett.* **25**, 3379–3382.
- Huntingford, C. and Cox, P. M. 2000. An analogue model to derive additional climate change scenarios from existing GCM simulations. *Climate Dyn.* **16**, 575–586.
- IPCC WG1 2004. Workshop report, IPCC Working Group I Workshop on Climate Sensitivity, Paris, France, 26–29 July 2004, 177 pp.
- Joshi, M., Shine, K., Ponater, M., Stuber, N., Sausen, R. and Li, L. 2003. A comparison of climate response to different radiative forcings in three general circulation models: towards an improved metric of climate change. *Climate Dyn.* **20**, 843–854.
- Kattenberg, A., Giorgi, F., Grassl, H., Meehl, G. A., Mitchell, J. F. B. and co-authors. 1996. Climate models – projections of future climate. *Climate change 1995. The Science of Climate Change* (eds J. T. Houghton, L. G. Meira Filho, B. A. Cllander, N. Harris, A. Kattenberg and K. Maskell). Cambridge University Press, Cambridge, 285–357.
- Kiktev, D., Sexton, D. M. H., Alexander, L. and Folland, C. K. 2003. Comparison of modeled and observed trends in indices of daily climate extremes. *J. Climate* **16**, 3560–3571.
- Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G. and co-authors. 2001. The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Amer. Meteor. Soc.* **82**, 247–268.
- Kittel, T. G. F., Giorgi, F. and Meehl, G. A. 1998. Intercomparison of regional biases and doubled CO₂-sensitivity of coupled atmosphere-ocean general circulation model experiments. *Climate Dyn.* **14**, 1–15.
- Knutson, T. R., Delworth, T. L., Dixon, K. W., Held, I. M., Lu, J. and co-authors. 2006. Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *J. Climate* **19**, 1624–1650.
- Knutti, R., Meehl, G. A., Allen, M. R. and Stainforth, D. A. 2006. Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Climate* **19**, 4224–4233.
- Kohfeld, K. E. and Harrison, S. P. 2000. How well can we simulate past climates? Evaluating the models using global palaeoenvironmental datasets. *Quaternary Sci. Rev.* **19**, 321–346.
- Lambert, S. J. and Boer, G. J. 2001. CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dyn.* **17**, 83–106.
- Lambert, F. H., Stott, P. A., Allen, M. R. and Palmer, M. A. 2004. Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.* **31**, L10203, doi:10.1029/2004GL019545.
- Lambert, F. H., Gillett, N. P., Stone, D. A. and Huntingford, C. 2005. Attribution studies of observed land precipitation changes with nine coupled models. *Geophys. Res. Lett.* **32**, L18704, doi:10.1029/2005GL023654.
- Levitus, S., Antonov, J., Boyer, T. P. and Stephens, C. 2000. Warming of the world ocean. *Science* **287**, 2225–2229.
- Levitus, S., Antonov, J. and Boyer, T. 2005. Warming of the world ocean, 1955–2003. *Geophys. Res. Lett.* **32**, L02604, doi:10.1029/2004GL021592.
- Lindzen, R. S. 1990. Some coolness concerning global warming. *Bull. Amer. Meteor. Soc.* **71**, 288–299.
- Lindzen, R. S., Chou, M.-D. and Hou, A. Y. 2001. Does the Earth have and adaptive infrared iris? *Bull. Amer. Meteor. Soc.* **82**, 417–432.
- Lopez, A., Tebaldi, C., New, M., Stainforth, D. A., Allen, M. R. and Kettleborough, J. A. 2006. Two approaches to quantifying uncertainty in global temperature changes. *J. Climate* **19**, 4785–4796.
- Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.
- Manabe, S. and Wetherald, R. T. 1987. Large-scale changes of soil wetness induced by an increase in atmospheric carbon dioxide. *J. Atmos. Sci.* **44**, 1211–1235.
- Manabe, S., Stouffer, R. J., Spelman, M. J. and Bryan, K. 1991. Transient response of a coupled ocean-atmosphere model to gradual changes of atmospheric CO₂. Part I: Annual mean response. *J. Climate* **4**, 785–818.
- McAvaney, B. J., Covey, C., Joussaume, S., Kattsov, V., Kitoh, A. and co-authors. 2001. Model evaluation. In: *Climate change 2001. The Scientific Basis* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 471–523.
- McGuffie, K. and Henderson-Sellers, A. 2005. *A Climate Modelling Primer*, 3rd Edition. John Wiley & Sons, Chichester, 296 pp.
- Meehl, G. A., Boer, G. J., Covey, C., Latif, M. and Stouffer, R. J. 2000. The Coupled Model Intercomparison Project (CMIP). *Bull. Amer. Meteor. Soc.* **81**, 313–318.
- Meehl, G. A., Washington, W. M., Arblaster, J. M. and Hu, A. 2004a. Factors affecting climate sensitivity in global coupled models. *J. Climate* **17**, 1584–1596.
- Meehl, G. A., Washington, W. M., Ammann, C. M., Arblaster, J. M., Wigley, T. M. L. and Tebaldi, C. 2004b. Combinations of natural and anthropogenic forcings in twentieth-century climate. *J. Climate* **17**, 3721–3727.

- Miller, R. L., Schmidt, G. A. and Shindell, D. T. 2006. Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report Models. *J. Geophys. Res.* **111**, D18101, doi: 10.129/2005JD006323.
- Mitchell, T. D. 2003. Pattern scaling: an examination of the accuracy of the technique for describing future climate. *Climatic Change* **60**, 217–242.
- Mitchell, J., Wilson, C. and Cunningham, W. 1987. On CO₂ climate sensitivity and model dependence of results. *Q. J. R. Meteor. Soc.* **113**, 293–322.
- Mitchell, J. F. B., Manabe, S., Tokioka, T. and Meleshko, V. 1990. Equilibrium climate change – and its implications for the future. In: *Climate Change. The IPCC Scientific Assessment* (eds J. T. Houghton, G. J. Jenkins and J. J. Ephraums). Cambridge University Press, Cambridge, United Kingdom, 131–172.
- Mitchell, J. F. B., Johns, T. C., Eagles, M., Ingram, W. J. and Davis, R. A. 1999. Towards the construction of climate change scenarios. *Climatic Change* **41**, 547–581.
- Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M. and New, M. 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). Tyndall Centre Working Paper 55, 30 pp.
- Mote, P. and O'Neill, A. (eds) 2000. *Numerical Modeling of the Global Atmosphere in the Climate System*. Kluwer Academic Publishers, Dordrecht, 517 pp.
- Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J. and co-authors. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772.
- Nakićenović, N. and Swart, R. (eds) 2000. *Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 599 pp.
- Neelin, J. D., Münnich, M., Su, H., Meyerson, J. E. and Holloway, C. E. 2006. Tropical drying trends in global warming models and observations. *Proc. Nat. Acad. Sci.* **103**, 6110–6115.
- Noh, Y., Cheon, W. G., Hong, S. Y. and Raasch, S. 2003. Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Boundary-Layer Meteorol.* **107**, 401–427.
- Palmer, T. N. and Räisänen, J. 2002. Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature* **415**, 514–517.
- Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T. and Leutbecher, M. 2005. Representing model uncertainty in weather and climate prediction. *Ann. Rev. Earth. Planet. Sci.* **33**, 163–193.
- Pan, Z., Christensen, J., Arritt, R., Gutowski, W., Takle, E. and Otieno, F. 2001. Evaluation of uncertainty in regional climate change simulations. *J. Geophys. Res.* **106**, 17735–17751.
- Petit, J. R., Jouzel, J., Raynaud, N. I., Barkov, N. I., Barnola, J. M. and co-authors. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436.
- Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S. and co-authors. 2004. Evaluating parameterizations in general circulation models: climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.* **85**, 1903–1915.
- Piani, C., Frame, D. J., Stainforth, D. A. and Allen, M. R. 2005. Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Lett.* **32**, L23825, doi:10.1029/2005GL024452.
- Pierce, D. W., Barnett, T. P., AchutaRao, K. M., Gleckler, P. J., Gregory, J. M. and Washington, W. M. 2006. Anthropogenic warming of the oceans: observations and model results. *J. Climate* **19**, 1873–1900.
- Räisänen, J. 2001. CO₂-induced climate change in CMIP2 experiments. Quantification of agreement and role of internal variability. *J. Climate* **14**, 2088–2104.
- Räisänen, J. and Palmer, T. N. 2001. A probability and decision model analysis of a multimodel ensemble of climate change simulations. *J. Climate* **14**, 3212–3226.
- Räisänen, J., Hansson, U., Ullerstig, A., Döscher, R., Graham, L. P. and co-authors. 2004. European climate in the late twenty-first century: regional simulations with two driving global models and two forcing scenarios. *Climate Dyn.* **22**, 13–31.
- Ramaswamy, V., Boucher, O., Haigh, J., Hauglustaine, D., Haywood, J. and co-authors. 2001. Radiative forcing of climate change. In: *Climate change 2001. The Scientific Basis* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 349–416.
- Ramaswamy, V., Schwarzkopf, M. D., Randel, W. J., Santer, B. D., Soden, B. J. and Stenchikov, G. L. 2006. Anthropogenic and natural influences in the evolution of lower stratospheric cooling. *Science* **311**, 1138–1141.
- Roeckner, E., Bengtsson, L., Feichter, J., Lelieveld, J. and Rodhe, H. 1999. Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. *J. Climate* **12**, 3004–3032.
- Rowell, D. P. and Jones, R. G. 2006. Causes and uncertainty of future summer drying over Europe. *Climate Dyn.* **27**, 281–299.
- Ruosteenoja, K., Carter, T. R., Jylhä, K. and Tuomenvirta, H. 2003. Future climate in world regions: and intercomparison of model-based projections for the new IPCC emissions scenarios. *The Finnish Environment* **644**, Finnish Environment Institute, Helsinki, Finland, 83 pp.
- Santer, B. D., Wigley, T. M. L., Schlesinger, M. E. and Mitchell, J. F. B. 1990. Developing climate scenarios from equilibrium GCM results. Report 47, Max-Planck Institut für Meteorologie, Hamburg, Germany, 29 pp.
- Santer, B. D., Wigley, T. M. L., Simmons, A. J., Källberg, P. W., Kelly, G. A. and co-authors. 2004. Identification of anthropogenic climate change using a second-generation reanalysis. *J. Geophys. Res.* **109**, D21104, doi:10.1029/2004JD005075.
- Sausen, R., Barthels, R. K. and Hasselmann, K. 1988. Coupled ocean-atmosphere models with flux correction. *Climate Dyn.* **2**, 154–163.
- Savijärvi, H. 1995. Error growth in a large numerical forecast system. *Mon. Wea. Rev.* **123**, 212–221.
- Schneider von Deimling, T., Held, H., Ganopolski, A. and Rahmstorf, S. 2006. Climate sensitivity estimated from ensemble simulations of glacial climate. *Climate Dyn.* **27**, 149–163.
- Schwartz, S. E. 2004. Uncertainty requirements in radiative forcing of climate change. *J. Air Waste Manage. Assoc.* **54**, 1351–1359.
- Selten, F. M., Branstator, G. W., Dijkstra, H. A. and Kliphuis, M. 2004. Tropical origins for recent and future Northern Hemisphere climate change. *Geophys. Res. Lett.* **31**, L21205, doi:10.1029/2004GL020739.
- Senior, C. A. and Mitchell, J. 2000. The time-dependence of climate sensitivity. *Geophys. Res. Lett.* **27**, 2685–2689.

- Siegenthaler, U., Stocker, T. F., Monnin, E., Lüthi, D., Schwander, J. and co-authors. 2005. Stable carbon cycle – climate relationship during the late Pleistocene. *Science* **310**, 1313–1317.
- Soden, B. J. and Held, I. M. 2006. An assessment of climate feedbacks in coupled ocean-atmosphere models. *J. Climate* **19**, 3354–3360.
- Soden, B. J., Jackson, D. L., Ramaswamy, V., Schwarzkopf, M. D. and Huang, X. 2005. The radiative signature of upper tropospheric moistening. *Science* **310**, 841–844.
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faui, N. and co-authors. 2005. Uncertainty in the predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403–406.
- Stott, P. A. and Kettleborough, J. A. 2002. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**, 723–726.
- Stott, P. A., Tett, S. F. B., Jones, G. J., Allen, M. R., Mitchell, J. F. B. and Jenkins, G. J. 2000. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137.
- Stott, P. A., Mitchell, J. F. B., Allen, M. R., Delworth, T. L., Gregory, J. M. and co-authors 2006a. Observational constraints on past attributable warming and predictions of future global warming. *J. Climate* **19**, 3055–3069.
- Stott, P. A., Kettleborough, J. A. and Allen, M. R. 2006b. Uncertainty in continental-scale temperature predictions. *Geophys. Res. Lett.* **33**, L02708, doi:10.1029/2005GL024423.
- Tebaldi, C., Smith, R., Nychka, D. and Mearns, L. O. 2005. Quantifying uncertainty in projections of regional climate change: A Bayesian Approach. *J. Climate* **18**, 1524–1540.
- Thorne, P. W., Parker, D. E., Christy, J. R. and Mears, C. A. 2005. Uncertainties in climate trends: lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.* **87**, 1437–1442.
- Trenberth, K. E. (ed.) 1992. *Climate System Modeling*. Cambridge University Press, Cambridge, 788 pp.
- Trenberth, K. E., Fasullo, J. and Smith, L. 2005. Trends and variability in column-integrated atmospheric water vapor. *Climate Dyn.* **24**, 741–758.
- Vose, R. S., Easterling, D. R. and Gleason, B. 2005. Maximum and minimum temperature trends for the globe: an update through 2004. *Geophys. Res. Lett.* **32**, L23822, doi:10.1029/2005GL024379.
- Watterson, I. G. 1998. An analysis of the global water cycle of present and doubled CO₂ climate simulated by the CSIRO general circulation model. *J. Geophys. Res.* **103**, 23113–23129.
- Webb, M. J., Senior, C. A., Sexton, D. M. H., Ingram, W. J., Williams, K. D. and co-authors. 2006. On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Climate Dyn.* **27**, 17–38.
- Whetton, P. H., England, M. H., O’Farrell, S. P., Watterson, I. G. and Pittock, A. B. 1996. Global comparison of the regional rainfall results of enhanced greenhouse coupled and mixed layer ocean experiments: Implications for climate change scenario development. *Climatic Change* **33**, 497–519.
- Wigley, T. M. L. and Schlesinger, M. E. 1985. Analytical solution for the effect of increasing CO₂ on global mean temperature. *Nature* **315**, 649–652.
- Wigley, T. M. L., Ammann, C. M., Santer, B. D. and Raper, S. C. B. 2005. Effect of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.* **110**, D09107, doi:10.1029/2004JD005557.
- Williamson, D. L. and Laprise, R. 2000. Numerical approximations for global atmospheric GCMs. In: *Numerical Modeling of the Global Atmosphere in the Climate System* (eds P. Mote and A. O’Neill). Kluwer Academic Publishers, Dordrecht, 147–219.
- Yin, J. H., 2005. A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.* **32**, L18701, doi:10.1029/2005GL023684.
- Zhang, X. and Walsh, J. E. 2006. Toward a Seasonally Ice-Covered Arctic Ocean: Scenarios from the IPCC AR4 Model Simulations. *J. Climate* **19**, 1730–1747.