Chapter 9: Evaluation of Climate Models

Coordinating Lead Authors: Gregory Flato (Canada), Jochem Marotzke (Germany)

Lead Authors: Babatunde Abiodun (South Africa), Pascale Braconnot (France), Sin Chan Chou (Brazil), William Collins (USA), Peter Cox (UK), Fatima Driouech (Morocco), Seita Emori (Japan), Veronika Eyring (Germany), Chris Forest (USA), Peter Gleckler (USA), Eric Guilyardi (France), Christian Jakob (Australia), Vladimir Kattsov (Russia), Chris Reason (South Africa), Markku Rummukainen (Sweden)

Contributing Authors: Alessandro Anav, Timothy Andrews, Johanna Baehr, Alejandro Bodas-Salcedo, Ping Chang, Aiguo Dai, Clara Deser, Bode Gbobaniyi, Fidel Gonzales-Rouco, Stephen Griffies, Alex Hall, Elizabeth Hunke, Tatiana Ilyina, Viatcheslav Kharin, Stephen A. Klein, Jeff Knight, Reto Knutti, Felix Landerer, Tatiana Pavlova, Florian Rauser, Mark Rodwell, Adam A. Scaife, John Scinocca, Hideo Shiogama, Jana Sillmann, Ken Sperber, David Stephenson, Bjorn Stevens, Mark Webb, Keith Williams, Tim Woollings, Shang-Ping Xie

Review Editors: Isaac Held (USA), Andy Pitman (Australia), Serge Planton (France), Zong-Ci Zhao (China)

Date of Draft: 16 December 2011

Notes: TSU Compiled Version

Table of Contents

Executive Summary

Climate models play an important role in climate research, enhancing our ability to understand past climate change and providing quantitative information about the future. Confidence in using climate models is based on careful evaluation of model performance, making use of increasingly comprehensive observationally-based data sets and well-designed model intercomparison activities. This chapter provides an assessment of climate model evaluation, focusing particularly on developments since the IPCC Fourth Assessment Report (AR4). A range of models are considered, including:

- coupled Atmosphere-Ocean General Circulation Models (AOGCMs), used in both long-term climate projection and shorter-term (seasonal to decadal) climate prediction;

- their extension to 'Earth System' Models (ESMs), in which representation of climatically important biogeochemical cycles are included;

- higher resolution, limited-area Regional Climate Models (RCMs), used extensively in downscaling global climate results to particular regions;

- Earth System Models of Intermediate Complexity (EMICs) used to undertake very long (e.g., millennial) climate simulations, or to provide large ensembles exploring parameter uncertainty.

The evaluation of climate models depends directly on the availability of high-quality observational data sets whose uncertainty is understood and quantified. These observational data have been described in earlier chapters. A particular advance since the AR4 has been in the area of model 'metrics' – that is, numerical measures of model performance reflecting the difference between a model and a corresponding observational estimate. These metrics allow more systematic evaluation of models and more concise presentation of evaluation results. This chapter will make extensive use of such metrics, along with more traditional presentation of 'error maps' and time series.

Another advance since the AR4 is the extensive use of 'satellite simulators' in climate models. This involves on-line calculations which provide output more directly comparable to remote sensing observations from satellites. This approach is particularly valuable in evaluating the representation of clouds and cloud processes in climate models.

The availability of carefully constructed multi-model experiments, notably the Coupled Model Intercomparison Projects (CMIP3 and CMIP5) and the Coordinated Regional Downscaling Experiment (CORDEX), allow for increasingly in-depth analysis of model results. The multi-model ensemble allows for some assessment of uncertainty in climate model capabilities in cases where suitable observations are not available, but more importantly it allows one to begin investigating the connection between particular model errors/biases and particular characteristics or process parameterisations in a model. This necessarily requires careful and extensive documentation of each model, something that has also improved significantly since the AR4.

The results presented here indicate that the ability of climate models to simulate historical climate, its change, and its variability, has improved in many important respects since the AR4. Particular examples include:

- biases in surface temperature and precipitation are typically smaller;

- the diurnal cycle of surface air temperature and the annual cycle of sea-ice extent is well simulated on average;

- some modes of climate variability, such as the North Atlantic Oscillation, are better represented;

- the new generation of ESMs are able to realistically simulate the annual cycle and spatial gradients of atmospheric CO_2, and the uptake of carbon, particularly by the ocean;

- regional climate models are able to add value to coarser-resolution global model results, providing realistic spatial detail and improved representation of climate extremes.

Of course, there are also many areas of model performance that remain to be improved. There is large inter-model spread and evident systematic bias in a number of important climate quantities. For example, models have problems simulating the mean temperature structure of the Tropical Atlantic Ocean, the diurnal cycle of precipitation, the Madden-Julien Oscillation, and clouds and cloud radiative effects. In some cases, model results are in general agreement with observations, but the observational uncertainty is sufficiently large as to render it impossible to make definitive statements about model quality.

An evaluation of models tends, necessarily, to focus on the identification of model weaknesses. This provides the ongoing challenge to model developers to probe the underlying processes involved and to improve their representation. Despite their shortcomings, contemporary climate and Earth system models are able to simulate many important aspects of past climate change and variability, including feedbacks and modes of variability that arise spontaneously in the coupled system.The errors and biases in models tend to be related to smaller-scale features and higher-order statistics such as correlations and teleconnections. Overall, climate models provide realistic simulations of the large-scale features of climate system and reproduce observed historical change with some fidelity. The climate sensitivity of current models has not changed dramatically from that of models assessed in the AR4, despite many improvements to the models' representation of physical processes.

9.1 Climate Models and their Characteristics

9.1.1 Scope and Overview of this Chapter

Climate models constitute the primary tools available for investigating the response of the climate system to various forcings, for making climate predictions on seasonal to decadal time scales, and for making projections of future climate over the coming century and beyond. It is crucial therefore to critically evaluate the performance of these models, both individually and collectively. The focus of this chapter is particularly on the models whose results will be used in the detection and attribution chapter and the projections chapters 10 through 12, and so this is necessarily an incomplete evaluation. In particular, we will draw heavily on model results collected as part of the Coupled Model Intercomparison Projects (CMIP3 and CMIP5 (Meehl et al., 2007a; Taylor, 2011)) as this constitutes a set of well-controlled and increasingly well-documented climate model experiments. Other intercomparison efforts, such as the Coordinated Regional Downscaling Experiment (CORDEX), dealing with regional climate models (RCMs), and those dealing with Earth System Models of intermediate complexity (EMICs) are also used. Results from earlier evaluations will be included so as to illustrate changes in model performance over time.

The direct approach to model evaluation is to compare observations with model output and analyze the resulting difference. This requires knowledge of the errors and uncertainties in the observations, which have been discussed in Chapters 2 through 5. Where possible, averages over the same time period in both models and observations will be compared, although for many quantities, only observationally-based estimates of the climatological mean are available. In cases where observations are lacking, we will resort to intercomparison of model results to provide some quantification of model 'uncertainty'.

After a more thorough discussion of the climate models and methods for evaluation in Sections 9.1 and 9.2, we describe the basic characterization of climate model experiments in Section 9.3, evaluate recent and longer-term records as simulated by climate models in Section 9.4, variability and extremes in Section 9.5, and downscaling and regional scale climate simulation in Section 9.6. We conclude with a discussion of the sources of model errors in Section 9.7 and the relation between model performance and the credibility of future climate projections in Section 9.8.

9.1.2 Overview of Models to be Evaluated

The models used in climate research range from simple energy balance models to complex Earth System Models using state of the art high-performance computing. The choice of model depends directly on the scientific question being addressed (Collins et al., 2006b; Held, 2005). Applications include simulating palaeoclimate, historical climate, predicting near-term climate variability and change on seasonal to decadal time scales, making projections of future climate change over the 21st century or more, and downscaling such projections to provide more detail at the regional and local scale. Computational cost is a factor in all of these, and so simplified models (with reduced complexity or spatial resolution) can be used when larger ensembles or longer integrations are required. Examples of the latter include exploration of parameter sensitivity or simulations of climate change on the millennial or longer time scale. Here, we provide a brief overview of the climate models evaluated in this chapter.

9.1.2.1 Atmosphere-Ocean General Circulation Models (AOGCMs)

AOGCMs were the "standard" climate models assessed in the AR4. Their primary function is to understand the dynamics of the physical components of the climate system (atmosphere, ocean, land, and sea-ice), and they are the workhorse for making projections based on future greenhouse gas and aerosol forcing. These models continue to be extensively used, and in particular are run (often at higher resolution) for seasonal to decadal climate prediction applications in which biogeochemical feedbacks from the carbon cycle are less important than they are in century-scale projections.

9.1.2.2 Earth System Models (ESMs)

ESMs are the current state-of-the-art climate models in the CMIP5, in terms of the extent to which the overall Earth system is represented. Compared to the AOGCMs that constituted the bulk of the models

assessed in the AR4, ESMs include representation of various biogeochemical cycles such as those involved in the carbon cycle, the sulphur cycle, or stratospheric ozone (Flato, 2011). These models provide the most comprehensive tools available for simulating past and future response of the climate system to external forcing, in which biogeochemical feedbacks play a potentially important role. These models require using the largest and fastest high-performance computing platforms available and typically require more than one month for a 100-year simulation.

9.1.2.3 Earth-System Models of Intermediate Complexity (EMICs)

EMICs attempt to include all relevant components of the earth-system, but often in an idealized manner or at lower resolution than the models described above. These models are applied to certain scientific questions such as understanding climate feedbacks on millennial time scales or exploring sensitivities in which long model integrations or large ensembles are required (Claussen et al., 2002; Petoukhov et al., 2005). This class of models often includes Earth system components not yet included in all ESMs (e.g., ice sheets). As computer power increases, this model class has continued to advance in terms of resolution and complexity.

9.1.2.4 Regional Climate Models (RCMs)

RCMs employ a limited area grid driven at the boundaries by output from a global climate model or reanalyses. The representations of climate processes are comparable to those in AOGCMs. The typical application of an RCM is to 'downscale' global model projections for some particular geographical region, in order to provide information on a higher resolution than what is attainable from many AOGCMs (Rummukainen, 2010).

9.1.3 The Path to Model Improvement

9.1.3.1 The Model Development Process

All of the comprehensive climate models introduced above are founded on well-known physical laws (e.g., energy and momentum conservation). Developing climate models involves two principal steps:

i) Finding the mathematical expressions for the system's physical laws. This requires theoretical work in deriving and simplifying the mathematical expressions that best describe the system;

ii) Implementing these mathematical expressions on a computer. This requires developing numerical methods that allow the solution of the mathematical expressions usually implemented on some form of grid such as the latitude-longitude-height grid for atmospheric models.

The application of complex climate models requires significant supercomputing resources. Limitations in those resources lead to additional constraints. Even when using the most powerful computers today, compromises need to be made in three main areas:

i) Numerical implementations allow for a choice of grid spacing, often referred to as "model resolution" (see Section 9.1.3.4). Higher model resolution generally leads to more accurate models but also to higher computational costs. Currently affordable climate model resolutions imply that certain processes are excluded from the numerical solutions and have to be represented through simple conceptual models usually referred to as parameterisations (e.g., carbon cycle or cloud processes, see Chapters 6 and 7).

ii) The climate system contains many processes the relative importance of which varies with the time-scale of interest. Hence compromises to include or exclude certain processes or components in a model must be made, with an increase in complexity leading to an increase in computational cost.

iii) The climate system is highly non-linear. A single model simulation therefore only represents one of the possible pathways the system might follow. To allow some evaluation of the resulting uncertainties it is necessary to carry out a number of simulations either with several models or using an ensemble of simulations with a single model, both increasing the computational cost.

Trade-offs amongst the various considerations outlined above depend on the model application and lead to the several classes of models introduced in Section 9.1.2.

The development of climate models takes place at several strongly connected levels illustrated in Figure 9.1. Having derived the numerical description of one of the model components (e.g., the atmosphere) it is necessary to develop representations of the processes that are excluded by the choice of model resolution. This process of parameterisation has become very complex (Jakob, 2010) and is often achieved by developing conceptual models of the process of interest in isolation. The complexity of each process representation is constrained by observations, computational resources and current knowledge (e.g., Randall et al., 2007). The next step in model development is the assembly and evaluation of single model components. For instance, the atmospheric component can be evaluated by prescribing sea surface temperature (Gates et al., 1999) or the ocean and land components by prescribing the atmospheric input (Barnier et al., 2006; Griffies et al., 2009). The final step of model development is the integration of all components into a full climate model and the evaluation of that model as a whole. The need for simple process representations in all model components introduces several parameters, which are adjusted using a variety of constraints ranging from observations of the parameter itself to well-known climate system characteristics such as global top-of-the-atmosphere radiative balance. This process is often referred to as model tuning (see Box 9.1).

[START BOX 9.1 HERE]

Box 9.1: Model Tuning and Evaluation

The global climate is an extremely complex system and there is no known set of equations that describes it completely. Rather climate models are comprised of a fundamental component, the part of the system described by established theory, and a non-fundamental component, whereby important processes are described as best one can (see Section 9.1.3), often through the use of empirically-derived but physically plausible relationships. This mix of fundamental and non-fundamental components in the description of the climate system leads to many of the key uncertainties in current climate models.

Due to constraints both in knowledge and computational resources there is a need to introduce a number of parameters in all model components, particularly related to the description of unresolved or poorly understood processes (see Section 9.1.3). Many of these parameters are not well-constrained by theory or data, although their rough determination is based on process-level understanding, observational constraints, and fine-scale modelling studies (see Chapter 7, Section 7.2 and Section 9.1.3).

Once various components have been assembled into a comprehensive model, a small subset of model parameters remain to be adjusted so that the model adheres to certain large-scale observational constraints. This final parameter adjustment procedure is usually referred to as ***model tuning***. As model tuning aims to match observed climate system behaviour, it is connected to judgments as to what constitutes a skilful representation of the Earth's climate. Some aspects of a simulation that one tunes toward are unquestionably important. For instance, maintaining the top-of-the-atmosphere energy balance in an unforced simulation is essential to prevent the climate system from drifting to an unrealistic state. Because modelling centres do not routinely describe how they tune their models the complete list of observational constraints toward which a particular model is tuned is generally not known; but experience suggests that tuning involves trade-offs, which keeps the list small, and is focused on globally integrated measures of skill related to budgets of energy, mass and momentum.

It bears emphasizing that model tuning is only the final step in the model development process, and this process as a whole is guided by an awareness of deficiencies in the simulation of current and past climate states (see Section 9.1.3). For instance whether or not one incorporates specific processes may be guided by deficiencies in previous model simulations, by emerging information about the potency of particular effects (e.g., Hoose et al., 2009), or the importance of some mode of variability (Kim et al., 2011a). Although these tuning steps complicate the a posteriori evaluation of models, the non-linear nature of the climate system (and the models that represent) it makes it impossible to tune many of the details of model solutions. Because this is true, it would be hard to construct models that produced any solution one liked. In particular,

every model that reasonably reproduces historical climate shows substantial warming as a result of increasing CO_2.

The use of all available data is both integral and important to the model development process, and this complicates the construction of critical tests. Nonetheless, by focusing on emergent quantities while discounting metrics clearly related to model tuning, it is possible to gain insight into model skill. The concurrent use of a large number of different model quantities, evaluation techniques, and performance metrics also helps ensure that the evaluation provides a stringent test of model quality. The application of sophisticated evaluation techniques that aim at identifying and testing key relationships or correlations between several quantities is another severe test for climate models, as those relationships are usually not considered in model tuning (see Sections 9.2.2 and 9.7.2).

In summary, model tuning is an intrinsic part of model development and arises naturally from the fact that climate simulations depend to a significant degree on the representation of poorly understood processes. Model tuning directly influences the evaluation of climate models in that it leads to grades of the severity of tests that models can undergo. The application of model tuning raises the question of whether climate models are a reliable source of knowledge for those aspects of the system toward which they were not tuned, for instance cloud feedbacks in a future warmer climate. Models are not tuned to match a particular future climate; they are tuned to faithfully reproduce the past. What emerges, however, is that models that plausibly reproduce the past universally produce significant warming under increasing greenhouse gas concentrations, and this fact underlies the broad consensus behind the results presented in this report.

[END BOX 9.1 HERE]

[INSERT FIGURE 9.1 HERE]

Figure 9.1: A sketch illustrating the interconnection of model components in Atmosphere-Ocean General Circulation Models (AOGCMs) and Earth system models (ESMs), and the way in which parameterisations are embedded in each component.

The increasingly modular nature of process-descriptions and model components has led to an increased use of shared software infrastructures (Collins et al., 2005; Valcke et al., 2006) and an increased sharing and exchange of model components. This sharing of components has consequences for evaluating the ensemble of the CMIP models, as not all of them can be considered completely independent (see Section 9.8). Table 9.1 provides an overview over the models used here, including an indication of where components are similar between models.

[INSERT TABLE 9.1 HERE]

Table 9.1: Salient features of the AOGCMs and ESMs participating in CMIP5. Column 1: identification (Model ID) along with the calendar year ('vintage') of the first publication for each model; Column 2: sponsoring institution(s), main reference(s) and flux correction implementation (*not yet described*); Subsequent Columns for each of the 8 CMIP5 realms: component name, code independence and main component reference(s). Additionally, there are standard entries for the Atmosphere realm: horizontal grid resolution, number of vertical levels, grid top (low or high top); and for the Ocean realm: horizontal grid resolution, number of vertical levels, top level, vertical coordinate type, ocean free surface type ("Top BC"). This table information was automatically extracted from the CMIP5 online questionnaire (http://q.cmip5.ceda.ac.uk/) as of 12 November 2011.

9.1.3.2 Parameterisations

As noted above, parameterisations are included in all model components to represent processes that cannot be explicitly resolved. Parameterisations are continuously developed in a process-oriented way (Section 9.2.2.2) and are evaluated both in isolation and in the context of the full model. The purpose of this section is to highlight recent developments in the parameterisations employed in each model component.

9.1.3.2.1 Atmosphere

Atmospheric models must parameterise a wide range of processes, including processes associated with atmospheric convection and clouds, cloud-microphysical and aerosol processes, boundary layer processes, as well as radiation and the treatment of unresolved gravity waves.

The treatment of atmospheric convection remains one of the most critical areas in atmospheric models. While there have been no major developments in the basic approach to this problem, there have been important refinements in existing convection parameterisations. A long-standing weakness of convection in climate models has been the lack of sensitivity of the development of convective clouds to their environment (Derbyshire et al., 2004). This has been a focus of development since the AR4, and has resulted in improved simulations of tropical variability (Bechtold et al., 2008; Chikira, 2010; Chikira and Sugiyama, 2010; Neale et al., 2008); Derbyshire et al., 2011). Another focus has been on improving of the transport of momentum in convection parameterisations (Richter et al., 2008). In many climate models, cumulus parameterizations now calculate the typical vertical velocity in cumulus updrafts to more realistically represent cloud microphysics and cloud droplet activation as well as the updraft dynamics. The work on an alternative approach to convection in climate models by using so-called super-parameterisations has progressed since the AR4 and continues to yield promising results, albeit at much increased computational cost (Demott et al., 2007, 2010; Khairoutdinov et al., 2008; Tao and Moncrieff, 2009; Tao et al., 2009; Zhu et al., 2009).

Previous assessments have highlighted the important role of cloud processes in modelled climate sensitivity. There have been some improvements in the underlying algorithms to determine the existence and structure of cloud fields, typically based on the use of probability density functions of thermodynamic variables (Watanabe et al., 2009). As cloud representations in climate models increasingly aim to represent the influence of aerosols on cloud evolution (see Chapter 7), there has also been considerable effort to improve the representation of cloud microphysical processes. This has led to upgrades in the treatment of atmospheric radiation modules (Rotstayn et al., 2010), so that the radiative effects of aerosols can be included in a physically consistent fashion. The treatment of the radiative effects of clouds has also seen significant development (Barker et al., 2008)

Improvements in representing the atmospheric boundary layer since the AR4 have focussed on basic boundary-layer processes, the representation of the stable boundary layer, and boundary layer clouds (Teixeira et al., 2008). Several global models have successfully adopted new approaches to the parameterization of shallow cumulus convection and moist boundary layer turbulence that acknowledge their close mutual coupling. One new development is the Eddy-Diffusivity-Mass-Flux (EDMF) approach (Neggers, 2009; Neggers et al., 2009; Siebesma et al., 2007). This approach, like the shallow cumulus scheme of (Park and Bretherton, 2009), determines the cumulus-base mass flux from the statistical distribution of boundary layer updraft properties, a conceptual advance over the ad-hoc closure assumptions used in the past. The realistic treatment of the stable boundary layer remains difficult (Beare et al., 2006; Cuxart et al., 2006; Svensson and Holtslag, 2009).

The influence of internal gravity waves on the general circulation and mass distribution of the troposphere and lower stratosphere has been well established by the success of early efforts to parameterise unresolved orographic gravity-wave drag (GWD) (e.g., Palmer et al., 1986; McFarlane, 1987). These initial parameterisations concentrated primarily on effects associated with the saturation (Lindzen, 1981) and critical-level interaction of freely propagating gravity waves. More recently, there have been efforts to develop more sophisticated parameterisations of orographically forced flows which include sources of low-level drag such as blocking, lee vortices, downslope windstorm flow, and trapped lee waves (e.g., Lott and Miller, 1997; Gregory et al., 1998; Scinocca and McFarlane, 2000).

The parameterisation of drag due to non-orographic gravity waves is becoming a common feature of GCMs that include the middle atmosphere (i.e., stratosphere and mesosphere). The basic wind and temperature structure of the middle atmosphere arises largely from a balance between radiative driving and (primarily non-orographic) GWD (Holton, 1983). The term non-orographic refers to the fact that the sources of these waves (e.g., convection and frontal dynamics) are non-stationary and so induce waves with non-zero horizontal phase speeds. In the stratosphere, such GWD is essential to the driving of both the quasi-biennial oscillation in the Tropics (Dunkerton, 1997) and the equator-to-pole residual circulation in the summer hemisphere (Alexander and Rosenlof, 1996). Over the past decade, the importance of a well-resolved stratosphere on tropospheric prediction and projection has become increasingly clear.

9.1.3.2.2 Ocean

Mesoscale and submesoscale eddy parameterisations

Ocean components in CMIP5 models generally have horizontal resolution that is too coarse to admit mesoscale eddies. Consequently, such models typically employ some version of the (Redi, 1982) neutral diffusion and (Gent and McWilliams, 1990) eddy advection parameterisation (see also Gent et al., 1995; McDougall and McIntosh, 2001). Since the AR4, the main research focus has been on how parameterised mesoscale eddy fluxes in the ocean interior interact with boundary-layer turbulence; some CMIP5 models have implemented such features (Gnanadesikan et al., 2007; Ferrari et al., 2008; Ferrari et al., 2010) and (Danabasoglu et al., 2008). Another focus concerns specification of the eddy diffusivity, with many more CMIP5 models employing flow dependent diffusivities than CMIP3 models. Both refinements to the eddy parameterisations are important for the mean state and the response to changing forcing, especially in the Southern Ocean (Boning et al., 2008; Farneti and Gent, 2011; Farneti et al., 2010; Gent and Danabasoglu, 2011; Hallberg and Gnanadesikan, 2006; Hofmann and Morales Maqueda, 2011).

In addition to mesoscale eddies, there has been a growing awareness since CMIP3 of the effects that submesoscale eddies and fronts play in restratifying the mixed layer, with Boccaletti et al. (2007), Fox-Kemper et al. (2008), and Klein and Lapeyre (2009) representative of this effort, and with the parameterisation of Fox-Kemper et al. (2011) used in some CMIP5 models.

Parameterisations of dianeutral transformation

There is an active research effort on the representation of dianeutral mixing associated with breaking gravity waves (MacKinnon et al., 2009), with this work adding rigor to the prototype abyssal tidal mixing parameterisation of (Simmons et al., 2004) now used in several climate models (e.g., Jayne, 2009). The transport of dense water down-slope with gravity currents (e.g., Legg et al., 2008; Legg et al., 2009) has also been the subject of focused efforts, with associated parameterizations making their way into some CMIP5 models (Danabasoglu et al., 2010; Jackson et al., 2008b; Legg et al., 2009).

Additional work has led to several CMIP5 models having interactive ocean biogeochemistry for their standard configuration, with impacts on sunlight penetration and upper ocean mixing in both open-ocean and sea-ice regions (Lengaigne et al., 2009; Lengaigne et al., 2007).

Ocean biogeochemical (OBGC) models

Oceanic uptake of CO_2 is highly variable and is determined by the interplay between the biogeochemical and physical processes in the ocean. Most CMIP5 OBGC models are based on so-called NPZD-type models that partition marine ecosystems into nutrients, plankton, zooplankton, and detritus. These models allow simulation of some of the important feedbacks between climate and oceanic CO_2 uptake, but are limited by the lack of marine ecosystem dynamics. Some efforts have been made to include more plankton groups or plankton functional types in the models (PFPs; Le Quere et al., 2005) with as-yet uncertain implications for Earth system response.

Ocean acidification and the associated decrease in calcification in many marine organisms provides a negative feedback on atmospheric CO_2 (Ridgwell et al., 2007). New-generation OBGC models therefore include various parameterisations of calcium carbonate ($CaCO_3$) production as a function of the saturation state of seawater with respect to calcite (Gehlen et al., 2007; Ilyina et al., 2009; Ridgwell et al., 2007) or pCO2 (Heinze, 2004). On centennial scales, deep-sea carbonate sediments neutralize atmospheric CO_2. A growing number of CMIP5 models include the sediment carbon reservoir, and progress has been made towards refined sediment representation in the models (Heinze et al., 2009).

9.1.3.2.3 Land

Land-surface properties such as vegetation, soil type, and the amount of water stored on the land as soil moisture, snow, and groundwater, all strongly influence climate, particularly through their effects on surface albedo and evapotranspiration. These climatic effects can be profound; for example, it has been suggested that changes in the state of the land-surface played an important part in the severity and length of the 2003 European drought (Fischer et al., 2007a).

The land-surface schemes employed in GCMs calculate the fluxes of heat, water, and momentum between the land and the atmosphere, and update the surface state variables such as soil moisture, soil temperature and snow-cover, that influence these fluxes. There has been a steady increase in the complexity of land-surface components on GCMs from the first generation soil “bucket” models employed in the 1970s

(Manabe, 1969) to fourth-generation schemes that attempt to model vegetation controls on transpiration through stomatal pores on their leaves (Cox et al., 1999; Sellers et al., 1996). However, even the more complex land surface schemes used in the AR4 suffered from obvious simplifications, such as the need to prescribe rather than simulate the vegetation cover, and a tendency to ignore lateral flows of water and sub-gridscale heterogeneity in soil moisture (Pitman, 2003).

Since the AR4 land-surface model development has focused on overcoming some of these limitations. A number of climate modelling groups now include some representation of sub-gridscale hydrology (Gedney and Cox, 2003; Oleson et al., 2008b) and most also include a large-scale river network model to route runoff to the appropriate ocean outflow points (Oki et al., 1999).

The evaluation of land-surface schemes is in principle more straightforward than for other components of climate models, because they can be tested easily in “standalone” or “offline” mode. The meteorological data required to drive land models is generally available and there is a growing amount of data for validation from flux towers (Baldocchi et al., 2001) and Earth Observation. International initiatives are now underway to develop benchmarking tools for land-surface models based on these copious observations (Randerson et al., 2009).

9.1.3.2.4 Ice

Most large-scale sea-ice processes are well understood and well represented in models (Hunke et al., 2011). For example, the basic thermodynamic description has been available for 40 years (Maykut and Untersteiner, 1971), and a relatively straightforward representation of sea-ice dynamics is nearly 35 years old (Hibler, 1979). Schemes like this capture the first-order behaviour of sea ice in the climate system. Since the AR4, in which there was a major advance associated with inclusion of sea-ice dynamics in most AOGCMs, progress in improving sea-ice components in climate models has apparently slowed. Sea ice model development is now addressing higher-order effects: (1) more precise descriptions of physical processes such as microstructure evolution and anisotropy; and (2) extensions of the model for “Earth system” simulations, by including biological and chemical species.

Sea ice dynamics

The Arctic and Antarctic sea ice packs are mixtures of open water, thin first-year ice, thicker multiyear ice, and even thicker pressure ridges. An essential aspect of sea ice thermodynamics is the variation of growth and melting rates for different ice thicknesses. Thin ice grows and melts more quickly than thicker ice. Similarly, thinner ice is more likely to undergo mechanical deformation than thicker ice. Most early sea ice models neglected sub-grid-scale thickness variations, but many models now include some representation of the thickness distribution, and a description of mechanical redistribution that converts thinner ice to thicker ice under convergence and shear.

In most models, ice area fraction, ice volume, and snow volume are advected horizontally. As in other model components, details of the advection scheme become particularly important when there are large spatial gradients and when certain physical constraints (such as positive-definiteness) must be respected. New approaches include more accurate , nearly monotonic schemes for conserved fields (e.g., ice area and volume), but not for tracers, (e.g., Vancoppenolle et al., 2009b; Lipscomb and Hunke, 2004).

Sea ice thermodynamics

Sea ice albedo has long been recognized as a critical aspect of the global heat balance. The average surface albedo on the scale of a climate model grid cell is (as on land) the result of a mixture of surface types: bare ice, melting ice, snow-covered ice, open water, etc. The parameterisation of surface albedo remains uncertain and is often tuned to produce a realistic simulation of sea ice extent, compensating for deficiencies in both atmosphere and ocean forcing, (e.g., Losch et al., 2011). Many sea ice models still use a relatively simple albedo parameterisation that specifies four albedo values: cold snow; warm, melting snow; cold, bare ice; and warm, melting ice. Other models use more complex formulations that take into account the ice and snow thickness, spectral band, and other parameters. Solar radiation may be distributed within the ice column assuming exponential decay (Beer's Law) or via more complex multiple-scattering radiative-transfer scheme (Briegleb and Light, 2007).

Melt ponds form in depressions on the surface of the ice and can drain through interconnected brine channels when the ice becomes warm and permeable. This flushing can effectively clean the ice of salt, nutrients, and other inclusions. There are several schemes for modelling melt ponds and surface melt. The simplest and most widely used does not involve tracking melt water, but rather decreases the ice surface albedo under warm, melting conditions, (e.g., Vancoppenolle et al., 2009b). More realistic melt pond schemes are under development (Flocco et al., 2010; Scott and Feltham, 2010).

Salinity affects thermodynamic properties of sea ice, and is used in the calculation of fresh water and salt exchanges at the ice-ocean interface. Some models allow the salinity to vary in time (Schramm et al., 1997), while others assume a salinity profile that is constant, (e.g., Bitz and Lipscomb, 1999). For LIM3, Vancoppenolle et al. (2009a) developed a simplified approach to simulate the desalination of Arctic sea ice as it grows and then transitions from first-year to multi-year ice. Another new thrust in climate modelling is the inclusion of chemistry and biogeochemistry (Piot and von Glasow, 2008a, 2008b; Zhao et al., 2008), with dependencies on the ice microstructure and salinity profile. Related work involves the vertical transport and cycling of quantities such as aerosols (Bailey et al., 2010) and gases (Nomura et al., 2010) that pass gradually through the ice and can modify oceanic or atmospheric chemistry.

One of the difficulties in evaluating the sea-ice component of a climate model is that errors arise from not only the sea-ice component itself, but also from errors in the atmosphere above and the ocean below (e.g., Bitz et al., 2002) and, because of the strong ice-albedo feedback, these errors are amplified.

9.1.3.3 New Components and Couplings: Biogeochemical Feedbacks and the Emergence of Earth System Modelling

9.1.3.3.1 Carbon Cycle

The omission of internally-consistent feedbacks between the physical, chemical, and biogeochemical processes in the climate system is an inherent feature of AOGCMs. The conceptual issue is that the physical climate controls the natural sources and sinks of CO_2 and CH_4, the two most important long-lived greenhouse gases (LLGHGs). ESMs incorporate many of the important biogeochemical processes, making it possible to simulate the evolution of radiatively active species based upon their emissions from natural and anthropogenic sources together with their interactions with the rest of the Earth system. Alternatively, when forced with specified concentrations, one can diagnose these sources (with feedbacks included). Given the large natural sources and sinks of CO_2 relative to its anthropogenic emissions, and given the primacy of CO_2 among anthropogenic GHGs, one of the most important enhancements is the addition of terrestrial and oceanic carbon cycles. These cycles have been incorporated into many models (Christian et al., 2010; Tjiputra et al., 2010) used to study the long-term evolution of the coupled Earth system under anthropogenic climate change (Schurgers et al., 2008).

9.1.3.3.2 Aerosols

The treatment of aerosols has advanced since the AR4. Many ESMs now include the basic features of the sulphur cycle and so represent both the direct effect of sulphate aerosols, along with some of the more complex indirect effects involving cloud droplet number and size. Further, several ESMs are currently capable of simulating the mass, number, size distribution, and mixing state of interacting multicomponent aerosols (Bauer et al., 2008). The incorporation of more physically complete representations of aerosols often improves the simulated climate under historical and present-day conditions, including the mean pattern and interannual variability in continental rainfall (Rotstayn et al., 2010). However, the representation of aerosols and their interaction with clouds and radiative transfer remains an important source of uncertainty in the simulation of historical climate change and climate sensitivity.

9.1.3.3.3 Methane cycle and permafrost

In addition to carbon dioxide, an increasing number of ESMs are also incorporating prognostic methane to quantify the feedbacks from changes in methane sources and sinks under a warming climate. Some models now simulate the evolution of permafrost carbon stock (Khvorostyanov et al., 2008a; Khvorostyanov et al., 2008b), and in some cases this is integrated with the representation of terrestrial and oceanic methane cycles (Volodin et al., 2010; Volodin, 2008a).

9.1.3.3.4 Dynamic global vegetation models and wildland fires

One of the potentially more significant effects of climate change is the alteration of the distribution, speciation and life cycle of vegetated ecosystems. In order to include these effects in projections of climate change, several dynamic global vegetation models (DGVMs) have been developed and deployed in AOGCMs and ESMs (Ostle et al., 2009) (Cramer et al., 2001; Sitch et al., 2008),. DGVMs can simulate the interactions among natural and anthropogenic drivers of global warming, the state of terrestrial ecosystems, and ecological feedbacks on further climate change. The incorporation of DGVMs has required considerable enhancement and improvement in coupled models to produce stable and realistic distributions of flora (Oleson et al., 2008b). The improvements include better treatments of surface, subsurface, and soil hydrological processes, the exchange of water with the atmosphere, and the discharge of water into rivers and streams. While the first DGVMs have been primarily coupled to the carbon cycle, the current generation of DGVMs are being extended to predict the ecological sources and sinks of other non-CO_2 trace gases including CH_4, N_2O, biogenic volatile organic compounds (BVOCs), and nitrogen oxides collectively known as NO_x (Arneth et al., 2010). BVOCs and NO_x can alter the lifetime of some GHGs and act as precursors for secondary organic aerosols (SOAs) and ozone. Disturbance of the natural landscape by fire has significant climatic effects through its impact on vegetation and air quality and through its emissions of greenhouse gases, aerosols, and aerosol precursors. Since the frequency of wildland fires increases rapidly with increases in ambient temperature, the effects of fires are projected to grow over the 21st century. The interactions of fires with the rest of the climate system are now being introduced into ESMs (Arora and Boer, 2005; Pechony and Shindell, 2009).

9.1.3.3.5 Land-use / land-cover change

The impacts of land-use and land-cover change (LULCC) on the environment and climate are explicitly included as part of the representative concentration pathways (RCPs) used for climate projections to be assessed in later chapters (Moss et al., 2010). Several important types of LULCC include effects of agriculture and changing agricultural practices, including the potential for widespread introduction of biofuel crops; the management of forests for preservation, wood harvest, and production of woody biofuel stock; and the global trends toward greater urbanization. ESMs include increasingly detailed treatments of crops and their interaction with the landscape (Arora and Boer, 2010; Smith et al., 2010a; Smith et al., 2010b), forest management (Bellassen et al., 2010; Bellassen et al., 2011), and the interactions between urban areas and the surrounding climate systems (Oleson et al., 2008a).

9.1.3.3.6 Chemistry-climate models

Stratospheric cooling and ozone recovery projected for the 21st century may affect the entire climate system, including the positions of the subtropical jets, atmospheric temperatures over Antarctica, the strength of the Brewer-Dobson circulation (Butchart et al., 2010; CCMVal, 2010; Son et al., 2008; Son et al., 2010; WMO, 2011b), and the shift in tropopause height. Chemistry-climate model (CCM) simulations of stratospheric ozone and related climate effects have been examined for common features through multi-model intercomparisons such as the first and second rounds of the CCM Validation (CCMVal) activity (CCMVal, 2010; Eyring et al., 2007; Son et al., 2008). CCMs are three-dimensional atmospheric climate models with fully coupled chemistry, i.e., where chemical reactions drive changes in atmospheric composition which in turn change the atmospheric radiative balance and hence dynamics. In general, these models have been operated with prescribed sea surface temperatures and sea ice concentrations rather than as a full AOGCM due to the computational expense of the reactive chemistry and the extension of the vertical domain through the stratosphere to the middle and upper atmosphere. Several of the stratospheric chemistry-climate models evaluated in CCMVal have been incorporated into ESMs and are part of the CMIP5 ensemble. Important chemistry-climate interactions have also been identified in tropospheric ozone projections for the 21st century. For example, tropospheric ozone may increase due to positive climate feedbacks such as an increased influx from the stratosphere, which then increases radiative forcing and thus impacts on climate. Several of the CMIP5 models currently simulate tropospheric chemistry interactively [tbc] , and those that include tropospheric and stratospheric chemistry can be used for internally consistent simulations of the interactions among stratospheric cooling, ozone recovery, and the rest of the climate system (Jöckel et al., 2006; Lamarque et al., 2011).

9.1.3.3.7 High-top/low-top global models

It is now widely accepted that in addition to the well-known effect that tropospheric circulation and climate change influence the stratosphere, stratospheric dynamics can in turn influence the surface climate

(CCMVal, 2010). To reduce uncertainties in the representation of climate variability on seasonal to multi-decadal timescales and to improve the representation of upper-troposphere dynamics, many climate models have the ability to include a fully resolved stratosphere with a model top above the stratopause. The subset of CMIP5 models with this so-called high-top configuration allows a multi-model comparison to the standard set of low-top models (i.e., those with a model top in the middle stratosphere).

9.1.3.3.8 Land ice sheets

The amount melt water that could be released from the Greenland and Antarctic ice sheets in response to climate change remains a major source of uncertainty in projections of sea-level rise. As recently as the AR4, the long-term response of these ice sheets to alterations in the surrounding atmosphere and ocean has been simulated using offline models. Several ESMs currently have the capability to have ice-sheet component models coupled to the rest of the climate system (Vizcaino et al., 2008), and idealized experiments suggest that the coupling causes accelerated melting of the Greenland ice sheet compared to passive coupling used in prior studies (Vizcaino et al., 2010).

9.1.3.3.9 New features in ocean-atmosphere coupling

Several new features have arisen in the coupling between the ocean and the atmosphere since AR4. The bulk formulae used to compute the turbulent fluxes of heat, water, and momentum at the air-sea interface have been revised. A number of models now consider the surface current when calculating surface wind shear and hence wind stress (e.g., Jungclaus et al., 2006; Luo et al., 2005). The coupling frequency has been increased in some cases to include the diurnal cycle, which was shown to improve SST bias in the tropical Pacific (Bernie et al., 2008; Ham et al., 2010). Several models now represent the coupling between the penetration of the solar radiation into the ocean and light-absorbing chlorophyll, with some implications on the representation of the mean climate and climate variability (Wetzel et al., 2006). This coupling is achieved either by prescribing the chlorophyll distribution from observations, or by computing the chlorophyll distribution with an ocean biogeochemical model (e.g., Arora et al., 2009). In the latter case there is a feedback between the solar radiation and the upper ocean characteristics that has an impact on the mean state and interannual variability in the Pacific (Lengaigne et al., 2007) or that modifies the seasonal melting on sea ice and the hydrological cycle of the Arctic (Lengaigne et al., 2009).

9.1.3.4 Resolution

9.1.3.4.1 Resolution of AOGCMs

Improved resolution in climate models (i.e., adopting a finer grid in the modelled atmosphere, ocean and other components) is expected to improve some aspects of model performance. Generally, improved resolution leads to better representation of finer scale structures, such as atmospheric and oceanic eddies and vertical stratification, as well as effects of finer scale topography, land-sea distribution and land cover. However, because of uncertainties in parameterisations and the way in which they scale with resolution, expected improvements are not always realized.

The typical horizontal resolution for current AOGCMs and ESMs is roughly 1 to 2 degrees for the atmospheric component and around 1 degree for the ocean (Table 9.1). Associated with increases in computational capacity, there has been some modest increase in model resolution since the AR4, especially for the near-term simulations (e.g., around 0.5 degree for the atmosphere in some cases). On the other hand, for the models used for the long-term simulations with interactive biogeochemistry, the resolution has not increased significantly due to the trade-off against higher complexity in such models.

In some cases, higher resolution may lead to a stepwise, rather than incremental, improvement in model performance (e.g., Roberts et al., 2004; Shaffrey et al., 2009). For example, oceanic models undergo a transition from laminar to turbulent when the computational grid contains more than one or two grid points per first baroclinic Rossby radius (i.e., finer than 50 km at low latitudes and 10 km at high latitudes) (McWilliams, 2008; Smith et al., 2000). Such mesoscale eddy-permitting ocean models better capture the large amount of energy contained in fronts, boundary currents, and time dependent eddy features (e.g., (McClean et al., 2006b). Models run at such resolution have been used for simulations of climate time-scales (decadal to centennial) and found to be promising, though much work remains before they are as mature as

the coarser models currently in general use (Bryan et al., 2007; Bryan et al., 2010; Farneti et al., 2010; McClean et al., 2011)

Similarly, atmospheric models with grids that allow the explicit representation of convective cloud systems (i.e., finer than a few km) avoid employing a convective cloud parameterisation, which has long been a source of uncertainty in climate modelling. For example, Miura et al. (2007) demonstrated that a Madden-Julian oscillation event, which is generally difficult to be realistically represented in current generation AOGCMs, is simulated well in a global cloud-system-permitting (3.5 km resolution) AGCM. However, this kind of simulation is limited to short simulations (typically up to a month) and not coupled to a dynamic ocean model, given current computational capacity. Moreover, a cloud-system-permitting model still depends on parameterisations for cloud microphysics and moist turbulence. Recent developments in this area are assessed in Chapter 7.

9.1.3.4.2 Downscaling methods

Regional climate models share many parameterisation and resolution issues with global climate models, though as a rule, regional climate models employ higher resolution than global climate models. Since the AR4, the typical regional climate model resolution has increased from around 50 km to around 25 km (e.g., Christensen et al., 2010). There are also regional climate models run at around 10 km resolution or higher (e.g., van Roosmalen et al., 2010; Kanada et al., 2008; Sasaki and Kurihara, 2008; Kusaka et al., 2010).

The quality of the AOGCM boundary conditions provided to the RCM is fundamental for the quality of downscaling results. Biases in the former inevitably lead to biases in the latter. Downscaling by means of an RCM is not meant to alter the larger-scale features that are resolved in the driving AOGCM (Sanchez-Gomez et al., 2009). Rather, the aim is a more detailed representation and simulation of climate processes, e.g., extremes. This is discussed further in conjunction of the 'value added' in Section 9.6.1.

9.2 Techniques for Assessing Model Performance

9.2.1 Objectives and Limitations

Systematic evaluation of models through comparisons with observations is a prerequisite to understanding and improving the representation of physical and biogeochemical processes and feedbacks. In addition, the identification and interpretation of the spread amongst state-of-the-art model simulations can serve as a means of quantifying uncertainty in cases where observations are lacking. The objective of climate model evaluation is to improve understanding of their strengths and weaknesses and to quantify improvements over time. An evaluation of the models with respect to how well they represent historical and present-day climate can also be used to guide the assessment of climate projections and their credibility.

In the AR4, the evaluation of climate models was mainly done somewhat qualitatively by comparing simulated and observed fields (e.g., time series or spatial maps). Since the AR4, performance metrics, which are statistical measures of agreement between a simulated and observed quantity (or covariability between quantities), have been more extensively used. Performance metrics derived from a variety of observationally-based diagnostics provide an objective synthesis and visualization of model performance (Cadule et al., 2010; Gleckler et al., 2008; Pincus et al., 2008; Waugh and Eyring, 2008) and enable quantitative assessment of model improvements, both for different versions of individual models and for community-wide collections used in international assessments (e.g., CMIP2 versus CMIP3, Reichler and Kim, 2008). These metrics can also be used to explore the value of weighting projections based on model performance, although for this purpose the need for process-oriented evaluation, especially for those processes that are related to known feedbacks, has been emphasized (Knutti et al., 2010b).

Despite these developments, quantitative assessment of climate model skill is still limited for a number of reasons. Unlike weather prediction models, which can be routinely tested, the evaluation of climate models must be undertaken across a range of much longer time scales. Climate model evaluation therefore requires the availability of long-term, consistent, error-characterized global and regional Earth observations (satellite and in situ) as well as accurate globally gridded reanalyses in the atmosphere, the ocean, or, ultimately, the coupled system. Since the AR4, the Earth Observation community has undertaken a large effort to develop consistent error-characterized data sets of selected Essential Climate Variables (ECVs). Observational

uncertainty can be included in model evaluation either by using error estimates provided with the observational data set, or by using more than one data set. The latter is a more common approach because many observational datasets are not accompanied by formal error estimates, but where multiple estimates exist they are often based on the same underlying measurements and therefore are not truly independent. Finally, the lack of long-term observations, observations for process-oriented model evaluation and observations in particular regions (e.g., polar areas, the upper troposphere / lower stratosphere (UTLS), and the deep ocean) remains an impediment.

9.2.2 New Developments in Model Evaluation Approaches

9.2.2.1 Evaluating the Overall Model Results

The most straightforward approach to evaluate models is to compare overall simulated fields (e.g., global distributions of temperature, precipitation etc.) with corresponding observations. For more quantitative comparison, statistical measures can be used (e.g., root-mean square error, centred and uncentred pattern correlations) (Abe et al., 2009; Annan and Hargreaves, 2010; Giorgi and Coppola, 2010; Gleckler et al., 2008; Knutti et al., 2010a; Raisanen et al., 2010; Reichler and Kim, 2008; Whetton et al., 2007). Many studies have also examined whether the performance of a model in reproducing past climate is related to the behaviour of the model in projecting future climate (see further discussion in Section 9.8). There are also some attempts to reduce redundancy of multiple performance metrics through statistical methods such as cluster analysis (Nishii et al., 2011; Yokoi et al., 2011).

Statistical techniques have also been applied to evaluate characteristics of a whole ensemble of models, instead of individual models in an ensemble. If each model is a random and independent sample from a distribution of possible models centred on the observations, the errors in the ensemble average are expected to converge to near zero as the ensemble size increases. Knutti et al. (2010a) have tested this hypothesis for the CMIP3 ensemble and found that the reduction of biases by averaging is slower than that expected, suggesting that there are some common biases across models. By contrast, Annan and Hargreaves (2010) showed that the biases in the ensemble mean converge to a value greater than zero as the ensemble size increases, based on a different statistical paradigm that the truth and ensemble members are drawn from the same distribution.

9.2.2.2 Isolating Processes

As discussed in Section 9.1.3.1 climate models heavily rely on processes parameterizations. It is therefore necessary to evaluate the representation of key processes both in the context of the full model, but also in isolation. A number of evaluation techniques to achieve this “process-isolation” have been developed.

One major stream of studies involves the so-called “regime-oriented” approach to process-evaluation. Instead of averaging model results in time (e.g., seasonal averages) or space (e.g., global averages), results are averaged within categories that describe physically important regimes of the system under study. Applications of this approach since AR4 include the use of circulation regimes (Bellucci et al., 2010; Bony and Dufresne, 2005; Brown et al., 2010b) or cloud regimes (Chen and Del Genio, 2009; Williams and Webb, 2009; Williams and Brooks, 2008). The importance of the regime-oriented approach lies in its ability to isolate processes that might be responsible for particular errors, and thereby directing more in-depth process studies (Jakob, 2010).

Another approach to process evaluation involves either the removal of a particular model component or process parameterisation from the host-model and the use of off-line simulations. The results of such simulations are compared to measurements from detailed field studies or to results from more sophisticated process models (Randall et al., 2003). Numerous important process-related data sets to support such evaluations have been collected since the AR4 (Illingworth et al., 2007; May et al., 2008; Redelsperger et al., 2006; Verlinde et al., 2007) and have been applied to the evaluation of climate model processes (Boone et al., 2009; Boyle and Klein, 2010; Hourdin et al., 2010; Xie et al., 2008). These studies are crucial to test the realism of the process formulations that underpin climate models, and they are assessed in Section 9.7.

9.2.2.3 Instrument Simulators

Satellites provide nearly global coverage, sampling all meteorological conditions. This makes them powerful tools for model evaluation. The standard approach of using satellite data is to convert the observed radiation information to *model-equivalents* through so-called retrievals (Stephens and Kummerow, 2007). Retrieved properties have been used in numerous studies to analyse the performance of clouds and precipitation in GCMs (Allan et al., 2007; Gleckler et al., 2008; Pincus et al., 2008). The main challenge using retrievals in model evaluation is that modelled and retrieved variables are difficult to define consistently due to limitations of the satellite sensors and the assumptions used in the retrievals. These limitations include sampling, poor vertical resolution, and the inability of satellite radiances to fully constrain the atmospheric state.

Instrument simulators allow more quantitative evaluation of climate models with satellite products. The approach calculates *observation-equivalents* from models using radiative transfer calculations to 'simulate' what the satellite products would provide if the satellite system were 'observing' the model. Microphysical assumptions (which differ from model to model) can be included in the simulators, avoiding retrieval inconsistencies. Since the AR4, the ISCCP simulator (Klein and Jakob, 1999; Webb et al., 2001; Yu et al., 1996) has continued to be used for model evaluation ((Wyant et al., 2009), (Chen and Del Genio, 2009), (Marchand et al., 2009), (Yokohata et al., 2010), often in conjunction with statistical techniques to separate model clouds into cloud regimes (e.g., Field et al., 2008; Williams and Webb, 2009; Williams and Brooks, 2008). New simulators for other satellite products have also been developed and are increasingly applied for model evaluation (Bodas-Salcedo et al., 2011). While often focussed on clouds and precipitation, the simulator approach has also been used successfully for other variables such as upper tropospheric humidity by comparing infrared or microwave satellite radiances to those computed from models (Allan et al., 2003; Brogniez et al., 2005; Iacono et al., 2003; Ringer et al., 2003; Zhang et al., 2008; Bodas-Salcedo et al., 2011; Brogniez and Pierrehumbert, 2007).

9.2.2.4 Paleoclimate Studies

Past climates offer a wide range of climatic states that can be used to test a model's response to different forcing (see Chapter 5); however this can be achieved only for periods with sufficient data coverage. Such data sets have been developed for the Last Glacial Maximum (21,000 years BP) and the mid-Holocene (6000 years BP), as part of the global ocean reconstruction from marine data (CLIMAP, 1981; Waelbroeck et al., 2009) and the Biome 6000 project (Prentice et al., 1998).

Paleo proxies, such as pollen or $\delta^{18}O$ in ice cores, are indirect measurements of climatic conditions (see Chapter 5). One approach to using such data is to compare a climate variable that best characterizes the major fluctuations of the proxy indicator. For example, the temperature of the coldest month or cumulative growing degree days are preferred to winter and summer temperature when comparing model outputs to estimates based on pollen records (Bartlein et al., 2010b). Recent work on marine proxies suggests that, depending on the region, the same proxy is not necessarily dominated by the same aspect of climate (seasonality, annual mean) (Jungclaus et al., 2010). In addition, different proxies record different local histories which must also to be accounted for in model-data comparisons (Leduc et al., 2010).

An alternative 'forward modelling' approach consists of simulating the proxy indicators themselves. This can be done by using either an off-line model or the inclusion of specific processes in the new generation of ESMs. For example, biome models have been used to simulated biome distributions for past climate conditions (Harrison et al., 1998). Some ESMs now include a dynamical vegetation module in their land surface scheme, so that the simulated vegetation can be directly compared to past vegetation (Braconnot et al., 2007d). Some models can be run with a representation of water isotopes which allows direct comparison of model output with isotopic measurements (LeGrande et al., 2006). In addition, there is growing interest in simulations of ocean tracers such as $\delta^{13}C$ of $\delta^{14}C$ that can provide additional comparisons with ocean proxies (Tagliabue et al., 2009).

9.2.2.5 Use of Data Assimilation and Initial Value Techniques

To be able to forecast the weather a few days ahead, knowledge of the present state of the atmosphere is of primary importance. In contrast, climate predictions and projections (see Chapter 11) simulate the *statistics* of weather seasons to centuries in advance. Despite their differences, both weather predictions and simulations of future climate are performed with very similar atmospheric model components. Climate models can be integrated as weather prediction models if they are initialised appropriately (Phillips et al., 2004) providing the advantage of testing some of the sub-grid scale physical processes that are parameterised in models on their characteristic time scales, without the complication of feedbacks altering the underlying state of the atmosphere (Boyle et al., 2005; Martin et al., 2010; Pope and Stratton, 2002; Williamson et al., 2005). It has been demonstrated that the application of a unified weather-climate modelling framework can provide significant advantages to model development (Martin et al., 2010).

Another approach to model evaluation involves the analysis of errors in the initial tendencies in a forecast (Klinker and Sardeshmukh, 1992). Data assimilation experiments have shown that this initial tendency methodology can be used to exclude certain model parameter values associated with high climate sensitivities (Rodwell and Palmer, 2007). Ensemble data assimilation provides another opportunity for a more systematic assessment of model error and can be used in the estimation of optimal model parameters (Koyama and Watanabe, 2010).

Longer time-scale initial value simulations, such as those for seasons and decades ahead, have also been shown to have potential as useful tools for climate model evaluation (Palmer et al., 2008). Decadal prediction in particular holds much promise to apply initial value techniques to the evaluation of ocean models. As with atmospheric data assimilation, it is evident that ocean data assimilation will provide a useful opportunity for the assessment of ocean processes at their characteristic timescales (Balmaseda et al., 2008; Bell et al., 2004).

9.2.2.6 Evaluation Techniques for RCMs

RCMs share many of the same evaluation techniques used for AOGCMs; however, there are a few techniques that are particular to RCM evaluation.

As AOGCMs often have various regional biases in their large-scale circulation (van Ulden and van Oldenborgh, 2006), biases in an RCM with boundary conditions taken from an AOGCM results may be due either to biases in the boundary conditions or to the representation of regional processes in the RCM itself (Deque et al., 2005; Déqué et al., 2007). The former can be circumvented by use of atmospheric reanalyses as boundary conditions. These so-called ‘perfect boundary condition’ experiments refer to the use of reanalyses as boundary conditions instead of AOGCM output (Christensen et al., 1997), and allow one to focus on particular anomalous weather or climate events in evaluating model performance.

Another evaluation technique specific to RCMs is the transferability experiment. This involves applying an RCM to a different region than the one it was initially developed for (Takle et al., 2007). In a typical transferability experiment RCMs are applied to multiple geographic regions, thus exposing the RCMs to different regional climates and exposing model biases (Takle et al., 2007).

A difficulty in the evaluation of RCMs, that is more acute than for global models, is the relative sparseness of observational networks in many regions (e.g., Nikulin et al. (2011), Driouech et al. (2009)). This has consequences for evaluating high resolution model results (Hofstra et al., 2010) since, for example, gridding station data affects the intensity of extremes.

9.2.2.7 Characterization of Model Uncertainty through Ensemble Approaches

Uncertainty in climate model simulations is a consequence of uncertainties in initial conditions, boundary conditions, parameter values, and structural uncertainties in the model design (Hawkins and Sutton, 2009; Knutti et al., 2010a; Tebaldi and Knutti, 2007a). Ensemble methods have been used extensively since the AR4 to understand the relative contributions of these sources of uncertainty. The methods employed are generally of two types: Multi-model Ensembles (MME) and Perturbed Physics Ensembles (PPE). The MME is created by gathering the existing model simulations from several climate modelling centres, whereas the

PPE involves only a single model. The merits of each are distinct and both can contribute to the model evaluation process.

Although the emphasis in this Chapter is on multi-model evaluation, ensembles constructed within a single-model framework can be useful for characterizing certain aspects of model uncertainty. Current methods to assess uncertainty from a single model include generating ensembles by perturbing parameters and then estimating the ability of each member to match specific observations or constraints. These ensemble-based methods have been used frequently in simpler models such as EMICs, (Forest et al., 2006, 2008; Forest et al., 2002; Knutti and Tomassini, 2008; Sokolov et al., 2009; Stott and Forest, 2007; Xiao et al., 1998) and are now being applied to more complex models (Brierley et al., 2010a; Collins et al., 2007; Collins et al., 2006a; Sanderson et al., 2008a; Stainforth et al., 2005).

While there is considerable evidence that a multi-model mean generally compares better with observations for a variety of diagnostics, because model errors tend to cancel (Gleckler et al., 2008; Pierce et al., 2009; Pincus et al., 2008; Reichler and Kim, 2008), the development of ensemble techniques for climate modelling is an active area of research that addresses several shortcomings of previous uncertainty assessments. For example, Knutti et al. (2010a) showed that averaging results from multiple model simulations leads to a loss of signal for precipitation change because models simulate similar overall patterns but slightly shifted in space. In addition, the development of climate models has occurred via sharing of specific model components and so certain lineages exist at most major modelling centres. This suggests that groups of models share biases, that the assumption of model independence is not correct, and therefore that the effective number of independent models is likely smaller than the actual number of models in the MME (Jun et al., 2008; Knutti, 2010; Knutti et al., 2010a; Pennell and Reichler, 2011; Tebaldi and Knutti, 2007a). By exploring the likelihoods of each model in reproducing historical climate, a likelihood-based ensemble can be derived to avoid this simplistic use of the MME (Sokolov et al., 2010; Tebaldi and Knutti, 2007a).

9.2.3 Overall Summary of Approach that will be Taken in this Chapter

The model evaluation in the following focuses primarily on the comparison of models with observations or observationally-based products. Exploitation of the most comprehensive set of observations necessitates an emphasis on recent decades, although older 20th century records and paleo data also play an important role. In some circumstances valuable insight into a model's behaviour can be achieved without observations (via analysis of intermodal differences), but we will use this approach sparingly.

A rational progression of such a broad scope evaluation begins with an examination of the large-scale features of the mean state in each of the model components (Section 9.4). This is followed by an evaluation of the ability of models to capture the dominant features of natural variability on observable time scales, including extreme behaviour of particular relevance to society (Section 9.5). This path of increasing focus takes us to more regional evaluation of model performance, including approaches to augment regional information with downscaling techniques (Section 9.6), and finally, more detailed process-oriented evaluation (Section 9.7).

Throughout our evaluation, we rely on routine diagnostic methods to compare model simulations with observations, such as spatial maps and space or time decompositions (e.g., zonal means or anomaly time series). As the evaluation focuses on increasing detail, a sampling of more sophisticated diagnostic approaches will also be exploited. To complement these diagnostics, we also rely on performance metrics to quantify the level of agreement between models and observations. Performance metrics provide an approach to succinctly summarize model performance and more concretely demonstrate what models simulate well, and where difficulties remain, quantify changes in model performance since the AR4, and illustrate the relative performance of different models.

While the development of increasingly realistic models remains a high priority in climate research, the need to better characterize the uncertainty in current models is rapidly becoming a comparable challenge. Efforts to formally describe model uncertainties are discussed in Section 9.7.3. There are a multitude of factors to consider, many of which are beyond the scope of this chapter (e.g., un-quantified uncertainties in external forcing or observations). The assessment in this chapter focuses on a multi-model perspective (e.g., CMIP3, CMIP5) and the inter-model spread between the individual models. The differences between the individual

models provide a lower bound estimate of model uncertainty. The error structure of model behaviour is however extremely complex (e.g., Santer et al., 2009), and it must be emphasized that the relative performance of the individual models can vary widely from one diagnostic/metric to another. The prospects for synthesizing this information to gauge the reliability of projections are addressed in Section 9.8.

9.3 Experimental Strategies in Support of Climate Model Evaluation

9.3.1 The Role of Model Intercomparisons

Gauging the extent to which climate models realistically simulate the Earth's climate and capture fundamental processes requires extensive comparisons with observations on a range of space and time scales. Organized model intercomparison projects (MIPs) serve a variety of purposes for the climate research community and typically include standard or "benchmark" experiments that represent critical tests of a model's ability to simulate the observed climate. When modelling centres perform a common experiment, it offers the possibility to compare their results not just with observations, but with other models as well. This "intercomparison" enables researchers to explore the various strengths and weakness of different models in a controlled setting. When modelling groups repeat the benchmark experiments over time, it is possible to determine how models improve as more realistic processes are incorporated.

All models suffer from errors, and model evaluation is a necessary step towards their identification. Benchmark MIP experiments offer a way to distinguish between the errors particular to an individual model from those which might be more universal. The resulting multi-model perspective provides the context for much of what follows.

9.3.2 Experimental Strategy for CMIP5

9.3.2.1 Structure of the Historical Experiments

In contrast to the CMIP3 ensemble of centennial-length simulations using AOGCMs assessed in AR4 (Meehl et al., 2007a), the CMIP5 collection also includes initialized decadal-length projections and long-term experiments using ESMs (Taylor, 2011) (Figure 9.2). The observable properties of the basic mean states from these experiments are evaluated against the historical data record in the next Section. This assessment addresses two principal requirements that climate models must satisfy in order to provide useful projections of climate change. First, since the effective climate sensitivity depends on the state of the climate system, it is necessary for climate models to reproduce the observed state as accurately as possible to minimize the effects of state-related errors on projections of future climate (Senior and Mitchell, 2000). Second, many relationships among climatic forcing, feedback, and response manifested in projections of future climate change can be tested using the observational record (Soden and Held, 2006). However, agreement with the observational record is a necessary but not sufficient condition to narrow the range of uncertainty in projections due to remaining uncertainties in historical forcing, recent trends in oceanic heat storage, and the coupled processes of the climate system.

Simulations of the atmosphere, ocean, sea-ice, and land surface are common to all three classes of experiments, and the basic states from these simulations are evaluated against the recent and historical record in Sections 9.4.1 through 9.4.4, respectively. The integration of chemical and biogeochemical cycles with the physical climate system is a general property of the ESMs included in the CMIP5 multi-model ensemble. The formulations and observational evaluations of the carbon and sulfur cycles in the ESMs are presented in Sections 9.4.5 and 9.4.6, respectively.

[INSERT FIGURE 9.2 HERE]

Figure 9.2: Left: Schematic summary of CMIP5 short-term experiments with tier 1 experiments (yellow background) organized around a central core (pink background). From Taylor et al. (2011), their Figure 2. Right: Schematic summary of CMIP5 long-term experiments with tier 1 and tier 2 experiments organized around a central core. Green font indicates simulations to be performed only by models with carbon cycle representations, and "E-driven" means "emission-driven". Experiments in the upper hemisphere either are suitable for comparison with observations or provide projections, whereas those in the lower hemisphere are either idealized or diagnostic in nature, and aim to provide better understanding of the climate system and model behaviour. From Taylor et al. (2011), their Figure 3.

9.3.2.2 Forcing of the Historical Experiments

Under the protocols adopted for CMIP5 and previous assessments, the transient climate experiments are conducted in three phases. The first phase covers the start of the modern industrial period through to the present-day corresponding to years 1850 to 2005 (van Vuuren et al.). The second phase covers the future, 2006 to 2100, and is described by a collection of Representative Concentration Pathways (Moss et al., 2010). The third phase is described by a corresponding collection of Extension Concentration Pathways (ECPs). The forcings for the first phase are relevant to the historical simulations evaluated in this Section and are described briefly here (with more details in Annex II).

In the CMIP3 experiments summarized in the AR4, the forcings used in each model contributed to the multi-model ensemble of 20th century experiments (known collectively as 20C3M) were left to the discretion of the individual modelling groups. By contrast, a comprehensive set of historical anthropogenic emissions and land-use and land-cover change (LULCC) data have been assembled for the AR5 experiments in order to produce a relatively homogeneous ensemble of historical simulations with common time-series of forcing agents.

For AOGCMs without chemical and biogeochemical cycles, the forcing agents are prescribed as a set of concentrations. The concentrations for GHGs and related compounds include CO_2, CH_4, N_2O, all fluorinated gases controlled under the Kyoto Protocol (HFCs, PFCs, and SF_6), and ozone depleting substances controlled under the Montreal Protocol (CFCs, HCFCs, Halons, CCl_4, CH_3Br, CH_3Cl). The concentrations for aerosol species include sulphate (SO_4), ammonium nitrate (NH_4NO_3), hydrophobic and hydrophilic black carbon, hydrophobic and hydrophilic organic carbon, secondary organic aerosols (SOA), and four size categories of dust and sea salt. For ESMs that include chemical and biogeochemical cycles, the forcing agents are prescribed as a set of emissions. The emissions include time-dependent spatially-resolved fluxes of CH_4, NO_X, CO, NH_3, black and organic carbon, and volatile organic carbon (VOCs). For models that treat the chemical processes associated with biomass burning, emissions of additional species such as C_2H_4O (acetaldehyde), C_2H_5OH (ethanol), C_2H_6S (dimethyl sulphide), and C_3H_6O (acetone) are also prescribed. Historical LULCC is described in terms of the time-evolving partitioning of land-surface area among cropland, pasture, primary land and secondary (recovering) land, including the effects of wood harvest and shifting cultivation, as well as land-use changes and transitions from/to urban land (Hurtt et al., 2009). These emissions data are aggregated from empirical reconstructions of grassland and forest fires (Mieville et al., 2010; Schultz et al., 2008); international shipping (Eyring et al., 2010); aviation (Lee et al., 2009); and sulphur (Smith et al., 2011b), black and organic carbon (Bond et al., 2007), and NO_X, CO, CH_4 and NMVOCs (Lamarque et al., 2010) contributed by all other sectors.

9.3.2.3 Relationship of Observational Initialization and Decadal Predictive Uncertainty

The CMIP5 archive includes a new class of decadal-prediction experiments (Meehl et al., 2009b) (Figure 9.2). The goal is to understand the relative roles of forced changes and internal variability in historical and near-term climate variables, and to assess the predictability that might be realized on decadal time scales. These experiments are comprised of two sets of hindcast and prediction ensembles with initial conditions spanning 1960 through 2005. The set of 10-year ensembles are initialized starting at 1960 in 5-year increments through the year 2005 while the 30-year ensembles are initialized at 1960, 1980, and 2005. Results from these experiments will be described in detail in Chapter 11; here we focus on evaluation of the models used in such predictions.

9.4 Simulation of Recent and Longer-Term Records in Global Models

9.4.1 Atmosphere

Many aspects of the atmosphere have been more extensively evaluated than other component models of the Earth system. One reason is the availability of observationally-based data with which to confront models. Near global data sets exist for energy fluxes at the top of the atmosphere, cloud cover, temperature, winds, moisture, total column ozone, and other important properties simulated by atmospheric models. Furthermore, promising new data sets are inspiring the diagnosis of models in innovative ways, for example using satellite

simulators for the evaluation of clouds (Section 9.2.2.3). In this section we evaluate the large-scale atmospheric behaviour with available atmospheric observational data.

9.4.1.1 Spatial Patterns of the Mean State

As discussed in Section 9.1, all component models of the Earth system are built upon fundamental principles such as the conservation of energy, momentum, and mass. For the atmospheric component, realistic simulation of the energy and water cycles is particularly important, and of obvious relevance to society. Surface temperature is perhaps the most routinely examined quantity in atmospheric models, and not simply because the surface is inhabited. The surface is influenced by many factors that must be adequately represented in order for a model to realistically capture the observed temperature distribution. The dominating external influence is incoming sunlight, but many aspects of the simulated climate play an important role in modulating regional temperature such as the presence of clouds and the complex interactions between the atmosphere and the underlying land, ocean, snow, ice, and biosphere.

The annual mean distribution of surface air temperature (at 2 meters) is shown in Figure 9.3 for the multi-model average of CMIP5 models presently available. The multi-model ensemble exhibits distinctive gradients of observed temperatures (not shown) that broadly decrease with latitude. The maximum annual mean temperatures of the western tropical Pacific and tropical Indian Ocean are also well represented by the models. A comparison of the multi-model average (middle panel) with observations (Jones et al., 1999) shows that in most areas the models agree with the observations to within 2°C, but there are several locations were the biases are much larger, particularly at elevations over the Himalayas and parts of both Greenland and Antarctica. The lower panel of Figure 9.3 shows the absolute bias of the individual CMIP5 models, providing similar information as the multi-model result but with the removal of the possibility of compensating errors across models.

[INSERT FIGURE 9.3 HERE]

Figure 9.3: Annual mean surface (2 meter) air temperature (K) for the period (1985–2005). Top panel: Multi-model ensemble (MME) constructed with 11 available AOGCMs used in the CMIP5 historical experiment. Middle panel shows the MME bias compared to observations (Jones, 1999). Bottom panel shows the average of the individual model absolute biases.

A first look of the seasonal performance of models can be obtained by examining the difference between extreme seasons (DJF and JJA). The top panel of Figure 9.3 shows the absolute difference between the DJF and JJA surface air temperature. This clearly demonstrates the much larger seasonal cycle over land, and particularly at higher latitudes. The middle panel shows the multi-model differences with observations (Jones et al., 1999). Over the oceans the models appear to slightly underestimate this measure of seasonal amplitude, particularly along the western boundary currents. Over land, the models tend to overestimate the temperature range with magnitudes noticeably larger than over the oceans. Observational uncertainty is important to consider in the higher latitude ocean regions.

[INSERT FIGURE 9.4 HERE]

Figure 9.4: Annual surface (2 meter) air temperature (K) range (DJF-JJA) for the period (1985–2005). Top panel shows updated observations from Jones (1999). Bottom panel shows bias in the multi-model ensemble (MME) constructed with 11 available AOGCMs used in the CMIP5 historical experiment.

Simulation of precipitation is a much tougher test for models as it depends heavily on processes that are not explicitly resolved, and must be parameterised (see Section 9.1). Figure 9.5 compares observationally based estimates of precipitation with the CMIP5 multi-model ensemble. Broad scale features in the observations (not shown) are evident in the MME, such as a maximum precipitation denoting the ITCZ just north of the equator in the central and eastern tropical Pacific, dry areas over the eastern subtropical ocean basins, and the minimum rainfall in Northern Africa (Dai, 2006). The bottom panel suggests that problem areas appear to persist in CMIP5, including the well-known biases in the structure of the tropical convergence zones in the western Pacific (see Section 9.4.2.3), which have been extensively analyzed in the CMIP3 simulations (e.g., Lin, 2007).

[INSERT FIGURE 9.5 HERE]

Figure 9.5: Annual mean precipitation for the period (1985–2005). Top panel: Multi-model ensemble (MME) constructed with 11 available AOGCMs used in the CMIP5 historical experiment. Middle panel shows the MME bias with updated observations (Adler et al., 2003). Bottom panel shows the average of the individual model absolute biases.

9.4.1.1.2 Atmospheric moisture, clouds, and radiation

The global annual-mean precipitable water is the measure of the total moisture content of the Earth's atmosphere. For AOGCMs in the CMIP3 ensemble, the values of precipitable water agreed with one another and with multiple estimates from the NCEP/NCAR and ERA meteorological reanalyses to within approximately 10% (Waliser et al., 2007). Modeling the vertical partitioning of water vapor is subject to greater uncertainty since the humidity profile is governed by a variety of hydrological processes, sub-grid vertical transport, and coupling between the boundary layer and free troposphere. In general, the models exhibit a significant dry bias of up to 25% in the boundary layer and a significant moist bias in the free troposphere of up to 100% (John and Soden, 2007). Upper tropospheric water vapor varies by a factor of three across the CMIP3 multi-model ensemble (Su et al., 2006). However, the models reproduce the gradients in free-tropospheric humidity between ascending and descending dynamical regimes and between convective-cloud-covered and cloud-free regions of the tropics to within 10% (Brogniez and Pierrehumbert, 2007). In addition, the relationship between tropospheric moisture and externally forced warming in the 20th century is consistent across the ensemble and is uncorrelated with the biases in the individual models (John and Soden, 2007).

The spatial patterns and annual cycle of the radiative flues at the top of the atmosphere represent some of the most important observable properties of the Earth system, and current models reproduce these patterns with considerable fidelity relative to the NASA CERES data sets (Pincus et al., 2008). This level of agreement is as expected since the spatial patterns and annual cycle of the radiative fluxes are governed primarily by the meridional gradient and seasonal cycle in solar insolation, both of which are reasonably reproduced by all the models in the CMIP3 ensemble. The models exhibit much less skill in reproducing either the spatial correlations or spatial variance in shortwave and longwave cloud radiative effects, although the skill of the individual climate models is comparable to that exhibited by the ECMWF reanalysis (Pincus et al., 2008).

Comparisons against surface fluxes show that, on average, the CMIP3 models overestimate the downward all-sky shortwave flux at the surface by 6 W m^{-2} and underestimate the corresponding downward longwave flux by –5.6 W m^{-2} (Wild, 2008). The resulting average error in the total downwelling radiant flux is 0.4 W m^{-2}. The correlation between the biases in the all-sky and clear-sky downwelling fluxes suggests that systematic errors in clear-sky radiative transfer calculations may be primary cause for these biases. This is consistent with an analysis of the global annual-mean estimates of clear-sky atmospheric absorption from the CMIP3 ensemble. Wild (2006) demonstrates that several CMIP3 models underestimate clear-sky shortwave absorption and hence overestimate surface insolation by up to 12 W m^{-2}. The underestimation of absorption can be attributed to the omission or underestimation of absorbing aerosols, in particular carbonaceous species, and to omission of weak-line absorption by water vapour, the predominant absorbing gas for shortwave radiation in the current climate (Wild et al., 2006). The net shortwave energy absorbed by the surface is set by the downwelling flux and the surface albedo. The mean surface albedo of 0.351 from the CMIP3 ensemble and the observationally derived albedo of 0.334 from the International Satellite Cloud Climatology Project (ISCCP) differ by much less than the standard deviation in surface albedo among the models (Wang et al., 2006).

Despite progress in the representation of cloud processes in GCMs (see Section 9.1.3.2.1 and Chapter 7), the simulation of clouds and their effect on radiative budgets remains a major challenge. Pincus et al. (2008) assessed the simulation of clouds and their radiative effects in the CMIP3 model ensemble and showed that significant errors remain. Figure 9.6 shows maps of deviations from observations in annual mean shortwave (top left), longwave (middle left) and net (bottom left) cloud radiative effect (CRE) for the CMIP3 multi-model mean. The Figure also shows zonal averages of the same quantities from observations (thick solid lines) and individual models (grey lines). The definition of CRE and observed mean fields for these quantities can be found in Chapter 7.

[INSERT FIGURE 9.6 HERE]

Figure 9.6: Annual mean errors in shortwave (top left), longwave (middle left) and net (bottom left) cloud radiative effect of the CMIP3 multi-model mean. Also shown are zonal averages of the absolute values of the same quantities

from observations (CERES ES-4 and ERBE S-4G, thick black lines) and individual models (thin grey lines). For a definition of cloud radiative effect and maps of its absolute values, see Chapter 7.

Models show large regional biases in CRE in the shortwave component, and these are particularly pronounced in the subtropics with an underestimation of the shortwave CRE in the stratocumulus regions and an overestimation in the trade cumulus regions. A large underestimation of the shortwave CRE is also evident over the sub-polar oceans of both hemispheres and the Nothern Hemisphere land areas. It is evident in the zonal mean graphs that the errors in shortwave CRE over the sub-polar and subtropical regions are common to all models, while errors in longwave CRE are smaller overall and vary in sign from model to model.

Subtropical clouds have been shown to be of great importance to a model's climate sensitivity (Bony and Dufresne, 2005; Dufresne and Bony, 2008; Williams and Webb, 2009). More in-depth analysis of several global and regional models (Karlsson et al., 2008; Teixeira et al., 2011) show that the intricacies of the interaction of boundary layer and cloud processes with the larger scale circulation systems that ultimately drive the observed subtropical cloud distribution remain poorly simulated. Several studies have also highlighted the potential importance and poor simulation of sub-polar clouds in the Arctic and Southern Oceans (Haynes et al., 2011; Karlsson and Svensson, 2010; Trenberth and Fasullo, 2010; Tsushima et al., 2006). Karlsson and Svensson (2010) showed that the CMIP3 models have great difficulties in simulating Arctic cloud properties. A particular challenge for models is the simulation of the correct phase of the cloud condensate and very few observations are available to evaluate models particularly with respect to their representation of cloud ice (Waliser et al., 2009). Process-oriented approaches to the evaluation of model clouds (Teixeira et al., 2011; Williams and Webb, 2009; Williams and Tselioudis, 2007; Williams and Brooks, 2008) are beginning to provide deeper insight into model errors and strategies for model improvement and are likely to lead to improved cloud representations in the medium-term.

In summary, there remain significant errors in the model simulation of clouds. It is very likely that these errors contribute significantly to the uncertainties in estimates of cloud feedbacks (see Section 9.7.4 and Chapter 7) and consequently in the climate change projections reported in Chapter 12.

9.4.1.1.3 Ozone

Ozone has been subject to a major perturbation in the stratosphere since the late 1970s due to anthropogenic emissions of ozone-depleting substances (ODSs), now successfully controlled under the Montreal Protocol and its Amendments and Adjustments. Trends in stratospheric ozone have important implications on surface climate (see further discussion in Section 9.4.1.3 and in Chapter 10) and so it is important to capture these in climate simulations. Figure 9.7 shows an evaluation of the mean total column ozone (1980–1999) from the CCMVal-2 models and the AC&C / SPARC ozone database to the NIWA database, and the multi-model mean agrees well with observations (panels c, d). As noted in Cionni et al. (2011), total column ozone over Antarctica in the AC&C / SPARC database is higher than in the NIWA database (panels e, f), because in this region the dataset is based only on the ozonesondes from the Syowa station located at 69°S. This station is not in the centre of the vortex but is close to the vortex edge and therefore the ozone measured there is occasionally indicative of midlatitude rather than polar air (Solomon et al., 2005).

[INSERT FIGURE 9.7 HERE]

Figure 9.7: September to November total column ozone climatology (1980–1999) from the CCMVal-2 multi-model mean (a) and the bias of it from the NIWA database (b). (c,d) same as (a,b), but for the AC&C / SPARC ozone database that was used as forcing in a subset of the CMIP5 model simulations. Ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere. Modified from Cionni et al. (2011).

Tropospheric ozone in the historical period has increased due to increases in ozone precursor emissions from anthropogenic activities. Since the AR4, a new emission dataset has been developed (Lamarque et al., 2010), which has led to some differences in tropospheric ozone burden compared to previous studies, mainly due to biomass burning emissions (Cionni et al., 2011; Lamarque et al., 2011; Lamarque et al., 2010), see Chapter 8. In general, tropospheric ozone is well simulated by CCMs with interactive tropospheric chemistry. For example, the historical tropospheric segment of the AC&C / SPARC ozone database used as forcing in a subset of CMIP5 models without interactive chemistry consists of a two model mean of the Community Atmosphere Model (Lamarque et al., 2010) and the NASA-GISS PUCCINI model (Shindell et al., 2006).

The geographical distribution and the annual cycle of this two-model mean compares well with a satellite climatology (2005–2009), although tropospheric column ozone is slightly lower than observed especially in the Southern Hemisphere. The vertical profiles of tropospheric ozone are also broadly consistent with ozonesondes and in-situ measurements, with some deviations in regions of biomass burning (Cionni et al., 2011).

9.4.1.1.4 Aerosols

In the RCPs adopted for CMIP5 (Moss et al., 2010), the geographic distribution and temporal evolution of emissions of sulphate precursors, in particular SO_2, are prescribed both for the simulations of the 20th century (Lamarque et al., 2010) and for the projections of future climate (e.g., Wise et al., 2009). Therefore differences among the multi-model ensemble derived from those models that calculate aerosols interactively should be due to differences in the modelled chemical production, transport, and removal of the sulphate species together with differences in the treatments of aerosol microphysical properties. Analogous experiments have been conducted using an ensemble chemical transport models with identical emissions of sulphate precursors as part of the AeroCom project (Textor et al., 2007) and compared against a corresponding ensemble with heterogeneous emissions. Intercomparison of the two ensembles shows that the intermodal differences in the heterogeneous ensemble are due primarily to the differences in model processes and transport rather than differences in emissions (Textor et al., 2007). Similar findings have been obtained from simulations of present-day conditions using a single chemical transport model and emissions data set run with three different operational meteorological analyses. In these simulations, the process parameterisations are identical although the meteorological fields driving the processes are not. Sulphate concentrations in the middle and upper troposphere and near the surface in the Northern Hemisphere close to anthropogenic source regions differ by a factor of three among the three simulations (Liu et al., 2007).

9.4.1.2 Quantifying Model Performance with Metrics

Performance metrics can be constructed to quantify what models simulate well and contrarily to demonstrate model performance deficiencies. As a simple example, Figure 9.8 illustrates how the pattern correlation between the observed and simulated climatological annual mean depends very much on the quantity examined. All CMIP3 models capture the mean surface temperature distribution quite well, with correlations of 0.95 and higher. Correlations for 200 hPa zonal winds and OLR are somewhat lower, although still mostly above 0.90. For precipitation however, the typical correlation between models and observations is below 0.8, and with considerable scatter. This example illustrates how fields associated with the large-scale atmospheric circulation (e.g., temperature and winds) agree more closely with observations than fields directly related to parameterisations (e.g., precipitation and clouds and their radiative effects).

[INSERT FIGURE 9.8 HERE]

Figure 9.8: Global annual mean climatology (1980–1999) pattern correlations between CMIP3 simulations and corresponding observations (see Table [9.x] for the default references for each field). Results for sea surface temperature and SW cloud radiative effects exclude data pole-ward of 50 degrees in both hemispheres. Individual model results are identified as dash marks. The green bars represent the average result for each variable.

Several studies have used performance metrics to examine the relative strengths and weaknesses of different CMIP3 models by comparing the mean state of multiple fields with available observations (e.g., Gleckler et al., 2008; Reichler and Kim, 2008; Pincus et al., 2008; Yokoi et al., 2011). Figure 9.9 (taken from Gleckler et al., 2008), depicts the space-time RMSE for the 1980–1999 climatological annual cycle of the historically forced CMIP3 simulations. For each of the fields examined, this “portrait plot” depicts model performance relative to the median of all model errors, with blue (red) shading indicating a model’s performance being better (worse) than the median result. In each case, two observational estimates are used to demonstrate the impact of the selection of reference data on the results. The results in this figure mixed. Some models consistently compare better with observations than others, some exhibit mixed performance, and some stand out as relatively poor performers. For most fields, the choice of the observational dataset does not substantially change the results, indicating that inter-model differences are substantially larger than the differences between the two reference datasets (which often rely on the same source of measurements). Nevertheless, it is important to recognize that these results can be sensitive to a variety of factors such as observational uncertainty, sampling errors (e.g., limited record length of observations), the spatial scale of comparison, the domain considered, and the choice of metric.

Another notable feature of Figure 9.9 is that in most cases the multi-model mean (and median) agree more favourably with the observations than any individual model. This apparent superiority of the multi-model mean has been previously recognized to hold for surface temperature and precipitation (Lambert and Boer, 2001) , but it is now evident that is applies for a broad range of climatological fields. Recent work has led to several possible explanations why the multi-model mean compares so well with observations. Pincus et al. (2008) argue that averaging across models smoothes results on shorter spatial scales where the models are known to be less reliable, but they conclude that a cancelation of compensating systematic errors is likely to be more important. An alternate hypothesis is that the behaviour of CMIP3 ensemble can be explained in the context of a statistical indistinguishable ensemble (Annan and Hargreaves, 2011; Annan and Hargreaves, 2010).

[INSERT FIGURE 9.9 HERE]

Figure 9.9: Relative error measures for 20th century CMIP3 models, based on the global annual cycle climatology (1980–1999) in the historical (20c3m) experiments. Treating each variable independently, the space-time RMSE is normalized by the median result across all models. A value of 0.3 indicates an error 30% larger than the median error, whereas –0.3 is 30% smaller than the median error. A diagonal splits each grid square showing the relative error with respect to both the primary (upper left triangle) and the alternate (lower right triangle) reference data sets. Taken from Gleckler et al. (2008). The two left hand columns depict the relative error for both the multi-model mean and median (which is distinct from the normalization by the median of individual models).

Correlations between the results for different fields in Figure 9.9 are known to exist, reflecting physical and relationships in the model formulations. (Gleckler et al., 2008; Yokoi et al., 2011) have applied cluster analysis methods in an attempt to reduce this redundancy. Starting from 43 multivariate RMSE and bias metrics, (Yokoi et al., 2011) identify 7 independent clusters. Approaches such as this may to lead to improved summaries of model performance.

Using a single measure of a models' ability to simulate the climatological annual cycle, (Reichler and Kim, 2008) quantified how errors were reduced in CMIP3 when compared to earlier generations of models. Use of a single skill score to gauge model performance can also be useful during the model development process in conjunction with the expert judgement of model developers. Construction of such an overall index is arbitrary however, and it is unclear to what extent it should be used to make any quantitative judgements about the relative performance of different models.

Performance metrics such as those discussed above are a typical first-step toward quantifying model agreement with observations and succinctly summarizing selected aspects of model performance. Confidence in metrics-based model evaluation is greatest when the metrics are relatively simple, statistically robust, and the results are not strongly dependent upon various analysis choices (Knutti et al., 2010b). Metrics are well suited for identifying outliers in various aspects of model performance, which once identified can be confirmed and further investigated with more in-depth analysis.

9.4.1.3 Long-Term Global-Scale Changes

As a precursor to our evaluation of the simulated variability in climate models (see Section 9.5), we first examine how well these models capture long-term changes evident in the observational record. The comparison of observed and simulated change is complicated by the fact that the simulation results depend on both model formulation and the time-varying external forcings imposed on the models (Allen et al., 2000; Santer et al., 2007). De-convolving the importance of model and forcing differences (e.g., indirect aerosol effects) in the historical simulations is an important topic that is addressed in Chapter 10.

Figure 9.10 compares the observational record of 20th century changes in global surface temperature to that simulated by each CMIP5 model. The frequency and magnitude of the interannual variability in most of these simulations is generally similar to that of the observations although there are several exceptions. The gradual warming evident in the observational record, particularly in the more recent decades, is also evident in the simulations although again there are some important differences among models. The interannual variations in the observations are noticeably larger than the multi-model ensemble because the averaging of individual model results acts to filter much of the variability simulated by the models. On the other hand, the episodic volcanic forcing that is applied to many of the models is evident in the multi-model agreement with

the observed cooling particularly after the 1991 eruption of Pinatubo. Because the interpretation of differences in model behaviour can be confounded by natural variability and forcing, some studies have attempted to identify and remove dominant factors such as ENSO and the impacts of volcanic eruptions (Fyfe et al., 2010). Efforts such as these can reduce trend uncertainties and thereby improve our ability to evaluate simulated changes with observations.

In summary, models broadly capture the observed historical changes in global surface temperature, and in particular the warming of recent decades. Both model formulation and the applied external forcings (see Chapter 10) influence this level of agreement.

[INSERT FIGURE 9.10 HERE]

Figure 9.10: Observed and simulated annual mean global average anomaly time series of surface air temperature. Lines (thin) show results from single simulations currently available for CMIP5. Thick black and red lines represent the observations and the multi-model mean respectively. Vertical grey bars represent times of major volcanic eruptions. Observational data are the HadCRUTT3v merged surface temperature, 2 meter of land and surface over the ocean. The current plot shows 2 meter temperature over the land and ocean for model simulations (to be updated with merged surface temperature). All anomalies are with respect to a 1961–1990 climatology.

Simulated changes in near surface specific humidity over land have been examined and found to be broadly consistent with observational estimates for the period 1973–1999 (Willett et al., 2010). In the Northern Hemisphere, the extratropical trend in most models is slightly smaller than the observed positive trend, whereas in the tropics the picture is less clear because of substantial inter-model differences, which can at least in part be attributed to large interannual variability. In the extratropics of the Southern Hemisphere there is no significant trend in the observations whereas most of the models have Southern Hemisphere trends similar to their northern counterparts. Given the sparse data network in the Southern Hemisphere this discrepancy may result from a combination of model errors and observational sampling uncertainty.

Several studies have focused on the ability of models to simulate observed trends in the free troposphere, in particular those in tropical latitudes. While some studies identify detectable discrepancies, with the models seemingly overestimating upper-tropospheric temperature trends in the tropics (Douglass et al., 2008; McKitrick et al., 2010; Christy et al., 2010; Bengtsson and Hodges, 2011; Fu et al., 2011), others found no statistically significant difference between the models and the observations once all uncertainties are accounted for (Santer et al., 2008; Thorne et al., 2011). It is evident that large uncertainties continue to make it difficult to evaluate the models' ability in simulating upper tropospheric tropical trends. First, there remain significant observational uncertainties (Chapter 2, Thorne et al., 2011; Mears et al., 2011). Second, the choice of metric and statistical method have been shown to crucially affect the conclusions, as demonstrated by (Santer et al., 2008) who found severe flaws in the statistics used by (Douglass et al., 2008). It has also been shown that the identification of trends in short records is severely affected by end-point issues (McKitrick et al., 2010; Santer et al., 2011; Thorne et al., 2011). For instance McKitrick et al. (2010) found a strong dependence of their conclusions on record length, and Thorne et al. (2011) found much better agreement between models and radiosonde observations when using the full radiosonde record instead of the shorter record that overlaps with satellite observations. In addition there are uncertainties in how the models are forced, in particular in the recent studies that used scenario simulations to represent the last decade from model simulations (Fu et al., 2011; McKitrick et al., 2010).

The coupling of the lower and upper troposphere in the tropics is achieved through convective and cloud processes, which are amongst the least certain in models. However, the amplification of surface temperature trends in the tropical upper troposphere has been shown to agree well between models and observations on seasonal and interannual time scales (Santer et al., 2005) and for the full radiosonde record (Thorne et al., 2011), lending some confidence in model behavior. While there are discrepancies between modeled and observed temperature trends in the upper tropical troposphere, observational uncertainty and contradictory analyses prevent a conclusive assessment of model quality.

Change in the global mean temperature of the lower stratosphere during the satellite era is characterized by substantial cooling occurring in two step-like transitions following two major volcanic eruptions (Section,

2.2.4, Figure 2.16). The chemistry climate models in CCMVal-2 are capable of simulating the observed evolution of stratospheric temperature since 1960 when forced with observed sea surface temperatures, observed emissions of long-lived greenhouse gases, ozone depleting substances and anthropogenic aerosol loadings, along with natural forcings like solar irradiance variations and volcanic aerosol (WMO, 2011a) . The set of CMIP3 climate models generally underestimate the observed lower stratosphere cooling trend even when both anthropogenic and natural forcings are included (Cordero and Forster, 2006). Further, about half of the CMIP3 models that include volcanic aerosol forcings significantly overestimate the warming in the upper stratosphere following tropical volcanic eruptions. While several studies have illustrated the importance of including both anthropogenic and natural forcings to reproduce the step-like cooling of the lower stratospheric temperature (Dall'Amico et al., 2010; Eyring et al., 2006; Ramaswamy et al., 2006), other studies suggest that an inadequate simulation of the Brewer Dobson Circulation and its changes (Lin et al., 2010), and lack of simulating the Quasi-Biennial Circulation in the tropical stratosphere (Dall'Amico et al., 2010) degrade the agreement between observed and simulated lower stratosphere trends especially on regional scales.

One important form of model evaluation involves the examination of co-variability between various quantities, and the resulting indication of correct simulation of various physical relationships. Several recent studies have examined the consistency of changes between lower tropospheric temperature and available precipitable water (Mears et al., 2007a), and surface temperature and relative humidity (Willett et al., 2010). Figure 9.11 (Mears et al., 2007b) shows the relationship between 12-year (1988–1999) linear trends in tropical precipitable water and lower tropospheric temperature for individual historical simulations and observations. The large distribution of trends in both quantities is expected from the short record length, with the impact of natural variability evident in the range of results from different realizations of the same model. It is difficult to judge which models more closely agree with the observations because the observational trends are also sensitive natural variability. However, as described by Mears et al. (2007b) it is clear that the scaling ratio between changes in these two quantities as simulated by the models is consistent with that of the observations. Interestingly, this study also found a similar scaling ratio for interannual time scales in both models and observations, suggesting that this relationship hold across a range of time scales.

[INSERT FIGURE 9.11 HERE]

Figure 9.11: Scatter plot of the variability of W as a function of the trend in W as a function of the TLT trend for the tropical oceans. Trends are calculated over the periods given in Table 1. In Figure 3a, UAH V5.2 and UAH V5.1 yield nearly identical results, so the UAH V5.2 data point is hidden. The lines shown bisect the two different linear fits obtained with first W, then TLT assumed to be the dependent variable. Isobe et al. (1990) show that this is a good method for finding an estimate of an underlying relationship in the presence of unknown measurement errors and or scatter that is not strictly related to measurement error, as is the case here. The climate model and reanalysis results are for the 1981–1999 period, while the satellite results are for the 1988–2006 period, so the trend results from the satellite data and the models and reanalysis cannot be directly compared. Also, for UAH V5.1, the calculations are performed over the 1988–2005 period when both SSM/I and UAH 5.1 data are available.

Since the AR4, there is increasing observational and modelling evidence that trends in Antarctic stratospheric ozone loss have contributed to changes in southern high-latitude climate (WMO, 2011b). Together with increasing GHG concentrations, the ozone hole has led to a poleward shift and strengthening of the Southern Hemisphere westerly tropospheric jet during summer, which has contributed to robust summertime trends in surface winds, observed warming over the Antarctic Peninsula, and cooling over the high plateau (Arblaster and Meehl, 2006; CCMVal, 2010; Perlwitz et al., 2008; Son et al., 2008; Son et al., 2010). These trends are well captured in chemistry-climate models (CCMs) with interactive stratospheric chemistry that have been extensively evaluated using a process-oriented approach (SPARC-CCMVal, 2010). They are also captured in CMIP3 models that prescribe time-varying ozone (Son et al., 2010); however, in a subset of these models prescribed ozone as a climatological zonal mean rather than a time-varying field. Several studies showed that models with prescribed climatological mean ozone were not able to simulate trends in surface climate correctly as a result of the missing ozone depletion (Fogt et al., 2009; Karpechko et al., 2008; Son et al., 2008; Son et al., 2010). For CMIP5, a continuous tropospheric and stratospheric vertically resolved ozone time series, from 1850 to 2099, has been generated to be used as forcing in global climate models that do not include interactive chemistry (Cionni et al., 2011). The CMIP5 simulations forced with this dataset capture the observed trend from 1960–2000 in total column ozone, see Figure 9.12. The resulting stratospherically induced changes on high latitude surface climate over the past decades including trends in the location of the 850 hPa jet and the Southern Annular Mode (SAM) are further discussed in

Chapter 10, Section 10.3. There is robust evidence, based on several studies, indicating that models with interactive or specified, time-varying stratospheric ozone are able to reproduce observed Antarctic surface climate trends, whereas models with fixed ozone are less realistic.

[INSERT FIGURE 9.12 HERE]

Figure 9.12: Time series of total column ozone over Antarctica (averaged from 60–90°S) from 1960 to 2000 for the CCMVal-2 multi-model mean (red line) and standard deviation (blue shaded area) in comparison to the AC&C / SPARC ozone database (green line) and observations from the NIWA database (black dots). Ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere. Modified from Cionni et al. (2011).

9.4.1.5 What do We Learn from Model-Data Comparisons for the Last Glacial Maximum, the Mid-Holocene, and the Last Millennium?

The Last Glacial Maximum (LGM) and mid-Holocene are benchmark periods to test the ability of climate models to represent a climate different from the modern one. We consider here results obtained from AOGCMs or EMICs, comparing new model results with results of previous phases of the Paleoclimate Modelling Intercomparison Project (PMIP, (Joussaume and Taylor, 1995). The LGM allows testing of the modelled climate response to the presence of a large ice-sheet in the northern hemisphere and to lower concentration of radiatively active trace gases, whereas the mid-Holocene tests the response to changes in seasonality of insolation in the northern Hemisphere (see Chapter 5). There is also interest in testing the ability of climate models to reproduce observed trends over more recent periods. In particular, the transition from the Medieval warm period (MWP) and the little ice-age (LIA) discussed in Chapter 5 offers a good test for climate models in a context where the climate is controlled more by natural forcing, such as volcanic and solar variability, and less by human activity (Jungclaus et al., 2010; Pongratz et al., 2009).

Figure 9.13 shows the most recent update of the continental dataset described by Bartlein et al. (2010a) and ocean datasets from MARGO (Waelbroeck et al., 2009) for annual mean LGM surface temperature and mid-Holocene precipitation, together with the ensemble mean of PMIP2 simulations for the corresponding variables. The Bartlein et al. (2010a) data set also includes information for different bioclimatic variables, such as the temperature of the coldest month, growing degree days or a moisture index. Both datatsets provide reconstruction uncertainties, considering sampling, dating, measurement errors, as well as uncertainties in the calibration that translate the proxy record to a climate variable. These two periods offer good signal to noise ratio, which allows for both qualitative and quantitative model-data comparisons. Hargreaves et al. (2011) compared PMIP simulations with MARGO SST reconstruction to show that, when the reconstructions uncertainties are considered, the PMIP2 ensemble is statistically reliable. They confirmed the conclusions from Kageyama et al. (2006) for the North Atlantic and of Otto-Bliesner et al. (2009) for the tropics. The LGM simulations reproduce the large scale patterns, but overestimate the tropical cooling and underestimate the change in SST gradient in mid-Latitude. The ensemble mean model bias is about 1 K with most values between –1 K and 2 K. In the North Atlantic the RMSE is about 1.7–1.9 K, except for one model (1.35 K). More regional assessments, show that models tend to underestimate the magnitude of the changes, including polar amplification (Masson-Delmotte et al., 2006; Zhang et al., 2010a); or the north-south temperature gradient over Europe both at the LGM (Ramstein et al., 2007) and at the mid-Holocene (Brewer et al., 2007; Davis and Brewer, 2009). Since the AR4 several studies have analysed in more depth the change in the mid-latitude westerlies in both hemispheres. In the southern hemisphere, the simulated change in atmospheric circulation is consistent with precipitation records in Patagonia and New Zealand, even though the differences between model results are large (Rojas and Moreno, 2011; Rojas et al., 2009). Comparisons of the PMIP1 atmosphere-only simulations and the PMIP2 coupled ocean atmosphere simulations confirm that the response of the ocean is needed to reproduce the observed changes (Braconnot et al., 2007d; Hargreaves et al., 2011).

These analyses all show that the CMIP3 and PMIP2 models can reproduce large scale features of climates that are different from the modern one, suggesting that they are likely to properly project future climate change. However the magnitude of the signal is underestimated in several regions, which comes either from lower than observed sensitivity or from missing feedbacks (Section 9.7).

[INSERT FIGURE 9.13 HERE]

Figure 9.13: Surface air temperature difference between the Last Glacial Maximum (21 ka) and today (top), and precipitation difference between the mid-Holocene (6 ka) and today (bottom), as shown by palaeo-environmental data (left) and the PMIP2 ensemble model simulations (right). The top uses reconstructed and simulated mean temperature of the coldest month (K) for the LGM, the bottom uses reconstructed and simulated mean annual precipitation (mm/day) for the MH. The land reconstructions are from Bartlein et al. (2010b) and the ocean reconstruction are from Waelbroeck et al. (2009). On the right figures the red line highlights the root-mean square of the inter-model differences.

9.4.2 Ocean

Accurate simulation of the ocean in climate models is essential for the correct estimation of transient ocean heat uptake and hence the transient climate response, ocean CO_2 uptake, sea level rise, and coupled modes of variability such as ENSO. In this Section we focus on the evaluation of model performance in simulating the mean state of ocean properties, surface fluxes and their impact on the simulation of ocean heat content and sea level, and tropical features of importance for climate variability. Simulations of both the recent (20th century mean and evolution) and more distant past are evaluated against available data.

9.4.2.1 Simulation of Mean Temperature and Salinity Structure

The zonal distribution of potential temperature and salinity (Figure 9.14) offers a first evaluation of the performance of climate models in simulating the different regions of the ocean (upper ocean, thermocline, deep ocean). Over most latitudes, at depths ranging from 200 m to 3,000 m, the CMIP3 multi-model mean zonally averaged ocean temperature is too warm. The maximum warm bias (about 2°C) is located in the region of the North Atlantic Deep Water (NADW) formation. Above 200 m, however, the CMIP3 multi-model mean is too cold, with maximum cold bias (more than 1°C) near the surface at mid-latitudes of the NH. Most models generally have an error pattern similar to the multi-model mean, indicating that the thermocline in the CMIP3 multi-model mean is too diffuse.

[INSERT FIGURE 9.14 HERE]

Figure 9.14: [PLACEHOLDER FOR SECOND ORDER DRAFT: Figure from AR4; to be redone from CMIP5 when results available.] Time-mean observed potential temperature (°C), zonally averaged over all ocean basins (labelled contours) and multi-model mean error in this field, simulated minus observed (colour-filled contours). The observations are from the 2004 World Ocean Atlas compiled by Levitus et al. (2005) for the period 1957 to 1990, and the model results are for the same period in the 20th-century simulations in the CMIP3 ensemble.

The simulation of sea surface temperature (SST) was evaluated in Section 9.4.1.1. The sea surface salinity (SSS) is more challenging to observe, even though the last decade has seen important and substantial improvements in the development of global salinity observations, such as those from the ARGO network (Nowlin et al., 2001; Roemmich and Gould, 2003; Roemmich and Argo Steering, 2009), see Chapter 3. Whereas SST is strongly constrained by air-sea interactions, the sources of SSS variations (surface forcing via evaporation minus precipitation, sea-ice formation/melt and river runoff) are only loosely related to the SSS itself, allowing errors to develop unchecked in coupled models. Further, there is evidence that a bias in ocean fresh water transport seen in various climate models may make the Atlantic Meridional Overturning Circulation (AMOC) overly stable in current models (Weber et al., 2007). An analysis based on twelve AOGCMs that simulated the 20th century in CMIP3, and did not use flux adjustments, showed that the near-global (60°N–60°S) mean SSS bias across the models lies between –0.8 and +0.3 psu (Waliser et al., 2011). Regional SSS biases are as high as ±2.5 psu. The most systematic biases include a saline bias in the tropical Pacific (Delcroix et al., 2010) and Bay of Bengal and a fresh bias along much of the southern subtropical oceans, and to a lesser degree in the western tropical Indian Ocean and North Atlantic. Comparisons of modeled versus observed estimates of evaporation minus precipitation suggest that model biases in surface freshwater flux play a role in some of these regions (e.g., double ITCZ in the East Pacific, (Lin, 2007)). Ocean advective processes can also play a role (Delcroix et al., 2010) but there is no clear and systematic origin of model biases.

Detailed assessments of the performance of coupled climate models in simulating hydrographic structure and variability are still relatively sparse. Two important regions, the Labrador and Irminger Seas and the Southern Ocean, have been investigated to some extent (de Jong et al., 2009) and (Sloyan and Kamenkovich, 2007). Eight CMIP3 models produced simulations of the intermediate and deep layers in the Labrador and

Irminger Seas that were generally too warm and saline, with biases up to 0.7 psu and 2.9°C. The biases arose because the convective regime was restricted to the upper 500 m; thus, intermediate water that in reality is formed by convection is, in the models, partly replaced by warmer water from the south. In the Southern Ocean, Subantarctic Mode Water (SAMW) and Antarctic Intermediate Water (AAIW), two water masses indicating very efficient ocean ventilation, were found to be well simulated in some models but not in others (Sloyan and Kamenkovich, 2007). McClean and Carman (2011) found biases in the properties of the North Atlantic mode waters and their formation rates in the CMIP3 models. For Subpolar Mode Water (SPMW), property biases principally involved salinity errors. For Subtropical Mode Water (STMW), property biases involved both salinity and temperature errors, while positioning of heat and water fluxes relative to the Gulf Stream and northwest Sargasso Sea influenced STMW formation rate. Deficiencies in STMW formation rate and volume produced a turnover time of 1–2 years, approximately half of that observed (Figure 9.15); these variations in mode water bulk properties imply variation in ocean heat storage and advection.

[INSERT FIGURE 9.15 HERE]

Figure 9.15: Sub-Tropical Mode Water (STMW) turnover time for various models compared with Kwon and Riser (2004); time is calculated by annual maximum volume divided by annual production. Values are means; error bars give ranges of one standard deviation. Square data symbols indicate those models with a distinct (if small) secondary water mass transformation rate peak corresponding to STMW formation. Triangular data symbols indicate those models with broad, diffuse, or indiscernible STMW formation peak (from McClean and Carman (2011).

9.4.2.2 Simulation of Sea Level and Ocean Heat Content

Steric and dynamical components of sea surface height (SSH) are simulated by the current generation of climate models and can be evaluated with high quality near-global satellite altimetry measurements. Yin et al. (2010) used routine metrics to evaluate the time mean spatial distribution of SSH in the CMIP3 simulations. A Taylor-diagram (Figure 9.16) of the mean annual cycle of SSH field in 17 CMIP3 models shows that the correlations with observations (Ducet et al., 2000) are relatively low compared to well-observed atmospheric quantities, with the CMIP3 ensemble mean outperforming any individual model. Most CMIP3 models cluster around a correlation between 0.55 and 0.8, with similar variability as seen in observations. A few models perform less well; these models also have a larger mean RMSE for the time-mean dynamic topography (Yin et al., 2010). Results from five CMIP5 models are currently available, and some of the new-generation models (e.g., MRI-CGCM) show clear improvement over CMIP3 versions. Improved simulation of SSH in eddy resolving ocean models has been demonstrated when compared to coarser resolution versions (McClean et al., 2006a). Chapter 13 provides a more extensive assessment of sea level changes in the CMIP3 and CMIP5 simulations including comparisons with century-scale historical estimates.

[INSERT FIGURE 9.16 HERE]

Figure 9.16: Taylor diagram of the dynamic sea-level height seasonal cycle climatology (1987–2000). The radial coordinate shows the standard deviation of the spatial pattern, normalised by the observed standard deviation. The azimuthal variable shows the correlation of the modelled spatial pattern with the observed spatial pattern. The root-mean square error is indicated by the dashed grey circles about the observational point. Analysis is for the global ocean, 50°S–50°N. The reference dataset is AVISO, a merged satellite product (Ducet et al., 2000), which is described in Chapter 3. Figure currently shows results for the CMIP3 models and the CMIP5 data currently available.

The evaluation of simulated ocean heat content (OHC) is more straightforward than sea level because OHC depends only on ocean temperature, whereas absolute changes in sea level also depend on physics that is only now being incorporated into global models (e.g., mass loss from large ice sheets). It is worth noting however, that global scale changes in OHC and steric sea level scale near linearly. About half of the historical simulations in CMIP3 (shown in Figure 9.17) included the effects of volcanic eruptions, and these simulations have been shown to yield changes in late 20th century OHC that correspond much more closely with observations than those that do not include volcanic forcing (Domingues et al., 2008; Gleckler et al., 2006). The differences between the three observational estimates are indicative of their uncertainty. It has also been demonstrated that the variability structure of simulated OHC agrees much better with observations when the model data is sampled consistently with available measurements, i.e., when and where actual measurements exist (AchutaRao et al., 2007).

[INSERT FIGURE 9.17 HERE]

Figure 9.17: Time series of observed and simulated (CMIP3) global ocean heat content (0–700 m) anomalies during the second half of the 20th Century. The three observational estimates (thick lines) are discussed in Chapter 3. Individual (one per model) simulations are shown, with solid lines for models that included volcanic forcings and dashed lines for those that did not. When updated with CMIP5 results, this figure may evolve into multiple panels, e.g., to depict averaging across multiple realizations to better capture trends and include results from historically forced ESMs.

9.4.2.3 Simulation of Circulation Features Important for Climate Response

9.4.2.3.1 Simulation of recent ocean circulation

Atlantic Meridional Overturning Circulation

The Atlantic Meridional Overturning Circulation (AMOC) plays a key role in present-days climate. It consists of northward transport of shallow warm water overlying a southward transport of deep cold water and is responsible for a considerable part of the northward oceanic heat transport. Direct observations of the AMOC would require basin-wide full-depth coverage of the meridional velocities. As these do not exist, AMOC estimates have had to be inferred from hydrographic measurements. Such estimates have been sporadically available over the last decades (e.g., Bryden et al., 2005; Lumpkin et al., 2008), indicating at 26°N a time-mean value of about 18 Sv with an observational uncertainty of ±6 Sv. Previously, climate models showed considerable spread in the time-mean strength of the AMOC, with about half of the models matching the observed estimate (Schmittner et al., 2005a; Schneider et al., 2007). However, AMOC estimates based on synoptic measurements represent sparse sampling (once every few years or decades), and are not capable of representing AMOC variability or long-term trends. Continuous AMOC monitoring was started at 26°N a few years ago (Cunningham et al., 2007); the four-year mean has been determined as 18.7 Sv with an error of ±2.1 Sv (Kanzow et al., 2010), which has permitted a much more stringent evaluation of climate models' ability to simulate the long-term AMOC. The ability of models to simulate this important circulation feature is tied to the credibility of simulated AMOC weakening during the 21st century because, at least in one EMIC, the weakening is significantly correlated with mean AMOC strength (Levermann et al., 2007). While the observed time series is still too short to analyze for long-term trends, the observational record now permits some comparison of observed and simulated AMOC variability (see Section 9.5.3.2).

Western boundary currents

The relatively low horizontal resolution of current AOGCMs leads to western boundary currents that are too weak and diffuse, and hence to biases in heat transport, SST, SSS and subtropical mode water formation (Kwon et al., 2010). Even models at the finer end of horizontal resolution (e.g., MRI.COM at roughly 1.1° by 0.5°) display this problem (Tsujino et al., 2011). Some improvements in eddy activity in the western boundary currents are noted by Farneti et al. (2010) in the GFDL CM2.4 coupled model that has an eddy permitting resolution (about 28 km near the equator increasing to about 10 km near the poles). In addition, models tend to underestimate the magnitudes of the covariance between the SST and the heat fluxes compared to re-analyses (Yu et al., 2011). Errors in the simulated time-mean state of the ocean lead to errors in the models capturing decadal variability of modes that are primarily ocean-driven (Jamison and Kravtsov, 2010). In the Southern Hemisphere, Sen Gupta et al. (2009) found considerable variations in the ability of the CMIP3 AOGCMs to represent both the meridional changes in the transports of the Agulhas, Brazil and East Australian Currents as well as in the latitude of maximum transport.

Southern Ocean circulation

The Southern Ocean is an important driver for the meridional overturning circulation and is closely linked to the zonally continuous Antarctic Circumpolar Current (ACC). The ACC has a typical transport through the Drake Passage of about 135 Sv (e.g., Cunningham et al., 2003). The ability of CMIP3 models to adequately represent Southern Ocean circulation and water masses seems to be affected by several factors (Russell et al., 2006). The most important appear to be the strength of the westerlies at the latitude of the Drake Passage, the heat flux gradient over this region, and the salinity gradient across the ACC down through the water column. Russell et al. (2006) emphasize this last factor, modulated by the upwelling of North Atlantic Deep Water (NADW) south of the ACC, as most strongly influencing the variations between models in ACC properties. Sen Gupta et al. (2009) noted several problems in these models with representing the circulation of the Southern Ocean; in particular, relatively small deficiencies in the position of the ACC lead to more obvious biases in the SST in the models. Although the models generally capture a strong circumpolar circulation and a Weddell Gyre that is located corrected and reasonably close to the observed transport, the Ross Gyre tends to be very weak in the models and located too far south in the model ensemble mean. At lower latitudes, the

Brazil/Malvinas/Falklands Confluence region is typically modelled too far to the north and offshore leading to regional temperature and salinity biases.

Rossby radius

The simulation of the first baroclinic Rossby radius in CMIP3 models was evaluated by (Sueyoshi and Yasuda, 2009). For fifteen models out of twenty, the average radius for the mid-latitude bands and the phase speed of long baroclinic Rossby waves is underestimated. A tendency is found for these quantities to be better reproduced in higher-resolution models than in lower-resolution models

9.4.2.3.2 Simulation of glacial ocean circulation

Reconstructions of the last glacial ocean circulation from sediment cores indicate that the regions of deep water formation in the North Atlantic were shifted southward, that the boundary between North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW) was substantially shallower than today, and that NADW formation was less intense (Curry and Oppo, 2005; Duplessy et al., 1988; McManus et al., 2004). This signal, although estimated from a limited number of sites, is robust. Evaluation of PMIP2 simulations, including both AOGCMs and EMICs, shows a wide range of model responses of the AMOC to LGM forcing (Weber et al., 2007), with some models reducing the strength of the AMOC and its extension at depth and other showing no change or an increase. The meridional density defined as the zonal and depth mean density at 20°S minus that at 25°N averaged over the lower 1000m provide a good criteria to compare model results (Weber et al., 2007). Otto-Bliesner et al. (2007) compared the results of 4 PMIP2 simulations with the deep ocean data from Adkins et al. (2002) (Figure 9.18). These models reproduce relatively well the modern deep ocean temperature-salinity (T-S) structure in the Atlantic basin. Greater differences between models occur for the LGM simulations, stressing large inter-model differences in LGM ocean heat and salt transports changes. All models show increased salinity, but only two of them produce a rather homogeneous temperature structure from north to south as observed. The sea-ice cover appears as a key factor in two of the models to explain the different behaviour. Theses results suggest that these processes are not well reproduced in climate models, stressing that the PMIP2 generation of climate models likely underestimated changes in deep-ocean water masses.

[INSERT FIGURE 9.18 HERE]

Figure 9.18: Temperature and salinity for modern (open symbols) and LGM (filled symbols) as estimated from data (with error bars) at ODP sites (Adkins et al., 2002) and predicted by the PMIP2 models. Site 981 triangles) is located in the North Atlantic (Feni Drift, 55_N, 15_W, 2184 m). Site 1093 (upside down triangles) is located in the South Atlantic (Shona Rise, 50_S, 6_E, 3626 m). OnlyCCSMincluded a 1 psu adjustment of ocean salinity at initialization to account for fresh water frozen into LGMice sheets; HadCM, MIROC, and ECBilt LGM predicted salinities have been adjusted to allow comparison. Show quantitatively how deep-ocean properties can be evaluated for both modern and palaeoclimate. From Otto-Bliesner et al. (2007).

9.4.2.4 Simulation of Surface Fluxes and Meridional Transports

9.4.2.4.1 Surface wind stress and meridional heat transport

The main ocean surface currents are wind-driven, and the zonal component of wind stress is particularly important. The annually averaged zonal mean zonal wind stress is reasonably well simulated by the CMIP3 models (Figure 9.19). At most latitudes, the reanalysis estimates lie within the range of model results. At middle to low latitudes, the CMIP3 model spread is relatively small and all the model results lie fairly close to the reanalysis, although near the equator this can occur through compensated zonal errors (Figure 9.20). At middle to high latitudes, the model-simulated wind stress maximum tends to lie equatorward of the reanalysis. This error is particularly large in the SH, a region where there is more uncertainty in the reanalysis. These wind stress errors in the CMIP3 control integrations may adversely affect aspects of the simulation such as oceanic heat and carbon uptake (Swart and Fyfe, 2011).

In steady state, the surface heat flux balances the convergence of ocean heat transport, which is therefore a convenient quantity for evaluation. North of 45°N, most CMIP3 model simulations transport too much heat northward when compared to the observational estimates used here (Figure 9.21), but there is uncertainty in the observations. At 45°N, for example, the model simulations lie much closer to the estimate of 0.6×10^{15} W obtained by Ganachaud and Wunsch (2003). From 45°N to the equator, most model estimates lie near or between the observational estimates shown. In the tropics and subtropical zone of the SH, most CMIP3

models underestimate the southward heat transport away from the equator. At middle and high latitudes of the SH, the observational estimates are more uncertain, and the model-simulated heat transports tend to surround the observational estimates.

[INSERT FIGURE 9.19 HERE]

Figure 9.19: Zonal mean zonal wind stress over the oceans in CMIP3 20th century simulations. [PLACEHOLDER FOR SECOND ORDER DRAFT: Final figure will include CMIP5 results and more observational estimates.]

[INSERT FIGURE 9.20 HERE]

Figure 9.20: (a) SST and (b) zonal wind stress along equator in the Indian, Pacific, and Atlantic Oceans for the CMIP3 and CMIP5 20th-century simulations. Observations are from HadISST1.1 for SST (Rayner et al., 2003) and ERSTAO for wind stress (Menkes et al., 1998).

[INSERT FIGURE 9.21 HERE]

Figure 9.21: [PLACEHOLDER FOR SECOND ORDER DRAFT: From AR4, to be redone using CMIP5 and more observational estimates.] Annual mean, zonally averaged oceanic heat transport implied by net heat flux imbalances at the sea surface, under an assumption of negligible changes in oceanic heat content. The observationally based estimate, taken from (Trenberth & Caron, 2001) for the period February 1985 to April 1989, derives from reanalysis products from the National Centers for Environmental Prediction (NCEP)/NCAR (Kalnay et al., 1996) and European Centre for Medium Range Weather Forecasts 40-year reanalysis (ERA40); (Uppala et al., 2005). The model climatologies are derived from the years 1980 to 1999 in the 20th-century simulations in the CMIP3.

9.4.2.5 Simulation of Tropical Mean State

9.4.2.5.1 Tropical Pacific Ocean

From CMIP1 through CMIP3, models have shown persistent biases in important properties of the mean state of the tropical Pacific, as reviewed recently by AchutaRao and Sperber (2002); Guilyardi et al. (2009b); and Randall et al. (2007). Among these properties are the mean thermocline depth and slope along the equator, the structure of the equatorial current system, and the equatorial cold tongue (Brown et al., 2010a; Reichler and Kim, 2008). Many of the processes leading to these biases have, in principle, been identified, such as too strong trade winds; a too diffusive thermocline; insufficient penetration of solar radiation; and too weak tropical instability waves (Lin, 2007; Meehl et al., 2001; Wittenberg et al., 2006). However, because of the strong interactions between these processes, it is challenging precisely to attribute the source of the resulting errors, although new approaches using the initial adjustment of seasonal hindcasts suggest that the equatorial wind stress may be at the origin of several errors (Vannière et al., 2011).

A particular problem in simulating the seasonal cycle in the tropical Pacific arises from the “double Intertropical Convergence Zone (ITCZ)”, defined as the appearance of a spurious ITCZ in the Southern Hemisphere and associated with excessive tropical precipitation (Lin, 2007). Further problems are too strong a seasonal cycle in simulated SST and winds in the eastern Pacific and the appearance of a spurious semi-annual cycle. The latter has been attributed to too weak a meridional asymmetry in the background state, possibly in conjunction with incorrect water vapour feedbacks (Guilyardi, 2006; Li and Philander, 1996; Timmermann et al., 2007; Wu et al., 2008).

A further persistent problem in AOGCMs is the too low cover of marine stratocumulus cloud in the eastern tropical Pacific, caused presumably by too weak coastal upwelling off South America and leading to a warm SST bias (Lin, 2007). Although the problem persists, there have been improvements in CMIP3 models (AchutaRao and Sperber, 2006; Reichler and Kim, 2008).

The equatorial undercurrent (EUC) is a major component of the tropical Pacific Ocean circulation. Even though EUC velocity in most CMIP3 models is sluggish relative to observations, it does not appear to impair other major components of the tropical circulation including upwelling and poleward transport (Karnauskas et al., 2011). These latter transports play a critical role in a theory for how the tropical Pacific may change under increased radiative forcing, i.e., the ocean dynamical thermostat mechanism (Clement et al., 1996; Seager and Murtugudde, 1997). These findings suggest that, in the mean, global climate models may not under-represent the role of equatorial ocean circulation, nor perhaps bias the balance between competing mechanisms for how the tropical Pacific might change in the future.

9.4.2.5.2 Tropical Atlantic Ocean

The current generation of climate models is plagued by severe biases in the tropical Atlantic Ocean, so severe that some of the most fundamental features – the east-west SST gradient and the eastward shoaling thermocline along the equator – cannot be reproduced by most of coupled climate models (e.g., Chang et al., 2008; Chang et al., 2007; Richter and Xie, 2008). In many models, the warm SST bias along the Benguela coast is in excess of 5°C and the Atlantic warm pool in the western basin is grossly underestimated. As in the Pacific, CMIP3 models suffer the double ITCZ syndrome in the Atlantic, with a southern ITCZ that is not observed in nature. Hypotheses for the complex Atlantic bias problem tend to draw on the fact that the Atlantic Ocean has a far smaller basin, and thus encourages a tighter and more complex land-atmosphere-ocean interaction. A recent study using a high-resolution coupled model suggests that the warm eastern equatorial Atlantic SST bias is more sensitive to the local rather than basin-wide trade wind bias and to a wet Congo basin instead of a dry Amazon – a finding that differs from previous GCM studies (Patricola et al., 2011). Recent ocean model studies show that warm subsurface temperature bias in the eastern equatorial Atlantic is a common feature to virtually all ocean models forced with "best estimated" surface momentum and heat fluxes, owing to problems in parameterization of vertical mixing (Hazeleger and Haarsma, 2005).

9.4.2.5.3 Tropical Indian Ocean

CMIP3 models simulate equatorial Indian Ocean climate reasonably well, though most models produce weak westerly winds and a flat thermocline on the equator. The models show a large spread in the modelled depth of the 20°C isotherm in the eastern equatorial Indian Ocean (Saji et al., 2006). The reasons are unclear but may be related to differences in the various model parameterisations of vertical mixing (Schott et al., 2009).

CMIP3 models generally simulate the Seychelles Chagos thermocline ridge in the Southwest Indian Ocean, a feature important for the Indian monsoon and tropical cyclone activity in this basin (Xie et al., 2002). The models, however, have significant problems in accurately representing its seasonal cycle because of the difficulty in capturing the asymmetric nature of the monsoonal winds over the basin, resulting in too weak a semi-annual harmonic in the local Ekman pumping over the ridge region compared to observations (Yokoi et al., 2009b).

Indian Ocean SST displays a basin-wide warming following El Niño (Klein et al., 1999). This Indian Ocean basin (IOB) mode peaks in boreal spring and persists through the following summer. Only about half of CMIP3 models capture this IOB mode, and the same models tend to simulate ENSO-forced ocean Rossby waves in the tropical South Indian Ocean (Saji et al., 2006), in support of a recent observational result that the IOB is not simply a thermodynamic response to ENSO but involves ocean dynamics and active ocean-atmosphere within the Indian Ocean basin (Du et al., 2009). In-depth analysis of one model (GFDL CM2.1) confirmed that slow propagation of ocean Rossby waves south of the equator anchors ocean-atmospheric patterns that persist IOB through the following summer (Zheng et al., 2011).

9.4.2.6 Summary

It is likely (robust evidence and medium agreement) that the ocean component of CMIP3 models simulate the essential processes at play during transient ocean heat uptake, ocean CO_2 uptake, sea level rise, and coupled modes of variability. There is robust evidence and high agreement that SST is well simulated, limited evidence and medium agreement that SSS is not correctly simulated, robust evidence and high agreement that the AMOC is simulated with mixed skill, limited evidence and medium agreement that western boundary currents are simulated with mixed skill, medium evidence and medium agreement that the meridional heat transport is simulated with mixed skill, robust evidence and high agreement that the tropical Pacific mean state is simulated with mixed skill, medium evidence and high agreement that the tropical Atlantic mean state is not correctly simulated, and low evidence and medium agreement that the tropical Indian Ocean is well simulated.

9.4.3 Sea Ice

Evaluation of AOGCM sea-ice component performance requires accurate information on ice concentration, thickness, velocity, salinity, snow cover and other factors. The most reliably measured characteristic of sea ice for model evaluation remains sea ice extent. Satellite passive microwave (PMW) sensors are the main data source for estimating sea ice extent and concentration. The accuracy of PMW retrieval algorithms has

been examined in a number of studies, (e.g., Meier and Stroeve, 2008); differences in total sea ice extent from different algorithms are as large as 1 million square kilometres. Most differences between PMW products tend to offset and thus trends and anomalies are generally in better agreement than the absolute extent (Kattsov et al., 2011).

Despite the significant differences between models, the CMIP3 and CMIP5 multi-model means of sea ice extent in both hemispheres agree reasonably well with observations. The simulated mean extent (calculated from all grid cells with an ice concentration above 15%) departs from observed values by up to roughly 1 million km^2 in the cold seasons both in the Northern and Southern hemispheres, over-simulating and under-simulating it, respectively (Figure 9.22). This difference is of the same order of magnitude as differences between various observational sea ice extent estimates. In many models, however, the regional distribution of sea ice is poorly simulated, even if the hemispheric extent is approximately correct. Notably, the CMIP5 multi-model ensemble demonstrates an improvement in simulation of the annual minimum extent in the Northern hemisphere, compared to CMIP3 model mean.

[INSERT FIGURE 9.22 HERE]

Figure 9.22: Mean sea ice extent (the ocean area within 15% sea ice concentration) seasonal cycle in the Northern (upper panel) and Southern (lower panel) hemispheres as simulated by the CMIP5 (blue line) and CMIP3 (black line) ensembles. The observed sea-ice extent cycles (1980–1999) are based on the Hadley Centre Sea Ice and Sea Surface Temperature – HadISST (Rayner et al., 2003) (red line) and the National Snow and Ice Data Center – NSIDC (Fetterer et al., 2002) (brown line) data sets. The shaded areas show the inter-model standard deviation for each ensemble.

There has been no dramatic increase in sophistication of sea ice treatment in CMIP5 AOGCMs compared to CMIP3. As a result, the improvement in simulating sea ice in the former, as a group, is not striking (compare Figure 9.23 with AR4 (Randall et al., 2007) Figure 8.10). Both annual maxima and minima in both hemispheres show a decrease compared to CMIP3, with few exceptions, particularly in the Northern North Atlantic in winter.

[INSERT FIGURE 9.23 HERE]

Figure 9.23: Sea ice distribution in the Northern Hemisphere (upper panels) and the Southern Hemisphere (lower panels) for March (left) and September (right). A) AR5 baseline climate (1986–2005) simulated by 14 of CMIP5 AOGCMs. For each 1° × 1° longitude-latitude grid cell, the figure indicates the number of models that simulate at least 15% of the area covered by sea ice. B) AR4 baseline climate (1980–1999) differences between 14 CMIP5 and 14 CMIP3 (AR4 (Randall et al., 2007) Figure 8.10) AOGCMs. For each 2.5° × 2.5° longitude-latitude grid cell, the figure indicates the difference in the number of CMIP5 and CMIP3 models that simulate at least 15% of the area covered by sea ice. The observed 15% concentration boundaries (red line) are based on the Hadley Centre Sea Ice and Sea Surface Temperature – HadISST data set (Rayner et al., 2003).

Compared to CMIP3, CMIP5 models better simulate the observed trend of September Arctic ice extent (Figure 9.24). This is likely related to the decreased bias in summer sea ice extent noted above (see e.g., Kattsov et al., 2011).

[INSERT FIGURE 9.24 HERE]

Figure 9.24: Arctic September sea-ice extent from observations (NSIDC, red line) (Fetterer et al., 2002), the 12 CMIP5 model multi-model ensemble mean (dark greenish line), the 12 CMIP3 model multi-model ensemble mean (black line), and one standard deviation ranges of the model estimates (bluish and grey shadings, correspondingly). Note that these are September means, not yearly minima.

Sea ice is a product of atmosphere-ocean interaction. There are a number of ways in which sea ice is influenced by and interacts with the atmosphere and ocean, and the nature and magnitude of associated feedbacks are still poorly quantified. As noted in the AR4, among the primary causes of biases in simulated sea ice extent, especially its geographical distribution, are problems with simulating high-latitude winds, ocean heat advection, vertical and horizontal mixing in the ocean. For example Koldunov et al. (2010) have shown, for a particular model, that significant ice thickness errors originate from biases in the atmospheric component. Such biases, common to many models, may be related to representation of processes specific to high-latitude regions (e.g., polar clouds) or processes not yet commonly included in models (e.g., deposition of carbonaceous aerosols on snow and ice).

The CMIP5 models appear to have limited abilities to generate sufficient unforced atmospheric variability (e.g., Koldunov et al., 2010) in the Arctic, but large-scale warming events similar to that observed from the 1920s through the 1940s are simulated better in CMIP5 compared to CMIP3. Some recent models are able to simulate rapid changes in the Arctic sea ice due mainly to natural variability (Holland et al., 2008).

9.4.4 Land Surface, Fluxes, and Hydrology

9.4.4.1 Snow Cover and Permafrost

The CMIP5 ensemble simulates snow-covered area reasonably well (Figure 9.25), however without any evident improvement compared to CMIP3. Individually, some models have been shown to improve simulation of terrestrial snow cover due to improvements in parameterization schemes (e.g., Lawrence et al., 2011) . Most models capture the observed decadal-scale variability over the 20th century. Large discrepancies remain in albedo for forested areas under snowy conditions, due to difficulties in determining the extent of masking of snow by vegetation.

[INSERT FIGURE 9.25 HERE]

Figure 9.25: Terrestrial snow-cover distribution in the Northern Hemisphere simulated by 9 CMIP5 AOGCMs for February. For each 1° × 1° longitude-latitude grid cell, the figure indicates the number of models that simulate at least 5 kg m^{-2} of snow water equivalent. The observed 20% concentration boundaries (red line) are based on the (Robinson and Frei, 2000) and cover the period 1986–2005. The annual mean 0°C isotherm at 3.3 m depth averaged across the 9 AOGCMs (yellow line) is a proxy for the permafrost boundary. Observed permafrost zonation in the Northern hemisphere (magenta dashed line) is based on (Nelson et al., 2002).

9.4.4.2 Soil Moisture and Land-Atmosphere Coupling

Soil moisture provides the land-surface with a memory of past anomalies in precipitation and surface radiation, and also influences future anomalies in these climate variables through its control over evaporation. The soil moisture-precipitation feedback is such that a dry anomaly reduces subsequent precipitation which tends to maintain the anomaly. Prior to the AR4 it became clear that the strength of the coupling between soil moisture and precipitation varied widely between climate models even though the pattern of land-atmosphere “hotspots” was broadly similar (Koster et al., 2004). Soil moisture has a particularly strong control on climate in semi-arid areas (Koster et al., 2004; Seneviratne et al., 2010). In some regions, such as the Sahel, land-atmosphere coupling may even be strong enough to support two alternative climate-vegetation states; one wet and vegetated, the other dry and desert-like.

Since the AR4 there have been a number of studies looking at the role of land-atmosphere coupling in the persistence of summer droughts (Fischer et al., 2007b), and high-temperature extremes (Hirschi et al., 2011b). Comparison of climate model simulations to observations suggests that the models correctly represent the soil-moisture impacts on temperature extremes in south-eastern Europe, but overestimate them in central Europe (Hirschi et al., 2011b). Climate change is expected to increase the extent of semi-arid areas on the globe, so representation of feedbacks involving soil moisture directly affect future-climate projections (Seneviratne et al., 2010).

9.4.4.3 Dynamic Global Vegetation and Nitrogen Cycling

At the time of the AR4 very few climate models included dynamic vegetation, with vegetation cover being prescribed and fixed in all but a handful of coupled climate-carbon cycle models (Friedlingstein et al., 2006). Dynamic Global Vegetation Models (DGVMs) certainly existed at the time of the AR4 (Cramer et al., 2001) but these were not typically incorporated in climate models. Since the IPCC AR4 there has been continual development of offline DGVMs, and many climate models incorporate dynamic vegetation in at least a subset of the runs submitted to CMIP5.

In the absence of nitrogen limitations on CO_2 fertilization, offline DGVMs agree qualitatively that CO_2 increase alone will tend to enhance carbon uptake on the land while the associated climate change will tend to reduce it. There is also good agreement on the degree of CO_2 fertilization in this limit of no nutrient limitation (Sitch et al., 2008). However, under more extreme emissions scenarios the responses of the DGVMs diverge markedly. Large uncertainties are associated with the responses of tropical and boreal

ecosystems to elevated temperatures and changing soil moisture status. Particular areas of uncertainty are the high-temperature response of photosynthesis (Galbraith et al., 2010), and the extent of CO_2 fertilization (Rammig et al., 2010) in Amazonian rainforest.

Most of the DGVMs used in the CMIP5 models continue to neglect nutrient-limitations on plant growth, even though these may significantly moderate the response of photosynthesis to CO_2 (Wang and Houlton, 2009). Recent extensions of two DGVMs to include nitrogen limitations to CO_2-fertilization improve the fit of these models to "Free-Air CO_2 Enrichment Experiments", and suggest that models without these limitations will most likely overestimate the land carbon sink in the nitrogen-limited mid and high latitudes (Thornton et al., 2007; Zaehle et al., 2010a). By contrast, tropical ecosystems are thought to be phosphorus rather than nitrogen limited.

9.4.4.4 Land-Use Change

Another major innovation in the land component of the climate models since the AR4 is the inclusion of land-use change. Changes in land-use associated with the spread of agriculture, urbanization and deforestation affects climate by altering the biophysical properties of the land-surface, such as its albedo, aerodynamic roughness and water-holding capacity (Bondeau et al., 2007; Bonan, 2008; Levis, 2010). Land-use change also contributes about 20% to global anthropogenic CO_2 emissions, and affects emissions of trace gases, and volatile organic compounds such as isoprene.

There has been significant progress in modeling the role of land cover change since the AR4 (Pielke et al., 2007), with the first systematic study demonstrating that large-scale land cover change directly and significantly affects regional climate (Pitman et al., 2009). However, climate models currently simulate rather different response of the climate even to the same imposed land-cover change (Pitman et al., 2009).

9.4.5 Carbon Cycle

9.4.5.1 Terrestrial Carbon Cycle Component Models

Current dynamic global vegetation models can reproduce the observed land-atmosphere fluxes of CO_2 to within 30% and can replicate the greater carbon uptake observed in the 1990s compared to the 1980s (Sitch et al., 2008). However, several coupled biogeochemistry/land-surface models underestimate the seasonal amplitude of CO_2 in the northern hemisphere by factors of 2 to 3 (Randerson et al., 2009). This conclusion is model dependent, however, as the fully-coupled Earth system models evaluated by Cadule et al (2010) exhibit much greater skill in simulating the amplitude of the seasonal cycle. The phasing of the annual cycle in CO_2 over northern latitudes is generally accurate, and the timing of observed spring drawdown of CO_2 is reproduced to within 1 month in the tropics with increasing phasing errors between 60ºN and 90ºN. Much larger phase errors emerge in some ESMs for remote regions near the South Pole (Cadule et al., 2010).

Accurate simulation of the Amazon is important for representing its buffering of atmospheric CO_2 and for projecting the effects of climate change on the amount of carbon stored in the Amazonian forests (Lewis et al., 2011). While two biogeochemical sub-models used in a particular ESM reproduced the gross primary productivity (GPP) of the Amazonian forests to within 14% of observational estimates (Lewis et al., 2011), the models overestimated the above-ground live biomass by 130 to 190% and underestimated soil carbon by 33 to 40% (Randerson et al., 2009). The overestimation of live biomass in the Amazon is attributable to parametric errors including excessive allocation of net primary productivity (NPP) to wood and underestimation of the flow of GPP to autotrophic respiration.

Wildland and human-induced fires contribute approximately 2.3 PgC yr^{-1} to the atmosphere based upon estimates for 1997–2004 (Randerson et al., 2009). Inadequate parameterisations of fires can lead to underestimation of this flux by a factor of 3 and to errors in its spatial and temporal variability caused by deforestation-linked fires and the effects of drought. Recent advances in parameterisations yield reasonably good agreement between simulated emissions and satellite-based retrievals on interannual timescales (Kloster et al., 2010).

9.4.5.2 Oceanic Carbon Cycle Component Models

Recent advances in the observational evaluation of ocean ecosystem-biogeochemical (OBGC) models include new diagnostic frameworks designed for quantitative multi-model intercomparisons (Doney et al., 2009) and protocols to evaluate the impact of ocean circulation on the marine carbon cycle, including export production, dissolved organic matter, and dissolved oxygen (Najjar et al., 2007). Similar error structures appear in the regional patterns and seasonal cycles of multiple independent variables dependant on the underlying physical ocean model. The findings support earlier studies that show that the empirical fidelity in the biological properties of ecosystem-biogeochemical models is contingent on corresponding levels of accuracy in properties of the simulated physical ocean system (Doney et al., 2009; Najjar et al., 2007), in particular the SSTs, mixed-layer depths (MLDs), upwelling rates, and vertical structure near the surface.

Evaluation of OBGC models has been performed with a focus on regional oceanic uptake of CO_2 (Roy et al., 2011), and a similar framework will be applied in the near future for evaluating CMIP5 models. Based on results already available, declining rates of net ocean CO_2 uptake observed in the temperate North Atlantic are broadly reproduced by historical OBGC model simulations (Thomas et al., 2008). These trends represent a superposition of interannual variability associated with the NAO and with secular trends in surface warming. The positive trend in observed sea-air CO_2 partial pressure differences between 1997 and 2004, which is indicative of reduced oceanic uptake or greater outflow of CO_2, is also simulated. However, models that have been evaluated against estimates of surface chlorophyll concentrations cannot reproduce the regime shifts observed in the Northern Atlantic since 1948 (Henson et al., 2009) or the broad-scale shifts from lower to higher biomass-normalized primary productivity between the 1980s and 1990s (Friedrichs et al., 2009). The greater skill in reproducing surface CO_2 fields compared to ecological variables including chlorophyll concentrations is consistent with the relative skills in these fields observed by Doney et al (2009). The errors in reproducing decadal regime shifts are due to challenges in modelling the phytoplankton community structure, the impact of the Gulf Stream on biological variability downstream, and transitions between ecological states (Henson et al., 2009).

9.4.5.3 The Carbon Cycle in Earth System Models

The transition from climate models to ESMs was motivated in part by the results from the first generation coupled climate-carbon cycle models, which suggested that feedbacks between the climate and the carbon cycle were uncertain but potentially very important in the context of 21st century climate change (Cox et al., 2000; Friedlingstein et al., 2001). The first generation models used in the Coupled Climate Carbon Cycle Model Intercomparison Project (C^4MIP) included both extended AOGCMS and EMICs. The C^4MIP experimental design involved running each model under a common emission scenario (SRES A1B) and calculating the evolution of the global atmospheric CO_2 concentration interactively within the model. The impacts of climate-carbon cycle feedbacks were diagnosed by carrying-out parallel "uncoupled" simulations in which increases in atmospheric CO_2 did not influence climate. Analysis of the C^4MIP runs highlighted: (a) a greater than 200ppmv range in the CO_2 concentration by 2100 due to uncertainties in climate-carbon cycle feedbacks, and (b) that the largest uncertainties were associated with the response of land ecosystems to climate and CO_2 (Friedlingstein et al., 2006).

For CMIP5 a different experimental design was proposed in which the core simulations use prescribed Representative Concentration Pathways (RCPs) of atmospheric CO_2 and other greenhouse gases (Moss et al., (2010)). Under such a prescribed CO_2 scenario, ESMs still calculate land and ocean carbon fluxes interactively, but these fluxes do not affect the evolution of atmospheric CO_2. Instead the modelled land and ocean fluxes, along with the prescribed increase in atmospheric CO_2, can be used to diagnose the "allowable" emissions of CO_2 consistent with the simulation (Miyama and Kawamiya, 2009; Arora et al., 2011). The allowable emissions for each model can then be evaluated against the best estimates of the actual historical CO_2 emissions. Parallel model experiments in which the carbon cycle does not respond to the simulated climate change (which are equivalent to the "uncoupled" simulations in C^4MIP) provide a means to diagnose climate-carbon cycle feedbacks in terms of their impact on the allowable emissions of CO_2 (Hibbard et al., 2007).

Figures 9.26 and 9.27 show results from the historical RCP simulations of the ESMs available in the CMIP5 archive. In each case these ESMs have provided both ocean and land CO_2 fluxes. In a major advance over

the C^4MIP experiments, which neglected land-use change, each of these models also simulate the affects of land-use change on both the carbon cycle and the biophysical properties of the land-surface (Bathiany et al., 2010). Figure 9.26 compares the simulations of global land carbon uptake (top left panel) and global ocean carbon uptake (top right panel) relative to 1960. Also shown are the estimates provided by the Global Carbon Project (GCP) which are derived from offline ocean carbon cycle models, measurements of atmospheric CO_2, and best estimates of the CO_2 fluxes from fossil fuels and land-use change (Le Quere et al., 2009).

[INSERT FIGURE 9.26 HERE]

Figure 9.26: Simulation of land carbon uptake (top left) and ocean carbon uptake (top right) in the CMIP5 Earth System Models (ESMs), for the period 1960–2005, relative to 1960. All of these models include the impact of land-use changes on land carbon storage. For comparison, the observation-based estimates provided by the Global Carbon Project ("GCP", Le Quere et al., 2009) are also shown as the dotted line. The bottom right panel shows the sum of the land and ocean uptake from 1900 to 2005, again relative to 1960.

[INSERT FIGURE 9.27 HERE]

Figure 9.27: Simulation of net land CO_2 flux (top left) and ocean CO_2 flux (top right) in the CMIP5 Earth system models (ESMs), for the period 1995–2005. In each panel the mean flux over the period is plotted on the x-axis, while the standard deviation of the annual fluxes is plotted on the y-axis. For comparison, the observation-based estimates provided by the Global Carbon Project ("GCP", Le Quere et al., 2009) are also shown as the dotted line.

The top left panel of Figure 9.26 shows the net change in land carbon storage, which arises from the response of the landscape to climate change and CO_2 increase, and also the net change in land carbon arising from land-use change. From 1960 to around 1989 the GCP data suggest the overall affect was little change in global land carbon storage, but this was followed by a 15 GtC accumulation of land carbon between 1989 and 2005. The ESMs simulate a wide-range of changes in land carbon storage from 1960 to 2005, of between 9 and 42 GtC. There is greater agreement concerning the uptake of CO_2 by the ocean between 1960 and 2004, with all models and the GCP data showing a monotonic increase in ocean carbon storage (top right panel of Figure 9.26). GCP data suggest an increase in ocean carbon storage of 97 GtC from 1960 to 2004, with the ESMs generally simulating a smaller uptake (60 to 84 GtC).

The GCP estimates of land and ocean uptake are themselves reliant on models, albeit ocean models forced by best estimates of the observed meteorology. However, the sum of land and ocean uptake is constrained by the known increase in atmospheric CO_2 and the estimated fossil fuel CO_2 emissions, without requiring the use of offline models of the ocean or land uptake. The lower panel of Figure 9.26 compares the sum of the land plus ocean uptake for each ESM for the period 1900 to 2005, to such an "inferred" land plus ocean uptake derived as the difference between the integrated fossil fuel emissions and the increase in atmospheric carbon. For comparison with the other panels of Figure 9.27 the change in carbon storage is given relative to 1960. In model experiments in which CO_2 is allowed to run free, the models that underestimate the land plus ocean uptake will tend to over-estimate the historical CO_2 rise, while the opposite will be the case for models that over-estimate the land plus ocean uptake.

Figure 9.27 compares the mean and standard deviation of the annual mean global land and ocean fluxes, for the more recent period 1995 to 2005. The mean ocean CO_2 sink is reasonably tightly clustered across the ESMs, ranging between 1.7 and 2.3 GtC yr^{-1}, which compares with the GCP estimate of 2.2 GtC yr^{-1}. Once again there is much less agreement on the mean land sink, with ESM simulations producing a range from 0 to 1.6 GtC yr^{-1}, in comparison to the GCP estimate of 0.7 GtC yr^{-1}. All ESMs and the GCP data agree that the year-to-year variability in the global land CO_2 flux is much larger than the variability in the global ocean CO_2 flux, by a factor of 5 to 15. In general the ESMs tend to overestimate the variability in the land flux compared to the GCP estimate, which may be indicative of an over-sensitivity of tropical land carbon to climate change in these models (see Section 9.8.3).

In summary, the ESMs evaluated in this report do reasonably well in reproducing the estimated uptake of CO_2 by the global ocean since 1960, but do less well in the simulation of the global land uptake. The additional complication of including the impacts of historical land-use change seems to be one of the main reasons for the range in the ESM simulations of global land carbon storage since 1960.

9.4.6 Sulfur Cycle

9.4.6.1 Recent Trends in Regional and Global Sulphate Burdens and Effects on Insolation

The historical emissions data used to drive the CMIP5 simulations of the 20th century reflect two recent trends in regional and global anthropogenic SO_2 emissions. During the last three decades, anthropogenic emissions of SO_2 from North American and Europe have declined due to the imposition of emission controls, while the emissions from Asia have increased. The combination of the European, North American, and Asians trends has yielded a global reduction in SO_2 emissions by 24% between 1987 and 2000.

The effects of these changes in emissions on the total atmospheric sulphate burden can be simulated using detailed chemical transport models (CTMs) forced with meteorological reanalyses. The results from these CTMs can then be used to evaluate the historical simulations of sulphate burdens from ESMs (Figure 9.28). The CTM calculations show that each 1% decrease in European emissions of SO_2 yields a 0.65% reduction in modeled sulphate burden while each 1% increase in Asian emissions yields a 0.88% increase in sulphate burdens. The reason is that emissions have generally moved southward to regions where the in-cloud oxidation process is less oxidant limited. In-cloud oxidation converts SO_2 to SO_4 and comprises 71% of the global sulphate production rate under present conditions.

The effects of sulphate and other aerosol species on surface insolation through direct and indirect forcing appear to be one of the principal causes of the “global dimming” between the 1950s and 1980s and subsequent “global brightening” in the last two decades (see Figure 9.29). This inference is supported by the correlative trends in aerosol optical depth and by trends in surface insolation under cloud-free conditions. Thirteen out of fourteen CMIP3 models examined by (Ruckstuhl and Norris, 2009) produce a transition from “dimming” to “brightening” that is consistent with the timing of the transition from increasing to decreasing global anthropogenic aerosol emissions.

[INSERT FIGURE 9.28 HERE]

Figure 9.28: The relative error in visible aerosol optical thickness (AOT) from the median of a subset of CMIP5 models’ historical simulations, relative to satellite retrievals of AOT. The figure was constructed following Kinne et al. (2006). The satellite AOT is from the MODIS instrument on the NASA Terra satellite from 2001 through 2005. The data version is MODIS 4; the model output is from CSIRO Mk3-6-0, GISS ER-2, HadGEM2-ES, IPSL CM5A-LR, and NorESM1-M.

[INSERT FIGURE 9.29 HERE]

Figure 9.29: Time series of the global oceanic-mean AOT from individual CMIP5 models’ historical simulations against the time series of global oceanic-mean AOT from the Global Aerosol Climatology Project (GACP). The figure is constructed following Mishchenko et al. (2007). The "brightening" trend shown with the straight green line is discussed in Chapter 7. The model output is from CSIRO Mk3-6-0, GISS ER-2, HadGEM2-ES, IPSL CM5A-LR, and NorESM1-M.

9.4.6.2 Principal Sources of Uncertainty in Projections of Sulphate Burdens

In contrast to the CMIP3 multi-model simulation ensemble, the CMIP5 ensemble is based upon a single internally consistent set of SO_4 concentrations and SO_2 emissions. The use of a single set of emissions removes an important, but not dominant, source of uncertainty in the AR5 simulations of the sulphur cycle. In experiments based upon a single chemistry-climate model with perturbations to both emissions and sulphur-cycle processes, uncertainties in emissions accounted for 53.3% the ensemble variance (Ackerley et al., 2009). The next largest source of uncertainty was associated with the wet scavenging of sulphate, which accounted for 29.5% of the intra-ensemble variance and represents the source/sink term with the largest relative range in the aerosol models evaluated by AeroCom (Faloona, 2009). Similarly, AeroCom simulations run with heterogeneous or harmonized emissions data sets yielded approximately the same intermodel standard deviation in sulphate burden of 25 Tg for both sets of experiments. These results show that a dominant fraction of the spread among the sulphate burdens produced by chemistry-climate models are due to differences in the treatment of chemical production, transport, and removal from the atmosphere (Liu et al., 2007; Textor et al., 2007).

Natural sources of sulphate from oxidation of natural dimethylsulphide (DMS) emissions from the ocean surface are not specified under the RCP protocol and therefore represent an additional source of uncertainty in the sulphur cycle simulated by the CMIP5 ensemble. In simulations of present-day conditions, DMS

emissions span a 5 to 95% confidence interval of 10.7 to 28.1 TgS yr^{-1} (Faloona, 2009). After chemical processing, DMS contributes between 18 to 42% of the global atmospheric sulphate burden and up to 80% of the sulphate burden over most the southern hemisphere (Carslaw et al., 2010). The effects from differences in DMS emissions and its subsequent oxidation to sulphate on sulphate burdens in the CMIP5 ensemble remain to be quantified.

9.5 Simulation of Variability and Extremes

9.5.1 Importance of Simulating Climate Variability

The ability of a model to simulate the mean climate, and the slow, externally-forced change in that mean state, is important and was evaluated in the previous Section. However, the ability to simulate climate variability, both unforced natural variability and forced variability (e.g., diurnal and seasonal cycles) is also important. This has implications for the signal-to-noise estimates inherent in climate change detection and attribution studies where low-frequency climate variability must be estimated, at least in part, from long control integrations of climate models. It also has implications for the ability of models to make quantitative projections of changes in climate variability and the statistics of extreme events under a warming climate. In many cases, the impacts of climate change will be experienced more profoundly in terms of the frequency, intensity or duration of extreme events (e.g., heat waves, droughts, extreme rainfall events). The ability to simulate climate variability is also central to the topic of climate prediction, since it is the ability to simulate the specific evolution of the varying climate system, beyond that due to external forcing, that provides useful predictive skill.

Evaluating model simulations of climate variability also provides a means to explore the representation of certain processes, such as the coupled processes underlying the El Niño Southern Oscillation (ENSO) and other important modes of variability. A model's representation of the diurnal or seasonal cycle – both of which represent responses to external (rotational or orbital) forcing – may also provide some insight into a model's 'sensitivity' and by extension, the ability to respond correctly to greenhouse gas, aerosol, volcanic and solar forcing.

In this Section we will also investigate the extent to which biases in the simulation of the mean climate and its long-term evolution (Section 9.4) are related to biases in variability, and we will explore to some extent model features, such as resolution, that may affect the simulation of variability, particularly aspects such as atmospheric blocking and convective precipitation events.

9.5.2 Diurnal-to-Seasonal Variability

9.5.2.1 Diurnal Cycles of Physical Climate Variables

The diurnally varying solar radiation received at a given location drives, through complex interactions with the atmosphere, land surface, and upper ocean, easily observable diurnal variations not only in surface and near-surface temperature, but also precipitation, low level stability and winds, and many other geophysical parameters. As the diurnal cycle of many climate variables depends on complex interactions of many different physical processes, the diurnal cycle provides insight into several aspects of model physics.

Coupled models capture the overall amplitude and phase of the diurnal cycle of surface air temperature (T) well over land, but tend to produce a weak diurnal cycle over the ocean because of a lack of diurnal variations in sea surface temperature (SST) (Figure 9.30). In recent simulations using an AOGCM that explicitly represents SST diurnal variations, some long-standing model biases, such as cold biases in the tropical Pacific, have been reduced (Bernie et al., 2008). However, most models still have difficulties in simulating the diurnal variations in SST due to coarse vertical resolution and the time lag in air-sea coupling in these models (Dai and Trenberth, 2004; Danabasoglu et al., 2006).

[INSERT FIGURE 9.30 HERE]

Figure 9.30: Composite diurnal cycle of surface air temperature from observations (black line) and CMIP3 models (coloured lines) averaged over land (left) and ocean (right) areas for three different zones. Adapted from Dai and Trenberth (2004).

The simulation of the diurnal cycle of precipitation over land is another significant challenge. Figure 9.31 shows that AOGCMs have difficulty simulating the diurnal cycle of warm-season precipitation over land, in particular when convective processes are involved, as is frequently the case for the tropics and extratropical land areas in summer. Many of the CMIP3 models tend to start moist convection prematurely and thus often rain too frequently at reduced intensity, resulting in a rainfall peak too early in the day (Dai, 2006) and the so-called "drizzling bias" (Dai, 2006; Stephens et al., 2010) that can have large adverse impacts on surface evaporation and runoff (Qian et al., 2006). Many models also produce too much convective rain but too little stratiform precipitation compared with satellite data (Dai, 2006).

[INSERT FIGURE 9.31 HERE]

Figure 9.31: Composite diurnal cycle precipitation from observations (black) and a subset of CMIP3 models (coloured lines) averaged over land (left) and ocean (right) areas for three different zones. Adapted from Dai (2006).

Several studies have attempted to identify the reasons for the poor model behaviour by carrying out sensitivity studies (Betts and Jakob, 2002; Zhang and Klein, 2010). Increased atmospheric resolution (Ploshay and Lau, 2010) as well as the use of the super-parameterisation approach (Khairoutdinov et al., 2005) or very high-resolution simulations of short duration (Sato et al., 2009) have shown significant promise for improvements in the simulation of the diurnal cycle of precipitation, although the physical reasons for these improvements remain poorly understood. An improved coupling between shallow and deep convection, as well as the inclusion of a description of density currents, has been shown to greatly improve the diurnal cycle of convection over tropical land in one model (Peterson et al., 2009). While the main focus in improving the diurnal cycle in models has been on deficiencies in cumulus convection, it is likely that model deficiencies in surface-atmosphere interactions and the planetary boundary layer also contribute to its poor representation.

While much of the focus of diurnal cycle studies has been on temperature and precipitation, other diurnal variations in have also been evaluated, such as those of surface pressure (Covey et al., 2011; Dai and Trenberth, 2004) humidity and cloudiness, low–level and tropospheric winds (e.g., Dai and Trenberth, 2004). The diurnal cycle of surface energy and water fluxes has also been extensively evaluated using a variety of observations and process modelling approaches. In general, most CMIP3 models are able to reproduce the observed surface pressure tides despite the low model tops in many of the models (Covey et al., 2011), and the associated diurnal variations in tropospheric wind are also broadly reproduced, though with relatively weak amplitudes over oceans based on limited analyses (Dai and Trenberth, 2004)

9.5.2.2 Intraseasonal Variability

Several features at the intraseasonal time scale influence or are influenced by the mean state or climate variability, and are strongly connected to the regional climate characteristics discussed in Chapter 14. In the following we concentrate on blocking events and the Madden Julian oscillation that appear to play an important role in mid latitudes and the tropics respectively.

9.5.2.2.1 Blocking and circulation regimes

During blocking weather regimes the prevailing midlatitude westerly winds and storm systems are interrupted by a local reversal of the zonal flow. Recent work has underlined the importance of blocking for the occurrence of extreme weather events (Buehler et al., 2011), yet climate models in the past have universally underestimated the occurrence of blocking. However, recent work has shown that very high resolution atmospheric AOGCMs can now simulate the observed level of blocking in both hemispheres, although in this case blocking in the North Pacific is in fact overestimated (Matsueda et al., 2009; Matsueda et al., 2010).

Since the AR4 there has been a renewed focus on the diagnostic methods used to characterize blocking. There are still important differences between methods (Barriopedro et al., 2010a), and the diagnosed blocking frequency can be very sensitive to details such as in the choice of latitude (Barnes et al., 2011). In particular, blocking indices based on the identification of reversed meridional gradients in quantities such as geopotential height can be sensitive to mean state biases in the models, so that the diagnosed biases in blocking reflect biases in the mean state rather than the level of variability in the models (Scaife et al., 2010).

In some models the mean state bias can explain most of the underestimation of blocking, but in other models a significant underestimation still remains (Scaife et al., 2010), reflecting problems with the model's simulation of variability (Barriopedro et al., 2010b). Other blocking indices use anomaly fields, rather than reversed absolute fields, to define blocking, and by these measures model skill can appear better (e.g., Sillmann and Croci-Maspoli, 2009). Recent work has confirmed the impression of a link between blocking events and stratospheric flow anomalies (Martius et al., 2009). This link mostly, but not exclusively, comprises blocking events perturbing the stratospheric flow through upward-propagating Rossby wave activity, and the observed links are shown to be quite well represented in a climate model with enhanced stratospheric resolution (Woollings et al., 2010d).

There is evidence that climate models can simulate the broad features of observed circulation regimes (Teng et al., 2007). There is observational evidence of regime behaviour in the variability of the North Atlantic eddy-driven jet stream (Woollings et al., 2010a), and the CMIP3 models show a range of skill in simulating this structure (Barnes and Hartmann, 2010). The CMIP3 models often underestimate the amplitude of low-frequency planetary wave variability (Lucarini et al., 2007), which is likely to contribute to biases in regime behaviour.

9.5.2.2.2 Madden Julian Oscillation

During the boreal winter the eastward propagating feature known as the Madden-Julian Oscillation (MJO) predominantly affects the deep tropics, while during the boreal summer there is also northward propagation over much of southern Asia (Annamalai and Sperber, 2005). Cassou (2008; Pan and Li, 2008) present evidence that the MJO controls part of the distribution and sequences of the four daily weather regimes defined over the North Atlantic–European region in winter, suggesting a link between the quality of the representation of the MJO in climate models and model ability to properly reproduce weather regimes. Previous assessments reported that most AOGCMs have difficulty in representing intraseasonal MJO variability with most models underestimating the strength and the coherence of convection and wind variability at MJO temporal and spatial scales (Lin and Li, 2008; Lin et al., 2006). Coupling with the ocean and convection schemes were highligted as important factors contributing to model deficiencies (Bernie et al., 2008). Simulation of the Madden-Julian Oscillation is still a challenge for climate models (Kim et al., 2009; Lin et al., 2006; Xavier et al., 2010), however, Sperber and Annamalai (2008) have shown that the CMIP3 models were able to simulate eastward propagating intraseasonal convection over the Indian Ocean. This represents an improvement over earlier models (Waliser et al., 2003), though it must be noted that only two of seventeen models were able simulate the observed northward propagation during boreal summer. As seen in Figure 9.32, using lead-lag temporal correlations of the two leading principal component time series, the maximum positive correlation and the time lag at which it occurs indicate that all of the CMIP3 models have less coherent eastward propagation than observed. There is a diverse representation of the time scale of the simulated MJO, and some models are incorrectly dominated by westward propagation.

[INSERT FIGURE 9.32 HERE]

Figure 9.32: Outgoing Longwave Radiation (OLR), 20–100 day filtered, from observations and each of the CMIP3 models' simulations of 20th-century climate is projected on the two leading Empiricol Orthogonal Functions (EOF's) of OLR that constitute the Madden-Julian Oscillation (MJO). Shown is the maximum positive correlation between the resulting MJO Principal Components (PC's) and the time lag at which it occurred for all winters (November-March). The maximum positive correlation is an indication of the coherence with which the MJO convection propagates from the Indian Ocean to the Maritime Continent/western Pacific and the time lag is approximately 1/4 of the period of the MJO. Most models have weaker coherence in the MJO propagation (smaller maximum positive correlation), and some have periods that are too short compared to observations. One CMIP3 model is not shown as its day of maximum positive correlation was –16, indicating that this model is incorrectly dominated by westward propagation. Constructed following (Sperber et al., 2005).

9.5.2.3 Large Scale Monsoon Rainfall and Circulation

The global monsoon is the dominant mode of annual variation in the tropics (Trenberth et al., 2000; Wang and Ding, 2008). Given the billions of people that fall under its influence, high fidelity simulation of the mean monsoon and its variability is of great importance (Sperber et al., 2010; Wang et al., 2006), and metrics of a models ability to simulate the monsoon domain and its intensity were introduced by (Wang and Ding, 2008). These measures are based on the annual range of precipitation using hemispheric summer minus winter values. The monsoon precipitation domain is defined where the annual range is >2.5 mm day^{-1},

and the monsoon precipitation intensity is the annual range/annual mean. These measures provide a large-scale view of the Earth's monsoon systems (Wang et al., 2011).

The monsoon precipitation domain and intensity from observations and CMIP3 models are shown in Figure 9.33. The subtropical monsoon domain is indicated by the grey-shaded contours, while the shading is the precipitation intensity, which indicates that summer rainfall dominates the annual mean, especially over the continents. The CMIP3 multi-model ensemble generally reproduces the observed spatial patterns but somewhat underestimates the extent and intensity, especially over Asia and North America. In terms of the threat score (a categorical metric (Wilks, 1995) which indicates how well a model simulates the monsoon precipitation domain) the best CMIP3 model outperforms the multi-model mean. Additionally, there is a strong disparity between the best and poorest CMIP3 models, with the latter failing to capture the monsoon precipitation domain over the Sahel, Central America, and Australia. The models' threat scores are smaller than the observational uncertainty, indicating the potential for improvement in the simulation of monsoon rainfall characteristics.

[INSERT FIGURE 9.33 HERE]

Figure 9.33: Monsoon precipitation intensity (shading, mm/day) and monsoon precipitation domain (lines) are shown for (a) observations from GPCP, (b) the CMIP3 multi-model mean, (c) the best model, and (d) the worst model in terms of the threat score for this diagnostic. The threat scores indicate how well the models represent the monsoon precipitation domain compared to the GPCP data. The threat score in panel (a) is between GPCP and CMAP rainfall to indicate observational uncertainty. A threat score of 1.0 would indicate perfect agreement between the two datasets. See Wang and Ding (2008); Wang et al. (2011); and Kim et al. (2011b) for details of the calculations..

Large variations of the monsoon systems have been recorded in paleo proxy records (see Chapter 5). They show for example that the boreal summer monsoon was stronger and penetrated further inland during the mid-Holocene, increasing monsoonal precipitation in western North America, northern Africa and China (Bartlein et al., 2010b; Zhao and Harrison, 2011). The representation of the northward shift of the rainbelt in the Sahel region has improved in the last generation of paleo-climate models, even though most models still underestimated the amount of precipitation north of 18°N (Braconnot et al., 2007d). Comparison with data over East Asia show that the PMIP2 simulations reproduce well the precipitation in China except for in the central region, but that the model spread is large (Wang et al., 2010). Evaluation of the southern-hemisphere monsoons is limited by lack of quantitative reconstructions, but initial results suggest that model skill in simulating these monsoons is limited (Zhao and Harrison, 2011)

9.5.3 Interannual-to-Centennial Variability

The mean climate is, by construction, the time average of the numerous scales at which the climate components vary. In addition to the annual and diurnal cycles, directly forced by the sun and described above, a number of other modes of variability arise from interactions (or feedbacks) between the various components on a number of time and space scales. Here we limit the scope to modes of variability whose timescale ranges from a few weeks (e.g., blocking regimes) to multi-decadal features that can modulate the trend arising from changes in GHGs. Most of these modes have a particular regional manifestation. The observational record is sometimes too short to fully evaluate the representation of variability in models and this motivates the use of re-analysis or proxies, even though these have their own limitation. In the following, we also emphasize recent research on the interactions between modes of variability via teleconnections, the processes involved, and model improvements since the AR4.

9.5.3.1 Global Surface Temparature Variability

Simulating climate variability on various time scales is important in applications of climate models to detection and attribution, and to predictions and projections of future climate. One approach to evaluation of model variability is to compare simulations of the last millennium to temperature reconstruction for this period (Chapter 5). Figure 9.34 shows the normalized spectra of last-millennium simulations run with various AOGCMs. These simulations include natural and anthropogenic forgings (solar, volcanism, greenhouse gases, land use) and so include aspects of forced and internal (unforced) variability. For comparison similar spectra are also provided for unforced variability arising from long pre-industrial control simulations. Significant differences between unforced and forced simulations are mainly found in the low frequency part of the spectra. This is due to the long term trend resulting from the combination of solar and

greenhouse gas forcing, and correspond to what is inferred from NH paleo reconstructions (Chapter 5). The lower spectral density found in the two MPI-ESM simulations results from an underestimation of the temperature trend, consistent with an underestimation of the carbon fluxes by the interactive carbon cycle (Jungclaus et al., 2010). The peak around 2–4 years which is particularly prominent in the MPI-ESM ensemble is due to strong ENSO variability and overly strong ENSO teleconnections in this model (Zhang et al., 2011; Joly et al., 2007b; Meehl et al., 2007b).

[INSERT FIGURE 9.34 HERE]

Figure 9.34: Power spectral density of NH temperature for a) several simulations of the last millennium performed with CMIP3-generation models (see Chapter 5) b) long pre-industrial simulations for a subset of the same models. In a) the model were all forced by the long-term evolution of the atmospheric trace gases, tropospheric aerosols (except ECHOG and CCSM3), solar irradiance, volcanism eruption (except for IPSL), even though from different reconstructions. The two MPI-ESM simulations differ by the magnitude of the change in the solar irradiance between the Little Ice Age and the present (0.1% in E1 instead of 0.25% in E2 and the other simulations) to better reflect the recent revised estimate by Solanki et al. (2004). A subset of simulations also includes the volcanic forcing or the evolution of land use (MPI-ESM and CNRM). In the MPI-ESM simulations the carbon cycle is interactive.

9.5.3.2 North Atlantic Oscillation and Annular Modes

This section assesses the ability of climate models to reproduce the North Atlantic Oscillation, the closely related Northern Annular Mode, and the Southern Annular Mode. Definition and interpretation of these modes is presented in Chapter 14.

Based on CMIP3 coupled model simulations, Gerber et al. (2008) confirmed the AR4 assessment that climate models are able to capture the broad spatial and temporal features of these modes as well as their main inter-hemispheric differences. While models successfully simulate the broad features of the NAM, there are substantial differences in the spatial patterns amongst individual models (Miller et al., 2006; Stephenson et al., 2006) especially in non-winter seasons (Stoner et al., 2009; Zhu and Wang, 2010). Climate models have a tendency to overestimate the teleconnection between the Atlantic and Pacific basins, so that patterns of variability tend to be more annular in character than in observations (Xin et al., 2008). Models substantially over-estimated the persistence on subseasonal and seasonal time scales, particularly during austral spring and summer, and showed much broader annual cycles than found in re-analyses for either hemisphere. The latter problem is particularly evident in the Northern Hemisphere where only the multi-model ensemble mean showed a robust annual cycle, although the time of peak activity was delayed by a month relative to that in re-analyses (Gerber et al., 2008). The unrealistically long timescale of jet variability is worse in models with particularly strong equatorward biases in the mean jet location, a result which has been found to hold in the North Atlantic and in the Southern Hemisphere (Barnes and Hartmann, 2010; Kidston and Gerber, 2010).

As described in theAR4, several climate models have been unable to simulate the observed level of multi-decadal variability in the NAO/NAM, in particular the strong positive trend over the latter half of the 20th century (Stephenson et al., 2006; Stoner et al., 2009). Underestimation of NAO trends can contribute substantially to underprediction of future warming in certain regions (Knutson et al., 2006). Scaife et al. (2009) showed that atmospheric GCMs forced with observed sea surface temperatures, sea-ice and radiative forcings are not able to simulate the strong NAO trend over the period 1965–1995. However, several coupled climate models do exhibit multi-decadal variability in unforced control simulations which is sometimes as large as the observed 50 year (Raible et al., 2005) and even 30 year trends (Selten et al., 2004; Semenov et al., 2008). Sampling variability may therefore be an explanation for the mismatch, but other explanations have also been suggested, so it is unclear to what extent the underestimation of late 20th century trends reflects real model shortcomings. While some studies suggest that greenhouse gas forcing could have played a role in this positive NAO trend (Paeth et al., 2008) model projections have a vertical structure of circulation change which is quite different from the NAO/NAM (Woollings, 2008). Scaife et al. (2005) showed that the trend can be reproduced reliably and repeatedly when the upper atmospheric winds are relaxed to the observed trend. Further evidence has emerged of the coupling of NAM variability between the troposphere and the stratosphere, and even models with improved stratospheric resolution appear to underestimate the vertical coupling (Morgenstern et al., 2010a). Furthermore, the representation of the stratosphere seems to have a direct bearing on the sign of the NAM response to anthropogenic forcing (Morgenstern et al., 2010a; Scaife et al., 2011). Improved representation of storms in higher-horizontal

resolution climate models has also been shown to improve model ability to simulate the NAO (Marti et al., 2010).

Dynamical processes such as blocking affect the shape of the NAO probability distribution. The daily NAO index from reanalyses has pronounced negative skewness, which is poorly represented in simulations from several climate models (Coppola et al., 2005; Woollings et al., 2010b). Climate models also poorly capture the asymmetry in persistence noted between positive and negative phases of the NAO (Woollings et al., 2010b). Similarly, the CMIP3 models tend to misrepresent the distribution in the latitude of the North Atlantic eddy-driven jet stream, a quantity closely related to the NAO. Specifically the jet latitude distribution tends to be unrealistically positively skewed in models, and this bias in skewness is associated with the equatorward bias in the mean jet latitude in the models (Barnes and Hartmann, 2010). In contrast to the SAM, the North Atlantic jet shows little relation between its mean bias and its response to forcing (Woollings and Blackburn, 2011).

While much of the literature remains focused on wintertime variability, the summertime equivalent of the NAO has been shown to have considerable influence on regional climate, although over a more limited region than in winter. Folland et al. (2009) tested the ability of two climate models to simulate the summer NAO, finding in general a good simulation of its main features, although in one of the models the summer NAO corresponds only to the second EOF.

There are also considerable biases in the Southern Hemisphere eddy-driven jet stream in the CMIP3 models and these appear to have a direct bearing on the magnitude of the SAM response to forcing (Barnes and Hartmann, 2010; Kidston and Gerber, 2010). In terms of spatial patterns, Raphael and Holland (2006) showed that coupled models produce a clear SAM but that there are relatively large differences between models in terms of the exact shape and orientation of this pattern. Karpechko et al. (2009) found that the CMIP3 models have problems in accurately representing the impacts of the SAM on SST, surface air temperature, precipitation and particularly sea-ice in the Antarctic region.

9.5.3.3 Atlantic Modes

9.5.3.3.1 AMOC variability

The spatial and temporal variability of the Atlantic Meridional Overturning Circulation (AMOC) has been sporadically observed (Chapter 3). Continuous AMOC time-series exist for latitues 41°N (reconstructions since 1993) and 26°N (direct observations since 2004) (Willis, 2010; Cunningham et al., 2010). At 26°N, model simulations show realistic variability for the total AMOC over the available observational record (Baehr et al., 2009; Marsh et al., 2009; Balan Sarojini et al., 2011).

Most AMOC observations (continuous or sporadic) estimate the total AMOC as the sum of a wind-driven component and an ocean interior, density-driven component. The wind-driven variability appears well represented to over-represented in different models, while the density-driven variability appears to be well represented to under-represented (Baehr et al., 2009; Balan Sarojini et al., 2011). The under-representation of the variability of the density-driven contribution might point to deficiencies in the simulation of hydrographic characteristics (Baehr et al., 2009). Some of these deficiencies in the simulation of the AMOC variability might improve at higher resolution (Marsh et al., 2009). Note that most models analyzed so far are too coarse to resolve eddies, which might play a role in the total transport variability (Kanzow et al., 2009; Wunsch, 2008).

Concerning the meridional coherence of the AMOC cell, models have suggested AMOC variability that is specific to individual ocean gyres (Baehr et al., 2009; Biastoch et al., 2008b; Bingham et al., 2007), something that has recently been confirmed by analysis of hydrographic data (Lozier et al., 2010).

9.5.3.3.2 Atlantic multi-decadal variability / AMO

The Atlantic Multidecadal Oscillation (AMO) is one of the principal modes of climate variability found in the instrumental climate record, with an apparent period of about 70 years and a pattern centred on the North Atlantic Ocean but with a near-global climatic reach (see Section 14.2.5). In the AR4, it was shown that a number of climate models produced AMO-like multidecadal variability in the North Atlantic Ocean linked to variability in the strength of the AMOC. Subsequent analyses have not changed this picture, with more

models showing Atlantic multidecadal variability. Despite this, detailed agreement is lacking, with, for example, simulated timescales ranging from 40–60 years (Park and Latif, 2010; Frankcombe et al., 2010), to a century or more (Msadek and Frankignoul, 2009; Menary et al., 2011). Models also tend to lack 'convergence' i.e., a model with good AMO characteristics often does not retain these characteristics as the model is upgraded to the next version (Hurrell et al., 2010). This is consistent with analyses of shorter-period variability (Farneti and Vallis, 2011) which show a marked sensitivity to oceanic parameters and mean state. Recent modelling does confirm the link to the overturning circulation, but models tend to differ on the mechanism generating multidecadal Atlantic variability. These include: coupled atmosphere-ocean interactions in the far North Atlantic (Msadek and Frankignoul, 2009), water mass exchange with the Arctic (Frankcombe et al., 2010), and advected tropical salinity feedbacks (Menary et al., 2011) (Figure 9.35).

The presence of AMO-like variability in unforced simulations has been taken to indicate the AMO has an internal origin within the climate system. This is further supported by the forced 20th century simulations in the CMIP3 multi-model dataset, which generally do not possess the observed sequence of AMO phases, thus implying the AMO is not a result of common forcing (Kravtsov and Spannagle, 2008; Knight, 2009; Ting et al., 2009). The indirect effects of sulphate aerosol may have been on average too small in the CMIP3 models, however, leading to an underestimation of 20th century tropical Atlantic temperature trends (Chang et al., 2011). Moreover, a more recent 20th century ensemble created using a more sophisticated aerosol treatment than was typically used in CMIP3 goes considerably further in reproducing historical AMO fluctuations (Booth et al., 2011). This would suggest that at least part of the AMO may in fact be forced, and that improved aerosol representations are important in representing the AMO in models. Further evidence for this comes from the fact that changes in atmospheric loading of African dust may also be a strong driver of multidecadal temperature variability in the tropical Atlantic (Evan et al., 2009), and could act as a positive feedback on the AMO (Foltz and McPhaden, 2008).

[INSERT FIGURE 9.35 HERE]

Figure 9.35: From top to bottom: SST composites using AMOC time series; precipitation composites using cross-equatorial SST difference time series; equatorial salinity composites using ITCZ-strength time series; subpolar-gyre depth-averaged salinity (top 800–1,000 m) using equatorial salinity time series; subpolar gyre depth averaged density using subpolar gyre depth averaged salinity time series. From left to right: HadCM3, MPI-ESM, and KCM. Black outlining signifies areas statistically significant at the 5% level for a two-tailed t test using the moving-blocks bootstrapping technique (Wilks, 1995) (Figure 3 from Menary et al. (2011)).

9.5.3.3.3 Tropical zonal and meridional modes

Atlantic Meridional Mode (AMM)

The AMM is the dominant mode of internannual variability in the tropical Atlantic in all seasons except for boreal summer, when the Atlantic Niño becomes slightly more prominent. The AMM is characterized by an anomalous meridional shift in the inter-tropical convergence zone (ITCZ) that is caused by a warming (cooling) of SSTs and a weakening (strengthening) of the easterly trade winds in the northern (southern) tropical Atlantic (Chiang and Vimont, 2004). Variations in the AMM have been shown to be related to principal variations in hurricane tracks over the North Atlantic (Xie et al., 2005; Smirnov and Vimont, 2011). Virtually all CMIP models simulate an AMM-like SST variability represented by the 2nd EOF in their 20th century climate simulations. However, most models underestimate the SST variance associated with the AMM, and position the SST anomaly over the North Tropical Atlantic too far equatorward. More problematic is the fact that the development of the AMM in many models is led by a zonal mode during boreal winter – a feature that is not observed in nature (Breugem et al., 2006). This spurious AMM behavior in the models is likely to be associated with the severe model biases in simulating the ITCZ.

Atlantic Niño

CMIP3 models have considerable difficulty simulating Atlantic Niño in their 20th century climate simulations. For many models the so-called 'Atl-3' SST index (20°W–0°W, 3°S–3°N) displays the wrong seasonality, with the maximum value in either DJF or SON instead of JJA as in observations (Breugem et al., 2006). Of the two models that capture the observed seasonality, one severely over-estimates the Atl-3 SST variance, while the other severely underestimates it. The models' inability to capture the observed Atlantic Niño activity is likely caused by mean biases in the region. Almost all of the CMIP3 models fail to simulate some of the most fundamental features of the equatorial Atlantic Ocean – the east-west equatorial SST gradient and the eastward shoaling thermocline (e.g., Richter and Xie, 2008).

9.5.3.4 Indo-Pacific Modes

9.5.3.4.1 El Niño-Southern Oscillation

ENSO properties in AOGCMs

The El Niño-Southern Oscillation (ENSO) phenomenon is the dominant mode of natural climate variability in the tropical Pacific on seasonal to interannual time scales (Wang and Picaut, 2004 and Chapter 14). The representation of ENSO in AOGCMs has steadily improved and now bears considerable similarity to observed ENSO properties (AchutaRao and Sperber, 2002; Guilyardi et al., 2009b; Randall et al., 2007). However, simulations of both background climate (time mean and seasonal cycle) and natural variability still exhibit serious systematic errors (Capotondi et al., 2006; Guilyardi, 2006; van Oldenborgh et al., 2005; Wittenberg et al., 2006; Stevenson et al., 2011; Dufresne et al., 2011; Watanabe et al., 2011), many of which can be traced back to deep convection, trade wind strength and cloud feedbacks (Braconnot et al., 2007a; Guilyardi et al., 2009a; L'Ecuyer and Stephens, 2007; Lloyd et al., 2010; Lloyd et al., 2009; Sun et al., 2009). Some models have been identified that perform particularly well (e.g., GFDL2.1 in CMIP3 and CNRM CM5 in CMIP5, (Kakitha et al., 2011)).

AOGCMs produce a variety of El Niño variability time scales (Figure 9.36): model spectra range from very regular near-biennial oscillations to spectra that are close to the observed 2 to 7 years. The amplitude of El Niño in AOGCMs ranges from less than half to more than double the observed amplitude (AchutaRao and Sperber, 2006; Guilyardi, 2006; Guilyardi et al., 2009b; van Oldenborgh et al., 2005), with CMIP5 models showing slightly less inter-model variability than was evident in CMIP3 (Figure 9.37). The observed seasonal phase locking – El Niño and La Niña anomalies tend to peak in boreal winter and are weakest in boreal spring – is often not captured by models, which either show little seasonal modulation or a phase locking to the wrong part of the annual cycle, although some models do show a tendency to have ENSO peak in boreal winter (Kakitha et al., 2011). All these biases combine to generate errors in ENSO amplitude, period, irregularity, skewness or spatial patterns (Guilyardi et al., 2009b; Leloup et al., 2008). Ohba et al. (2010) separately investigate the simulated transition process of a warm-phase and a cold-phase ENSO in the CMIP3 models. Some of the models reproduce the features of the observed transition process of El Niño/La Niña, whereas most models fail to concurrently reproduce the process during both phases. A few recent studies suggest the existence of several types of ENSO (e.g., Central Pacific vs. East Pacific CP/EP). Even though the observational record is still insufficient to fully conclude (see Chapter 14), several studies attempted to detect this distinction in the CMIP3 multi-model ensemble with mixed results (Yeh et al., 2009; Ham and Kug, 2011; Guilyardi, 2006; Lengaigne and Vecchi, 2010).

[INSERT FIGURE 9.36 HERE]

Figure 9.36: [PLACEHOLDER FOR SECOND ORDER DRAFT: Figure from AR4; to be updated.] Maximum entropy power spectra of surface air temperature averaged over the NINO3 region (i.e., 5°N to 5°S, 150°W to 90°W) for (a) the CMIP3 models and (b) the CMIP2 models. Note the differing scales on the vertical axes and that ECMWF reanalysis in (b) refers to the European Centre for Medium Range Weather Forecasts (ECMWF) 15-year reanalysis (ERA15) as in (a). The vertical lines correspond to periods of two and seven years. The power spectra from the reanalyses and for SST from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) version 1.1 data set are given by the series of solid, dashed and dotted black curves. Adapted from AchutaRao and Sperber (2006) and Sperber.

[INSERT FIGURE 9.37 HERE]

Figure 9.37: ENSO metrics comparing CMIP3 and CMIP5 [PLACHOLDER FOR SECOND ORDER DRAFT: To be updated with more CMIP5 results.]

ENSO decadal variability

CMIP3 models display a wide range of skill in simulating the interdecadal variability of ENSO (Lin, 2007). The models can be categorized into three groups: those that show an oscillation with a constant period shorter than the observed ENSO period, and sometimes with a constant amplitude; those that do not produce statistically significant peaks in the ENSO frequency band, but usually produces one or two prominent peaks (episodes) at period longer than 6 years; and those that display significant interdecadal variability of ENSO in both amplitude and period. Among them, only the MPI model reproduces the observed eastward shift of the westerly anomalies in the low-frequency regime.

Given the short observational record (both for ENSO and for the external forcing fields) (Li et al., 2011b; Wittenberg, 2009; Deser et al., 2011b), the complexity and diversity of the paradigms and processes involved (Wang and Picaut, 2004), and the shortcoming of current state-of-the-art models (van Oldenborgh et al., 2005; Leloup et al., 2008; Guilyardi et al., 2009b), understanding and evaluating ENSO in models remains a considerable challenge.

9.5.3.4.2 Indian Ocean Dipole

The Indian Ocean zonal dipole mode (IOD) is an important source of variability (Saji et al., 1999; Webster et al., 1999) as is, at higher latitudes, the subtropical SST dipole mode (Behera and Yamagata, 2001) which appears to be part of a hemispheric response to tropical atmospheric forcing (Fauchereau et al., 2003; Hermes and Reason, 2005). Most CMIP3 AOGCMs are able to reproduce the general features of the IOD, including its phase lock onto the July-November season (Saji et al., 2006). The SST anomalies in models, however, tend to show too strong a westward extension along the equator in the eastern Indian Ocean. IOD amplitude varies a great deal among CMIP3 models, an inter-model variation that can be explained by the difference in the strength of the simulated Bjerknes feedback among zonal SST gradient, wind and thermocline depth on the equator (Lin et al., 2011).

Many models simulate the observed correlation between IOD and ENSO. This correlation varies substantially in value among models, and this inter-model variation is not tied to the amplitude of ENSO (Saji et al., 2006). Models that simulate a deeper thermocline off Sumatra also tend to show a larger correlation between indices of ENSO and the IOD than do models with a shallower thermocline. A subset of CMIP3 models show a spurious correlation with ENSO following the decay of ENSO events, instead of during the ENSO developing phase, possibly due to erroneous representation of oceanic pathways connecting the equatorial Pacific and Indian Oceans (Cai et al., 2011).

9.5.3.4.3 The Quasi-Biennial Oscillation (QBO)

Significant progress has been made in recent years to model and understand the impacts of the QBO (Baldwin et al., 2001). More models now reproduce a QBO in climate simulations. Some of these employ high vertical and horizontal resolution (Takahashi, 1999; Kawatani et al., 2011), while others use parameterised wave spectra to circumvent the need for such high resolution (Scaife et al., 2000; Giorgetta et al., 2002; McLandress, 2002). These model results are consistent with recent observations which confirm that small scale gravity waves carry a large proportion of the momentum flux which drives the QBO (Sato and Dunkerton, 1997; Ern and Preusse, 2009). Many features of the QBO such as its width and phase asymmetry also appear spontaneously in these simulations due to internal dynamics (Dunkerton, 1991; Scaife et al., 2002; Haynes, 2006). Some of the QBO effects on the extratropical climate (Holton and Tan, 1980; Hamilton, 1998; Naoe and Shibata, 2010) as well as ozone (Butchart et al., 2003; Shibata and Deushi, 2005) are also reproduced. Subsequent influences on the Arctic/North Atlantic Oscillation have also been suggested from observational and modelling studies (Thompson et al., 2002; Boer and Hamilton, 2008; Marshall and Scaife, 2009).

9.5.3.4.4 Pacific Decadal Variability (PDO)

The “Pacific Decadal Oscillation” (PDO) refers to the leading Empirical Orthogonal Function (EOF) of monthly Sea Surface Temperature (SST) anomalies over the North Pacific (north of 20°N) from which globally-averaged SST anomalies have been subtracted (Mantua et al., 1997). It exhibits anomalies of one sign along the west coast of North America, and of opposite sign over the western and central North Pacific. Although the PDO time series exhibits considerable decadal variability, it is difficult to ascertain whether there are any robust spectral peaks given the relatively short observational record (Minobe, 1997, 1999; Deser et al., 2004; Pierce, 2001). The ability of climate models to represent the PDO, in particular its spatial pattern, temporal characteristics, and association with the tropical Indo-Pacific has been assessed by Stoner et al. (2009). Their results indicate that approximately half of the CMIP3 models simulate a realistic spatial pattern and general temporal behaviour of the PDO (e.g., enhanced variance at low frequencies); however, spectral peaks are consistently higher in frequency than those suggested by the short observational record. The models’ PDO correlations with SST anomalies in the tropical Indo-Pacific are strongly underestimated by the CMIP3 models (Lienert et al., 2011; Deser et al., 2011a).

9.5.3.4.5 Tropical Ocean Decadal Variability

Pacific Subtropical Cells (STCs) are the shallow meridional cells in which water flows out of the tropics within the surface layer, subduct in the subtropics, flows equatorward in the thermocline and upwells in the equatorial ocean (Blanke and Raynaud, 1997; McCreary and Lu, 1994). The STCs provide a pathway by which extra-tropical atmospheric variability can force tropical variability. Observational studies have shown that these wind driven cells are major drivers of SST change in the tropical Pacific (McPhaden and Zhang, 2002), where a decrease in tropical Pacific SST is significantly correlated with a spin-up of the STCs and an increase in SST with a spin-down. Several studies have shown that this relationship is absent from the CMIP3 climate model simulations of the 19th-20th centuries (Zhang and McPhaden, 2006). Hence the full impact of a weakening of the Walker Circulation with climate change (Vecchi et al.; Vecchi et al., 2006b) may not be fully accounted for. (Solomon and Zhang, 2006) suggest that the CMIP3 coupled models may be reproducing the observed *local* ocean response to changes in forcing but inadequately reproduce the *remote* STC-forcing of the tropical Pacific due to the underestimate of extratropical winds that force these ocean circulations.

9.5.3.5 Teleconnections

In general terms, teleconnections characterize the response of the climate system in one location to forcings such as SST anomaly patterns in another. SST variability provides a significant forcing of atmospheric teleconnection response and drives a significant portion of the climate variability over land (Goddard and Mason, 2002; Shin et al., 2010). Although local forcings and feedbacks can play an important role (Pitman et al., 2010; Section 9.6.1), the simulation of land surface temperatures and precipitation requires accurate predictions of SST patterns (Compo and Sardeshmukh, 2009; Shin et al., 2010). As examples, Goddard et al. (2009) explored the effect of observational uncertainty in the SST patterns on regional climate change. Shin et al. (2010) examined the difference between CMIP3 modelled SST trends and observations, and found that the skill of regional predictions over land could be improved with improved skill in tropical SST patterns. These results imply that the uncertainty in the atmospheric teleconnection response to idealized forcings can be used to assess atmospheric models (Barsugli et al., 2006; Shin et al., 2010; Li et al., 2011c). Specific modes of variability typically driven by SST patterns are discussed in more detail in the following.

9.5.3.5.1 Pacific North American Pattern

The majority of CMIP3 models simulate a realistic spatial structure of the PNA pattern in wintertime (Stoner et al., 2009). The temporal variability of the PNA generally resembles a white-noise process with an e-folding time scale of about 10 days (Johnson and Feldstein, 2010), and some year-to-year autocorrelation reflecting influences of ENSO and the PDO (Deser et al., 2004). This temporal behavior is generally captured by the CMIP3 models, although the level of year-to-year autocorrelation varies according to the strength of the simulated ENSO and PDO (Stoner et al., 2009).

9.5.3.5.2 Pacific South America Pattern

The Pacific South America (PSA) pattern is one of the major atmospheric teleconnection patterns from the tropical Indo-Pacific into the mid- and high latitude South America and South Atlantic region (Colberg et al., 2004; Mo and White, 1985). It also appears to act as a link between ENSO and the generation of subtropical basin modes in the Southern Hemisphere (Hermes and Reason, 2005). (Vera and Silvestri, 2009) showed that there was a considerable range in the ability of CMIP3 models to represent these wave trains. Furthermore, different models had different abilities depending on the season.

9.5.3.5.3 ENSO – West African Monsoon (WAM)

A regression of the WAM precipitation index with global SSTs reveal two major teleconnections (Fontaine and Janicot, 1996). The first mode highlights the strong influence of ENSO. The second mode reveals a relationship between the SST in the Gulf of Guinea and the northward migration of the monsoon rainbelt over the West African continent. Most CMIP3 20th century simulations show a single dominant Pacific teleconnection, which is, however, of the wrong sign for half of the models (Joly et al., 2007a). Only one model shows a significant second mode, emphasizing the GCMs' difficulty in simulating the response of the African rainbelt to Atlantic SST anomalies that are not synchronous with Pacific anomalies.

9.5.3.5.4 *Mid-latitude jet position*

Westerly jet streams occur in both hemispheres and are associated with storm tracks that can play a dominant role in regional climate variability and change. Early studies noted an increase in the strength of the westerly jets and a poleward shift in their location under increasing greenhouse gases due to the increase in meridional temperature gradient between the rapidly warming upper troposphere in the tropics and cooling lower stratosphere in the extratropics. Indeed, observations suggest that recent decades saw a large poleward shift in the subtropical jets associated with widening of the tropics (Seidel et al., 2008). However, more recently, it has been noted that the degree of poleward shifting of the jets may be systematically affected by model error (Kidston and Gerber, 2010) while higher vertical resolution models indicate that the previous consistency between models may still not indicate robust response due to shared limitations in resolution (Morgenstern et al., 2010b; Scaife et al., 2011). In the southern hemisphere, a similar response of a poleward shift in the southern hemisphere jet and associated stormtracks has been observed in recent decades.

9.5.3.6 *Carbon Cycle Variability*

9.5.3.6.1 *Interannual variability in terrestrial sources and sinks of carbon*

Both coupled biogeochemistry/land models evaluated by Randerson et al (2009) reproduce the interannual variability in land fluxes during 1988–2004 when compared against Atmospheric Tracer Transport Model Intercomparison Project (TRANSCOM). The models are significantly and positively correlated with the time series of annual-mean fluxes and explain between 43% and 53% of the fluctuations in TRANSCOM. The models produce year-to-year variability that agrees to within 30% with the interannual standard deviation from TRANSCOM of 1.0 PgC yr^{-1}. Over the longer time period spanning 1860 to 2002, the inclusion of nitrogen cycling and deposition on global carbon sequestration accounts for less than 20% of recent changes in annual NPP due to atmospheric composition and climate (Zaehle et al., 2010b).

When these components are linked to fully coupled Earth system models, these models tend to overestimate the long-term trend in global-mean atmospheric CO_2 concentrations (Cadule et al., 2010). The simulation of various types of interannual variability, including the oscillations in CO_2 associated with the positive and negative phases of ENSO and long-term trends in the seasonal amplitude, is moderately skilful although some models examined by Cadule et al. (2010) exhibit essentially no skill on this metric.

9.5.3.6.2 *Internal and interannual variability in ocean sources and sinks of carbon*

Most of the ocean biogeochemical GCMs that have been compared against depth-integrated primary productivity (PP) underestimate the observed interannual variance in PP with discrepancies frequently exceeding a factor of two (Friedrichs et al., 2009). The majority of this error is contributed by the infrequent occurrence of low PP values (<0.2 g C m^{-2} d^{-1}) in the models relative to the observations. Pattern correlations between temporal anomalies in PP estimated from satellite ocean-colour data and from the GCMs are generally in the range of 0.5 to 0.6. On a global scale, temporal variability in PP is contributed primarily by tropical oceans where stronger stratification and higher SSTs lead yield negative PP anomalies (Schneider et al., 2008). In analysis of a limited sample of biogeochemical GCMs, Schneider et al (2008) identify only one model that emulates the inverse relationship between PP and SST inferred from satellite ocean-colour data over low-latitude oceans. Reproduction of this inverse relationship is a necessary but not sufficient condition for projecting the effects of more El Nino-like conditions with higher SSTs and stronger stratification on the uptake of CO_2 by low-latitude oceans.

9.5.3.7 *Summary*

In summary, the assessment of interannual to interdecadal variability in climate models presents a varied picture. This is either due to the short reliable observational record available to assess the models or to the diversity of model behavior, or to both. There is medium evidence and medium agreement that the NAO and Northern Annual Mode is well simulated, and low evidence and medium agreement that the Southern Annual Mode is simulated with mixed skill. There is low evidence and medium agreement that the AMOC variability and the Atlantic Multi-decadal Variability are simulated with mixed skill. There is medium evidence and high agreement that the Atlantic Meridional Mode is not correctly simulated and low evidence and medium agreement that the Atlantic Niño is not correctly simulated. There is medium evidence and high agreement that ENSO is simulated with mixed skill and medium evidence and medium agreement that the Indian Ocean Dipole is simulated with mixed skill. There is medium evidence and medium agreement that

the QBO is well simulated by the models. Finally, there is low evidence and medium agreement that the Pacific Decadal Oscillation is simulated with mixed skill, low evidence and low agreement that the Pacific North American pattern is well simulated and that the Pacific South American pattern is simulated with mixed skill.

9.5.4 Extreme Events

Extreme events are realizations of the tail of probability distribution functions of natural variability of climate or weather. They are higher-order statistics and thus generally expected to be more difficult to realistically represent in climate models. Shorter time scale extreme events are often associated with smaller scale spatial structure, which cannot be captured by coarse resolution models but better represented as the resolution of a model increases. In AR4, it was concluded that the models could simulate the statistics of extreme events better than expected from generally coarse resolutions of the models at that time, especially for temperature extremes (Randall et al., 2007).

The IPCC has conducted an assessment of extreme events in the context of climate change -- the Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) (IPCC, 2012). Although there is no comprehensive climate model evaluation with respect to extreme events in SREX, there is some consideration of model performance taken into account in assessing uncertainties in projections.

9.5.4.1 Extreme Temperature

Since AR4, the evaluation of CMIP3 models in terms of extreme events has been extended. Kharin et al. (2007) have comprehensively evaluated the ability of models to reproduce 20-year return values of annual extremes of near-surface temperature and daily precipitation amounts. They found that the CMIP3 models simulated present-day warm extremes reasonably well on the global scale, as compared to estimates from reanalysis. The model discrepancies in simulating cold extremes were found to be generally larger than those for warm extremes, especially in sea ice-covered areas.

Some studies have compared modelled and observed historical trends of temperature extremes. As noted in SREX (Seneviratne et al., 2012), recent detection and attribution studies have shown that models tend to overestimate the observed warming of warm temperature extremes and underestimate the warming of cold extremes in the second half of 20th century (Christidis et al., 2011; Zwiers et al., 2011) [See also Chapter 10]. Moreover, Meehl et al. (2009a) found that a particular climate model (CCSM3) overestimated the ratio of daily record high maximum temperatures to record low minimum temperatures averaged across the USA compared with the observed value, implying that the model tends to overestimate the increase in record high temperatures, which is consistent with the bias suggested by detection and attribution studies.

As a preliminary overview of results from new generation models, Figure 9.38 shows relative error estimates based on the approach by (Gleckler et al., 2008) of available CMIP5 models for various extreme indices over land, which shows that the performance of simulated extreme events, in terms of temporal mean magnitude and spatial distribution, varies with models and indices.

[INSERT FIGURE 9.38 HERE]

Figure 9.38: Portrait diagram display (Gleckler, Taylor, & Doutriaux, 2008) of relative error metrics for the CMIP5 temperature and precipitation indices compared to ERA40 (top) and NCEP (bottom) re-analyses for the base period 1961–1990. Only land areas are considered. Top row in each diagram indicates the mean RMSE across all indices for a particular model. The indices marked with a * are bias-corrected (indices are calculated using the bias-corrected minimum and maximum temperature time series but using the temperature thresholds estimated from the re-analysis, so that the bias in simulated temperature variability can be assessed). Indices shown are 'Warm/Cold spell duration' (twsd/tcsd), 'Warm/Cool days' (tx90p/tx10p), 'Warm/Cool nights' (tn90p/tn10p), 'Min/Max T2MAX' (txmin/txmax), 'Min/Max T2MIN' (tnmin/tnmax), 'Ice/Summer/Frost days' (txid/txsu/tnfd), 'Tropical nights' (tn20), 'Growing season length' (tgsl), 'Diurnal temperature range' (tdtr), 'Consecutive wet/dry days' (pxcwd/pxcdd), 'Max 5-day/1-day precipitation amount' (px5d/px1d), 'Simple daily intensity index' (psdii), 'Annual total wet-day precipitation' (prtot), 'Extremely/Very wet days' (pr99p/pr95p) and 'Number of very heavy/heavy precipitation days' (p20mm/p10mm) from Klein Tank et al. (2009). (Haylock et al., 2008).

9.5.4.2 Extreme Precipitation

Kharin et al. (2007) concluded that precipitation extremes simulated by CMIP3 models were plausible in the extratropics, but uncertainties in the Tropics were very large, both in the models and the available observational datasets. Simulated precipitation extremes are known to be highly resolution dependent. Growing evidence has shown that high-resolution models (50 km or finer atmospheric horizontal resolution) can capture the magnitude of extreme precipitation realistically (Sakamoto et al., 2011; Wehner et al., 2010). See also Section 9.7.3.3.

Since the AR4, more attention has been devoted to a higher temporal resolution precipitation extremes, e.g., hourly precipitation extremes, rather than daily. Lenderink and Meijgaard (2008) have analysed a 99-year record of hourly precipitation observations from a station in the Netherlands and found that the temperature dependence of hourly precipitation intensity was greater than that expected from the Clausius-Clapeyron relation in higher temperature regimes, although the cause was not clear (Haerter and Berg, 2009; Lenderink and van Meijgaard, 2009). This feature was partly (only for very high extremes) reproduced by a 25 km resolution regional climate model.

Similar to temperature extremes, studies comparing modelled and observed historical trends of precipitation extremes have found that the observed intensification of extreme precipitation over Northern Hemisphere land areas in the second half of the 20th century was consistently simulated by climate models, but the simulated amplitude of intensification was smaller than observed (Min et al., 2011) (See also Chapter 10). Related to this, it has been pointed out from comparisons between CMIP3 models and satellite-based dataset of precipitation intensity that models underestimate the temperature dependence of precipitation intensity over the tropical ocean (Allan and Soden, 2008) and at a global scale (Liu et al., 2009). These studies consistently suggest that CMIP3 models tend to underestimate the intensification of extreme precipitation with temperature increase.

9.5.4.3 Other Extremes

With regard to tropical cyclones, it was concluded in the AR4 that high-resolution AGCMs produced generally good simulation of their frequency and distribution, but underestimated their intensity (Randall et al., 2007). Since then, Mizuta et al. (2011) have shown that a newer version of MRI-AGCM with improved parameterizations but with the same resolution as the previous version (20 km) simulates tropical cyclones as intense as those observed with improved distribution as well. This implies that 20 km atmospheric resolution might be enough for simulating realistic intensity of tropical cyclones if adequate parameterizations are used. Another remarkable finding since AR4 is that the observed year-to-year counts of Atlantic hurricanes can be well simulated by AGCMs driven only by observed sea surface temperature (Larow et al., 2008; Zhao et al., 2009). This finding is particularly important for physical understanding, detection, attribution and future projection of hurricane count changes, but also encouraging in the context of model evaluation.

One of the important extreme events at longer timescale (months or longer) is drought, which is caused by variability of both precipitation and evaporation. Sheffield and Wood (2008) found that AOGCMs in the CMIP3 ensemble simulated large-scale droughts in 20th century reasonably well, although the frequency of long-term (more than 12-month duration) droughts were overestimated.

9.5.4.4 Summary

There is limited evidence but high agreement that models in CMIP3 ensemble simulate 20-year extreme temperatures reasonably well, but tend to overestimate the observed warming of warm temperature extremes, underestimate the observed warming of cold temperature extremes and underestimate the observed intensification of extreme precipitation. The evidence is growing, though still rather limited, and agreement is high that AGCMs driven by observed sea surface temperature can simulate the observed year-to-year counts of Atlantic hurricanes.

9.6 Downscaling and Simulation of Regional-Scale Climate

In the above sections, climate model evaluation is discussed in terms of the simulation of different components of the climate system as well as climate averages, variability and extremes. This section addresses climate model evaluation for geographical regions, which both complements the previous sections and provides the basis for the use of regional-scale climate projections in climate change impact research.

Regional climate scale information can be extracted from the AOGCMs. However, many important features concern finer scales than those that are resolved well in the AOGCMs. This has led to the development and use of different regional downscaling methods. The most notable of such methodologies are statistical downscaling and dynamical downscaling (a.k.a. regional climate modelling). Use is also made of variable-resolution AOGCMs and high resolution AGCMs.

9.6.1 Regional-Scale Simulation by AOGCMs

In general, the performance of AOGCMs on continental and sub-continental scales varies with region, metric and climate aspect (e.g., Perkins et al., 2007; Overland et al., 2011; Xu et al., 2010). Van Haren et al. (2011 (submitted)) showed that both the CMIP3 models and RCMs forced with these AOGCMs largely fail to capture recent precipitation trends in Europe. The authors concluded that the cause was biases in the CMIP3 models in atmospheric circulation in winter and SST in summer. In contrast the same RCMs, when forced with reanalyses, fared much better. Similar results have been reported earlier by e.g., van Ulden and van Oldenborgh (2006). Räisänen (2007) noted that AOGCMs do better for temperature than precipitation and MSLP in capturing trends over the past 50 years.

An example of evaluation of a regional feature from AOGCMs is net precipitation (P-E) over Antarctica. It is is an important component of sea level rise in that water storage on the ice sheet effectively withdraws water from the ocean and hence lowers sea level (discussed in more detail in Chapter 13). Raphael and Holland (2006) found that CMIP3 models underestimate moisture transport into the high-latitudes which is hard to reconcile with the realistic simulation of P-E. Detailed analyses of 15 CMIP3 models (Uotila et al., 2007) suggest that AOGCMs generally reproduce late 20th century Antarctic P-E quite well, with a multi-model mean P-E of 184 mm yr^{-1} over the period 1979–2000, as compared to observationally-based range of 150–190 mm yr^{-1}. The same analyses showed that a 5-model subset of the CMIP3 models chosen on the basis of their ability to reproduce the observed near surface circulation over Antarctica, yielded a mean P-E of 171 mm/yr. The multi-model ensemble indicates a small increase (<1%) in P-E over the period 1979–2000, but individual models simulate either increases or decreases (Uotila et al., 2007). It is likely that current AOGCMs are able to simulate realistic P-E over Antarctica, but further analysis of newer model results will be required to determine if this is a robust result.

The question of which scales AOGCMs should be expected to model skillfully remains a timely one (Masson and Knutti, 2011; Räisänen and Ylhäisi, 2011) Appropriate smoothing scales were found to vary from the grid-point scale to around 2000 km, depending on the variable, and region in question. Pitman et al. (2010) argued that the lack of representation of phenomena and processes that have characteristic regional distribution limits the regional-scale performance of AOGCMs. Examples are fire, aspects of land cover change and vegetation's emissions of biogenic VOCs (BVOC).

Since AR4, the resolution of AOGCMs has generally improved (Section 9.1.3.4), but typically still limits representation of many sub-continental and regional climate features. The overall performance of the AOGCM simulations of regional-scale temperature and precipitation is illustrated in Figures 9.2 and 9.4. The general cold bias is now smaller in the CMIP5 models than in the CMIP3 MME. Average RMSEs are still quite similar, and the largest biases persist in high latitude regions, the eastern parts of the tropical basins and some of the major mountaineous regions. A more detailed comparison of the annual cycle of temperature and precipitation for different regions, simulated by CMIP5 models, is provided in Figure 9.39.

[INSERT FIGURE 9.39 HERE]

Figure 9.39: Mean annual cycle of temperature (a) and precipitation (b) from CMIP5 GCM (dotted lines) historical runs and CRU (solid lines) data for the indicated areas. Average is taken over land points over the period 1979 to 2005. Units are mm/day for precipitation and °C for temperature.

9.6.2 Regional Climate Downscaling

9.6.2.1 Recent Developments of Statistical Methods

Statistical or empirical statistical downscaling (SD) techniques involve deriving empirical relationships linking large-scale atmospheric variables (predictors) and local/regional climate variables (predictands), which then can be applied to equivalent predictors from AOGCM simulations. The availability of long observational data sets can be a limiting factor in regions with sparse observations. The development of SD since the AR4 has been quite vigourous and has been reviewed by Fowler et al. (2007) and Maraun et al. (2010b).

Many state-of-the-art approaches now combine different approaches (van Vliet, 2011; Vrac and Naveau, 2008). An increasing number of studies focus on extremes, particulary precipitation and use methods based on extreme value theory (e.g., Wang and Zhang, 2008; Vrac and Naveau, 2008). SD methods have also been used to project statistical attributes of the predictands, instead of raw values (e.g., Tolika et al., 2008). Techniques have also been developed to consider multiple climatic variables simultaneously, in order to preserve some physical consistency and/or preserve sub-grid correlations and variability (e.g., Zhang and Georgakakos, 2011). Another development is the application of SD methods to RCMs (e.g., Boé et al., 2007; Déqué, 2007; Paeth, 2011; Quintana Seguí et al., 2010; van Vliet et al., 2011). A growing number of studies address model-structure uncertainties (e.g., Fowler et al., 2007; Najac et al., 2009; Prudhomme and Davies, 2009; Burton et al., 2010; Lewis and Lamoureux, 2010; Maraun et al., 2010b). Finally, SD methods have been extended to new variables such as snowpack (e.g., Teutschbein and Wetterhall, 2011), wind (e.g., Najac et al., 2009; Goubanova et al., 2011), pan-evaporation (e.g., Chu et al., 2010) and hurricanes (Emanuel et al., 2008).

9.6.2.2 Recent Developments of Dynamical Methods

Since the AR4, the development of RCMs has included an increase in the number of regions studied, longer simulations, higher model resolution, coordinated experiments as well as extension of model evaluation to additional processes (Rummukainen, 2010; Wang et al., 2004). This has facilitated the evaluation of simulated variability, trends and extremes. For example, Yhang and Hong (2008) tested improvements of land surface, boundary layer and cumulus parameterisation schemes in a specific RCM, which helped to improve the simulation of the East Asian summer monsoon. Dorn et al. (2009) showed that, in a coupled RCM for the Arctic, substantial improvements in sea ice simulation could be obtained by implementation of improved ice growth, ice albedo and snow cover descriptions. Samuelsson et al. (2010) showed that by including a lake model in an RCM, the effect of the collective presence of the lakes in Europe on the air temperature over adjacent land could be captured. There are now also more coupled regional climate models with interactive ocean and sea ice (Dorn et al., 2009; Doscher et al., 2010; Smith et al., 2011a; Somot et al., 2008; Smith et al., 2011a).

9.6.3 Fidelity of Downscaling Methods and Value Added

9.6.3.1 Assessment of Skill

Studies that compare different SD techniques and/or SD and dynamical downscaling approaches (e.g., Schmidli et al. (2007), Maurer and Hidalgo (2008) suggest that the downscaling skill depends on the location, the AOGCM used as RCM boundary conditions or as SD predictors, season, parameter, etc. When it comes to statistical downscaling, the statistical stationarity hypothesis (do the statistical relationship inferred from historical data remain valid under future climate change?) remains an assumption that cannot be directly tested. Some recent studies have proposed ways to examine its validity using RCM outputs (e.g., Vrac et al., 2008; Driouech et al., 2010) or using long series of observations (e.g., Schmith, 2008). For the time being, the validity of the stationarity hypothesis remains an unsettled issue.

Comparison of modelled and observed climatological statistics has a long history as an evaluation method of for downscaling techniques. Examples of more recent metrics for statistical downscaling evaluation relate to intensities (e.g., Ning et al., 2011; Tryhorn and DeGaetano, 2011), to temporal behaviour (e.g.,Brands et al., 2011; May, 2007; Timbal and Jones, 2008), Maraun et al. (2010a), and to physical processes (Lenderink and

Meijgaard (2008), Maraun et al. (2010b)). In addition to climate variables as such, SD capabilities are increasingly being examined through secondary variables like river discharge and stream flow, which pertains to coherency between variables and/or their spatial autocorrelation (e.g., Boé et al. (2007), Teutschbein and Wetterhall (2011)). In the case of RCMs, process-oriented evaluation, i.e., model skill at simulating processes (e.g., fluxes rather than state variables) has become more common. For example, Sasaki and Kurihara (2008) examined the ability of an RCM to reproduce observed relationships between precipitation and elevation. Driouech et al. (2010) showed that a specific variable-resolution AGCM reproduced rather well the observed link between north Atlantic weather regimes and local precipitation. Döscher et al. (2010) found that a coupled RCM reproduced empirical relationships between Arctic sea ice extent and sea ice thickness characteristics and NAO variations. Hirschi et al. (2011a) found that a number of RCMs reproduce observed relationships between soil moisture and extreme temperature, albeit with some overestimation.

There is considerable evidence that the representation of the land surface and land-atmosphere coupling is an area that continues to require attention. In some regions, RCMs tend to dry out the soil too effectively at high temperatures, suggesting that projected changes in warm summertime conditions including heat waves may be subject to systematic biases that change with increasing average warming (Christensen et al., 2008; Kostopoulou et al., 2009). Both RCMs and GCMs exhibit a similar bias for present-day conditions for regions characterised by pronounced summer time drying. This is illustrated in Figure 9.40 for the Mediterranean region based on the RCMs participating in ENSEMBLES and the GCMs in CMIP3. Many models have a tendency towards an enhanced warm bias in the warmer months of the year. In terms of climate change projections, part of the warming signal in these regions could be due to model bias instead of being a part of the real effect (Boberg and Christensen, 2011).

Land-surface and atmosphere coupling is particularly important for monsoon regions (see also van den Hurk and van Meijgaard, 2010), for example the West African monsoon region (e.g., Boone et al., 2010; Druyan et al., 2010). These studies found that different RCMs display biases in their simulated latent heat fluxes and precipitation as well as a general tendency to place the monsoon too far to the north and to have it mistimed. (Lucas-Picher et al., 2011) explored the skill of four RCMs in simulating the Indian Monsoon and found that these reanalyses-forced RCMs, tended to exhibit a warm bias in the north and correspondingly biases in MSLP. The representation of the monsoon was in turn affected by the biases in the regional temperature and pressure gradients. Whereas monsoon onset was generally captured, its termination was not. The regional distribution of precipitation also varied between the RCMs in question.

[INSERT FIGURE 9.40 HERE]

Figure 9.40: Ranked modelled versus observed monthly mean temperature for a Mediterranean subregion, for the 1961–2000 period. The RCM data (panel a) are from Christensen et al. (2008) and are adjusted to get a zero mean in model temperature with respect to the diagonal. The GCM data in panel b are from CMIP3 and adjusted to get a zero mean in model temperature with respect to the diagonal. Figure after Boberg and Christensen (2011).

A fundamental issue is how evaluation findings on the performance of a downscaling method translate into skill in downscaling future climate change projections. Buser et al. (2009) looked at the effect of assumptions regarding biases in RCMs. They found that the projected summertime warming in the European Alpine region, obtained from a combination of several RCMs, came out as 3.4°C and 5.4°C for assumptions of constant bias and constant relationship, respectively. Projected changes for the wintertime warming were not similarly sensitive. Buser et al. (2010) found, however, that the consideration of these two bias assumptions together, leads to overall lower projection of summer and autumn warming, but larger winter warming in different parts of Europe, as simulated by several RCMs. They traced the reason for this back to deficiencies in how the RCMs' reproduce interannual temperature variability in the different seasons.

Even though RCMs are constrained by their boundary conditions, they still generate internal variability. This varies with the synoptic situation, season, and choice of model domain (Alexandru et al., 2007; Leduc and Laprise, 2009; Nikiema and Laprise, 2010; Rapaić et al., 2010; Xue et al., 2007). In general, when the regional model boundary is close to the area of interest, the downscaling is strongly constrained by the driving boundary data. For example, in a so-called Big-brother experiment, Leduc et al. (2011) showed that the variance of the small-scale transient eddies of wind can be underestimated in smaller domains and at upper levels over North America, the effect being larger in winter than in summer (Leduc and Laprise,

2009). While the smaller scale features of precipitation improved in larger domains, the time correlation of precipitation decreased. Internal variability may lead to an RCM showing deviations from the corresponding driving data for specific situations and short periods, but these effects are generally found to be small in climate applications (Laprise et al., 2008; Separovic et al., 2008). Internal variability can be constrained by spectral nudging or other techniques (Misra, 2007); however, it is not well-established that such nudging is needed (Veljovic et al., 2010). Spectral nudging may also lead to deterioration of some features, e.g., precipitation extremes (Alexandru et al., 2009). Nevertheless, RCMs do simulate small-scale climate features that are absent in the lateral boundary conditions. Indeed, due to their more detailed representation of scales and processes, an RCM may have effects that travel up-scale and improve aspects of the larger-scale features, as shown by e.g., Lorenz and Jacob (2005) and Inatsu and Kimoto (2009), but may also cause some degradation (Castro et al., 2005; Laprise et al., 2008). These kinds of findings can provide further insights for use in the development of high-resolution global models.

There has been general development of evaluation methodologies for downscaling, and this has provided more insight into the fidelity of downscaling methods in capturing features and processes. The skill of regional climate downscaling depends both on the downscaling method itself and the quality of the driving data. A key emerging finding (medium agreement, limited evidence) is that some biases that are evident in model evaluation may change with time and/or climate regime, which is important to consider in the interpretation of projections.

9.6.3.2 Value Added

A downscaling method should provide useful additional information compared to the driving AOGCM (Laprise et al., 2008). While not guaranteed, studies do rather consistently indicate that added value arises, and this is most obvious for regions with variable topography and fine structure in land-sea distribution, lake area, etc. Some specific examples of added value by downscaling are improved simulation of convective precipitation (Rauscher et al., 2010), near-surface temperature (Feser, 2006), near-surface temperature and wind (Kanamaru and Kanamitsu, 2007) and maximum daily precipitation (Kanada et al., 2008). Winterfeldt and Weisse (2009) found that coastal wind characteristics were improved by dynamical downscaling and the same was found for European storm damage estimates in Donat et al. (2010). Fox-Rabinovitz et al. (2008) found that global simulations with a stretched grid providing higher resolution for a specific part of the globe had reduced errors compared to uniform lower resolution. Déqué et al. (2010) found that the simulated mean climate over the high-resolution portion of a stretched grid model was similar to that produced by a global model with the same high resolution everywhere, reinforcing the notion that this kind of downscaling can provide information of comparable quality to a much more computationally expensive global high-resolution model. However, there is not a one-to-one relationship between increased resolution and model fidelity. For example, Woollings et al. (2010c) investigated the effect of different spatial and temporal resolution of Atlantic SST used as boundary conditions in an RCM. They found that a higher spatial resolution improved the simulation Atlantic storm tracks, but that a higher temporal resolution led to some degradation. Walther et al. (2011) showed how increasing resolution clearly reduces biases in precipitation, perhaps because of interplay between the resolution and convection parameterisation. Figure 9.41 provides an illustration how the geographical patterns of RCM-simulated precipitation have been found to improve with resolution. In Walther et al. (2011), the simulated peak timing and amplitude of the diurnal precipitation cycle as well as the frequency of light precipitation improved more when going from 12km to the 6km resolution, that when going from 50 to 25km or from 25 to 12km. Rojas (2006) similarly found a non-linear relationship between simulation quality and resolution. In the latter case, however, there was more improvement when increasing RCM resolution from 135km to 45km that when going from 45km to 15km. This was despite the highly variable regional topography in the domain considered.

[INSERT FIGURE 9.41 HERE]

Figure 9.41: Summer seasonal mean (JJA, 1987–2008) in Southern Norway gridded observational precipitation with 1 km resolution from Met.no and RCM-simulated precipitation with boundary conditions from the ERA40 reanalysis and ECMWF operational analysis (top row). The RCM has been run at four different resolutions ranging from 50 to 6 km. Differences between the simulated precipitation and the gridded observations aggregated from 1 km to respectively 50, 25, 12 and 6 km grids are shown in the bottom row. The model runs are from Walther et al. (2011).

More formalised approaches to quantify aspects of added value have been attempted, for example by spatial filtering and more spatially-explicit skill measures (Feser, 2006; Kanamitsu and DeHaan, 2011).

Performance-based ranking of RCMs has also been explored. Christensen et al. (2010) suggested that this could apply to the ability of RCMs to simulate extremes, mesoscale features, trends, aspects of variability and consistency with the driving boundary conditions. Some of their metrics led to striking differentiation among RCMs (Lenderink, 2010), whereas others did not. The latter may imply general skilfulness of models, but also be an artefact of the choice of the metric. Nevertheless, Coppola et al. (2010) and Kjellström et al. (2010) demonstrated that weighted sets of RCMs outperform sets without weighting in terms of bias and RMSE of temperature and precipitation, although the extent to which this is a general result remains to be investigated.

The overall finding from many specific studies is that downscaling does add value, especially for regions with highly-variable topography and for small-scale phenomena. The evidence comes from a variety of distinct studies, rather than some larger set of coordinated experiments, so there is high agreement, but medium evidence.

9.7 Sources of Model Errors and Uncertainty

9.7.1 Approach to Linking Process Understanding and Model Performance

The previous Sections have dealt with the ability of climate models to simulate recent and longer-term records, variability and extremes, and regional-scale climate. We have assessed this capability both by comparing model solutions against observations and by evaluating inter-model spread, the latter being a minimum-level estimate for model uncertainty. The current Section 9.7 assesses *why* models show errors and spread. This identification is crucial not only for understanding why models fail to reproduce observations, but also for diagnosing whether models obtain the right answer for the right reason.

Error in model results can be conceptually subdivided into “modelling error”, caused by the difference between model formulation and physical process, and “approximation error”, caused by the difference between true model solution and numerical approximation (Oden and Prudhomme, 2002). No general framework exists for diagnosing modelling error. In contrast, for approximation error in geophysical fluid dynamics a general framework has just been formulated (Rauser et al., 2011), but application has so far been restricted to a shallow-water model. When we assess the causes of errors in *current* climate models, we thus cannot build on a general conceptual framework and must instead rely on more ad-hoc approaches, governed by practicality.

Since the AR4 significant progress has been made in understanding climate model error and spread. Section 9.7.2 assesses process-oriented evaluation, meaning that not only the end result is of interest (for example, change in climate sensitivity caused by a change in cloud parameterisation) but the entire causal process chain. Section 9.7.3 considers targeted numerical experiments, devoted in to: the application of climate models in weather-forecasting mode (“Transpose AMIP”, Section 9.7.3.1), useful because some important model errors manifest themselves within days; the simulation of key periods in the past, useful because models are applied in configurations for which they have not been tuned (Section 9.7.3.2); the effect of high spatial resolution, important because more of the processes are shifted from where we are uncertain about representation (unresolved) to where we know the underlying equations (resolved), thus reducing “modelling error” (Section 9.7.3.3); and perturbed-physics ensembles in which uncertain model parameters are varied systematically (Section 9.7.3.4). Section 9.7.4 assesses the rich literature since AR4 devoted to understanding the feedbacks causing spread in climate sensitivity, which we expect still to be a large source of spread in projected climate change. Section 9.7 ends with a synthesis, linking the lessons learned from the process-based analysis assessed here to the model errors assessed earlier.

9.7.2 Process Oriented Evaluation

Our focus in previous Sections has been on the comparison of observed and simulated climate variables at the global scale. These routine tests provide a valuable summary of overall model performance but their broad scope does not isolate specific processes or feedbacks believed to be of importance for realistic simulation of climate change.

Process-oriented evaluation is often applied over limited areas to focus on particular processes or phenomena (e.g., monsoons, deep convection) and the objective is to quantify how well a specific process, feedback, or phenomenon is represented in a model. In contrast to overall evaluation, process-oriented evaluation is based on the understanding of the individual processes or mechanisms involved, hence ensuring the phenomena are correctly represented for the right reasons, and not via error compensation. Moreover, it is sometimes possible to relate the ability of a model to represent a process in the current climate and the credibility of the same model to project the change in the process in the future climate (Boe et al., 2009a; Boe et al., 2009b; Eyring et al., 2007; Hall and Qu, 2006). This feature of process-oriented evaluation is crucial for understanding model spread in future projections and constraining future projections using observed data. See Section 9.8 for more discussion.

Isolating the processes involved in representing ENSO variability is one example of this kind of evaluation. Jin et al. (2006) and Kim and Jin (2010b) performed a linear-feedback analysis on the SST equation and identified five different feedbacks affecting the Bjerknes (or BJ) index, which in turn characterizes ENSO stability. Kim and Jin (2010a) applied this process-based analysis of ENSO to the CMIP3 multi-model ensemble and demonstrated a significant positive correlation between ENSO amplitude and the BJ index. When respective components of the BJ index obtained from the coupled models were compared with those from observations, it was revealed that most coupled models underestimated the negative thermal damping feedback and the positive zonal advective and thermocline feedbacks.

Process oriented evaluation is also useful for evaluating the ability of a model to simulate other regional phenomena (e.g., Inoue and Ueda, 2009; Nishii et al., 2009; Yokoi et al., 2009a). For example, Nishii et al. (2009) have established a relationship between the seasonal march of storm-track activity over the Far East and the occurrence of the first spring storm with strong southerly winds over Japan. They evaluated the ability of each model in the CMIP3 ensemble to simulate the particular seasonal march of the storm-track activity.

Evaluation of chemistry-climate simulations has also made use of process-based approaches. In contrast to most of the previous studies that applied performance metrics, the focus of the SPARC-CCMVal (2010) report was, as in Waugh and Eyring (2008), on quantitatively evaluating important processes rather than the quantity of interest itself (in this case was stratospheric ozone). This is a key aspect of the CCMVal model evaluation concept which aims to identify the sources of model errors and to avoid cases where an ozone performance metric may look good because of compensating errors in the underlying processes (Eyring et al., 2005). Chemical and radiative processes in the CCMs were assessed in the SPARC CCMVal report, and the upper troposphere / lower stratosphere (UTLS) has been subsequently examined (Gettelman et al., 2010; Hegglin et al., 2010). The identification of model deficiencies through a process-oriented evaluation has led to quantifiable improvements in particular models from the first to the second round of CCMVal (e.g., transport, inorganic chlorine abundance, tropical tropopause temperatures) and to a much better understanding of the strengths and weaknesses of CCMs. The quantitative evaluation has also allowed identification of remaining common systematic errors in the simulation of tropical lower stratospheric temperature and water vapour, details of the Antarctic polar vortex and the ozone hole, and the representation of the quasi-biennial oscillation (SPARC-CCMVal, 2010).

The evaluation of clouds in models has seen a significant increase in process-oriented approaches (Chen and Del Genio, 2009; Williams and Webb, 2009; Williams and Tselioudis, 2007; Williams and Brooks, 2008) many of which are making use of definitions of cloud regimes using satellite data and evaluate the models' performance in capturing both the spatial distribution and occurrence of cloud regimes as well as the radiative, microphysical and precipitation properties for each regime. (Williams and Brooks, 2008) have also successfully made use of initial value techniques and showed that errors in climate simulations can be successfully identified using such approaches. The regime-oriented evaluation of clouds provides a powerful tool to both inform studies that aim at isolating processes (see Section 9.2.2.2) and to derive process oriented performance metrics (Williams and Webb, 2009).

9.7.3 Targeted Experiments

9.7.3.1 Transpose AMIP

It is well understood that the source of many errors in climate model simulations can be traced to uncertainties in the parameterisation of sub-grid scale processes in the atmosphere. These processes, including clouds, convection and turbulence, are computed from the large-scale (i.e., resolved-scale) state of the atmosphere. Furthermore, differences in the simulation of these processes account for much of the spread between models in their climate change projections (Soden and Held, 2006; Webb et al., 2006). These processes have inherent timescales that are considerably shorter than those associated with the evolution of the large-scale state; therefore it is valuable to test their response when initialized with an observed large-scale state, before errors in the processes can substantially alter the large-scale. Such tests are performed by initializing the atmosphere portion of a climate model with a global analysis of the large-scale atmospheric state from a numerical weather prediction center – essentially running the climate model in 'weather forecast mode' (see also Section 9.2.2.5). From examination of the first few days of model simulations one can identify which errors in climate simulation are due to errors in the parameterisation of these processes and which errors result from longer time scale feedbacks of these processes with the large-scale state of the atmosphere or other component models of Earth's climate.

The increasing use of this technique by the climate modeling community has led to an intercomparison project entitled Transpose-AMIP. From many recent studies with individual models, as well as the long-term experience at modeling centers with both climate and weather prediction goals (Martin et al., 2010), it is anticipated that Transpose-AMIP will be valuable in identifying the source of model errors and assessing confidence in the ability of climate models to simulate the fast processes. A recent example shows that biases in tropical precipitation associated with the double Intertropical Convergence Zone result from longer-time scale interactions with the large-scale state (Phillips et al., 2004). For phenomena with time-scales of weeks, such as soil moisture and the Madden Julian Oscillation, errors in simulated precipitation are related to whether the model in climate mode can successfully simulate the fast processes associated with these phenomena (Boyle et al., 2008; Klein et al., 2006; Martin et al., 2010). Errors in cloud properties are present from very early on in a forecast in at least one model (Williams and Brooks, 2008), although this was not the case in another model (Zhang et al., 2010b). The Transpose-AMIP methodology also allows model developers to test new parameterisations against advanced process observations that are only available for limited locations and times (Bodas-Salcedo et al., 2008; Boyle and Klein, 2010; Hannay et al., 2009; Williamson and Olson, 2007; Williamson et al., 2005; Xie et al., 2008).

9.7.3.2 Simulation of Key Periods in the Past

Comparison of model results for the LGM, the mid-Holocene and the Last millennium also help to identify some of the feedbacks that explain differences between model results. Sensitivity experiments have also been used to estimate the uncertainties arising from the experimental set up or from systematic model biases.

The comparison between the PMIP1 ensemble simulations with atmosphere alone models and PMIP2 ensemble with coupled ocean-atmosphere models confirmed that the feedbacks from the ocean improved model-data comparisons in several regions (Braconnot et al., 2007d). The impact of the SST biases in the control simulation on the simulated mid-holocene asian monsoon was investigated by (Ohgaito and Abe-Ouchi, 2009). Using the MIROC3.2 model coupled to a slab ocean model they mimic the SST produced by the different PMIP2 pre-industrial simulation. They show that the *pattern* of the preindustrial SST has larger impact than the *magnitude of the bias* on mid-Holocene simulations of the Asian monsoon. However, the representation of atmospheric processes such as convection seems to dominate the model spread in this region.

The sensitivity of the simulated change in AMOC at the LGM to different aspects of fresh water forcing, including river runoff, precipitation minus evaporation or ice calving has been reported in several publications (Kageyama et al., 2009). They show that the transition from an active circulation to a shut down is sensitive to small changes in these fluxes. A correct simulations of the temperature and salinity fields is thus required in high latitude which imposes a realistic treatment of the river runoff and of its change (Alkama et al., 2008).

Vegetation and CO_2 feedbacks are also crucial in simulations of the LGM (Woillez et al., 2011). Biases in vegetation contribute to model errors in polar amplification of past climate change (O'ishi and Abe-Ouchi, 2011) and errors related to coupling between vegetation and soil moisture are implicated in intermodel

spread and excess aridity in continental regions (Wohlfahrt et al., 2008; Braconnot et al., 2007b; Dallmeyer et al., 2010; Wang et al., 2008; Harrison and Prentice, 2003).

9.7.3.3 Sensitivity to Resolution, High-Resolution GCMs

Impacts of improved resolution on model performance (or, equivalently, the role of insufficient resolution as a source of model error) can be tested by comparing simulations from the same model (or a component of the model) run at different resolutions.

Since the AR4, several studies have investigated the resolution dependence of various aspects of model performance. Roeckner et al. (2006) compared seasonal mean climate simulated by an AGCM (ECHAM5) run at different horizontal and vertical resolutions. They found that at lower vertical resolution (19 levels) there was no consistent improvement with increasing horizontal resolution, while at higher vertical resolution (31 levels) model error decreased monotonically with increasing horizontal resolution. Wehner et al. (2010) showed that an AGCM (CAM2) run at around 50 km horizontal resolution realistically reproduced observed intensity of extreme daily precipitation over the continental United States, while lower resolution runs severely underestimated the intensity. Improvements of precipitation intensity and distributions especially related to fine-scale orography have also been found in AOGCMs with enhanced atmospheric horizontal resolutions to around 50 km (Delworth et al., 2011; Gent et al., 2010; Sakamoto et al., 2011) and in AGCMs run at similar resolution (Lau and Ploshay, 2009) or finer (20 km) (Kitoh et al., 2011). Improvements are also found in diurnal cycle of tropical precipitation with higher resolution (Ploshay and Lau, 2010).

As for oceanic components, although in principle solving the same equations, ocean models that resolve eddies simulate an ocean that is quite distinct from that simulated by an ocean model that does not. This point is well illustrated by the European Drakkar Project, wherein a hierarchy of global ocean-ice model configurations (with resolutions from two degrees to 1/12 degree) has been used to systematically study modifications arising from refining the ocean resolution (allowing better representation of the ocean mesoscale, and more accurate land-sea boundaries) (e.g., Penduff et al., 2010; Penduff et al., 2011). A similar hierarchy was documented for the Southern Ocean in Hallberg and Gnanadesikan (2006). Certain indices, such as sea level variability and eddy kinetic energy, are improved with refined resolution (Hallberg and Gnanadesikan, 2006; Penduff et al., 2010; Penduff et al., 2011). Furthermore, Biastoch et al (2008a) emphasizes the importance of resolving Agulhas eddies to properly simulate Atlantic decadal variability. Notably, some high latitude watermass transformation processes may degrade due to inadequate eddy parameterizations in refined, though not fully eddy resolving, simulations (Gulev et al., 2007). A relatively high resolution (around 1/2 degree, not eddy permitting) ocean model can also realistically simulate the tropical cyclone-induced cooling of sea surface temperature (Vincent et al., 2011).

Marti et al. (2010) increased the atmospheric resolution of an AOGCM (IPSL CM4) and found improvements in storm-tracks and the North Atlantic oscillation over the standard model. The impact of the higher atmospheric resolution also extends to improvements in the Atlantic meridional overturning circulation in the ocean and ocean-atmosphere dynamical coupling in the tropics. However, in their case, the improved dynamics in the tropics resulted in a too large ENSO amplitude and somewhat deteriorated performance in that aspect. Adopting much higher resolutions, mesoscale eddying ocean simulations (roughly 10 km or finer) coupled to fine scale (order 50 km or finer) atmospheric models exhibit vigorous frontal scale air-sea interactions between sea surface temperature and winds, such as those seen in high-resolution satellite observations, with some of these interactions impacting on processes important for water mass formation (Bryan et al., 2010). In the tropics, a hierarchy of GFDL coupled climate models with varying grid resolution (Delworth et al., 2011) point to the dual importance of atmosphere and ocean resolution for ameliorating biases associated with the double ITCZ occurring in the tropical Pacific (Lin, 2007). Other studies utilizing AOGCMs with eddy permitting ocean also report improvements in aspects such as tropical instability waves and coastal upwelling (Roberts et al., 2009; Sakamoto et al., 2011; Shaffrey et al., 2009). An enhanced atmospheric resolution also contributes to the improved coastal climate by better simulated stratocumulus clouds and wind patterns associated with more realistic topography (Gent et al., 2010; Shaffrey et al., 2009).

In summary, growing evidence after AR4 confirms that improvement of resolution is a promising way to improve various aspects of model performance. Particularly, there is much evidence and high agreement in the literature for improvements in precipitation intensity and distribution with increased atmospheric horizontal resolution to around 50 km or finer, and for improvements in ocean currents and air-sea interactions with ocean eddy permitting simulations. However, it should be noted that improved resolution sometimes causes a degradation of performance in some large-scale aspects, as in the case of ENSO amplitude in Marti et al. (2010), since there are still substantial uncertainties in parameterizations in the models.

9.7.3.4 RCMs

Compared to AOGCMs, coordinated RCM experiments and ensembles have only recently become commonplace, and now exist for Europe (Christensen et al., 2010), North America (Gutowski et al., 2010), South America (Menendez et al., 2010), Africa (Druyan et al., 2010; Paeth et al., 2011; Ruti et al., 2011), and Asian regions (Feng and Fu, 2006). As specific RCMs have often been developed for specific regions, an application of a model for different regions provides means for additional model evaluation. Consequently, these developments enable new studies for improved characterization of uncertainty due to model formulation, as well as exploration of performace-based metrics (Christensen et al., 2010).

Transferability experiments are a category of coordinated RCM experiments. They involve running RCMs for different regions, which and may help to expose shortcomings that are not evident in applications for the "home region" of the model. Many of the coordinated RCM ensembles mentioned above include models predominantly developed for different regions. There are also specific transferability studies, which indeed suggest that RCMs exhibit different skill for different regions, e.g., for simulated temperature and precipitation as well surface energy fluxes, (Takle et al., 2007; Gbobaniyi et al., 2011). In all, the most common conclusion from RCM intercomparison studies is, that no single model outperforms the others, which supports the usefulness of the multi-model approach also in the case of RCMs in model evaluation and, by extension, in climate projections.

9.7.3.5 Perturbed Physics Ensembles

Perturbed physics ensembles (PPE) have been developed to evaluate the sensitivity of a single model to uncertain model parameters, and to evaluate the range of model parameters consistent with observed climate records. Both goals have been addressed using full-complexity component models and reduced-complexity EMICs (see below). Owing to the computational requirements, full-complexity AOGCMs have only recently begun to be employed within the PPE framework (Brierley et al., 2010b; Collins et al., 2007; Sanderson et al., 2010).

Initial work with PPEs was undertaken in the EMIC community to sample model response uncertainty and calibrate parameters so as to reproduce climate change observations (Forest et al., 2008; Forest et al., 2002; Knutti et al., 2002) This approach provides estimates of joint distributions of model parameters that typically correspond to climate system properties. As a model evaluation tool, these joint distributions are then used to assess uncertainty in models for which creating ensembles are prohibitively expensive, namely, AOGCMs and ESMs. Key model diagnostics such as climate sensitivity, ocean carbon uptake, or aerosol forcing are analyzed in both EMICs and ESMs; the joint distributions from the EMIC calibration provide a measure of uncertainties in the ESMs given their distribution within the calibrated estimate of the model diagnostics (Forest et al., 2008; Forest et al., 2002; Knutti et al., 2002; Sokolov et al., 2010; Stott and Forest, 2007; Tebaldi and Knutti, 2007b; Xiao et al., 1998).

Several PPE ensembles constructed with the Hadley Centre climate model (HadCM3) have been compared with the multi-model ensembles of CMIP3 and CFMIP (Collins et al., 2010). For many variables the range of errors in the PPE ensembles is comparable to that found in the CMIP3 ensemble. In the PPE experiments, the systematic component (i.e., common to all members) of the total error is larger than the random component (unique to individual members). As a result, the ensemble average does not yield better agreement with observations than the individual members, in contrast to the often superior MME average (Reichler and Kim, 2008). However, there is evidence that the experimental design of a PPE can be controlled to more closely mimic the multi-model case where the magnitude of the random and systematic

errors is comparable. This kind of comparison between the error structure of MMEs and PPEs can help improve understanding of the fundamental differences between the two, and may possibly lead to a better characterization of model uncertainty.

9.7.4 Climate Sensitivity and Climate Feedbacks

9.7.4.1 Equilibrium Climate Sensitivity, Idealised Radiative Forcing, and Transient Climate Response in the CMIP5 Ensemble

Climate sensitivity, the equilibrium change in global-mean surface temperature after doubling the atmospheric CO_2 concentration relative to pre-industrial, is the most important single measure of climate response because the response to a CO_2 increase of many other climate variables scales according to the increase in global-mean surface temperature (Meehl et al., 2007b). Diagnosing the climate sensitivities of the CMIP5 models is therefore crucial for understanding the similarities and differences between different models' responses in climate projections.

The method of diagnosing climate sensitivity within CMIP5 has changed considerably since AR4 (Randall et al., 2007). There, an atmospheric GCM was coupled to a non-dynamic mixed-layer (slab) ocean model; in equilibrium the implied ocean heat transport convergence was diagnosed to yield the observed SST. CO_2 concentration was then doubled, and the model with fixed implied ocean heat transport convergence integrated to a new equilibrium. While computationally efficient, this method had the disadvantage of employing a different model from the AOGCM used for the historical simulations and climate projections. However, in the few comparisons that were made between the equilibrium climate sensitivity of an AOGCM and it corresponding slab version, the agreement was within 10% or even less (Boer and Yu, 2003; Danabasoglu and Gent, 2009; Li et al., 2011a). Nevertheless, in CMIP5 it was decided to diagnose climate sensitivity directly from the AOGCMs used elsewhere in CMIP5, following (Gregory et al., 2004). The CO_2 concentration is instantaneously quadrupled and kept constant for 150 years of simulation; both climate sensitivity and radiative forcing (see below) are then diagnosed from a linear fit of the energy balance

$$N = F - \alpha \Delta T \tag{9.1}$$

where N is the instantaneous radiative imbalance at the top of the atmosphere, F the radiative forcing (see Chapter 7), α the inverse of the climate sensitivity parameter (see Glossary), and ΔT the perturbation in global-mean surface temperature. If F and α are constant, a linear fit of (Equation 9.1) to the AOGCM output displaying N against ΔT in principle yields climate sensitivity and F as the intercept of the linear fit with the N=0-axis and ΔT=0-axis, respectively. Because in CMIP5 the CO_2 concentration is quadrupled and not doubled, F and climate sensitivity are obtained by a division by two from the intercepts, assuming a logarithmic dependence of F and climate sensitivity on CO_2 concentration (Manabe and Bryan, 1985).The method employed in CMIP5 obviates the need to maintain a separate slab version of the AOGCM, but introduces uncertainties arising from assuming constant F and α (Boer and Yu, 2003; Williams et al., 2008), from the possibility that feedbacks might be different during the transient and the equilibrium phase (Yokohata et al., 2008), and from assuming a strictly logarithmic dependence of climate response on the CO_2 concentration, for which there is evidence (Manabe and Bryan, 1985) but at least two counterexamples, although with a lower-resolution model than most of those assessed in AR4 or used in CMIP5 (Gregory et al., 2004; Li et al., 2011a). A rigorous comparison of the method of (Gregory et al., 2004) against the equilibrium climate response of an AOGCM to CO_2 increase (quadrupling) found the method of (Gregory et al., 2004) to be accurate to within 10% (Li et al., 2011a).

An alternative method to using (9.1) for diagnosing radiative forcing is to keep SST fixed while quadrupling the CO_2 concentration and then to diagnose the radiative imbalance at the top of the atmosphere (Hansen et al., 2005). Both methods are used in CMIP5; the difference in the outcome gives a measure of the methodological uncertainty. To obtain comparability with the climate sensitivity, both estimates of radiative forcing are likewise divided by 2 to obtain the result equivalent to CO_2 doubling.

The third important quantity in this context is the transient climate response, defined as the global-mean surface temperature change, averaged over the 20-year period centred on the time of CO_2 doubling in an experiment in which the CO_2 concentration is increased by 1% compound per year. The transient climate

response is smaller than the equilibrium climate sensitivity because ocean heat uptake delays surface warming; hence, different ocean heat uptake in different models also contributes to a spread in transient climate response.

Based on the methods just outlined, Table 9.2 shows the diagnosed radiative forcing for CO_2 doubling, the equilibrium climate sensitivity, and the transient climate response from the CMIP5 ensemble. The two estimates of radiative forcing agree with each other to within 5% for four models, although for three models the deviation is around 10%, pointing to nonlinearities or deviations from the basic assumptions. However, the spread between methods for diagnosing radiative forcing is much less than spread among models for radiative forcing, which is around 40%. Equilibrium climate sensitivity and transient climate response vary by a factor of two or slightly less, respectively. The model spread in climate sensitivity, from 2.1 K to 4.6 K, is nearly indistinguishable from that in AR4, and while every model whose heritage can be traced to AR4 shows some change in climate sensitivity, there is no discernible systematic tendency in how climate sensitivities changed from AR4 to CMIP5. This broad similarity between AR4 and CMIP5 and the good agreement between different methods where they were applied to the same atmospheric GCM indicate that the uncertainty in methodology is minor compared to the overall spread in climate sensitivity. The change in transient climate response from AR4 to CMIP5 is generally of the same sign but of smaller magnitude, compared to the change in climate sensitivity.

Table 9.2: Climate sensitivity estimates from the AOGCMs (see Table 9.1 for model details). The entries were calculated based on the CMIP5 archive and according to (Hansen et al., 2005) for radiative forcing, fixed SST; (Gregory et al., 2004) for radiative forcing, regression, and equilibrium climate sensitivity; and from the 20-year mean centred on the year of CO_2 doubling in a 1% per year increase experiment for the transient climate response. Notice that the entries in columns 2–4 were obtained by dividing the original results, which were obtained for CO_2 quadrupling, by two.

Model	Radiative forcing ($W\ m^{-2}$)		Equilibrium Climate Sensitivity (K)	Transient Climate Response (K)
	Fixed SST	Regression		
CanESM2	3.68	3.84	3.69	2.4
CNRM-CM5	n.a.	3.72	3.25	2.1
CSIRO-Mk3-6-0	3.10	2.59	4.08	1.8
HadGEM2-ES	3.50	2.92	4.58	2.5
INM-CM4	3.12	2.98	2.08	1.3
IPSL-CM5A-LR	3.25	3.10	4.13	2.0
MIROC-ESM	n.a.	n.a.	n.a.	2.2
MIROC5	n.a.	4.13	2.72	1.5
MPI-ESM-LR	4.32	4.09	3.63	n.a.
MRI-CGCM3	3.60	3.25	2.60	1.6
NorESM1-M	n.a.	3.11	2.80	1.4

9.7.4.2 Model Evaluation of EMICs and Constraints on Climate Sensitivity

Earth-system models of intermediate complexity (EMICs) are used in Chapters 6, 10 and 12. In Chapter 6, carbon-cycle components of EMICs are discussed with specific attention to feedbacks, and are used to analyze permissible emissions for stabilization scenario RCP4.5. In Chapter 10, EMICs are used to estimate the probability distributions of equilibrium climate sensitivity (ECS). In Chapter 12, EMICs are used to explore commitment scenarios that extend the RCP transient forcing scenarios to year 3000. These scenarios explore climate stabilization and long-term targets in which either radiative forcings are set to a constant value (i.e., constant concentrations) or emissions are set to zero after 2300.

Because EMICs are calibrated to reproduce present-day climate, model evaluation requires testing them in alternative scenarios (e.g., paleoclimate) or comparing directly with results from transient scenarios from AOGCMs (e.g., Forest et al., 2008; Meinshausen et al., 2009). When results from the latter are combined with probability distributions of ECS or TCR, the MME for AOGCMs/ESMs can be assessed against the observationally constrained distributions. Some EMICs have been modified to include ice-sheets (UVic 2.9, CLIMBER-2.4) and ocean sediment models (DCESS, UVic 2.9, Bern3D-LPJ). Plattner et al. (2008) presented results of an EMIC intercomparison project that explored stabilization scenarios and concluded that the EMICs compared favorably over the 2000 to 2100 period with the AOGCM results. This implies that the models are well-suited for simulations extending over the future millenium. Attributes of current EMICs are provided in Table 9.3.

[INSERT TABLE 9.3 HERE]

Table 9.3: Features of Earth System Models of Intermediate Complexity (EMICs).

9.7.4.3 Role of Cloud Feedbacks in Climate Sensitivity

Cloud feedbacks represent one of the main causes for the range in climate sensitivity across multi-model ensembles of AOGCMs. The spread due to inter-model differences in cloud feedbacks is approximately 3 times larger than the spread contributed by feedbacks due to variations in water vapor, lapse-rate, and ocean heat uptake (Dufresne and Bony, 2008), and is a primary factor governing the range of climate sensitivity across 18 models in the CMIP3 ensemble (Volodin, 2008b). Differences between the equilibrium sensitivity to $2 \times CO_2$ and the transient climate response to 1% CO_2 yr^{-1} at the time of doubling are due primarily to the differences in the shortwave cloud feedback between the two experiments (Yokohata et al., 2008). In perturbed ensembles of the Hadley Centre Atmospheric Model coupled to a slab ocean (HadSM3) and the Model for Interdisciplinary Research on Climate (MIROC3.2), the primary factor contributing to the spread in equilibrium climate sensitivity in both ensembles is the low-level shortwave cloud feedback (Yokohata et al., 2010). Changes in the sign of low-cloud feedbacks to increased CO_2 forcing also explain the lower sensitivity of the new MIROC5 model relative to the prior version MIROC3.2 (Watanabe et al., 2010).

Application of radiative kernel techniques to multiple models forced by doubled CO_2 show that while changes in cloud forcing can be either positive or negative, the cloud feedbacks are generally positive or near neutral (Shell et al., 2008; Soden et al., 2008). All of the models examined in a multi-thousand member ensemble of AOGCMs constructed by parameter perturbations also have net positive or neutral cloud feedbacks (Sanderson et al., 2010). This finding is consistent with the modeled and measured relationships between SSTs and top-of-atmosphere radiative fluxes, which suggest that interannual cloud variations act as a positive feedback in the current climate (Chung et al., 2010b).

Over the north-east Pacific, decadal-scale fluctuations in surface and satellite-based measurements of low-level cloud cover are significantly negatively correlated with variations in SST (Clement et al., 2009). This negative correlation is consistent with a positive low-cloud feedback in this region operating on decadal time scales. Models that reproduce this negative correlation and other relationships between cloud cover and regional meteorological conditions simulate a positive low-cloud feedback over much of the Pacific basin (Clement et al., 2009).

Analyses of the tendencies in cloud condensate when multiple models are subjected to a CO_2 increase shows that inter-model differences in cloud response are attributable to different parameterisations of ice sedimentation processes (Ogura et al., 2008). In experiments with perturbed physics ensembles of AOGCMs, the parameterisation of icefall speed also emerges as one of the most important determinants of climate sensitivity (Sanderson et al., 2010; Sanderson et al., 2008b).

9.7.4.3 Relationships Among Forcings, Feedbacks, and Climate Sensitivity

Despite the range in equilibrium sensitivity of 2.1°C to 4.4°C for AR4 AOGCMs, these models reproduce the global surface air temperature anomaly of 0.76°C over 1850–2005 to within 25% relative error. The relatively small range of transient climate response suggests that there is another mechanism in the models that counteracts the relatively large range in sensitivity, and that mechanism appears to be a systematic negative correlation across the multi-model ensemble between climate sensitivity and anthropogenic forcing (Anderson et al., 2010; Kiehl, 2007; Knutti, 2008). The effect of eliminating this compensation between forcing and feedback could range from relatively minor (Knutti, 2008) to major expansion in the range of equilibrium climate sensitivity to 2.1°C–4.4°C (Huybers, 2010).

9.7.4.3.1 Role of humidity and lapse rate feedbacks in climate sensitivity

Correlations between coincident variations in SST and clear-sky outgoing longwave radiation (OLR) provide estimates on the rate of radiative damping of SST fluctuations. Modelled values for clear-sky damping are internally consistent across the AR4 multi-model ensemble and are a good approximation of the empirical damping rate obtained from SST data and satellite observations of clear-sky OLR (Chung et al., 2010a). The modelled and observationally derived damping rates are consistent with a strong positive correlation between SST and water vapour on regional to global scales. The relationship of fluctuations in SST and upper-tropospheric humidity can be derived directly from the Atmospheric Infrared Sounder (AIRS), and the results show that a typical AGCM is capable of reproducing the positive rate of increase in specific humidity with increased SST of 10%–25%°C^{-1} (Gettelman and Fu, 2008).

9.7.4.3.2 Role of oceanic heat uptake and other oceanic processes in climate Sensitivity

Atmospheric feedbacks derived using the radiative kernel technique from perturbed-physics AOGCM ensembles are relatively insensitive to perturbations of ocean parameters (Sanderson et al., 2010). In perturbed-physics ensembles with alterations to parameters governing three key ocean processes, the effects of the perturbations on the ocean heat uptake and transient climate response are relatively small (Collins et al., 2007). The key ocean processes perturbed in these experiments include isopycnal and vertical diffusivity and the structure of the mixed layer adjacent to the ocean surface.

9.7.4.3.3 Sources of uncertainty in modelled climate sensitivity

Objective methods for perturbing uncertain model parameters to optimize performance relative to a set of observational metrics have shown a tendency toward an increase in the mean and a narrowing of the spread of estimated climate sensitivity (Jackson et al., 2008a). This tendency is opposed by the effects of gaining better knowledge regarding structural biases shared across a multi-model ensemble. Determination that there is a nonzero probability of shared structural biases tends to reduce the mode of the sensitivity distribution towards lower values while simultaneously the tail of the distribution towards larger sensitivities (Lemoine, 2010). Roe and Baker (2007) suggest that symmetrically distributed (e.g., Gaussian) uncertainties in feedbacks lead to inherently asymmetrically distributed uncertainties in climate sensitivity with increased probability in extreme positive values of the sensitivity. Roe and Baker (2007) conclude that this relationship makes it extremely difficult to reduce uncertainties in climate sensitivity through incremental improvements in the specification of feedback parameters. Subsequent analysis suggests that this finding and the underlying relationship between uncertainties in feedbacks and climate sensitivity artefacts both of their statistical formulation (Hannart et al., 2009) and their linearization (Zaliapin and Ghil, 2010).

Using a Bayesian framework to analyse perturbed physics experiments using a slab-ocean GCM, Sanderson et al (2008b) and Rougier et al (2009) find that the rate of cloud entrainment is the single most important source of uncertainty in AOGCM sensitivity. An additional source of uncertainty in equilibrium sensitivity is apparently an inherent feature of the idealized experiments used to derive it. These experiments involve instantaneously increasing (usually doubling) the concentrations of CO_2 and then monitoring the rate at which radiative equilibrium is restored or estimating the asymptotic equilibrated surface temperature increase. The instantaneous increase induces very rapid atmospheric and terrestrial adjustments analogous to the semi-direct effects of aerosols including adjustments to the cloud field, tropospheric lapse rate and humidity and snow cover (Andrews and Forster, 2008; Gregory and Webb, 2008). These findings have highlighted the importance of separating the fast responses that depend on (instantaneous) changes in forcing and the feedbacks that follow the much slower adjustments in ocean temperature.

9.7.4.4 Comparison Between Future Climate and Last Glacial Maximum

The AR4 reported on attempts to relate the simulated LGM changes in tropical SST to global climate sensitivity, providing a range of acceptable climate sensitivity values (Hegerl et al., 2007). These studies tested either ensemble simulations with varying parameters or the PMIP2 multi-model ensemble (Crucifix, 2006). (Edwards et al., 2007) summarized these studies (Figure 9.42) and discussed the value added and limitations of combining constraints on past climate and present climate, such as done by (Annan and Hargreaves, 2006). LGM Temperature changes in the tropics and in Antarctica have been shown to scale well with climate sensitivity (Hargreaves et al., 2007), because the signal is mostly dominated by the CO_2 forcing in these regions (Braconnot et al., 2007c) (Figure 9.33). The analogy between the LGM climate sensitivity and future climate sensitivity is however not perfect. In ensemble simulations with the MIROC model coupled to a slab ocean model the LGM cooling and the warming induced by a doubling of CO_2 are not symmetrical (Hargreaves et al., 2007). Differences in the cloud radiative feedback are at the origin of this asymmetry (Yoshimori et al., 2009).

[INSERT FIGURE 9.42 HERE]

Figure 9.42: (a) Comparison of simulated and observed changes in annual mean temperature, LGM compared to modern and ocean compared to land, (b) simulated relationship between regional cooling in the tropics and over eastern Antarctica and global cooling. This figure is adapted from (Crucifix, 2006; Cunningham et al., 2009; Kageyama et al., 2006; Masson-Delmotte et al., 2006; Otto-Bliesner et al., 2009) using the PMIP2 dataset (Braconnot et al., 2007d). In a) the colour dots represent the different model results for the two regions, and the large crosses the estimates for the ocean and land surface data from (Waelbroeck et al., 2009) and (Bartlein et al., 2010b) respectively. In b) the different points represent different model results and the hatched bars the estimates with error bars for the (Waelbroeck et al., 2009) SST reconstruction and East Antarctica air temperature reconstruction from ice cores (Masson-Delmotte et al., 2006).

The overall feature that emerges is that model tend to underestimate the polar amplification, which is a feature also seen in other climatic contexts (Masson-Delmotte et al., 2010). Note that the comparison over the Antarctic ice sheet is complex because there are possible inconsistency between the ice sheet topography use to force the model and the ice core indication (Masson-Delmotte et al., 2006). Nevertheless the underestimation is also shown from independent land and ocean data when comparing the cooling in the tropics and in high latitudes. The ratio between the change in temperature over land and over the ocean is rather similar in different models, resulting mainly from differences in the hydrological cycle over land and ocean (Sutton et al., 2007). This ratio is similar for future climate projections and LGM simulations (Cunningham et al., 2009). The LGM data do not support the model simulations with equivalent cooling over land and ocean.

Although the LGM climate does not provide a direct assessment of climate sensitivity, it helps to infer the realism of model behaviour. In particular, the evaluation of LMG simulations indicate that it is likely that the polar amplification, and thereby the equator to pole latitudinal gradient, is underestimated in climate change experiments and that the models that do not produce the proper ratio of land/sea temperature changes do not reproduce proper contrast of evaporation over land and ocean.

9.8 Relating Model Performance to Credibility of Model Applications

9.8.1 Overall Assessment of Model Performance

This chapter has quantitatively assessed the performance of individual CMIP5 models as well as the multi-model mean in comparison to observations. A wide range of skills was obtained showing that there is a large variation in the ability of the models to simulate essential climate variables (Cadule et al., 2010; Connolley and Bracegirdle, 2007; Gleckler et al., 2008; Macadam et al., 2010; Pincus et al., 2008; Reichler and Kim, 2008), underlying key processes (Waugh and Eyring, 2008; Williams and Webb, 2009), and climate phenomena (Guilyardi et al., 2009b; Stoner et al., 2009). The large variation in skill occurs both for different performance metrics applied to a single model as well as for the same performance metric applied to different models. No model scores high or low in all metrics, but some models perform substantially better in comparison to others. The assessment has also shown that some classes of models, e.g., those with higher horizontal resolution, higher model top or a more complete representation of the carbon cycle, aerosols or

chemistry, agree better with observations in selected processes, phenomena or ECVs than others [PLACEHOLDER FOR SECOND ORDER DRAFT: tbc when more CMIP5 output available].

Figure 9.8.1 provides a synthesis of key findings with respect to how well models (CMIP3 and CMIP5 in some cases) represent important features of the climate in the 20th century. The figure makes use of the calibrated language for uncertainty assessments as defined in Mastrandrea et al. (2011). The y-axis refers to the level of agreement between available model evaluation studies, while the assessment on the x-axis weighs the available evidence for this judgment, including number of studies and quality of observational data. Generally, evidence is most robust when there are multiple, independent studies that evaluate multiple models using high-quality observations. The level of agreement is high if multiple studies come to the same conclusions or if there is only one study. However, in the latter case the evidence is low since multiple independent lines of evidence are missing. Overall, confidence increases towards the top-right corner as suggested by the increasing strength of shading. In addition, this figure has a third color coded dimension that assesses how well the models perform compared to observations. The figure shows that several important aspects of the climate are simulated well by contemporary models (blue entries), with varying level of evidence and agreement. For example, global mean surface air temperature (SAT) is simulated well and has been evaluated in multiple studies using high-quality observations, so is placed in the upper right corner of the figure. In contrast, the diurnal cycle of SAT is also simulated well but the evidence and agreement among studies is only medium. On the other hand, the models are showing mixed results or still have problems simulating other important aspects of the climate, as signified by the grey or red text in the figure. A description that explains the expert judgment for each of the key climate features presented in Figure 9.8.1 can be found in the body of this chapter, with a concrete link to the specific sections given in the figure caption.

[INSERT FIGURE 9.43 HERE]

Figure 9.43: Summary of the findings of Chapter 9 with respect to how well the CMIP3 models simulate important features of the climate of the 20th century [PLACEHOLDER FOR SECOND ORDER DRAFT: Will be updated with CMIP5 models]. Confidence in the assessment increases towards the top-right corner as suggested by the increasing strength of shading. Features that current state-of-the-art AOGCMS and ESMs simulate well, show mixed results, or have problems representing are shown in blue, grey, and red, respectively. The figure highlights the following key features (subject to revisions), with the sections that back up the assessment added in brackets:

Annual cycle SIE:	Annual cycle Arctic and Antarctic Sea Ice Extent (Section 9.4.3)
AMOC:	Atlantic Meridional Overturning Circulation (Section 9.4.2.6)
Antarctic P-E:	Antarctic Precipitation minus Evaporation (Section 9.6.1)
Circulation regimes:	Blocking events and others circulation regimes (Section 9.5.2.2)
Clouds and CRE:	Clouds and Cloud Radiative Effects (Section 9.4.1)
ENSO:	El Niño Southern Oscillation (Section 9.5.3.7)
Global Monsoon:	see Section 9.4.2
Global Scale P:	Global scale precipitation (Section 9.4.1)
Meridional heat transport:	see Section 9.4.2.6
MJO:	Madden Julian Oscillation (Section 9.5.2.2)
NAO:	Northern Annual Mode (Section 9.5.3.7)
OHC:	Ocean Heat Content (Section 9.4.2)
SAF:	Snow albedo feedback (Sections 9.8.3)
SAO:	Southern Annual Mode (Section 9.5.3.7)
SAT:	Surface Air Temperature (Section 9.4.1)
SIE:	Sea Ice Extent (Sections 9.4.3 and 9.8.3)
SSIE:	September SIE (Sections 9.4.3 and 9.8.3)
SSS:	Sea Surface Salinity (Section 9.4.2.6)
SST:	Sea Surface Temperature (Sections 9.4.1 and 9.4.2.6)
Trends in T and P Extremes:	Trends in temperature and precipitation extremes (Section 9.5.4)
Trop Atlantic / Pacific MS:	Tropical Atlantic / Pacific Mean State (Section 9.4.2)
Trop Indian Ocean MS:	Tropical Indian Ocean Mean State (Section 9.4.2.6)
Upper ocean heat uptake:	see Section 9.4.2
UTTT:	Upper tropospheric temperature trends (Section 9.4.1)
WBC:	Western Boundary Current (Section 9.4.2.6)

[PLACEHOLDER FOR SECOND ORDER DRAFT: This figure is a preliminary version subject to revisions; it will be updated with more CMIP5 results and will possibly be separated into different panels that show: Panel 1: Climatologies and trends; Panel 2: Variability including extremes; Panel 3: Regional performance.]

9.8.2 Implications of Model Evaluation for Climate Change Detection and Attribution

Climate models are developed, tuned and evaluated based on historical climate conditions. Their evaluation is therefore of direct use by detection and attribution studies using a combination of observation records and model studies to assess anthropogenic signature on recent climate trends. The key aspects for detection/attribution studies in terms of model evaluation are that models accurately reproduce climate variability and patterns. Biases in magnitude or forcing fingerprint are less important. For instance, the detection and attribution study by Santer et al. (2009) shows that the anthropogenic water vapour fingerprint is insensitive to current model uncertainties, and is governed by basic physical processes that are well-represented in CMIP3 models.

Several statements in the AR4 Chapter 9 (Hegerl et al., 2007) related the conclusions of D&A studies to the quality of the simulated climate. They indicated in particular that anthropogenic changes were detected on temperature rather accurately at large scale and continental scale, but that there was less confidence in the understanding of forced changes in other variables surface as surface pressure and precipitation. In addition difficulties were highlighted in attributing temperature changes on smaller than continental scales and shorter than 50 year time scales, because of model accuracy and misrepresentation of some of the key patterns.

The representation of climate variability at different time and space scales in CMIP5 cautions that care must be taken when considering DA studies at the regional or more local scales. Indeed models still suffer from major biases in the representation of the intraseasonal to multidecadal variability that affects regional features (Sections 9.5.2 and 9.5.3). Recent studies of climate extremes provides further assessment that models have some skill in this regard, however, most models do not reproduce the right balance between cold and warm extremes. They also confirm that resolution affects the confidence that can be placed in the analyses of extreme in precipitation (Section 9.5.4).

9.8.3 Implications of Model Evaluation for Model Applications to Future Climate

The ability of a climate model to reproduce past climate and its variability is a necessary, but not sufficient, condition for reliable projections of future change. Certainly the ability to realistically simulate the response to historical changes in climate forcing (between contemporary and paleo, or transient changes over the 20th century) provides some reassurance that projections of future change are credible. However, future climate forcing drives the climate system outside of the observed range. The ability of climate models to realistically reproduce observed climate processes, variability and interrelationships contributes to our confidence in their ability to simulate future change in spite of the excursion into ‘unknown territory’. The application of models to climate prediction on seasonal to interannual time scales (discussed in Chapter 11) provides some modest ability to directly verify climate model predictions. Nevertheless, direct evaluation of long-term future climate projections is limited to inferences drawn from past performance.

The collection of contemporary climate models is rather inhomogeneous, with some models performing better in some regards, and less well in others. This raises the question whether at least for some applications the reliability of climate projections can be improved by weighting the models according to their ability to reproduce observed climate. In weather and seasonal forecasting a large range of skill measures is routinely applied and sophisticated methods to combine multiple model results have been shown to be superior to simple multi-model mean averages (Stephenson et al., 2005). However, demonstrating the advantages of a weighted multi-model mean of climate projections remains difficult due to the longer time-scales and challenges to establish conclusive links between model performance in the current climate and the response to climate change forcings.

Several studies have started to explore the value of weighting based on the models’ ability to simulate observed climate (Christensen et al., 2010; Connolley and Bracegirdle, 2007; Knutti et al., 2010a; Murphy et al., 2007; Pierce et al., 2009; Raisanen et al., 2010; Scherrer, 2010; Schmittner et al., 2005a; Waugh and Eyring, 2008). In general, only small differences between the weighted and unweighted multi-model mean were found, while the standard deviation in the weighted mean was smaller. Other approaches have employed statistical techniques mostly based on a Bayesian approach in which prior distributions of model simulations are weighted by their ability to reproduce present day climatological variables and trends to produce posterior predictive distributions of climate variables (Furrer et al., 2007; Tebaldi and Knutti,

2007a). Perturbed physics ensembles in which perturbations are made to the parameters in a single modelling structure have also been explored (Murphy et al., 2007), see also Chapter 12, Section 12.4.1.

[INSERT FIGURE 9.44 HERE]

Figure 9.44: *Left:* Scatter plot of simulated springtime snow albedo feedback ($\Delta\alpha_s/\Delta T_s$) values in climate change (ordinate) versus simulated springtime $\Delta\alpha_s/\Delta T_s$ values in the seasonal cycle (abscissa) in transient climate change experiments with 17 AOGCMs from CMIP3 (α_s and T_s are surface albedo and surface air temperature, respectively). From Hall and Qu (2006); [update with CMIP5 data; show CMIP3 in different colour in addition]. *Right:* Constraint on the climate sensitivity of land carbon in the Tropics (30°N-30°S) from interannual variability in the growth-rate of global atmospheric CO_2. This version is based on C^4MIP GCMs (black labels), and three land carbon "physics ensembles" with HadCM3 (red labels). The y-axis is calculated over the period 1960–2099 inclusive, and the y-axis is calculated over the period 1960-2010 inclusive. In both cases the temperature used is the mean (land+ocean) temperature over 30°N-30°S. The vertical grey band shows the estimated sensitivity of the observed global CO_2 growth-rate to the observed tropical mean temperature.

There are several encouraging examples of "emergent constraints", where the large inter-model variations in mean climate, past trends, or seasonal variability are well correlated with comparably large inter-model variations in aspects of the model projections (Boe et al., 2009a; Boe et al., 2009b; Eyring et al., 2007; Hall and Qu, 2006; Mahlstein and Knutti, 2010). In the archetypal example of an emergent constraint, Hall and Qu (2006) showed that inter-model variations in the contemporary seasonal cycle of snow cover are strongly correlated with comparably large inter-model variations in snow albedo feedback under future climate change, see Figure 9.44 (de Jong et al., 2009; Hall and Qu, 2006). Since the seasonal variation in snow-cover is reasonably well-known from observations, the intermodel relationship provides a means to transform this observation into a constraint on the strength of the snow-albedo feedback under climate change. The right panel of Figure 9.44 shows an alternative approach based-on an emergent constraint between the sensitivity of tropical land carbon to warming (i.e., without CO_2 fertilization effects) and the sensitivity of the interannual variability of the growth-rate of atmospheric CO_2 to the interannual variability in the annual mean tropical temperature (30°N–30°S). The y-axis of this plot is essentially the γ_L of Friedlingstein et al. (2006), but for the tropics only. The x-axis of this plot is the regression of the anomaly in the atmospheric CO_2 growth-rate on the tropical temperature anomaly, for each model. Variants of this regression, using the NINO3 index rather than the mean tropical temperature, have been published on a number of previous occasions (e.g., Jones and Cox, 2005). The strong relationship between these two variables is consistent with the fact that the interannual variability in the CO_2 growth-rate is known to be dominated by the response of the tropical land to climatic anomalies, associated particularly with ENSO. The interannual sensitivity of the CO_2 growth-rate to tropical temperature can be estimated from observational data. The intermodel relationship therefore provides a means to transform this into a constraint on the sensitivity of tropical land carbon to tropical climate change.

[INSERT FIGURE 9.45 HERE]

Figure 9.45: Projected decline of Arctic sea ice area with increasing global temperature. Shaded areas depict the uncertainty range (red based on observations from 1980–2007 and blue from 1960–2010). The time period in the legend indicates the time window that is used to estimate polar amplification. The models are calibrated to start at the current observational point (1980–2007) and show points for sea ice larger than 1.0 million km^2. Warming in 2090–2099 and associated uncertainties for three SRES non-intervention emission scenarios from (IPCC, 2007) are indicated at the bottom. From Mahlstein and Knutti (2011).

Another area where model weighting has been explored is Arctic September sea ice extent, where a large decline has been observed since 1979, that was severely underestimated by the CMIP3 models (Stroeve et al., 2007). Boe et al. (2009b) found a good correlation between past and future sea ice extent which they argued is physically based since they are both determined to a large extend by the initial sea ice thickness. In a second step they weighted the projections of September Arctic sea ice extent by using only those models that simulate a trend close to that observed and found that, unlike in the unweighted CMIP3 multi-model mean, the Arctic is likely to be ice-free before the end of the 21st century under the SRES A1B scenario. Other studies used a correlation between the change in global mean or Arctic surface temperature with changes in the sea ice extent to narrow the uncertainty in climate projections for this specific application (Mahlstein and Knutti, 2011; Zhang, 2010). As an example, Figure 9.45 shows a scenario independent estimate of predicted Arctic sea ice extent based on a recalibrated ensemble of CMIP3 models using observations over 28 years. The calibrated sea ice extrapolation based on observations since 1980 suggests a

most likely threshold for a near ice free Arctic in September of about 2°C above present, about half of the value the CMIP3 multi-model mean suggests. However, open issues remain. An appropriate weighting is only possible if the model biases are much larger than the estimated variability (Weigel et al., 2010), and natural variability in Arctic sea ice extent is large (Kay et al., 2011; Winton, 2011). In addition, Blanchard-Wrigglesworth et al. (2011) found that the predictability of Arctic sea ice extent beyond three years is dominated by climate forcing rather than initial conditions. These issues need to be considered in the recalibration and weighting of multi-model ensembles, in particular when observed trends are relatively short. The weighting of Arctic sea ice extent therefore remains an open area of research, see further discussion in Chapter 12.

The examples above demonstrate that progress has been made in relating some processes to key climate feedbacks. One challenge associated with the interpretation of the range of model results is that of model dependency associated with shared assumptions, codes, and datasets. Several recent studies suggest that the effective number of independent models in the CMIP3 ensemble is much smaller (Annan and Hargreaves, 2011; Jun et al., 2008; Masson and Knutti, 2011; Pennell and Reichler, 2011). In particular, models developed at the same centre tend to be similar. Any multi-model mean, whether weighted or not by agreement with observations, is influenced by the initial sample of models and their interdependency.

All models suffer from a multitude of systematic errors, and the importance of these errors to the reliability of projections is not well understood. The implications of this on the above emergent constraints are unkown. Ultimately, our confidence in model projections is built upon the demonstration of how well models represent a wide range of processes on various spatial and temporal scales (Eyring et al., 2005; Knutti et al., 2010b), especially those related to important feedbacks in the Earth's climate system. Hence the thorough evaluation of climate models as carried out in this chapter appears prudent to guide the assessment of model quality.

[START FAQ 9.1 HERE]

FAQ 9.1: Are Climate Models Getting Better, and How Would We Know?

Climate models are extremely complex pieces of software that simulate, with as much fidelity as possible, the marvellously complex interactions between the atmosphere, ocean, land surface, and ice, the global ecosystem, and a variety of chemical and biological processes. Complexity in such models has certainly increased over the years and in that sense current Earth System models are vastly 'better' than the models available at the time of the previous IPCC Assessments. Current models also operate at much higher spatial resolution (i.e., they resolve much finer-scale detail) owing to the continuing increase in available computing resources. Today's models have also benefitted from the past two decades of research into various climate processes, more comprehensive observations, and generally improved scientific understanding. Overall, climate models of today are better than their predecessors. However, as every bit of added complexity also introduces new sources of error and new interactions between model components that may, perhaps only temporarily, degrade the overall simulation of the climate system.

Quantifying model performance is the primary objective of Chapter 9, and corresponding chapters have appeared in all of the previous IPCC Working Group I reports. Reading back over these earlier assessments provides a general sense of the improvements that have been made. However, past reports have typically provided a rather broad survey of model performance (either by showing differences between model-calculated versions of some climate quantity and some corresponding observational estimate). Inevitably some models perform better for certain climate variables, but no individual model clearly emerges as 'the best' overall. Recently, there has been progress in computing various 'performance metrics' whose aim is to synthesize model performance relative to a range of different observations in a simple numerical 'score' (e.g., (Gleckler et al., 2008; Murphy et al., 2004). Of course, the definition of such a score, how it is computed, the observations that are used (which are themselves uncertain to some extent), and the manner in which various scores are combined are all important and will all affect the end result. Nevertheless, if the metric is computed consistently, one can use it to compare different generations of models. (Reichler and Kim, 2008) demonstrated that, at least for the particular 'performance index' they computed, there was a steady improvement in models participating in the series of Coupled Model Intercomprison Projects: CMIP1

included models from the mid 1990s; CMIP2 included models from around 2000; and CMIP3 from about 2005. Their showed that, although each generation exhibited a range in model performance, the average model performance index improved steadily between each generation, with even the poorest performing model in a later generation performing on par with the mean model of the previous generation. A summary of model performance over time is shown in FAQ 9.1, Figure 1.

So, yes, climate models are getting better, and we can demonstrate this with quantitative performance metrics. The issue that remains is that model performance can only be evaluated relative to past observations. In order to have confidence in future projections made with such models, it is considered necessary that past climate, its variability and change, be well simulated. But this may not sufficient. Whereas weather predictions, seasonal climate predictions and to some degree paleoclimate simulations, can be evaluated against observations, climate projections spanning a century or more of the future cannot. However, given that climate models are based on verifiable physical principles, and are able to reproduce many important aspects of past response to external forcing, and they are considerd to be able to provide a scientifically sound preview of the climate to come, given a climate forcing scenario.

[PLACEHOLDER FOR SECOND ORDER DRAFT: FAQ 9.1, Figure 1 will be created based on results from all the past CMIP intercomparisons --- this is to be determined based on analyses still to be done and results that are not yet available. A ‘mocked up’ figure as a placeholder has been provided.]

[INSERT FAQ 9.1, FIGURE 1 HERE]

FAQ 9.1, Figure 1: Quantitative examination of model skill as measured in the three recent phases of CMIP (CMIP2, CMIP3 and CMIP5). The RMSE is normalized in each case by the observational standard deviation to facilitate comparison across variables. Results are shown for global precipitation and surface air temperature. [PLACEHOLDER FOR SECOND ORDER DRAFT: This figure very preliminary; will be updated as additional CMIP5 simulations become available. Additional fields may also be included in future renditions.] Redrafted from (P. Gleckler, K. Taylor, and C. Doutriaux, 2008) and updated with CMIP5 results.

[END FAQ 9.1 HERE]

References

Abe, M., H. Shiogama, J. C. Hargreaves, J. D. Annan, T. Nozawa, and S. Emori, 2009: Correlation between Inter-Model Similarities in Spatial Pattern for Present and Projected Future Mean Climate. *Sola*, **5**, 133-136.

AchutaRao, K., and K. Sperber, 2002: Simulation of the El Niño Southern Oscillation: results from the coupled model intercomparison project. *Clim. Dyn.*, **19**, 191-209.

——, 2006: ENSO simulations in coupled ocean-atmosphere models: are the current models better? *Clim. Dyn.*, **27**, 1-16.

AchutaRao, K. M., et al., 2007: Simulated and observed variability in ocean temperature and heat content. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 10768-10773.

Ackerley, D., E. J. Highwood, and D. J. Frame, 2009: Quantifying the effects of perturbing the physics of an interactive sulfur scheme using an ensemble of GCMs on the climateprediction.net platform. *Journal of Geophysical Research-Atmospheres*, **114**.

Adkins, J. F., K. McIntyre, and D. P. Schrag, 2002: The salinity, temperature, and delta O-18 of the glacial deep ocean. *Science*, **298**, 1769-1773.

Adler, R. F., et al., 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeor.*, **4**.

Alexander, M. J., and K. H. Rosenlof, 1996: Nonstationary gravity wave forcing of the stratospheric zonal mean wind. *Journal of Geophysical Research-Atmospheres*, **101**, 23465-23474.

Alexandru, A., R. de Elia, and R. Laprise, 2007: Internal variability in regional climate downscaling at the seasonal scale. *Monthly Weather Review*, DOI 10.1175/MWR3456.1. 3221-3238.

Alexandru, A., R. de Elia, R. Laprise, L. Separovic, and S. Biner, 2009: Sensitivity Study of Regional Climate Model Simulations to Large-Scale Nudging Parameters. *Monthly Weather Review*, DOI 10.1175/2008MWR2620.1. 1666-1686.

Alkama, R., M. Kageyama, G. Ramstein, O. Marti, P. Ribstein, and D. Swingedouw, 2008: Impact of a realistic river routing in coupled ocean-atmosphere simulations of the Last Glacial Maximum climate. *Climate Dynamics*, **30**, 855-869.

Allan, R. P., and B. J. Soden, 2008: Atmospheric warming and the amplification of precipitation extremes. *Science*, **321**, 1481-1484.

Allan, R. P., M. A. Ringer, and A. Slingo, 2003: Evaluation of moisture in the Hadley Centre climate model using simulations of HIRS water-vapour channel radiances. *Quarterly Journal of the Royal Meteorological Society*, **129**, 3371-3389.

Allan, R. P., A. Slingo, S. F. Milton, and M. E. Brooks, 2007: Evaluation of the Met Office global forecast model using Geostationary Earth Radiation Budget (GERB) data. *Quarterly Journal of the Royal Meteorological Society*, **133**, 1993-2010.

Allen, M., P. Stott, J. Mitchell, R. Schnur, and T. Delworth, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*. 617-620.

Anderson, B. T., J. R. Knight, M. A. Ringer, C. Deser, A. S. Phillips, J. H. Yoon, and A. Cherchi, 2010: Climate forcings and climate sensitivities diagnosed from atmospheric global circulation models. *Climate Dynamics*, **35**, 1461-1475.

Andrews, T., and P. M. Forster, 2008: CO2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophysical Research Letters*, **35**.

Annamalai, H., and K. R. Sperber, 2005: Regional heat sources and the active and break phases of boreal summer intraseasonal (30-50 day) variability. *Journal of the Atmospheric Sciences*, **62**, 2726-2748.

Annan, J., and J. Hargreaves, 2011: Understanding the CMIP3 Multimodel Ensemble. *Journal of Climate*, **24**, 4529-4538.

Annan, J. D., and J. C. Hargreaves, 2006: Using multiple observationally-based constraints to estimate climate sensitivity. *Geophysical Research Letters*, **33**, -.

——, 2010: Reliability of the CMIP3 ensemble. *Geophysical Research Letters*, **37**, 5.

Arblaster, J., and G. Meehl, 2006: Contributions of external forcings to southern annular mode trends. *Journal of Climate*. 2896-2905.

Arneth, A., et al., 2010: From biota to chemistry and climate: towards a comprehensive description of trace gas exchange between the biosphere and atmosphere. *Biogeosciences*, **7**, 121-149.

Arora, V. K., and G. J. Boer, 2005: Fire as an interactive component of dynamic vegetation models. *Journal of Geophysical Research-Biogeosciences*, **110**.

——, 2010: Uncertainties in the 20th century carbon budget associated with land use change. *Global Change Biology*, **16**, 3327-3348.
Arora, V. K., et al., 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, **38**, L05805, doi:05810.01029/02010GL046270.
Arora, V. K., et al., 2009: The Effect of Terrestrial Photosynthesis Down Regulation on the Twentieth-Century Carbon Budget Simulated with the CCCma Earth System Model. *Journal of Climate*, **22**, 6066-6088.
Baehr , J., A. Stroup, and J. Marotzke, 2009: Testing concepts for continuous monitoring of the meridional overturning circulation in the South Atlantic. *Ocean Modelling*, **29** 147-153.
Baehr, J., S. Cunnningham, H. Haak, P. Heimbach, T. Kanzow, and J. Marotzke, 2009: Observed and simulated estimates of the meridional overturning circulation at 26.5 N in the Atlantic. *Ocean Science Discussions*, **6**, 1333-1367.
Bailey, D., M. Holland, E. Hunke, B. Lipscomb, B. Briegleb, B. Bitz, and J. Schramm, 2010: Community ice code (cice) user's guide, version 4.0, released with cesm1.0. National Center for Atmospheric Research.
Balan Sarojini, B., et al., 2011: High frequency variability of the Atlantic meridional overturning circulation. *Ocean Science Discussions*, **8**, 219-246.
Baldocchi, D., et al., 2001: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, **82**, 2415-2434.
Baldwin, M. P., et al., 2001: The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179-229.
Balmaseda, M. A., A. Vidard, and D. L. T. Anderson, 2008: The ECMWF Ocean Analysis System: ORA-S3. *Monthly Weather Review*, **136**, 3018-3034.
Barker, H. W., J. N. S. Cole, J. J. Morcrette, R. Pincus, P. Raisaenen, K. von Salzen, and P. A. Vaillancourt, 2008: The Monte Carlo Independent Column Approximation: An assessment using several global atmospheric models. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1463-1478.
Barnes, E. A., and D. L. Hartmann, 2010: Influence of eddy-driven jet latitude on North Atlantic jet persistence and blocking frequency in CMIP3 integrations. *Geophys. Res. Lett.*, **37**, L23802-.
Barnes, E. A., J. Slingo, and T. Woollings, 2011: A methodology for the comparison of blocking climatologies across indices, models and climate scenarios. *Climate Dynamics*, **in press**.
Barnier, B., et al., 2006: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dyn.*, **56**, 543-567.
Barriopedro, D., R. GarcÃ-a-Herrera, and R. Trigo, 2010a: Application of blocking diagnosis methods to General Circulation Models. Part I: a novel detection scheme. 1373-1391.
Barriopedro, D., R. GarcÃ-a-Herrera, J. GonzÃ¡lez-Rouco, and R. Trigo, 2010b: Application of blocking diagnosis methods to General Circulation Models. Part II: model simulations. 1393-1409.
Barsugli, J. J., S.-I. Shin, and P. D. Sardeshmukh, 2006: Sensitivity of global warming to the pattern of tropical ocean warming. *Climate Dynamics*, **27**, 483-492.
Bartlein, P. J., et al., 2010a: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics*, 10.1007/s00382-010-0904-1.
Bartlein, P. J., et al., 2010b: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics*, 10.1007/s00382-010-0904-1.
Bathiany, S., M. Claussen, V. Brovkin, T. Raddatz, and V. Gayler, 2010: Combined biogeophysical and biogeochemical effects of large-scale forest cover changes in the MPI earth system model. *Biogeosciences*, **7**, 1383-1399.
Bauer, S. E., et al., 2008: MATRIX (Multiconfiguration Aerosol TRacker of mIXing state): an aerosol microphysical module for global atmospheric models. *Atmospheric Chemistry and Physics*, **8**, 6003-6035.
Beare, R., et al., 2006: An Intercomparison of Large-Eddy Simulations of the Stable Boundary Layer. *Boundary-Layer Meteorology*, **118**, 247-272-272.
Bechtold, P., et al., 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1337-1351.
Behera, S. K., and T. Yamagata, 2001: Subtropical SST dipole events in the southern Indian ocean. *Geophysical Research Letters*, **28**, 327-330.

Bell, M. J., M. J. Martin, and N. K. Nichols, 2004: Assimilation of data into an ocean model with systematic errors near the equator. *Quarterly Journal of the Royal Meteorological Society*, **130**, 873-893.
Bellassen, V., G. Le Maire, J. F. Dhote, P. Ciais, and N. Viovy, 2010: Modelling forest management within a global vegetation model Part 1: Model structure and general behaviour. *Ecological Modelling*, **221**, 2458-2474.
Bellassen, V., G. le Maire, O. Guin, J. F. Dhote, P. Ciais, and N. Viovy, 2011: Modelling forest management within a global vegetation model-Part 2: Model validation from a tree to a continental scale. *Ecological Modelling*, **222**, 57-75.
Bellucci, A., S. Gualdi, and A. Navarra, 2010: The Double-ITCZ Syndrome in Coupled General Circulation Models: The Role of Large-Scale Vertical Circulation Regimes. *Journal of Climate*, **23**, 1127-1145.
Bengtsson, L., and K. Hodges, 2011: On the evaluation of temperature trends in the tropical troposphere. *Climate Dynamics*, **36**, 419-430.
Bernie, D. J., E. Guilyardi, G. Madec, J. M. Slingo, S. Woolnough, and J. Cole, 2008: Impact od resolving the diurnal cycle in an ocean-atmosphere GCM. Part 2: A fiurnally coupled CGCM. *Climate Dynamics*, **31**, 909-925.
Betts, A. K., and C. Jakob, 2002: Study of the diurnal cycle of convective precipitation over Amazonia using a single column model. *J. Geophys. Res.*, **107**, 4732.
Biastoch, A., C. W. Boening, and J. R. E. Lutjeharms, 2008a: Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation. *Nature*, **456**, 489-492.
Biastoch, A., C. W. BÃ¶ning, J. Getzlaff, J.-M. Molines, and G. Madec, 2008b: Causes of Interannual-Decadal Variability in the Meridional Overturning Circulation of the Midlatitude North Atlantic Ocean. *Journal of Climate*, **21**, 6599-6615.
Bingham, R. J., C. W. Hughes, V. Roussenov, and R. G. Williams, 2007: Meridional coherence of the North Atlantic meridional overturning circulation. *Geophys. Res. Lett.*, **34**, L23606-.
Bitz, C. M., and W. H. Lipscomb, 1999: An energy-conserving thermodynamic sea ice model for climate study. *Journal of Geophysical Research. Oceans*, **104**, 15,669-615,677.
Bitz, C. M., J. C. Fyfe, and G. M. Flato, 2002: Sea ice response to wind forcing from AMIP models. *Journal of Climate*, **15**, 522-536.
Blanchard-Wrigglesworth, E., C. Bitz, and M. Holland, 2011: Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, **38**, -.
Blanke, B., and S. Raynaud, 1997: Kinematics of the Pacific Equatorial Undercurrent: An Eulerian and Lagrangian Approach from GCM Results. *Journal of Physical Oceanography*, **27**, 1038-1053.
Boberg, F., and J. H. Christensen, 2011: Confidence in Lowered Southern Europe Temperature Projections due to Model Deficiencies. *Nature Climate Change*, **In review.**
Boccaletti, G., R. Ferrari, and B. Fox-Kemper, 2007: Mixed layer instabilities and restratification. *Journal of Physical Oceanography*, **37**, 2228-2250.
Bodas-Salcedo, A., M. Webb, M. Brooks, M. Ringer, K. Williams, S. Milton, and D. Wilson, 2008: Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities. *Journal of Geophysical Research-Atmospheres*, ARTN D00A13, DOI 10.1029/2007JD009620. -.
Bodas-Salcedo, A., et al., 2011: COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, **92**, 1023-1043.
Boe, J., A. Hall, and X. Qu, 2009a: Deep ocean heat uptake as a major source of spread in transient climate change simulations. *Geophysical Research Letters*, ARTN L22701, DOI 10.1029/2009GL040845. -.
Boe, J., A. Hall, and X. Qu, 2009b: September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience*, DOI 10.1038/NGEO467. 341-343.
Boe, J., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, **27**, 1643-1655.
Boer, G. J., and B. Yu, 2003: Climate sensitivity and climate state. *Climate Dynamics*, **21**, 167-176.
Boer, G. J., and K. Hamilton, 2008: QBO influence on extratropical predictive skill. *Climate Dynamics*, **31**, 987-1000.
Bonan, G. B., 2008: Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, **320**, 1444-1449.
Bond, T. C., et al., 2007: Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850-2000. *Global Biogeochemical Cycles*, **21**, 16.

Bondeau, A., P. C. Smith, S. Zaehle, S. Schaphoff, W. Lucht, W. Cramer, and D. Gerten, 2007: Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, **13**, 679-706.

Boning, C. W., A. Dispert, M. Visbeck, S. R. Rintoul, and F. U. Schwarzkopf, 2008: The response of the Antarctic Circumpolar Current to recent climate change. *Nature Geoscience*, **1**, 864-869.

Bony, S., and J. L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models - art. no. L20806. *Geophysical Research Letters*, **32**, 20806-20806.

Boone, A., et al., 2009: THE AMMA LAND SURFACE MODEL INTERCOMPARISON PROJECT (ALMIP). *Bulletin of the American Meteorological Society*, **90**, 1865-1880.

Boone, A. A., I. Poccard-Leclercq, Y. K. Xue, J. M. Feng, and P. de Rosnay, 2010: Evaluation of the WAMME model surface fluxes using results from the AMMA land-surface model intercomparison project. *Climate Dynamics*, **35**, 127-142.

Booth, B., P. Halloran, and N. Dunstone, 2011: External forcing of North Atlantic Sea Surface Temperatures. *Nature (submitted)*.

Boyle, J., and S. A. Klein, 2010: Impact of horizontal resolution on climate model forecasts of tropical precipitation and diabatic heating for the TWP-ICE period. *J. Geophys. Res.*, **115**, D23113.

Boyle, J., S. Klein, G. Zhang, S. Xie, and X. Wei, 2008: Climate Model Forecast Experiments for TOGA COARE. *Monthly Weather Review*, **136**, 808-832.

Boyle, J. S., et al., 2005: Diagnosis of Community Atmospheric Model 2 (CAM2) in numerical weather forecast configuration at Atmospheric Radiation Measurement sites. *J. Geophys. Res.*, **110**, D15S15.

Braconnot, P., F. Hourdin, S. Bony, J. Dufresne, J. Grandpeix, and O. Marti, 2007a: Impact of different convective cloud schemes on the simulation of the tropical seasonal cycle in a coupled oceanâ□“atmosphere model. *Climate Dynamics*, **29**, 501-520.

——, 2007b: Impact of different convective cloud schemes on the simulation of the tropical seasonal cycle in a coupled ocean-atmosphere model. *Climate Dynamics*, **29**, 501-520.

Braconnot, P., et al., 2007c: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 2: feedbacks with emphasis on the location of the ITCZ and mid- and high latitudes heat budget. *Climate of the Past*, **3**, 279-296.

Braconnot, P., et al., 2007d: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 1: experiments and large-scale features. *Climate of the Past*, **3**, 261-277.

Brands, S., J. Taboada, A. Cofino, T. Sauter, and C. Schneider, 2011: Statistical downscaling of daily temperatures in the NW Iberian Peninsula from global climate models: validation and future scenarios. *Climate Research*, **48**, 163-176.

Breugem, W. P., W. Hazeleger, and R. J. Haarsma, 2006: Multimodel study of tropical Atlantic variability and change. *Geophysical Research Letters*, **33**.

Brewer, S., J. Guiot, and F. Torre, 2007: Mid-Holocene climate change in Europe: a data-model comparison. *Climate of the Past*, **3**, 499-512.

Briegleb, B. P., and B. Light, 2007: A Delta-Eddington multiple scattering parameterization for solar radiation in the sea ice component of the Community Climate System Model. National Center for Atmospheric Research.

Brierley, C., M. Collins, and A. Thorpe, 2010a: The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. *Climate Dynamics*, DOI 10.1007/s00382-008-0486-3. 325-343.

Brierley, C. M., M. Collins, and A. J. Thorpe, 2010b: The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. *Climate Dynamics*, **34**, 325-343.

Brogniez, H., and R. T. Pierrehumbert, 2007: Intercomparison of tropical tropospheric humidity in GCMs with AMSU-B water vapor data. *Geophysical Research Letters*, **34**.

Brogniez, H., R. Roca, and L. Picon, 2005: Evaluation of the distribution of subtropical free tropospheric humidity in AMIP-2 simulations using METEOSAT water vapor channel data. *Geophys. Res. Lett.*, **32**, L19708.

Brovkin, V., J. Bendtsen, M. Claussen, A. Ganopolski, C. Kubatzki, V. Petoukhov, and A. Andreev, 2002: Carbon cycle, vegetation, and climate dynamics in the Holocene: Experiments with the CLIMBER-2 model. *Global Biogeochemical Cycles*, **16**.

Brown, J., A. Fedorov, and E. Guilyardi, 2010a: How well do coupled models replicate ocean energetics relevant to ENSO? *Climate Dynamics*. 1-12.

Brown, J. R., C. Jakob, and J. M. Haynes, 2010b: An Evaluation of Rainfall Frequency and Intensity over the Australian Region in a Global Climate Model. *Journal of Climate*, **23**, 6504-6525.

Bryan, F. O., M. W. Hecht, and R. D. Smith, 2007: Resolution convergence and sensitivity studies with North Atlantic circulation models. Part I: The western boundary current system. *Ocean Modelling*, **16**, 141-159.
Bryan, F. O., R. Tomas, J. M. Dennis, D. B. Chelton, N. G. Loeb, and J. L. McClean, 2010: Frontal Scale Air-Sea Interaction in High-Resolution Coupled Climate Models. *Journal of Climate*, **23**, 6277-6291.
Bryden, H. L., H. R. Longworth, and S. A. Cunningham, 2005: Slowing of the Atlantic meridional overturning circulation at 25N. *Nature*, **438**, 655-657.
Buehler, T., C. C. Raible, and T. F. Stocker, 2011: The relationship of winter season North Atlantic blocking frequencies to extreme cold or dry spells in the ERA-40. *Tellus A*, **63**, 212-222.
Burton, A., H. Fowler, S. Blenkinsop, and C. Kilsby, 2010: Downscaling transient climate change using a Neyman-Scott Rectangular Pulses stochastic rainfall model. *Journal of Hydrology*, **381**, 18-32.
Buser, C., H. Kunsch, and C. Schar, 2010: Bayesian multi-model projections of climate: generalization and application to ENSEMBLES results. *Climate Research*, DOI 10.3354/cr00895. 227-241.
Buser, C. M., H. R. Kunsch, D. Luthi, M. Wild, and C. Schar, 2009: Bayesian multi-model projection of climate: bias assumptions and interannual variability. *Climate Dynamics*, **33**, 849-868.
Butchart, N., A. A. Scaife, J. Austin, S. H. E. Hare, and J. R. Knight, 2003: Quasi-biennial oscillation in ozone in a coupled chemistry-climate model. *J. Geophys. Res.*, **108**, ACL14-11-ACL14-ACL14-10.
Butchart, N., et al., 2010: Chemistry-Climate Model Simulations of Twenty-First Century Stratospheric Climate and Circulation Changes. *Journal of Climate*, **23**, 5349-5374.
Cadule, P., et al., 2010: Benchmarking coupled climate-carbon models against long-term atmospheric CO2 measurements. *Global Biogeochemical Cycles*, **24**.
Cai, W., A. Sullivan, and T. Cowan, 2011: Interactions of ENSO, the IOD, and the SAM in CMIP3 Models. *Journal of Climate*, **24**, 1688-1704.
Calov, R., A. Ganopolski, V. Petoukhov, M. Claussen, and R. Greve, 2002: Large-scale instabilities of the Laurentide ice sheet simulated in a fully coupled climate-system model. *Geophysical Research Letters*, **29**.
Capotondi, A., A. Wittenberg, and S. Masina, 2006: Spatial and temporal structure of Tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modelling*, **15**, 274–298.
Carslaw, K. S., O. Boucher, D. V. Spracklen, G. W. Mann, J. G. L. Rae, S. Woodward, and M. Kulmala, 2010: A review of natural aerosol interactions and feedbacks within the Earth system. *Atmospheric Chemistry and Physics*, **10**, 1701-1737.
Cassou, C., 2008: Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. *Nature*, **455**, 523-527.
Castro, C. L., R. A. Pielke, and G. Leoncini, 2005: Dynamical downscaling: Assessment of value retained and added using the regional atmospheric modeling system (RAMS). *Journal of Geophysical Research-Atmospheres*, **110**.
CCMVal, S., 2010: SPARC Report on the Evaluation of Chemistry-Climate Models.
Chang, C. Y., S. Nigam, and J. A. Carton, 2008: Origin of the springtime westerly bias in equatorial Atlantic surface winds in the Community Atmosphere Model version 3 (CAM3) simulation. *Journal of Climate*, **21**, 4766-4778.
Chang, C. Y., J. A. Carton, S. A. Grodsky, and S. Nigam, 2007: Seasonal climate of the tropical Atlantic sector in the NCAR community climate system model 3: Error structure and probable causes of errors. *Journal of Climate*, **20**, 1053-1070.
Chang, C. Y., J. C. H. Chiang, M. F. Wehner, A. R. Friedman, and R. Ruedy, 2011: Sulfate Aerosol Control of Tropical Atlantic Climate over the Twentieth Century. *Journal of Climate*, **24**, 2540-2555.
Chen, Y. H., and A. D. Del Genio, 2009: Evaluation of tropical cloud regimes in observations and a general circulation model. *Climate Dynamics*, **32**, 355-369.
Chiang, J. C. H., and D. J. Vimont, 2004: Analogous Pacific and Atlantic meridional modes of tropical atmosphere-ocean variability. *Journal of Climate*, **17**, 4143-4158.
Chikira, M., 2010: A Cumulus Parameterization with State-Dependent Entrainment Rate. Part II: Impact on Climatology in a General Circulation Model. *Journal of the Atmospheric Sciences*, **67**, 2194-2211.
Chikira, M., and M. Sugiyama, 2010: A Cumulus Parameterization with State-Dependent Entrainment Rate. Part I: Description and Sensitivity to Temperature and Humidity Profiles. *Journal of the Atmospheric Sciences*, **67**, 2171-2193.
Christensen, J., F. Boberg, O. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, **35**, -.

Christensen, J., E. Kjellstrom, F. Giorgi, G. Lenderink, and M. Rummukainen, 2010: Weight assignment in regional climate models. *Climate Research*, DOI 10.3354/cr00916. 179-194.
Christensen, J., B. Machenhauer, R. Jones, C. Schar, P. Ruti, M. Castro, and G. Visconti, 1997: Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions. *Climate Dynamics*, **13**, 489-506.
Christian, J. R., et al., 2010: The global carbon cycle in the Canadian Earth system model (CanESM1): Preindustrial control simulation. *Journal of Geophysical Research-Biogeosciences*, **115**.
Christidis, N., P. A. Stott, and S. J. Brown, 2011: The Role of Human Activity in the Recent Warming of Extremely Warm Daytime Temperatures. *Journal of Climate*, **24**, 1922-1930.
Christy, J. R., et al., 2010: What Do Observational Datasets Say about Modeled Tropospheric Temperature Trends since 1979? *Remote Sensing*, **2**, 2148-2169.
Chu, J., J. Xia, C. Xu, and V. Singh, 2010: Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theoretical and Applied Climatology*, **99**, 149-161.
Chung, E. S., D. Yeomans, and B. J. Soden, 2010a: An assessment of climate feedback processes using satellite observations of clear-sky OLR. *Geophysical Research Letters*, **37**.
Chung, E. S., B. J. Soden, and B. J. Sohn, 2010b: Revisiting the determination of climate sensitivity from relationships between surface temperature and radiative fluxes. *Geophysical Research Letters*, **37**.
Cionni, I., et al., 2011: Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing. *Atmospheric Chemistry and Physics Discussion*, **11**, 10875-10933.
Claussen, M., et al., 2002: Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate Dynamics*, **18**, 579-586.
Clement, A. C., R. Burgman, and J. R. Norris, 2009: Observational and Model Evidence for Positive Low-Level Cloud Feedback. *Science*, **325**, 460-464.
Clement, A. C., R. Seager, M. A. Cane, and S. E. Zebiak, 1996: An ocean dynamical thermostat. *Journal of Climate*, **9**, 2190-2196.
CLIMAP, 1981: Seasonal reconstructions of the Earth's surface at the last glacial maximum. geological Society of America, Boulder, Colorado. Map Series Technical Report MC-36.
Colberg, F., C. J. C. Reason, and K. Rodgers, 2004: South Atlantic response to El Nino Southern Oscillation induced climate variability in an ocean general circulation model. *Journal of Geophysical Research-Oceans*, **109**.
Collins, M., C. M. Brierley, M. MacVean, B. B. B. Booth, and G. R. Harris, 2007: The sensitivity of the rate of transient climate change to ocean physics perturbations. *Journal of Climate*, **20**, 2315-2320.
Collins, M., B. Booth, G. Harris, J. Murphy, D. Sexton, and M. Webb, 2006a: Towards quantifying uncertainty in transient climate change. *Climate Dynamics*, DOI 10.1007/s00382-006-0121-0. 127-147.
Collins, M., B. Booth, B. Bhaskaran, G. Harris, J. Murphy, D. Sexton, and M. Webb, 2010: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Climate Dynamics*, 10.1007/s00382-010-0808-0. 1-30.
Collins, N., et al., 2005: Design and implementation of components in the Earth system modeling framework. *International Journal of High Performance Computing Applications*, **19**, 341-350.
Collins, W. D., et al., 2006b: The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**, 2122-2143.
Compo, G. P., and P. D. Sardeshmukh, 2009: Oceanic influences on recent continental warming. *Climate Dynamics*, **32**, 333-342.
Connolley, W., and T. Bracegirdle, 2007: An antarctic assessment of IPCC AR4 coupled models. *Geophysical Research Letters*, ARTN L22505, DOI 10.1029/2007GL031648. -.
Coppola, E., F. Kucharski, F. Giorgi, and F. Molteni, 2005: Bimodality of the North Atlantic Oscillation in simulations with greenhouse gas forcing. *Geophys. Res. Lett.*, **32**, L23709-.
Coppola, E., F. Giorgi, S. Rauscher, and C. Piani, 2010: Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models. *Climate Research*, DOI 10.3354/cr00940. 121-134.
Cordero, E. C., and P. M. D. Forster, 2006: Stratospheric variability and trends in models used for the IPCC AR4. *Atmospheric Chemistry and Physics*, **6**, 5369-5380.
Covey, C., A. G. Dai, D. Marsh, and R. S. Lindzen, 2011: The Surface-Pressure Signature of Atmospheric Tides in Modern Climate Models. *Journal of the Atmospheric Sciences*, **68**, 495-514.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell, 2000: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**, 184-187.
Cox, P. M., R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree, and J. Smith, 1999: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, **15**, 183-203.
Cramer, W., et al., 2001: Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357-373.
Crucifix, M., 2006: Does the Last Glacial Maximum constrain climate sensitivity? *Geophysical Research Letters*, **33**, -.
Cunningham, S., et al.: The present and future system for measuring the Atlantic meridional overturning circulation and heat transport. [Available online at http://www.oceanobs09.net/blog/?p=69.]
——, 2010: The present and future system for measuring the Atlantic meridional overturning circulation and heat transport. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2)*, Venice, Italy, 21-25 September 2009, ESA Publication.
Cunningham, S. A., S. G. Alderson, B. A. King, and M. A. Brandon, 2003: Transport and variability of the Antarctic Circumpolar Current in Drake Passage. *Journal of Geophysical Research-Oceans*, **108**.
Cunningham, S. A., et al., 2007: Temporal Variability of the Atlantic Meridional Overturning Circulation at 26.5°N. *Science*, **317**, 935-938.
Curry, W. B., and D. W. Oppo, 2005: Glacial water mass geometry and the distribution of delta C-13 of Sigma CO2 in the western Atlantic Ocean. *Paleoceanography*, **20**, -.
Cuxart, J., et al., 2006: Single-Column Model Intercomparison for a Stably Stratified Atmospheric Boundary Layer. *Boundary-Layer Meteorology*, **118**, 273-303-303.
Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, **19**, 4605-4630.
Dai, A., and K. E. Trenberth, 2004: The diurnal cycle and its depiction in the Community Climate System Model. *Journal of Climate*, **17**, 930-951.
Dall'Amico, M., L. J. Gray, K. H. Rosenlof, A. A. Scaife, K. P. Shine, and P. A. Stott, 2010: Stratospheric temperature trends: impact of ozone variability and the QBO. *Climate Dynamics*, **34**, 381-398.
Dallmeyer, A., M. Claussen, and J. Otto, 2010: Contribution of oceanic and vegetation feedbacks to Holocene climate change in monsoonal Asia. *Climate of the Past*, **6**, 195-218.
Danabasoglu, G., and P. R. Gent, 2009: Equilibrium Climate Sensitivity: Is It Accurate to Use a Slab Ocean Model? *Journal of Climate*, **22**, 2494-2499.
Danabasoglu, G., R. Ferrari, and J. C. McWilliams, 2008: Sensitivity of an ocean general circulation model to a parameterization of near-surface eddy fluxes. *Journal of Climate*, **21**, 1192-1208.
Danabasoglu, G., W. G. Large, and B. P. Briegleb, 2010: Climate impacts of parameterized Nordic Sea overflows. *Journal of Geophysical Research-Oceans*, **115**.
Danabasoglu, G., W. G. Large, J. J. Tribbia, P. R. Gent, B. P. Briegleb, and J. C. McWilliams, 2006: Diurnal coupling in the tropical oceans of CCSM3. *Journal of Climate*, **19**, 2347-2365.
Davis, B. A. S., and S. Brewer, 2009: Orbital forcing and role of the latitudinal insolation/temperature gradient. *Climate Dynamics*, **32**, 143-165.
de Jong, M. F., S. S. Drijfhout, W. Hazeleger, H. M. van Aken, and C. A. Severijns, 2009: Simulations of Hydrographic Properties in the Northwestern North Atlantic Ocean in Coupled Climate Models. *Journal of Climate*, **22**, 1767-1786.
Delcroix, T., G. Alory, S. Cravatte, T. Correge, and M. J. McPhaden, 2010: A gridded sea surface salinity data set for the tropical Pacific with sample applications (1950-2008). *Deep Sea Research Part I: Oceanographic Research Papers*.
Delworth, T. L., et al., 2011: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*. submitted.
Demott, C. A., D. A. Randall, and M. Khairoutdinov, 2007: Convective precipitation variability as a tool for general circulation model analysis. *Journal of Climate*, **20**, 91-112.
——, 2010: Implied Ocean Heat Transports in the Standard and Superparameterized Community Atmospheric Models. *Journal of Climate*, **23**, 1908-1928.
Deque, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, **57**, 16-26.
Deque, M., 2010: Regional climate simulation with a mosaic of RCMs. *Meteorologische Zeitschrift*, **19**, 259-266.

Deque, M., et al., 2005: Global high resolution versus Limited Area Model climate change projections over Europe: quantifying confidence level from PRUDENCE results. *Climate Dynamics*, **25**, 653-670.
Déqué, M., et al., 2007: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. 53-70.
Derbyshire, S. H., A. V. Maidens, S. F. Milton, R. A. Stratton, and M. R. Willett, 2011: Adaptive detrainment in a convective parametrization. *Quarterly Journal of the Royal Meteorological Society*. n/a-n/a.
Derbyshire, S. H., I. Beau, P. Bechtold, J. Y. Grandpeix, J. M. Piriou, J. L. Redelsperger, and P. M. M. Soares, 2004: Sensitivity of moist convection to environmental humidity. *Quarterly Journal of the Royal Meteorological Society*, **130**, 3055-3079.
Deser, C., A. S. Phillips, and J. W. Hurrell, 2004: Pacific interdecadal climate variability: Linkages between the tropics and the North Pacific during boreal winter since 1900. *Journal of Climate*, **17**, 3109-3124.
Deser, C., A. S. Phillips, V. Bourdette, and H. Teng, 2011a: Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, 10.1007/s00382-010-0977-x.
Deser, C., et al., 2011b: ENSO and Pacific Decadal Variability in Community Climate System Model Version 4. *J. Climate (in press)*.
Domingues, C., J. Church, N. White, P. Gleckler, S. Wijffels, P. Barker, and J. Dunn, 2008: Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, DOI 10.1038/nature07080. 1090-U1096.
Donat, M., G. Leckebusch, S. Wild, and U. Ulbrich, 2010: Benefits and limitations of regional multi-model ensembles for storm loss estimations. *Climate Research*, DOI 10.3354/cr00891. 211-225.
Doney, S. C., et al., 2009: Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *Journal of Marine Systems*, **76**, 95-112.
Dorn, W., K. Dethloff, and A. Rinke, 2009: Improved simulation of feedbacks between atmosphere and sea ice over the Arctic Ocean in a coupled regional climate model. *Ocean Modelling*, **29**, 103-114.
Doscher, R., K. Wyser, H. E. M. Meier, M. W. Qian, and R. Redler, 2010: Quantifying Arctic contributions to climate predictability in a regional coupled ocean-ice-atmosphere model. *Climate Dynamics*, **34**, 1157-1176.
Douglass, D., J. Christy, B. Pearson, and S. Singer, 2008: A comparison of tropical temperature trends with model predictions. *International Journal of Climatology*, DOI 10.1002/joc.1651. 1693-1701.
Driouech, F., M. Deque, and A. Mokssit, 2009: Numerical simulation of the probability distribution function of precipitation over Morocco. *Climate Dynamics*, **32**, 1055-1063.
Driouech, F., M. Deque, and E. Sanchez-Gomez, 2010: Weather regimes-Moroccan precipitation link in a regional climate change simulation. *Global and Planetary Change*, **72**, 1-10.
Druyan, L. M., et al., 2010: The WAMME regional model intercomparison study. *Climate Dynamics*, **35**, 175-192.
Du, Y., S.-P. Xie, G. Huang, and K. Hu, 2009: Role of Air-Sea Interaction in the Long Persistence of El Nino-Induced North Indian Ocean Warming. *Journal of Climate*, **22**, 2023-2038.
Ducet, N., P. Y. Le Traon, and G. Reverdin, 2000: Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2. *Journal of Geophysical Research-Oceans*, **105**, 19477-19498.
Dufresne, J.-L., and a. co-authors, 2011: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Submitted ed., Clim. Dyn.
Dufresne, J. L., and S. Bony, 2008: An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. *Journal of Climate*, **21**, 5135-5144.
Dunkerton, T. J., 1991: NONLINEAR PROPAGATION OF ZONAL WINDS IN AN ATMOSPHERE WITH NEWTONIAN COOLING AND EQUATORIAL WAVEDRIVING. *Journal of the Atmospheric Sciences*, **48**, 236-263.
Dunkerton, T. J., 1997: The role of gravity waves in the quasi-biennial oscillation. *J. Geophys. Res.*, **102**, 26053-26076.
Duplessy, J. C., N. J. Shackleton, R. Fairbanks, L. Labeyrie, D. Oppo, and N. Kallel, 1988: Deep water source variation during the last climatic cycle and their impact on th global deep water circulation. *Paleoceanography*, **3**, 343-360.
Eby, M., K. Zickfeld, A. Montenegro, D. Archer, K. J. Meissner, and A. J. Weaver, 2009: Lifetime of Anthropogenic Climate Change: Millennial Time Scales of Potential CO(2) and Surface Temperature Perturbations. *Journal of Climate*, **22**, 2501-2511.
Edwards, N., and R. Marsh, 2005: Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model. *Climate Dynamics*, **24**, 415-433.

Edwards, T. L., M. Crucifix, and S. P. Harrison, 2007: Using the past to constrain the future: how the palaeorecord can improve estimates of global warming. *Progress in Physical Geography*, **31**, 481-500.
Emanuel, K., R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming - Results from downscaling IPCC AR4 simulations. *Bulletin of the American Meteorological Society*, **89**, 347-+.
Ern, M., and P. Preusse, 2009: Wave fluxes of equatorial Kelvin waves and QBO zonal wind forcing derived from SABER and ECMWF temperature space-time spectra. *Atmospheric Chemistry and Physics*, **9**, 3957–3986.
Evan, A. T., D. J. Vimont, A. K. Heidinger, J. P. Kossin, and R. Bennartz, 2009: The Role of Aerosols in the Evolution of Tropical North Atlantic Ocean Temperature Anomalies. *Science*, **324**, 778-781.
Eyring, V., et al., 2010: Transport impacts on atmosphere and climate: Shipping. *Atmospheric Environment*, **44**, 4735-4771.
Eyring, V., et al., 2005: A strategy for process-oriented validation of coupled chemistry-climate models. *Bulletin of the American Meteorological Society*, DOI 10.1175/BAMS-86-8-1117. 1117-+.
Eyring, V., et al., 2006: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *Journal of Geophysical Research-Atmospheres*, **111**.
Eyring, V., et al., 2007: Multimodel projections of stratospheric ozone in the 21st century. *Journal of Geophysical Research-Atmospheres*, ARTN D16303, DOI 10.1029/2006JD008332. -.
Faloona, I., 2009: Sulfur processing in the marine atmospheric boundary layer: A review and critical assessment of modeling uncertainties. *Atmospheric Environment*, **43**, 2841-2854.
Fanning, A. F., and A. J. Weaver, 1996: An atmospheric energy-moisture balance model: climatology, interpentadal climate change, and coupling to an ocean general circulation model. *J. Geophys. Res.*, **101**, 15111-15128.
Farneti, R., and G. K. Vallis, 2011: Mechanisms of interdecadal climate variability and the role of ocean-atmosphere coupling. *Climate Dynamics*, **36**, 289-308.
Farneti, R., and P. R. Gent, 2011: The effects of the eddy-induced advection coefficient in a coarse-resolution coupled climate model. *Ocean Modelling*, **39**, 135-145.
Farneti, R., T. L. Delworth, A. J. Rosati, S. M. Griffies, and F. R. Zeng, 2010: The Role of Mesoscale Eddies in the Rectification of the Southern Ocean Response to Climate Change. *Journal of Physical Oceanography*, **40**, 1539-1557.
Fauchereau, N., S. Trzaska, Y. Richard, P. Roucou, and P. Camberlin, 2003: Sea-surface temperature co-variability in the southern Atlantic and Indian Oceans and its connections with the atmospheric circulation in the southern hemisphere. *International Journal of Climatology*, **23**, 663-677.
Feng, J., and C. Fu, 2006: Inter-comparison of 10-year precipitation simulated by several RCMs for Asia. *Advances in Atmospheric Sciences*, DOI 10.1007/s00376-006-0531-2. 531-542.
Ferrari, R., J. C. McWilliams, V. M. Canuto, and M. Dubovikov, 2008: Parameterization of eddy fluxes near oceanic boundaries. *Journal of Climate*, **21**, 2770-2789.
Ferrari, R., S. M. Griffies, A. J. G. Nurser, and G. K. Vallis, 2010: A boundary-value problem for the parameterized mesoscale eddy transport. *Ocean Modelling*, **32**, 143-156.
Feser, F., 2006: Enhanced Detectability of Added Value in Limited-Area Model Results Separated into Different Spatial Scales. *Monthly Weather Review*, **134**, 2180-2190.
Fetterer, F., K. Knowles, W. Meier, and M. Savoie, 2002: Sea Ice Index. National Snow and Ice data Center.
Fichefet, T., and M. A. Morales Maqueda, 1997: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res.*, **102**, 12,646.
Field, P. R., A. Gettelman, R. B. Neale, R. Wood, P. J. Rasch, and H. Morrison, 2008: Midlatitude Cyclone Compositing to Constrain Climate Model Behavior Using Satellite Observations. *Journal of Climate*, **21**, 5887-5903.
Fischer, E. M., S. I. Seneviratne, D. Lüthi, and C. Schär, 2007a: Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophys. Res. Lett.*, **34**, L06707.
Fischer, E. M., S. I. Seneviratne, D. Luthi, and C. Schar, 2007b: Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*, **34**.
Flato, G., 2011: Earth system models: an overview. John Wiley and Sons, 783-800.
Flocco, D., D. L. Feltham, and A. K. Turner, 2010: Incorporation of a physically based melt pond scheme into the sea ice component of a climate model. *J. Geophys. Res.*, **115**, C08012.
Fogt, R. L., J. Perlwitz, A. J. Monaghan, D. H. Bromwich, J. M. Jones, and G. J. Marshall, 2009: Historical SAM Variability. Part II: Twentieth-Century Variability and Trends from Reconstructions, Observations, and the IPCC AR4 Models. *Journal of Climate*, **22**, 5346-5365.

Folland, C. K., J. Knight, H. W. Linderholm, D. Fereday, S. Ineson, and J. W. Hurrell, 2009: The Summer North Atlantic Oscillation: Past, Present, and Future. *Journal of Climate*, **22**, 1082-1103.
Foltz, G. R., and M. J. McPhaden, 2008: Trends in Saharan dust and tropical Atlantic climate during 1980-2006. *Geophysical Research Letters*, **35**.
Fontaine, B., and S. Janicot, 1996: Sea Surface Temperature Fields Associated with West African Rainfall Anomaly Types. *Journal of Climate*, **9**, 2935-2940.
Forest, C. E., P. H. Stone, and A. P. Sokolov, 2006: Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophysical Research Letters*, **33**.
——, 2008: Constraining climate model parameters from observed 20th century changes. *Tellus Series a-Dynamic Meteorology and Oceanography*, **60**, 911-920.
Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**, 113-117.
Fowler, H., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, **27**, 1547-1578.
Fox-Kemper, B., R. Ferrari, and R. Hallberg, 2008: Parameterization of mixed layer eddies. Part I: Theory and diagnosis. *Journal of Physical Oceanography*, **38**, 1145-1165.
Fox-Kemper, B., et al., 2011: Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate simulations. *Ocean Modelling*, **39**, 61-78.
Fox-Rabinovitz, M., J. Cote, B. Dugas, M. Deque, J. McGregor, and A. Belochitski, 2008: Stretched-grid Model Intercomparison Project: decadal regional climate simulations with enhanced variable and uniform-resolution GCMs. *Meteorology and Atmospheric Physics*, DOI 10.1007/s00703-008-0301-z. 159-177.
Frankcombe, L. M., A. von der Heydt, and H. A. Dijkstra, 2010: North Atlantic Multidecadal Climate Variability: An Investigation of Dominant Time Scales and Processes. *Journal of Climate*, **23**, 3626-3638.
Friedlingstein, P., et al., 2001: Positive feedback between future climate change and the carbon cycle. *Geophysical Research Letters*, **28**, 1543-1546.
Friedlingstein, P., et al., 2006: Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison. *Journal of Climate*, **19**, 3337-3353.
Friedrichs, M. A. M., et al., 2009: Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *Journal of Marine Systems*, **76**, 113-133.
Fu, Q., S. Manabe, and C. M. Johanson, 2011: On the warming in the tropical upper troposphere: Models versus observations. *Geophys. Res. Lett.*, **38**, L15704.
Furrer, R., R. Knutti, S. Sain, D. Nychka, and G. Meehl, 2007: Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis. *Geophysical Research Letters*, ARTN L06711, DOI 10.1029/2006GL027754. -.
Fyfe, J. C., N. P. Gillett, and D. W. J. Thompson, 2010: Comparing variability and trends in observed and modelled global-mean surface temperature. *Geophysical Research Letters*, **37**.
Fyke, J. G., A. J. Weaver, D. Pollard, M. Eby, L. Carter, and A. Mackintosh, 2011: A new coupled ice sheet/climate model: description and sensitivity to model physics under Eemian, Last Glacial Maximum, late Holocene and modern climate conditions. *Geoscientific Model Development*, **4**, 117-136.
Galbraith, D., P. E. Levy, S. Sitch, C. Huntingford, P. Cox, M. Williams, and P. Meir, 2010: Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change. *New Phytologist*, **187**, 647-665.
Ganachaud, A., and C. Wunsch, 2003: Large-scale ocean heat and freshwater transports during the World Ocean Circulation Experiment. *Journal of Climate*, **16**, 696-705.
Gangsto, R., F. Joos, and M. Gehlen, 2011: Sensitivity of pelagic calcification to ocean acidification. *Biogeosciences*, **8**, 433-458.
Gates, W. L., et al., 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, **80**, 29-55.
Gbobaniyi, E. O., B. J. Abiodun, M. A. Tadross, B. C. Hewitson, and W. J. Gutowski, 2011: The coupling of cloud base height and surface fluxes: a transferability intercomparison. *Theoretical and Applied Climatology*, **106**, 189-210.
Gedney, N., and P. M. Cox, 2003: The Sensitivity of Global Climate Model Simulations to the Representation of Soil Moisture Heterogeneity. *Journal of Hydrometeorology*, **4**, 1265-1275.

Gehlen, M., R. Gangsto, B. Schneider, L. Bopp, O. Aumont, and C. Ethe, 2007: The fate of pelagic CaCO3 production in a high CO2 ocean: a model study. *Biogeosciences*, **4**, 505-519.
Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, **20**, 150-155.
Gent, P. R., and G. Danabasoglu, 2011: Response to Increasing Southern Hemisphere Winds in CCSM4. *Journal of Climate*, **24**, 4992-4998.
Gent, P. R., J. Willebrand, T. J. McDougall, and J. C. McWilliams, 1995: PARAMETERIZING EDDY-INDUCED TRACER TRANSPORTS IN OCEAN CIRCULATION MODELS. *Journal of Physical Oceanography*, **25**, 463-474.
Gent, P. R., S. G. Yeager, R. B. Neale, S. Levis, and D. A. Bailey, 2010: Improvements in a half degree atmosphere/land version of the CCSM. *Climate Dynamics*, **34**, 819-833.
Gerber, E. P., L. M. Polvani, and D. Ancukiewicz, 2008: Annular mode time scales in the Intergovernmental Panel on Climate Change Fourth Assessment Report models. *Geophys. Res. Lett.*, **35**, L22707-.
Gettelman, A., and Q. Fu, 2008: Observed and simulated upper-tropospheric water vapor feedback. *Journal of Climate*, **21**, 3282-3289.
Gettelman, A., et al., 2010: Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends. *Journal of Geophysical Research-Atmospheres*, ARTN D00M08, DOI 10.1029/2009JD013638. -.
Giorgetta, M. A., E. Manzini, and E. Roeckner, 2002: Forcing of the Quasi-Biennial Oscillation from a broad spectrum of atmospheric waves. *Geophysical Research Letters*, **29**.
Giorgi, F., and E. Coppola, 2010: Does the model regional bias affect the projected regional climate change? An analysis of global model projections. *Clim. Change*, **100**, 787-795.
Gleckler, P., K. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *Journal of Geophysical Research-Atmospheres*, ARTN D06104, DOI 10.1029/2007JD008972. -.
Gleckler, P., K. AchutaRao, J. Gregory, B. Santer, K. Taylor, and T. Wigley, 2006: Krakatoa lives: The effect of volcanic eruptions on ocean heat content and thermal expansion. *Geophysical Research Letters*, ARTN L17702, DOI 10.1029/2006GL026771. -.
Gnanadesikan, A., S. M. Griffies, and B. L. Samuels, 2007: Effects in a climate model of slope tapering in neutral physics schemes. *Ocean Modelling*, **16**, 1-16.
Goddard, L., and S. J. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Climate Dynamics*, **19**, 619-631.
Goddard, L., D. G. DeWitt, and R. W. Reynolds, 2009: Practical implications of uncertainty in observed SSTs. *Geophysical Research Letters*, **36**.
Goubanova, K., V. Echevin, B. Dewitte, F. Codron, K. Takahashi, P. Terray, and M. Vrac, 2011: Statistical downscaling of sea-surface wind over the Peru-Chile upwelling region: diagnosing the impact of climate change from the IPSL-CM4 model. *Climate Dynamics*, **36**, 1365-1378.
Gregory, D., G. J. Shutts, and J. R. Mitchell, 1998: A new gravity-wave-drag scheme incorporating anisotropic orography and low-level wave breaking: Impact upon the climate of the UK Meteorological Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **124**, 463-493.
Gregory, J., and M. Webb, 2008: Tropospheric adjustment induces a cloud component in CO2 forcing. *Journal of Climate*, **21**, 58-71.
Gregory, J. M., et al., 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, **31**.
Griffies, S. M., et al., 2009: Coordinated Ocean-ice Reference Experiments (COREs). *Ocean Modelling*, **26**, 1-46.
Guilyardi, E., 2006: El Niño - mean state - seasonal cycle interactions in a multi-model ensemble. *Clim. Dyn.*, **26**, 229-348.
Guilyardi, E., P. Braconnot, F. F. Jin, S. T. Kim, M. Kolasinski, T. Li, and I. Musat, 2009a: Atmosphere Feedbacks during ENSO in a Coupled GCM with a Modified Atmospheric Convection Scheme. *Journal of Climate*, **22**, 5698-5718.
Guilyardi, E., et al., 2009b: UNDERSTANDING EL NINO IN OCEAN-ATMOSPHERE GENERAL CIRCULATION MODELS Progress and Challenges. *Bulletin of the American Meteorological Society*, **90**, 325-+.
Gulev, S. K., B. Barnier, J.-M. Molines, T. Penduff, and J. Chanut, 2007: Impact of spatial resolution on simulated surface water mass transformations in the Atlantic. *Ocean Modelling*, **19**, 138-160.

Gupta, A. S., A. Santoso, A. S. Taschetto, C. C. Ummenhofer, J. Trevena, and M. H. England, 2009: Projected changes to the southern hemisphere ocean and sea ice in the IPCC AR4 climate models. *Journal of Climate*, 10.1175/2008jcli2827.1. 3047-3078.

Gutowski, W., et al., 2010: Regional Extreme Monthly Precipitation Simulated by NARCCAP RCMs. *Journal of Hydrometeorology*, DOI 10.1175/2010JHM1297.1. 1373-1379.

Haerter, J. O., and P. Berg, 2009: Unexpected rise in extreme precipitation caused by a shift in rain type? *Nature Geoscience*, **2**, 372-373.

Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, ARTN L03502, DOI 10.1029/2005GL025127. -.

Hallberg, R., and A. Gnanadesikan, 2006: The role of eddies in determining the structure and response of the wind-driven southern hemisphere overturning: Results from the Modeling Eddies in the Southern Ocean (MESO) project. *Journal of Physical Oceanography*, **36**, 2232-2252.

Ham, Y.-G., J. S. Kug, I. S. Kang, F. F. Jin, and A. Timmermann, 2010: Impact of diurnal atmospher-ocean coupling on tropical climate simulations using a coupled GCM. *Climate Dynamics*, **34**, 905-917.

Ham, Y. G., and J. S. Kug, 2011: How well do current climate models simulate two types of El Nino? , published online ed., Climate Dynamics.

Hamilton, K., 1998: Effects of an imposed Quasi-Biennial Oscillation in a comprehensive troposphere-stratosphere-mesosphere general circulation model. *Journal of the Atmospheric Sciences*, **55**, 2393–2418.

Hannart, A., J. L. Dufresne, and P. Naveau, 2009: Why climate sensitivity may not be so unpredictable. *Geophysical Research Letters*, **36**.

Hannay, C., et al., 2009: Evaluation of Forecasted Southeast Pacific Stratocumulus in the NCAR, GFDL, and ECMWF Models. *Journal of Climate*, **22**, 2871-2889.

Hansen, J., et al., 2005: Efficacy of climate forcings. *Journal of Geophysical Research-Atmospheres*, **110**.

Hargreaves, J. C., A. Abe-Ouchi, and J. D. Annan, 2007: Linking glacial and future climates through an ensemble of GCM simulations. *Climate of the Past*, **3**, 77-87.

Hargreaves, J. C., A. Paul, R. Ohgaito, A. Abe-Ouchi, and J. D. Annan, 2011: Are paleoclimate model ensembles consistent with the MARGO data synthesis? *Climate of the Past*, **7**, 917-933.

Harrison , S., and C. Prentice, 2003: Climate and CO_2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. *Global Change Biology*, **9**, 983-1004.

Harrison, S. P., et al., 1998: Intercomparison of Simulated Global Vegetation Distributions in Response to 6 kyr BP Orbital Forcing. *Journal of Climate*, **11**, 2721-2742.

Hawkins, E., and R. Sutton, 2009: The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, **90**, 1095-1107.

Haylock, M., N. Hofstra, A. Tank, E. Klok, P. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research-Atmospheres*, **113**, -.

Haynes, J. M., C. Jakob, W. B. Rossow, G. Tselioudis, and J. Brown, 2011: Major Characteristics of Southern Ocean Cloud Regimes and Their Effects on the Energy Budget. *Journal of Climate*, **24**, 5061-5080.

Haynes, P. H., 2006: The Latitudinal Structure of the QBO. *Quarterly Journal of the Royal Meteorological Society*, **124**, 2645-2670.

Hazeleger, W., and R. J. Haarsma, 2005: Sensitivity of tropical Atlantic climate to mixing in a coupled ocean-atmosphere model. *Climate Dynamics*, **25**, 387-399.

Hegerl, G. C., et al., 2007: Understanding and Attributing Climate Change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to theFourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis,K.B. Averyt, M. Tignor and H.L. Miller, Ed., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 665-775.

Hegglin, M., et al., 2010: Multimodel assessment of the upper troposphere and lower stratosphere: Extratropics. *Journal of Geophysical Research-Atmospheres*, ARTN D00M09, DOI 10.1029/2010JD013884. -.

Heinze, C., 2004: Simulating oceanic CaCO3 export production in the greenhouse. *Geophysical Research Letters*, **31**.

Heinze, C., I. Kriest, and E. Maier-Reimer, 2009: Age offsets among different biogenic and lithogenic components of sediment cores revealed by numerical modeling. *Paleoceanography*, **24**.

Held, I. M., 2005: The gap between simulation and understanding in climate modeling. *Bulletin of the American Meteorological Society*, **86**, 1609-+.
Henson, S. A., D. Raitsos, J. P. Dunne, and A. McQuatters-Gollop, 2009: Decadal variability in biogeochemical models: Comparison with a 50-year ocean colour dataset. *Geophysical Research Letters*, **36**.
Hermes, J. C., and C. J. C. Reason, 2005: Ocean model diagnosis of interannual coevolving SST variability in the South Indian and South Atlantic Oceans. *Journal of Climate*, **18**, 2864-2882.
Hibbard, K. A., G. A. Meehl, P. M. Cox, and P. Friedlingstein, 2007: A Strategy for Climate Change Stabilization Experiments. *Eos Trans.* , **88**.
Hibler, W. D., 1979: A dynamic thermodynamic sea ice model. *Journal of Physical Oceanography*, **9**, 815-846.
Hirschi, M., et al., 2011a: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, DOI 10.1038/NGEO1032. 17-21.
Hirschi, M., et al., 2011b: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, **4**, 17-21.
Hofmann, M., and M. A. Morales Maqueda, 2011: The response of Southern Ocean eddies to increased midlatitude westerlies: A non-eddy resolving model study. *Geophysical Research Letters*, **38**, L03605, doi:03610.01029/02010GL045972.
Hofstra, N., M. New, and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, **35**, 841-858.
Holland, P. R., A. Jenkins, and D. M. Holland, 2008: The response of ice shelf basal melting to variations in ocean temperature. *Journal of Climate*, **21**, 2558-2572.
Holton, J. R., 1983: THE INFLUENCE OF GRAVITY-WAVE BREAKING ON THE GENERAL-CIRCULATION OF THE MIDDLE ATMOSPHERE. *Journal of the Atmospheric Sciences*, **40**, 2497-2507.
Holton, J. R., and H. C. Tan, 1980: THE INFLUENCE OF THE EQUATORIAL QUASI-BIENNIAL OSCILLATION ON THE GLOBAL CIRCULATION AT 50 MB. *Journal of the Atmospheric Sciences*, **37**, 2200-2208.
Hoose, C., J. E. Kristjansson, T. Iversen, A. Kirkevag, O. Seland, and A. Gettelman, 2009: Constraining cloud droplet number concentration in GCMs suppresses the aerosol indirect effect. *Geophysical Research Letters*, **36**, 10.1029/2009gl038568.
Hourdin, F., et al., 2010: AMMA-MODEL INTERCOMPARISON PROJECT. *Bulletin of the American Meteorological Society*, **91**, 95-+.
Hunke, E. C., W. H. Lipscomb, and A. K. Turner, 2011: Sea Ice Models for Climate Study: Retorspective and New Directions. *Journal of Glaciology*, **56**, 1162-1172.
Hurrell, J. W., et al., 2010: Decadal Climate Prediction: Opportunities and Challenges. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2)*, Venice, Italy, 21-25 September 2009, ESA Publication
Hurtt, G. C., et al., 2009: Harmonization of global land-use scenarios for the period 1500-2100 for IPCC-AR5. *iLEAPS Newsletter*, **7**, 6-8.
Huybers, P., 2010: Compensation between Model Feedbacks and Curtailment of Climate Sensitivity. *Journal of Climate*, **23**, 3009-3018.
Iacono, M. J., J. S. Delamere, E. J. Mlawer, and S. A. Clough, 2003: Evaluation of upper tropospheric water vapor in the NCAR Community Climate Model (CCM3) using modeled and observed HIRS radiances. *J. Geophys. Res.*, **108**, 4037.
Illingworth, A. J., et al., 2007: Cloudnet. *Bull. Amer. Meteor. Soc.*, **88**, 883-898.
Ilyina, T., R. E. Zeebe, E. Maier-Reimer, and C. Heinze, 2009: Early detection of ocean acidification effects on marine calcification. *Global Biogeochemical Cycles*, **23**.
Inatsu, M., and M. Kimoto, 2009: A Scale Interaction Study on East Asian Cyclogenesis Using a General Circulation Model Coupled with an Interactively Nested Regional Model. *Monthly Weather Review*, DOI 10.1175/2009MWR2825.1. 2851-2868.
Inoue, T., and H. Ueda, 2009: Evaluation for the Seasonal Evolution of the Summer Monsoon over the Asian and Western North Pacific Sector in the WCRP CMIP3 Multi-model Experiments. *Journal of the Meteorological Society of Japan*, **87**, 539-560.
IPCC, 2012: *IPCC WGI/WGII Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX).* in press pp.

Jackson, C. S., M. K. Sen, G. Huerta, Y. Deng, and K. P. Bowman, 2008a: Error Reduction and Convergence in Climate Prediction. *Journal of Climate*, **21**, 6698-6709.

Jackson, L., R. Hallberg, and S. Legg, 2008b: A parameterization of shear-driven turbulence for ocean climate models. *Journal of Physical Oceanography*, **38**, 1033-1053.

Jakob, C., 2010: ACCELERATING PROGRESS IN GLOBAL ATMOSPHERIC MODEL DEVELOPMENT THROUGH IMPROVED PARAMETERIZATIONS Challenges, Opportunities, and Strategies. *Bulletin of the American Meteorological Society*, **91**, 869-+.

Jamison, N., and S. Kravtsov, 2010: Decadal Variations of North Atlantic Sea Surface Temperature in Observations and CMIP3 Simulations. *Journal of Climate*, **23**, 4619-4636.

Jayne, S. R., 2009: The Impact of Abyssal Mixing Parameterizations in an Ocean General Circulation Model. *Journal of Physical Oceanography*, **39**, 1756-1775.

Jin, F. F., S. T. Kim, and L. Bejarano, 2006: A coupled-stability index for ENSO. *Geophys. Res. Let.*, **33**, L23708.

Jöckel, P., et al., 2006: The atmospheric chemistry general circulation model ECHAM5/MESSy1: consistent simulation of ozone from the surface to the mesosphere. *Atmospheric Chemistry and Physics*, **6**, 5067-5104.

John, V., and B. Soden, 2007: Temperature and humidity biases in global climate models and their impact on climate feedbacks. *Geophysical Research Letters*, ARTN L18704, DOI 10.1029/2007GL030429. -.

Johnson, N. C., and S. B. Feldstein, 2010: The Continuum of North Pacific Sea Level Pressure Patterns: Intraseasonal, Interannual, and Interdecadal Variability. *Journal of Climate*, **23**, 851-867.

Joly, M., A. Voldoire, H. Douville, P. Terray, and J. F. Royer, 2007a: African monsoon teleconnections with tropical SSTs in a set of IPCC4 coupled models. *Clim. Dyn.*, 1-32.

——, 2007b: African monsoon teleconnections with tropical SSTs: validation and evolution in a set of IPCC4 simulations. *Climate Dynamics*, **29**, 1-20.

Jones, C. D., and P. M. Cox, 2005: On the significance of atmospheric CO2 growth rate anomalies in 2002-2003. *Geophysical Research Letters*, **32**.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its variations over the last 150 years. *Rev. Geophys.*, **37**, 173-199.

Joussaume, S., and K. E. Taylor, 1995: Status of the Paleoclimate Modeling Intercomparison Project. *in Proceedings of the first international AMIP scientific conference, WCRP-92, Monterey, USA*. 425-430.

Jun, M., R. Knutti, and D. Nychka, 2008: Spatial Analysis to Quantify Numerical Model Bias and Dependence: How Many Climate Models Are There? *Journal of the American Statistical Association*, DOI 10.1198/016214507000001265. 934-947.

Jungclaus, J. H., et al., 2006: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *Journal of Climate*, **19**, 3952-3972.

Jungclaus, J. H., et al., 2010: Climate and carbon-cycle variability over the last millennium. *Climate of the Past*, **6**, 723-737.

Kageyama, M., J. Mignot, D. Swingedouw, C. Marzin, R. Alkama, and O. Marti, 2009: Glacial climate sensitivity to different states of the Atlantic Meridional Overturning Circulation: results from the IPSL model. *Climate of the Past*, **5**, 551-570.

Kageyama, M., et al., 2006: Last Glacial Maximum temperatures over the North Atlantic, Europe and western Siberia: a comparison between PMIP models, MARGO sea-surface temperatures and pollen-based reconstructions. *Quaternary Science Reviews*, **25**, 2082-2102.

Kakitha, K., Y. Peings, P. Terray, and H. Douville, 2011: Indian Summer Monsoon and its relationship with ENSO as simulated by the two French CMIP5 coupled GCMs. submitted ed.

Kanada, S., M. Nakano, S. Hayashi, T. Kato, M. Nakamura, K. Kurihara, and A. Kitoh, 2008: Reproducibility of Maximum Daily Precipitation Amount over Japan by a High-resolution Non-hydrostatic Model. *SOLA*, **4**, 105-108.

Kanamaru, H., and M. Kanamitsu, 2007: Fifty-seven-year California Reanalysis Downscaling at 10 km (CaRD10). Part II: Comparison with North American Regional Reanalysis. *Journal of Climate*, **20**, 5572-5592.

Kanamitsu, M., and L. DeHaan, 2011: The Added Value Index: A new metric to quantify the added value of regional models. *Journal of Geophysical Research-Atmospheres*, **116**, -.

Kanzow, T., et al., 2009: Basinwide Integrated Volume Transports in an Eddy-Filled Ocean. *Journal of Physical Oceanography*, **39**, 3091-3110.

Kanzow, T., et al., 2010: Seasonal variability of the Atlantic meridional overturning circulation at 26.5°N. *Journal of Climate*, **23**, 5678–5698.
Karlsson, J., and G. Svensson, 2010: The simulation of Arctic clouds and their influence on the winter surface temperature in present-day climate in the CMIP3 multi-model dataset. *Clim. Dyn.*, DOI 10.1007/s00382-00010-00758-00386.
Karlsson, J., G. Svensson, and H. Rodhe, 2008: Cloud radiative forcing of subtropical low level clouds in global models. *Climate Dynamics*, **30**, 779-788.
Karnauskas, K., G. Johnson, and R. Murtugudde, 2011: An Equatorial Ocean Bottleneck in Global Climate Models. revised ed., J. Climate.
Karpechko, A., N. Gillett, G. Marshall, and A. Scaife, 2008: Stratospheric influence on circulation changes in the Southern Hemisphere troposphere in coupled climate models. *Geophysical Research Letters*, ARTN L20806, DOI 10.1029/2008GL035354. -.
Karpechko, A. Y., N. P. Gillett, G. J. Marshall, and J. A. Screen, 2009: Climate impacts of the southern annular mode simulated by the CMIP3 models. (vol 22, pg 3751, 2009). *Journal of Climate*, **22**, 6149-6150.
Kattsov, V. M., et al., 2011: Arctic sea-ice change: a grand challenge of climate science. *Journal of Glaciology*, **56**, 1115-1121.
Kawatani, Y., K. Hamilton, and S. Watanabe, 2011: The quasi-biennial oscillation in a double CO2 climate. *Journal of the Atmospheric Sciences*, doi: 10.1175/2010JAS3623.1.
Kay, J., M. Holland, and A. Jahn, 2011: Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world. *Geophysical Research Letters*, **38**, -.
Khairoutdinov, M., C. DeMott, and D. Randall, 2008: Evaluation of the simulated interannual and subseasonal variability in an AMIP-Style simulation using the CSU multiscale modeling framework. *Journal of Climate*, **21**, 413-431.
Khairoutdinov, M. F., D. A. Randall, and C. DeMott, 2005: Simulations of the Atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *Journal of the Atmospheric Sciences*, **62**, 2136-2154.
Kharin, V. V., F. W. Zwiers, X. B. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate*, **20**, 1419-1444.
Khvorostyanov, D. V., G. Krinner, P. Ciais, M. Heimann, and S. A. Zimov, 2008a: Vulnerability of permafrost carbon to global warming. Part I: model description and role of heat generated by organic matter decomposition. *Tellus Series B-Chemical and Physical Meteorology*, **60**, 250-264.
Khvorostyanov, D. V., P. Ciais, G. Krinner, S. A. Zimov, C. Corradi, and G. Guggenberger, 2008b: Vulnerability of permafrost carbon to global warming. Part II: sensitivity of permafrost carbon stock to global warming. *Tellus Series B-Chemical and Physical Meteorology*, **60**, 265-275.
Kidston, J., and E. P. Gerber, 2010: Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *Geophys. Res. Lett.*, **37**, L09708-.
Kiehl, J. T., 2007: Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, **34**.
Kim, D., A. H. Sobel, E. D. Maloney, E. D. Frierson, and I.-S. Kang, 2011a: A Systematic Relationship between Intraseasonal Variability and mean State Bias in AGCM Simulations *Journal of Climate*.
Kim, D., et al., 2009: Application of MJO Simulation Diagnostics to Climate Models. *Journal of Climate*, **22**, 6413-6436.
Kim, H.-J., K. Takata, B. Wang, M. Watanabe, M. Kimoto, T. Yokohata, and T. Yasunari, 2011b: Global Monsoon, El Niño, and Their Interannual Linkage Simulated by MIROC5 and the CMIP3 CGCMs. *Journal of Climate*, **24**, 5604-5618.
Kim, S., and F.-F. Jin, 2010a: An ENSO stability analysis. Part II: results from the twentieth and twenty-first century simulations of the CMIP3 models. *Climate Dynamics*. 1-19.
——, 2010b: An ENSO stability analysis. Part I: results from a hybrid coupled model. *Climate Dynamics*. 1-15.
Kinne, S., et al., 2006: An AeroCom initial assessment - optical properties in aerosol component modules of global models. *Atmospheric Chemistry and Physics*, **6**, 1815-1834.
Kitoh, A., S. Kusunoki, and T. Nakaegawa, 2011: Climate change projections over South America in the late 21st century with the 20 and 60 km mesh Meteorological Research Institute atmospheric general circulation model (MRI-AGCM). *Journal of Geophysical Research-Atmospheres*, **116**.

Kjellstrom, E., F. Boberg, M. Castro, J. Christensen, G. Nikulin, and E. Sanchez, 2010: Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Research*, DOI 10.3354/cr00932. 135-150.

Klein, P., and G. Lapeyre, 2009: The Oceanic Vertical Pump Induced by Mesoscale and Submesoscale Turbulence. *Annual Review of Marine Science*, **1**, 351-375.

Klein, S. A., and C. Jakob, 1999: Validation and sensitivities of frontal clouds simulated by the ECMWF model. *Monthly Weather Review*, **127**, 2514-2531.

Klein, S. A., B. J. Soden, and N. C. Lau, 1999: Remote sea surface temperature variations during ENSO: Evidence for a tropical atmospheric bridge. *Journal of Climate*, **12**, 917-932.

Klein, S. A., X. Jiang, J. Boyle, S. Malyshev, and S. Xie, 2006: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.*, **33**, L18805.

Klinker, E., and P. D. Sardeshmukh, 1992: The Diagnosis of Mechanical Dissipation in the Atmosphere from Large-Scale Balance Requirements. *Journal of the Atmospheric Sciences*, **49**, 608-627.

Kloster, S., et al., 2010: Fire dynamics during the 20th century simulated by the Community Land Model. *Biogeosciences*, **7**, 1877-1902.

Knight, J. R., 2009: The Atlantic Multidecadal Oscillation Inferred from the Forced Climate Response in Coupled General Circulation Models. *Journal of Climate*, **22**, 1610-1625.

Knutson, T. R., et al., 2006: Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *Journal of Climate*, **19**, 1624-1651.

Knutti, R., 2008: Why are climate models reproducing the observed global surface warming so well? *Geophysical Research Letters*, **35**.

——, 2010: The end of model democracy? *Clim. Change*, **102**, 395-404.

Knutti, R., and L. Tomassini, 2008: Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophysical Research Letters*, ARTN L09701, DOI 10.1029/2007GL032904. -.

Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**, 719-723.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010a: Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate*, **23**, 2739-2758.

Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P. J. Gleckler, B. Hewitson, and L. Mearns, 2010b: Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections. *In: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, and P.M. Midgley (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland.*

Koldunov, N. V., D. Stammer, and J. Marotzke, 2010: Present-day Arctic sea ice variability in the coupled ECHAM5/MPI-OM model. *Journal of Climate*, **23**, 2520–2543.

Koster, R. D., et al., 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138-1140.

Kostopoulou, E., K. Tolika, I. Tegoulias, C. Giannakopoulos, S. Somot, C. Anagnostopoulou, and P. Maheras, 2009: Evaluation of a regional climate model using in situ temperature observations over the Balkan Peninsula. *Tellus Series a-Dynamic Meteorology and Oceanography*, DOI 10.1111/j.1600-0870.2009.00389.x. 357-370.

Koyama, H., and M. Watanabe, 2010: Reducing Forecast Errors Due to Model Imperfections Using Ensemble Kalman Filtering. *Monthly Weather Review*, **138**, 3316-3332.

Kravtsov, S., and C. Spannagle, 2008: Multidecadal climate variability in observed and modeled surface temperatures. *Journal of Climate*, **21**, 1104-1121.

Kusaka, H., T. Takata, and Y. Takane, 2010: Reproducibility of Regional Climate in Central Japan Using the 4-km Resolution WRF Model. *Sola*, **6**, 113-116.

Kwon, Y. O., and S. C. Riser, 2004: North Atlantic Subtropical Mode Water: A history of ocean-atmosphere interaction 1961–2000. *Geophysical research letters*, **31**, L19307-L19307.

Kwon, Y. O., M. A. Alexander, N. A. Bond, C. Frankignoul, H. Nakamura, B. Qiu, and L. Thompson, 2010: Role of the Gulf Stream and Kuroshio-Oyashio Systems in Large-Scale Atmosphere-Ocean Interaction: A Review. *Journal of Climate*, **23**, 3249-3281.

L'Ecuyer, T., and G. Stephens, 2007: The Tropical Atmospheric Energy Budget from the TRMM Perspective. Part II: Evaluating GCM Representations of the Sensitivity of Regional Energy and Water Cycles to the 1998–99 ENSO Cycle. *J. Climate*, **20**, 4548-4571.

Lamarque, J.-F., et al., 2011: Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways. *Clim. Change*, **109**, 191-212.

Lamarque, J. F., et al., 2010: Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmospheric Chemistry and Physics*, **10**, 7017-7039.

Lambert, S., and G. Boer, 2001: CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dynamics*, **17**, 83-106.

Laprise, R., et al., 2008: Challenging some tenets of Regional Climate Modelling. *Meteorology and Atmospheric Physics*, **100**, 3-22.

Larow, T. E., Y. K. Lim, D. W. Shin, E. P. Chassignet, and S. Cocke, 2008: Atlantic basin seasonal hurricane simulations. *Journal of Climate*, **21**, 3191-3206.

Lau, N.-C., and J. J. Ploshay, 2009: Simulation of Synoptic- and Subsynoptic-Scale Phenomena Associated with the East Asian Summer Monsoon Using a High-Resolution GCM. *Monthly Weather Review*, **137**, 137-160.

Lawrence, D. M., et al., 2011: Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, **3**, 2011MS000045.

Le Quere, C., et al., 2005: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, **11**, 2016-2040.

Le Quere, C., et al., 2009: Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, **2**, 831-836.

Leduc, G., R. Schneider, J. H. Kim, and G. Lohmann, 2010: Holocene and Eemian sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry. *Quaternary Science Reviews*, **29**, 989-1004.

Leduc, M., and R. Laprise, 2009: Regional climate model sensitivity to domain size. *Climate Dynamics*, **32**, 833-854.

Leduc, M., R. Laprise, Moretti-Poisson, and M. JP, 2011: Sensitivity to domain size of mid-latitude summer simulations with a regional climate model. *Climate Dynamics*, 10.1007/s00382-011-1008-2.

Lee, D. S., et al., 2009: Aviation and global climate change in the 21st century. *Atmospheric Environment*, **43**, 3520-3537.

Legg, S., L. Jackson, and R. W. Hallberg, 2008: Eddy-resolving modeling of overflows. *Eddy resolving ocean models*, 177 ed., M. Hecht, and H. Hasumi, Eds., American Geophysical Union, 63-82.

Legg, S., et al., 2009: IMPROVING OCEANIC OVERFLOW REPRESENTATION IN CLIMATE MODELS The Gravity Current Entrainment Climate Process Team. *Bulletin of the American Meteorological Society*, **90**, 657-+.

LeGrande, A. N., et al., 2006: Consistent simulations of multiple proxy responses to an abrupt climate change event. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 837-842.

Leloup, J., M. Lengaigne, and J.-P. Boulanger, 2008: Twentieth century ENSO characteristics in the IPCC database. *Clim. Dyn.*, **30**, 277-291.

Lemoine, D. M., 2010: Climate Sensitivity Distributions Dependence on the Possibility that Models Share Biases. *Journal of Climate*, **23**, 4395-4415.

Lenderink, G., 2010: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. *Climate Research*, DOI 10.3354/cr00946. 151-166.

Lenderink, G., and E. Van Meijgaard, 2008: Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, **1**, 511-514.

——, 2009: Unexpected rise in extreme precipitation caused by a shift in rain type? *Nature Geoscience*, **2**, 373-373.

Lengaigne, M., and G. A. Vecchi, 2010: Contrasting the termination of moderate and extreme El Nio events in coupled general circulation models. *Climate Dynamics*, **35**, 299-313.

Lengaigne, M., G. Madec, L. Bopp, C. Menkes, O. Aumont, and P. Cadule, 2009: Vio-physical feedbacks in the Arctic ocean using an Earth system model. *Geophysical Research Letters*, **36**.

Lengaigne, M., C. Menkes, O. Aumont, T. Gorgues, L. Bopp, J.-M. André, and G. Madec, 2007: Influence of the oceanic biology on the tropical Pacific climate in a coupled general circulation model. *Climate Dynamics*, **28**, 503-516.

Levermann, A., J. Mignot, S. Nawrath, and S. Rahmstorf, 2007: The role of Northern sea ice cover for the weakening of the thermohaline circulation under global warming. *Journal of Climate*, **20**, 4160-4171.
Levis, S., 2010: Modeling vegetation and land use in models of the Earth System. *Wiley Interdisciplinary Reviews: Climate Change*, **1**, 840–856.
Levitus, S., J. Antonov, and T. Boyer, 2005: Warming of the world ocean, 1955-2003. *Geophysical Research Letters*, **32**.
Lewis, S. L., P. M. Brando, O. L. Phillips, G. M. F. van der Heijden, and D. Nepstad, 2011: The 2010 Amazon Drought. *Science*, **331**, 554-554.
Lewis, T., and S. Lamoureux, 2010: Twenty-first century discharge and sediment yield predictions in a small high Arctic watershed. *Global and Planetary Change*, **71**, 27-41.
Li, C., J.-S. von Storch, and J. Marotzke, 2011a: Deep-ocean heat uptake and equilibrium climate response. *Climate Dynamics*. Submitted.
Li, J., S.-P. X. and A. Mestas-Nunez, E. R. C. and Gang Huang, R. D'Arrigo, F. Liu, J. Ma, and X. Zheng, 2011b: Interdecadal Modulation of ENSO Amplitude During the Last Millennium. *Nature Climate Change*.
Li, T., and G. H. Philander, 1996: On the annual cycle in the eastern equatorial Pacific. *J. Climate*, **9**, 2986-2998.
Li, W., C. E. Forest, and J. J. Barsugli, 2011c: Comparing two methods on estimating the sensitivity of regional climate simulation to tropical SST anomalies. *J. Geophys. Res. submitted*.
Lienert, f., J. C. Fyfe, and W. J. Merryfield, 2011: Do Climate Models Capture the Tropical Influences on North Pacific sea surface temperature variability? *Journal of Climate*, 10.1175/JCLI-D-11-00205.1.).
Lin, A. L., and T. Li, 2008: Energy Spectrum Characteristics of Boreal Summer Intraseasonal Oscillations: Climatology and Variations during the ENSO Developing and Decaying Phases. *Journal of Climate*, **21**, 6304-6320.
Lin, J.-L., 2007: The Double-ITCZ Problem in IPCC AR4 Coupled GCMs: Ocean-Atmosphere Feedback Analysis. *Journal of Climate*, **20**, 4497-4525.
Lin, J. L., et al., 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *Journal of Climate*, **19**, 2665-2690.
Lin, L., W. Yu, and T. Li, 2011: Dynamic and Thermodynamic Air–Sea Coupling Associated with the Indian Ocean Dipole Diagnosed from 23 WCRP CMIP3 Models. *Journal of Climate*, **24**, 4941–4958.
Lin, P., Q. A. Fu, S. Solomon, and J. M. Wallace, 2010: Temperature Trend Patterns in Southern Hemisphere High Latitudes: Novel Indicators of Stratospheric Change (vol 22, pg 6325, 2009). *Journal of Climate*, **23**, 4263-4280.
Lindzen, R. S., 1981: TURBULENCE AND STRESS OWING TO GRAVITY-WAVE AND TIDAL BREAKDOWN. *Journal of Geophysical Research-Oceans and Atmospheres*, **86**, 9707-9714.
Lipscomb, W. H., and E. C. Hunke, 2004: Modeling sea ice transport using incremental remapping. *Monthly Weather Review*, **132**, 1341-1354.
Liu, S. C., C. B. Fu, C. J. Shiu, J. P. Chen, and F. T. Wu, 2009: Temperature dependence of global precipitation extremes. *Geophysical Research Letters*, **36**, 4.
Liu, X. H., et al., 2007: Uncertainties in global aerosol simulations: Assessment using three meteorological data sets. *Journal of Geophysical Research-Atmospheres*, **112**.
Lloyd, J., E. Guilyardi, and H. Weller, 2010: The role of atmosphere feedbacks during ENSO in the CMIP3 models. Part II: using AMIP runs to understand the heat flux feedback mechanisms. *Climate Dynamics*. 1-22.
Lloyd, J., E. Guilyardi, H. Weller, and J. Slingo, 2009: The role of atmosphere feedbacks during ENSO in the CMIP3 models. *Atmospheric Science Letters*, **10**, 170-176.
Lorenz, P., and D. Jacob, 2005: Influence of regional scale information on the global circulation: A two-way nesting climate simulation. *Geophysical Research Letters*, **32**, -.
Losch, M., D. Menemenlis, J.-M. Campin, P. Heimbach, and C. Hill, 2011: On the formulation of sea-ice models. Part 1: Effects of different solver implementations and parameterizations. *Ocean Modelling*.
Lott, F., and M. J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, **123**, 101-127.
Lozier, M. S., V. Roussenov, M. S. C. Reed, and R. G. Williams, 2010: Opposing decadal changes for the North Atlantic meridional overturning circulation. *Nature Geosci*, **3**, 728-734.
Lucarini, V., S. Calmanti, A. Dellâ Aquila, P. Ruti, and A. Speranza, 2007: Intercomparison of the northern hemisphere winter mid-latitude atmospheric variability of the IPCC models. 829-848.

Lucas-Picher, P., et al., 2011: Can regional climate models represent the Indian Monsoon? *J. Hydrometeor.*, **12**, 849-868.
Lumpkin, R., K. G. Speer, and K. P. Koltermann, 2008: Transport across 48Â°N in the Atlantic Ocean. *Journal of Physical Oceanography*, **38**, 733-752.
Luo, J. J., S. Masson, E. Roeckner, G. Madec, and T. Yamagata, 2005: Reducing climatology bias in an ocean-atmosphere CGCM with improved coupling physics. *Journal of Climate*, **18**, 2344-2360.
Macadam, I., A. Pitman, P. Whetton, and G. Abramowitz, 2010: Ranking climate models by performance using actual values and anomalies: Implications for climate change impact assessments. *Geophysical Research Letters*, ARTN L16704, DOI 10.1029/2010GL043877. -.
MacKinnon, J., et al., 2009: Using global arrays to investigate internal-waves and mixing. *OceanObs09: Sustained Ocean Observations and Information for Society*, Venice, Italy, ESA.
Mahlstein, I., and R. Knutti, 2010: Regional climate change patterns identified by cluster analysis. *Climate Dynamics*, **35**, 587-600.
Mahlstein, I., and R. Knutti, 2011: September Arctic sea ice predicted to disappear near 2°C global warming above present. *Journal of Geophysical Research-Atmospheres*, **submitted**.
Manabe, S., 1969: CLIMATE AND THE OCEAN CIRCULATION1. *Monthly Weather Review*, **97**, 739-774.
Manabe, S., and K. Bryan, 1985: CO2-induced change in a coupled ocean-atmosphere model and its paleoclimatic implications. *J. Geophys. Res.*, **90**, 1689-1707.
Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, **78**, 1069-1079.
Maraun, D., H. Rust, and T. Osborn, 2010a: Synoptic airflow and UK daily precipitation extremes Development and validation of a vector generalised linear model. *Extremes*, **13**, 133-153.
Maraun, D., et al., 2010b: PRECIPITATION DOWNSCALING UNDER CLIMATE CHANGE: RECENT DEVELOPMENTS TO BRIDGE THE GAP BETWEEN DYNAMICAL MODELS AND THE END USER. *Rev. Geophys.*, **48**, -.
Marchand, R., N. Beagley, and T. P. Ackerman, 2009: Evaluation of Hydrometeor Occurrence Profiles in the Multiscale Modeling Framework Climate Model Using Atmospheric Classification. *Journal of Climate*, **22**, 4557-4573.
Marsh, R., et al., 2009: Recent changes in the North Atlantic circulation simulated with eddy-permitting and eddy-resolving ocean models. *Ocean Modelling*, **28**, 226-239.
Marshall, A., and A. A. Scaife, 2009: Impact of the Quasi-Biennial Oscillation on seasonal forecasts. *J. Geophys. Res.*, **114**, D18110.
Marti, O., et al., 2010: Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. *Climate Dynamics*, **34**, 1-26.
Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, S. Ineson, T. Reichler, and J. Kim, 2010: Analysis and Reduction of Systematic Errors through a Seamless Approach to Modeling Weather and Climate. *Journal of Climate*, **23**, 5933-5957.
Martius, O., L. M. Polvani, and H. C. Davies, 2009: Blocking precursors to stratospheric sudden warming events. *Geophys. Res. Lett.*, **36**, L14806-.
Masson-Delmotte, V., et al., 2010: EPICA Dome C record of glacial and interglacial intensities. *Quaternary Science Reviews*, **29**, 113-128.
Masson-Delmotte, V., et al., 2006: Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints (vol 26, pg 513, 2006). *Climate Dynamics*, **27**, 437-440.
Masson, D., and R. Knutti, 2011: Spatial-Scale Dependence of Climate Model Performance in the CMIP3 Ensemble. *Journal of Climate*, **24**, 2680-2692.
Mastrandrea, M. D., et al., 2011: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC).
Matsueda, M., R. Mizuta, and S. Kusunoki, 2009: Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric global circulation model. *J. Geophys. Res.*, **114**, D12114-.
Matsueda, M., H. Endo, and R. Mizuta, 2010: Future change in Southern Hemisphere summertime and wintertime atmospheric blockings simulated using a 20-km-mesh AGCM. *Geophys. Res. Lett.*, **37**, L02803-.
Matsumoto, K., K. S. Tokos, A. R. Price, and S. J. Cox, 2008: First description of the Minnesota Earth System Model for Ocean biogeochemistry (MESMO 1.0). *Geoscientific Model Development*, **1**, 1-15.

Maurer, E., and H. Hidalgo, 2008: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences*, **12**, 551-563.
May, P. T., J. H. Mather, G. Vaughan, K. N. Bower, C. Jakob, G. M. McFarquhar, and G. G. Mace, 2008: The Tropical Warm Pool International Cloud Experiment. *Bulletin of the American Meteorological Society*, **89**, 629-645.
May, W., 2007: The simulation of the variability and extremes of daily precipitation over Europe by the HIRHAM regional climate model. *Global and Planetary Change*, **57**, 59-82.
Maykut, G. A., and N. Untersteiner, 1971: Some results from a time dependent thermodynamic model of sea ice. *J. Geophys. Res.*, **76**, 1550-1575.
McClean, J., M. Maltrud, and F. Bryan, 2006a: Measures of the fidelity of eddying ocean models. *Oceanography*, **19**.
McClean, J. L., and J. C. Carman, 2011: Investigation of IPCC AR4 coupled climate model North Atlantic modewater formation. *Ocean Modelling*, **40**, 14–34.
McClean, J. L., M. E. Maltrud, and F. O. Bryan, 2006b: Measures of the fidelity of eddying ocean models. *Oceanography*, **19**, 104-117.
McClean, J. L., et al., 2011: A prototype two-decade fully-coupled fine-resolution CCSM simulation. *Ocean Modelling*, **39**, 10-30.
McCreary, J. P., and P. Lu, 1994: Interaction between the Subtropical and Equatorial Ocean Circulations: The Subtropical Cell. *Journal of Physical Oceanography*, **24**, 466-497.
McDougall, T. J., and P. C. McIntosh, 2001: The temporal-residual-mean velocity. Part II: Isopycnal interpretation and the tracer and momentum equations. *Journal of Physical Oceanography*, **31**, 1222-1246.
McFarlane, N. A., 1987: THE EFFECT OF OROGRAPHICALLY EXCITED GRAVITY-WAVE DRAG ON THE GENERAL-CIRCULATION OF THE LOWER STRATOSPHERE AND TROPOSPHERE. *Journal of the Atmospheric Sciences*, **44**, 1775-1800.
McKitrick, R., S. McIntyre, and C. Herman, 2010: Panel and multivariate methods for tests of trend equivalence in climate data series. *Atmospheric Science Letters*, **11**, 270-277.
McLandress, C., 2002: Interannual Variations of the Diurnal Tide in the mesosphere induced by a zonal mean wind oscillation in the tropics. *Geophysical Research Letters*, **29**.
McManus, J. F., R. Francois, J. M. Gherardi, L. D. Keigwin, and S. Brown-Leger, 2004: Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature*, **428**, 834-837.
McPhaden, M. J., and D. X. Zhang, 2002: Slowdown of the meridional overturning circulation in the upper Pacific Ocean. *Nature*, **415**, 603-608.
McWilliams, J. C., 2008: The nature and consequences of oceanic eddies. *Ocean Modeling in an eddying regime*, M. Hecht, and H. Hasumi, Eds., American Geophysical Union, 5-15.
Mears, C., A., B. Santer, D., F. Wentz, J., K. Taylor, E., and M. Wehner, F., 2007a: Relationship between temperature and precipitable water changes over tropical oceans. *GEOPHYSICAL RESEARCH LETTERS*, **34**.
Mears, C. A., F. J. Wentz, P. Thorne, and D. Bernie, 2011: Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique. *J. Geophys. Res.*, **116**, D08112.
Mears, C. A., B. D. Santer, F. J. Wentz, K. E. Taylor, and M. F. Wehner, 2007b: Relationship between temperature and precipitable water changes over tropical oceans. *Geophysical Research Letters*, **34**.
Meehl, G. A., C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, 2009a: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U. S. *Geophysical Research Letters*, **36**.
Meehl, G. A., P. R. Gent, J. M. Arblaster, B. L. Otto-Bliesner, E. C. Brady, and A. Craig, 2001: Factors that affect the amplitude of El Niño in global coupled climate models. *Clim. Dyn.*, **17**, 515.
Meehl, G. A., et al., 2007a: The WCRP CMIP3 multimodel dataset - A new era in climate change research. *Bulletin of the American Meteorological Society*, **88**, 1383-+.
Meehl, G. A., et al., 2007b: Global Climate Projections. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, et al., Eds., Cambridge University Press, 747-845.
Meehl, G. A., et al., 2009b: DECADAL PREDICTION Can It Be Skillful? *Bulletin of the American Meteorological Society*, **90**, 1467-1485.

Meier, W. N., and J. Stroeve, 2008: Comparison of sea ice extent and ice edge location estimates from passive microwave and enhanced-resolution scatterometer data, **48**, 65-70.
Meinshausen, M., et al., 2009: Greenhouse-gas emission targets for limiting global warming to 2 degrees C. *Nature*, **458**, 1158-1162.
Meissner, K. J., A. J. Weaver, H. D. Matthews, and P. M. Cox, 2003: The role of land surface dynamics in glacial inception: a study with the UVic Earth System Model. *Climate Dynamics*, **21**, 515-537.
Menary, M., W. Park, K. Lohmann, M. Vellinga, D. Palmer, M. Latif, and J. H. Jungclaus, 2011: A multimodel comparison of centennial Atlantic meridional overturning circulation variability. *Climate Dynamics*, 10.1007/s00382-011-1172-4.
Menendez, C., M. de Castro, A. Sorensson, J. Boulanger, and C. M. Grp, 2010: CLARIS Project: towards climate downscaling in South America. *Meteorologische Zeitschrift*, DOI 10.1127/0941-2948/2010/0459. 357-362.
Menkes, C., et al., 1998: Impact of TAO vs. ERS wind stresses onto simulations of the tropical Pacific Ocean during the 1993-1998 period by the OPA OGCM. *Climatic Impact of Scale Interaction for the Tropical Ocean-Atmosphere System - Euroclivar Workshop No.13*, 46--48.
Mieville, A., et al., 2010: Emissions of gases and particles from biomass burning during the 20th century using satellite data and an historical reconstruction. *Atmospheric Environment*, **44**, 1469-1477.
Miller, R. L., G. A. Schmidt, and D. T. Shindell, 2006: Forced annular variations in the 20th century intergovernmental panel on climate change fourth assessment report models. *Journal of Geophysical Research-Atmospheres*, **111**.
Min, S. K., X. B. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 376-379.
Minobe, S., 1997: A 50-70 year climatic oscillation over the North Pacific and North America. *Geophysical Research Letters*, **24**, 683-686.
——, 1999: Resonance in bidecadal and pentadecadal climate oscillations over the North Pacific: Role in climatic regime shifts. *Geophysical Research Letters*, **26**, 855-858.
Mishchenko, M. I., et al., 2007: Long-term satellite record reveals likely recent aerosol trend. *Science*, **315**, 1543-1543.
Misra, V., 2007: Addressing the issue of systematic errors in a regional climate model. *Journal of Climate*, **20**, 801-818.
Miura, H., M. Satoh, T. Nasuno, A. T. Noda, and K. Oouchi, 2007: A Madden-Julian Oscillation event realistically simulated by a global cloud-resolving model. *Science*, **318**, 1763-1765.
Miyama, T., and M. Kawamiya, 2009: Estimating allowable carbon emission for CO(2) concentration stabilization using a GCM-based Earth system model. *Geophysical Research Letters*, **36**.
Mizuta, R., et al., 2011: Climate simulations using MRI-AGCM3.2 with 20-km grid. *Journal of the Meteorological Society of Japan*. submitted.
Mo, K. C., and G. H. White, 1985: TELECONNECTIONS IN THE SOUTHERN-HEMISPHERE. *Monthly Weather Review*, **113**, 22-37.
Montoya, M., A. Griesel, A. Levermann, J. Mignot, M. Hofmann, A. Ganopolski, and S. Rahmstorf, 2005: The earth system model of intermediate complexity CLIMBER-3 alpha. Part 1: description and performance for present-day conditions. *Climate Dynamics*, **25**, 237-263.
Morgenstern, O., et al., 2010a: Anthropogenic forcing of the Northern Annular Mode in CCMVal-2 models. *J. Geophys. Res.*, **115**, D00M03-.
Morgenstern, O., et al., 2010b: Anthropogenic forcing of the Northern Annular Mode in CCMVal-2 models. *Journal of Geophysical Research-Atmospheres*, **115**.
Moss, R. H., et al., 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747-756.
Msadek, R., and C. Frankignoul, 2009: Atlantic multidecadal oceanic variability and its influence on the atmosphere in a climate model. *Climate Dynamics*, **33**, 45-62.
Müller, W. A., and E. Roeckner, 2006: ENSO impact on midlatitude circulation patterns in future climate change projections. *Geophysical Research Letters*, **33**.
Murphy, J., B. Booth, M. Collins, G. Harris, D. Sexton, and M. Webb, 2007: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, DOI 10.1098/rsta.2007.2077. 1993-2028.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, -772.

Najac, J., J. Boe, and L. Terray, 2009: A multi-model ensemble approach for assessment of climate change impact on surface winds in France. *Climate Dynamics*, **32**, 615-634.

Najjar, R. G., et al., 2007: Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2). *Global Biogeochemical Cycles*, **21**.

Naoe, H., and K. Shibata, 2010: Equatorial quasi-biennial oscillation influence on northern winter extratropical circulation. *Journal of Geophysical Research-Atmospheres*, **115**.

Neale, R. B., J. H. Richter, and M. Jochum, 2008: The Impact of Convection on ENSO: From a Delayed Oscillator to a Series of Events. *Journal of Climate*, **21**, 5904-5924.

Neggers, R. A. J., 2009: A Dual Mass Flux Framework for Boundary Layer Convection. Part II: Clouds. *Journal of the Atmospheric Sciences*, **66**, 1489-1506.

Neggers, R. A. J., M. Kohler, and A. C. M. Beljaars, 2009: A Dual Mass Flux Framework for Boundary Layer Convection. Part I: Transport. *Journal of the Atmospheric Sciences*, **66**, 1465-1487.

Nelson, F. E., O. A. Anisimov, and N. I. Shiklomanov, 2002: Climate change and hazard zonation in the circum-Arctic permafrost regions. *Natural Hazards*, **26**, 203–225.

Nikiema, O., and R. Laprise, 2010: Diagnostic budget study of the internal variability in ensemble simulations of the Canadian RCM. *Climate Dynamics*, 10.1007/s00382-010-0834-y. 1-25.

Nikulin, G., E. Kjellstrom, U. Hansson, G. Strandberg, and A. Ullerstig, 2011: Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus Series a-Dynamic Meteorology and Oceanography*, DOI 10.1111/j.1600-0870.2010.00466.x. 41-55.

Ning, L., M. E. Mann, R. Crane, and T. Wagener, 2011: Probabilistic Projections of Climate Change for the Mid-Atlantic Region of the United States - Validation of Precipitation Downscaling During the Historical Era. *Journal of Climate*, **Early online release**.

Nishii, K., T. Miyasaka, Y. Kosaka, and H. Nakamura, 2009: Reproducibility and Future Projection of the Midwinter Storm-Track Activity over the Far East in the CMIP3 Climate Models in Relation to "Haru-Ichiban" over Japan. *Journal of the Meteorological Society of Japan*, **87**, 581-588.

Nishii, K., et al., 2011: Relationship of the reproducibility of multiple variables among global climate models. *Journal of the Meteorological Society of Japan*. submitted.

Nomura, D., H. Yoshikawa-Inoue, T. Toyota, and K. Shirasawa, 2010: Effects of snow, snowmelting and refreezing processes on air-sea-ice CO2 flux. *Journal of Glaciology*, **56**, 262-270.

Nowlin, W. D., M. Briscoe, N. Smith, M. J. McPhaden, D. Roemmich, P. Chapman, and J. F. Grassle, 2001: Evolution of a sustained ocean observing system. *Bulletin of the American Meteorological Society*, **82**, 1369-1376.

O'ishi, R., and A. Abe-Ouchi, 2011: Polar amplification in the mid-Holocene derived from dynamical vegetation change with a GCM. *Geophysical Research Letters*, **38**.

Oden, J. T., and S. Prudhomme, 2002: Estimation of modeling error in computational mechanics. *Journal of Computational Physics*, **182**, 496-515.

Ogura, T., S. Emori, M. J. Webb, Y. Tsushima, T. Yokohata, A. Abe-Ouchi, and M. Kimoto, 2008: Towards understanding cloud response in atmospheric GCMs: The use of tendency diagnostics. *Journal of the Meteorological Society of Japan*, **86**, 69-79.

Ohba, M., D. Nohara, and H. Ueda, 2010: Simulation of Asymmetric ENSO Transition in WCRP CMIP3 Multimodel Experiments. *Journal of Climate*, **23**, 6051-6067.

Ohgaito, R., and A. Abe-Ouchi, 2009: The effect of sea surface temperature bias in the PMIP2 AOGCMs on mid-Holocene Asian monsoon enhancement. *Climate Dynamics*, **33**, 975-983.

Oki, T., T. Nishimura, and P. Dirmeyer, 1999: Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). *J. Meteorol. Soc. Jap.*, **77**, 235-255.

Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, and C. S. B. Grimmond, 2008a: An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities. *Journal of Applied Meteorology and Climatology*, **47**, 1038-1060.

Oleson, K. W., et al., 2008b: Improvements to the Community Land Model and their impact on the hydrological cycle. *Journal of Geophysical Research-Biogeosciences*, **113**.

Ostle, N. J., et al., 2009: Integrating plant-soil interactions into global carbon cycle models. *Journal of Ecology*, **97**, 851-863.

Otto-Bliesner, B. L., et al., 2007: Last Glacial Maximum ocean thermohaline circulation: PMIP2 model intercomparisons and data constraints. *Geophysical Research Letters*, **34**, -.
Otto-Bliesner, B. L., et al., 2009: A comparison of PMIP2 model simulations and the MARGO proxy reconstruction for tropical sea surface temperatures at last glacial maximum. *Climate Dynamics*, **32**, 799-815.
Overland, J., M. Wang, N. Bond, J. Walsh, V. Kattsov, and W. Chapman, 2011: Considerations in the Selection of Global Climate Models for Regional Climate Projections: The Arctic as a Case Study. *Journal of Climate*, **24**, 1583-1597.
Paeth, H., 2011: Postprocessing of simulated precipitation for impact research in West Africa. Part I: model output statistics for monthly data. *Climate Dynamics*, **36**, 1321-1336.
Paeth, H., M. Rauthe, and S.-K. Min, 2008: Multi-model Bayesian assessment of climate change in the northern annular mode. *Global and Planetary Change*, **60**, 193-206.
Paeth, H., et al., 2011: Progress in regional downscaling of west African precipitation. *Atmospheric Science Letters*, 10.1002/asl.306. n/a-n/a.
Palmer, T. N., G. J. Shutts, and R. Swinbank, 1986: ALLEVIATION OF A SYSTEMATIC WESTERLY BIAS IN GENERAL-CIRCULATION AND NUMERICAL WEATHER PREDICTION MODELS THROUGH AN OROGRAPHIC GRAVITY-WAVE DRAG PARAMETRIZATION. *Quarterly Journal of the Royal Meteorological Society*, **112**, 1001-1039.
Palmer, T. N., F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, 2008: Toward seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bulletin of the American Meteorological Society*, **89**, 459-470.
Pan, L. L., and T. Li, 2008: Interactions between the tropical ISO and midlatitude low-frequency flow. *Climate Dynamics*, **31**, 375-388.
Parekh, P., F. Joos, and S. A. Muller, 2008: A modeling assessment of the interplay between aeolian iron fluxes and iron-binding ligands in controlling carbon dioxide fluctuations during Antarctic warm events. *Paleoceanography*, **23**.
Park, S., and C. S. Bretherton, 2009: The University of Washington Shallow Convection and Moist Turbulence Schemes and Their Impact on Climate Simulations with the Community Atmosphere Model. *Journal of Climate*, **22**, 3449-3469.
Park, W., and M. Latif, 2010: Pacific and Atlantic multidecadal variability in the Kiel Climate Model. *Geophysical Research Letters*, **37**.
Patricola, C. M., M. Li, Z. Xu, P. Chang, R. Saravanan, and J.-S. Hsieh, 2011: An Investigation of Tropical Atlantic Bias in a High-Resolution Coupled Regional Climate Model. submitted ed.
Pechony, O., and D. T. Shindell, 2009: Fire parameterization on a global scale. *Journal of Geophysical Research-Atmospheres*, **114**.
Penduff, T., et al., 2010: Impact of global ocean model resolution on sea-level variability with emphasis on interannual time scales. *Ocean Science*, **6**, 269-284.
Penduff, T., et al., 2011: Sea-level expression of intrinsic and forced ocean variabilities at interannual time scales. *Journal of Climate*, **24**, 5652-5670.
Pennell, C., and T. Reichler, 2011: On the Effective Number of Climate Models. *Journal of Climate*, 10.1175/2010JCLI3814.1.
Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *Journal of Climate*, **20**, 4356-4376.
Perlwitz, J., S. Pawson, R. Fogt, J. Nielsen, and W. Neff, 2008: Impact of stratospheric ozone hole recovery on Antarctic climate. *Geophysical Research Letters*, ARTN L08714, DOI 10.1029/2008GL033317. -.
Peterson, T. C., et al., 2009: State of the Climate in 2008. *Bulletin of the American Meteorological Society*, **90**, S1-S196.
Petoukhov, V., A. Ganopolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf, 2000: CLIMBER-2: a climate system model of intermediate complexity. Part I: model description and performance for present climate. *Climate Dynamics*, **16**, 1-17.
Petoukhov, V., et al., 2005: EMIC Intercomparison Project (EMIP-CO2): comparative analysis of EMIC simulations of climate, and of equilibrium and transient responses to atmospheric CO2 doubling. *Climate Dynamics*, **25**, 363-385.
Phillips, T. J., et al., 2004: Evaluating Parameterizations in General Circulation Models: Climate Simulation Meets Weather Prediction. *Bulletin of the American Meteorological Society*, **85**, 1903-1915.

Pielke, R. A., et al., 2007: An overview of regional land-use and land-cover impacts on rainfall. *Tellus Series B-Chemical and Physical Meteorology*, **59**, 587-601.
Pierce, D. W., 2001: Distinguishing coupled ocean-atmosphere interactions from background noise in the North Pacific. *Progress in Oceanography*, **49**, 331-352.
Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 8441-8446.
Pincus, R., C. P. Batstone, R. J. P. Hofmann, K. E. Taylor, and P. J. Glecker, 2008: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *Journal of Geophysical Research-Atmospheres*, **113**.
Piot, M., and R. von Glasow, 2008a: The potential importance of frost flowers, recycling on snow, and open leads for ozone depletion events. *Atmospheric Chemistry and Physics*, **8**, 2437–2467.
——, 2008b: The potential importance of frost flowers, recycling on snow, and open leads for ozone depletion events. *Atmospheric Chemistry and Physics*, **8**, 2437–2467.
Pitman, A., A. Arneth, and L. Ganzeveld, 2010: Review. Regionalizing global climate models. *International Journal of Climatology*, **Published online**.
Pitman, A. J., 2003: The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, **23**, 479-510.
Pitman, A. J., et al., 2009: Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. *Geophysical Research Letters*, **36**.
Plattner, G. K., et al., 2008: Long-term climate commitments projected with climate-carbon cycle models. *Journal of Climate*, **21**, 2721-2751.
Ploshay, J. J., and N.-C. Lau, 2010: Simulation of the Diurnal Cycle in Tropical Rainfall and Circulation during Boreal Summer with a High-Resolution GCM. *Monthly Weather Review*, **138**, 3434-3453.
Pongratz, J., T. Raddatz, C. H. Reick, M. Esch, and M. Claussen, 2009: Radiative forcing from anthropogenic land cover change since AD 800. *Geophysical Research Letters*, **36**.
Pope, and Stratton, 2002: The processes governing horizontal resolution sensitivity in a climate model. *Climate Dynamics*, **19**, 211-236-236.
Prentice, I. C., S. P. Harrison, D. Jolly, and J. Guiot, 1998: The climate and biomes of Europe at 6000 yr BP: Comparison of model simulations and pollen-based reconstructions. *Quaternary Science Reviews*, **17**, 659-668.
Prudhomme, C., and H. Davies, 2009: Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate. *Clim. Change*, **93**, 177-195.
Qian, T. T., A. Dai, K. E. Trenberth, and K. W. Oleson, 2006: Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations. *Journal of Hydrometeorology*, **7**, 953-975.
Raible, C. C., T. F. Stocker, M. Yoshimori, M. Renold, U. Beyerle, C. Casty, and J. Luterbacher, 2005: Northern Hemispheric trends of pressure indices and atmospheric circulation patterns in observations, reconstructions, and coupled GCM simulations. *Journal of Climate*.
Raisanen, J., 2007: How reliable are climate models? *Tellus Series a-Dynamic Meteorology and Oceanography*, **59**, 2-29.
Raisanen, J., and J. S. Ylhaisi, 2011: How Much Should Climate Model Output Be Smoothed in Space? *Journal of Climate*, **24**, 867-880.
Raisanen, J., L. Ruokolainen, and J. Ylhaisi, 2010: Weighting of model results for improving best estimates of climate change. *Climate Dynamics*, **35**, 407-422.
Ramaswamy, V., M. D. Schwarzkopf, W. J. Randel, B. D. Santer, B. J. Soden, and G. L. Stenchikov, 2006: Anthropogenic and natural influences in the evolution of lower stratospheric cooling. *Science*, **311**, 1138-1141.
Rammig, A., et al., 2010: Estimating the risk of Amazonian forest dieback. *New Phytologist*, **187**, 694-706.
Ramstein, G., M. Kageyama, J. Guiot, H. Wu, C. Hely, G. Krinner, and S. Brewer, 2007: How cold was Europe at the Last Glacial Maximum? A synthesis of the progress achieved since the first PMIP model-data comparison. *Climate of the Past*, **3**, 331-339.
Randall, D. A., M. F. Khairoutdinov, A. Arakawa, and W. W. Grabowski, 2003: Breaking the Cloud Parameterization Deadlock. *Bulletin of the American Meteorological Society*, **84**, 1547-1564.
Randall, D. A., et al., 2007: Climate Models and Their Evaluation. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, et al., Eds., Cambridge University Press,, 589-662.

Randerson, J. T., et al., 2009: Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Global Change Biology*, **15**, 2462-2484.
Rapaić, M., M. Leduc, and R. Laprise, 2010: Evaluation of the internal variability and estimation of the downscaling ability of the Canadian Regional Climate Model for different domain sizes over the north Atlantic region using the Big-Brother experimental approach. *Climate Dynamics*, 10.1007/s00382-010-0845-8. 1-23.
Raphael, M. N., and M. M. Holland, 2006: Twentieth century simulation of the southern hemisphere climate in coupled models. Part 1: large scale circulation variability. *Climate Dynamics*, **26**, 217-228.
Rauscher, S. A., E. Coppola, C. Piani, and F. Giorgi, 2010: Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Climate Dynamics*, **35**, 685-711.
Rauser, F., P. Korn, and J. Marotzke, 2011: Predicting goal error evolution from near-initial information: a learning algorithm. *Journal of Computational Physics*, **230**, 7284-7299.
Rayner, N. A., et al., 2003: Global analysis of sea surface temperature, sea ice, and night marine air temperature since the late ninteeth century. *J. Geophys. Res.*, **108**, doi:10,1029/2002JD002670.
Redelsperger, J.-L., C. D. Thorncroft, A. Diedhiou, T. Lebel, D. J. Parker, and J. Polcher, 2006: African Monsoon Multidisciplinary Analysis: An International Research Project and Field Campaign. *Bulletin of the American Meteorological Society*, **87**, 1739-1746.
Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography*, **12**, 1154-1158.
Reichler, T., and J. Kim, 2008: How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, **89**, 303-311.
Richter, I., and S.-P. Xie, 2008: On the origin of equatorial Atlantic biases in coupled general circulation models. *Climate Dynamics*, **31**, 587-598.
Richter, J. H., F. Sassi, R. R. Garcia, K. Matthes, and C. A. Fischer, 2008: Dynamics of the middle atmosphere as simulated by the Whole Atmosphere Community Climate Model, version 3 (WACCM3). *Journal of Geophysical Research-Atmospheres*, **113**.
Ridgwell, A., I. Zondervan, J. C. Hargreaves, J. Bijma, and T. M. Lenton, 2007: Assessing the potential long-term increase of oceanic fossil fuel CO2 uptake due to CO2-calcification feedback. *Biogeosciences*, **4**, 481-492.
Ringer, M. A., J. M. Edwards, and A. Slingo, 2003: Simulation of satellite channel radiances in the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **129**, 1169-1190.
Ritz, S. P., T. F. Stocker, and F. Joos, 2011: A Coupled Dynamical Ocean-Energy Balance Atmosphere Model for Paleoclimate Studies. *Journal of Climate*, **24**, 349-375.
Roberts, M. J., et al., 2004: Impact of an eddy-permitting ocean resolution on control and climate change simulations with a global coupled GCM. *Journal of Climate*, **17**, 3-20.
Roberts, M. J., et al., 2009: Impact of Resolution on the Tropical Pacific Circulation in a Matrix of Coupled Models. *Journal of Climate*, **22**, 2541-2556.
Robinson, D. A., and A. Frei, 2000: Seasonal variability of northern hemisphere snow extent using visible satellite data. *Professional Geographer*, **51**, 307–314.
Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, **133**, 129-146.
Roe, G. H., and M. B. Baker, 2007: Why is climate sensitivity so unpredictable? *Science*, **318**, 629-632.
Roeckner, E., et al., 2006: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *Journal of Climate*, **19**, 3771-3791.
Roemmich, D., and W. J. Gould, 2003: The future of climate observations in the global ocean. *Sea Technology*, **44**, 10-+.
Roemmich, D., and T. Argo Steering, 2009: Argo THE CHALLENGE OF CONTINUING 10 YEARS OF PROGRESS. *Oceanography*, **22**, 46-55.
Rojas, M., 2006: Multiply Nested Regional Climate Simulation for Southern South America: Sensitivity to Model Resolution. *Monthly Weather Review*, **134**, 2208-2223.
Rojas, M., and P. I. Moreno, 2011: Atmospheric circulation changes and neoglacial conditions in the Southern Hemisphere mid-latitudes: insights from PMIP2 simulations at 6 kyr. *Climate Dynamics*, **37**, 357-375.
Rojas, M., et al., 2009: The Southern Westerlies during the last glacial maximum in PMIP2 simulations. *Climate Dynamics*, **32**, 525-548.

Rotstayn, L. D., et al., 2010: Improved simulation of Australian climate and ENSO-related rainfall variability in a global climate model with an interactive aerosol treatment. *International Journal of Climatology*, **30**, 1067-1088.

Rougier, J., D. M. H. Sexton, J. M. Murphy, and D. Stainforth, 2009: Analyzing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments. *Journal of Climate*, **22**, 3540-3557.

Roy, T., et al., 2011: Regional impacts of climate change and atmospheric CO2 on future ocean carbon uptake: A multi-model linear feedback analysis. *Journal of Climate*, **24**, 2300-2318.

Ruckstuhl, C., and J. R. Norris, 2009: How do aerosol histories affect solar "dimming" and "brightening" over Europe?: IPCC-AR4 models versus observations. *Journal of Geophysical Research-Atmospheres*, **114**.

Rummukainen, M., 2010: State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews: Climate Change*, **1**, 82-96.

Russell, J. L., R. J. Stouffer, and K. W. Dixon, 2006: Intercomparison of the Southern Ocean circulations in IPCC coupled model control simulations. *Journal of Climate*, **19**, 4560-4575.

Ruti, P. M., et al., 2011: The West African climate system: a review of the AMMA model inter-comparison initiatives. *Atmospheric Science Letters*, 10.1002/asl.305. n/a-n/a.

Saji, N. H., S. P. Xie, and T. Yamagata, 2006: Tropical Indian Ocean variability in the IPCC twentieth-century climate simulations. *Journal of Climate*, **19**, 4397-4417.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360-363.

Sakamoto, T. T., et al., 2011: MIROC4h – a new high-resolution atmosphere-ocean coupled general circulation model. *Journal of Meteorological Society of Japan*. submitted.

Samuelsson, P., E. Kourzeneva, and D. Mironov, 2010: The impact of lakes on the European climate as simulated by a regional climate model. *Boreal Environment Research*, **15**, 113-129.

Sanchez-Gomez, E., S. Somot, and M. Deque, 2009: Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961-2000. *Climate Dynamics*, **33**, 723-736.

Sanderson, B., et al., 2008a: Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *Journal of Climate*, DOI 10.1175/2008JCLI1869.1. 2384-2400.

Sanderson, B. M., K. M. Shell, and W. Ingram, 2010: Climate feedbacks determined using radiative kernels in a multi-thousand member ensemble of AOGCMs. *Climate Dynamics*, **35**, 1219-1236.

Sanderson, B. M., C. Piani, W. J. Ingram, D. A. Stone, and M. R. Allen, 2008b: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Climate Dynamics*, **30**, 175-190.

Santer, B., et al., 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc. Natl. Acad. Sci. U. S. A.*, DOI 10.1073/pnas.0901736106. 14778-14783.

Santer, B., et al., 2007: Identification of human-induced changes in atmospheric moisture content. *Proc. Natl. Acad. Sci. U. S. A.*, DOI 10.1073/pnas.0702872104. 15248-15253.

Santer, B., et al., 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *International Journal of Climatology*, DOI 10.1002/joc.1756. 1703-1722.

Santer, B., et al., 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, DOI 10.1126/science.1114867. 1551-1556.

Santer, B. D., et al., 2011: Separating Signal and Noise in Atmospheric Temperature Changes: The Importance of Timescale. *J. Geophys. Res.*, **116**.

Sasaki, H., and K. Kurihara, 2008: Relationship between Precipitation and Elevation in the Present Climate Reproduced by the Non-hydrostatic Regional Climate Model. *SOLA*, **4**, 109-112.

Sato, K., and T. J. Dunkerton, 1997: Estimates of momentum flux associated with equatorial Kelvin and gravity waves. *Journal of Geophysical Research-Atmospheres*, **102**, 26247-26261.

Sato, T., H. Miura, M. Satoh, Y. N. Takayabu, and Y. Q. Wang, 2009: Diurnal Cycle of Precipitation in the Tropics Simulated in a Global Cloud-Resolving Model. *Journal of Climate*, **22**, 4809-4826.

Scaife, A. A., N. Butchart, C. D. Warner, and R. Swinbank, 2002: Impact of a spectral gravity wave parameterization on the stratosphere in the met office unified model. *Journal of the Atmospheric Sciences*, **59**, 1473-1489.

Scaife, A. A., J. R. Knight, G. K. Vallis, and C. K. Folland, 2005: A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophysical Research Letters*, **32**.

Scaife, A. A., T. Woollings, J. Knight, G. Martin, and T. Hinton, 2010: Atmospheric Blocking and Mean Biases in Climate Models. *Journal of Climate*, **23**, 6143-6152.
Scaife, A. A., N. Butchart, C. D. Warner, D. Stainforth, W. Norton, and J. Austin, 2000: Realistic Quasi-Biennial Oscillations in a simulation of the global climate. *Geophysical Research Letters*, **27**, 3481-3484.
Scaife, a. A., et al., 2011: Climate Change and Stratosphere-Troposphere Interaction. *Clim. Dyn.*, 10.1007/s00382-011-1080-7.
Scaife, A. A., et al., 2009: The CLIVAR C20C project: selected twentieth century climate events. *Climate Dynamics*, **33**, 603-614.
Scherrer, S. C., 2010: Present-day interannual variability of surface climate in CMIP3 models and its relation to future warming. *International Journal of Climatology*, doi: 10.1002/joc.2170.
Schmidli, J., C. Goodess, C. Frei, M. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith, 2007: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research-Atmospheres*, **112**, -.
Schmith, T., 2008: Stationarity of regression relationships: Application to empirical downscaling. *Journal of Climate*, **21**, 4529-4537.
Schmittner, A., M. Latif, and B. Schneider, 2005a: Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations. *Geophys. Res. Lett.*, **32**, L23710-.
Schmittner, A., A. Oschlies, X. Giraud, M. Eby, and H. L. Simmons, 2005b: A global model of the marine ecosystem for long-term simulations: Sensitivity to ocean mixing, buoyancy forcing, particle sinking, and dissolved organic matter cycling. *Global Biogeochemical Cycles*, **19**.
Schneider, B., M. Latif, and A. Schmittner, 2007: Evaluation of different methods to assess model projections of the future evolution of the Atlantic meridional overturning circulation. *Journal of Climate*, **20**, 2121-2132.
Schneider, B., et al., 2008: Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models. *Biogeosciences*, **5**, 597-614.
Schott, F. A., S.-P. Xie, and J. P. McCreary, Jr., 2009: Indian Ocean circulation and climate variability. *Rev. Geophys.*, **47**, -.
Schramm, J. L., M. M. Holland, J. A. Curry, and E. E. Ebert, 1997: Modeling the thermodynamics of a sea ice thickness 1. Sensitivity to ice thickness resolution. *J. Geophys. Res.*, **102**, 23,079-023,091.
Schultz, M. G., et al., 2008: Global wildland fire emissions from 1960 to 2000. *Global Biogeochemical Cycles*, **22**.
Schurgers, G., U. Mikolajewicz, M. Groger, E. Maier-Reimer, M. Vizcaino, and A. Winguth, 2008: Long-term effects of biogeophysical and biogeochemical interactions between terrestrial biosphere and climate under anthropogenic climate change. *Global and Planetary Change*, **64**, 26-37.
Scinocca, J. F., and N. A. McFarlane, 2000: The parametrization of drag induced by stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2353-2393.
Scott, F., and D. L. Feltham, 2010: A model of the three-dimensional evolution of Arctic melt ponds on first-year and multiyear sea ice. *J. Geophys. Res.*, **115**, C12064.
Seager, R., and R. Murtugudde, 1997: Ocean dynamics, thermocline adjustment, and regulation of tropical SST. *Journal of Climate*, **10**, 521-534.
Segui, P., A. Ribes, E. Martin, F. Habets, and J. Boe, 2010: Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins. *Journal of Hydrology*, **383**, 111-124.
Seidel, D. J., Q. Fu, W. J. Randel, and T. J. Reichler, 2008: Widening of the tropical belt in a changing climate. *Nature Geoscience*, **1**, 21-24.
Sellers, P. J., et al., 1996: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. *Journal of Climate*, **9**, 676-705.
Selten, F. M., G. W. Branstator, H. A. Dijkstra, and M. Kliphuis, 2004: Tropical origins for recent and future Northern Hemisphere climate change. *Geophys. Res. Lett.*, **31**, L21205-.
Semenov, V. A., M. Latif, J. H. Jungclaus, and W. Park, 2008: Is the observed NAO variability during the instrumental record unusual? *Geophys. Res. Lett.*, **35**, L11701-.
Seneviratne, S., et al., 2012: Changes in Climate Extremes and their Impacts on the Natural Physical Environment. *IPCC WGI/WGII Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX)*, in press.
Seneviratne, S. I., et al., 2010: Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, **99**, 125-161.

Senior, C. A., and J. F. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophysical Research Letters*, **27**, 2685-2688.
Separovic, L., R. De Elia, and R. Laprise, 2008: Reproducible and Irreproducible Components in Ensemble Simulations with a Regional Climate Model. *Monthly Weather Review*, DOI 10.1175/2008MWR2393.1. 4942-4961.
Shaffer, G., and J. L. Sarmiento, 1995: Biogeochemical Cycling in the Global Ocean .1. A New, Analytical Model with Continuous Vertical Resolution and High-Latitude Dynamics. *Journal of Geophysical Research-Oceans*, **100**, 2659-2672.
Shaffer, G., S. M. Olsen, and J. O. P. Pedersen, 2008: Presentation, calibration and validation of the low-order, DCESS Earth System Model (Version 1). *Geoscientific Model Development*, **1**, 17-51.
Shaffrey, L. C., et al., 2009: UK HiGEM: The New UK High-Resolution Global Environment Model-Model Description and Basic Evaluation. *Journal of Climate*, **22**, 1861-1896.
Sheffield, J., and E. F. Wood, 2008: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Climate Dynamics*, **31**, 79-105.
Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the radiative kernel technique to calculate climate feedbacks in NCAR's Community Atmospheric Model. *Journal of Climate*, **21**, 2269-2282.
Shibata, K., and M. Deushi, 2005: Radiative effect of ozone on the quasi-biennial oscillation in the equatorial stratosphere. *Geophysical Research Letters*, **32**.
Shin, S. I., D. Sardeshmukh, and K. Pegion, 2010: Realism of local and remote feedbacks on tropical sea surface temperatures in climate models. *Journal of Geophysical Research-Atmospheres*, **115**.
Shindell, D., et al., 2006: Simulations of preindustrial, present-day, and 2100 conditions in the NASA GISS composition and climate model G-PUCCINI. *Atmos. Chem. Phys.*, **6**, 4427-4459.
Siebesma, A. P., P. M. M. Soares, and J. Teixeira, 2007: A Combined Eddy-Diffusivity Mass-Flux Approach for the Convective Boundary Layer. *Journal of the Atmospheric Sciences*, **64**, 1230-1248.
Sillmann, J., and M. Croci-Maspoli, 2009: Present and future atmospheric blocking and its impact on European mean and extreme climate. *Geophys. Res. Lett.*, **36**, L10702-.
Simmons, H. L., S. R. Jayne, L. C. St Laurent, and A. J. Weaver, 2004: Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Modelling*, **6**, 245-263.
Sitch, S., et al., 2003: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, **9**, 161-185.
Sitch, S., et al., 2008: Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Global Change Biology*, **14**, 2015-2039.
Six, K. D., and E. MaierReimer, 1996: Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles*, **10**, 559-583.
Sloyan, B. M., and I. V. Kamenkovich, 2007: Simulation of Subantarctic Mode and Antarctic Intermediate Waters in climate models. *Journal of Climate*, **20**, 5061-5080.
Smirnov, D., and D. J. Vimont, 2011: Variability of the Atlantic Meridional Mode during the Atlantic Hurricane Season. *Journal of Climate*, **24**, 1409-1424.
Smith, B., P. Samuelsson, A. Wramneby, and M. Rummukainen, 2011a: A model of the coupled dynamics of climate, vegetation and terrestrial ecosystem biogeochemistry for regional applications. *Tellus Series a-Dynamic Meteorology and Oceanography*, DOI 10.1111/j.1600-0870.2010.00477.x. 87-106.
Smith, P. C., N. De Noblet-Ducoudre, P. Ciais, P. Peylin, N. Viovy, Y. Meurdesoif, and A. Bondeau, 2010a: European-wide simulations of croplands using an improved terrestrial biosphere model: Phenology and productivity. *Journal of Geophysical Research-Biogeosciences*, **115**.
Smith, P. C., P. Ciais, P. Peylin, N. De Noblet-Ducoudre, N. Viovy, Y. Meurdesoif, and A. Bondeau, 2010b: European-wide simulations of croplands using an improved terrestrial biosphere model: 2. Interannual yields and anomalous CO2 fluxes in 2003. *Journal of Geophysical Research-Biogeosciences*, **115**.
Smith, R. D., M. E. Maltrud, F. O. Bryan, and M. W. Hecht, 2000: Numerical simulation of the North Atlantic Ocean at 1/10 degrees. *Journal of Physical Oceanography*, **30**, 1532-1561.
Smith, S. J., J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. Delgado Arias, 2011b: Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.*, **11**, 1101-1116.
Soden, B. J., and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean-atmosphere models. *Journal of Climate*, **19**, 3354-3360.
Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *Journal of Climate*, **21**, 3504-3520.

Sokolov, A. P., C. E. Forest, and P. H. Stone, 2010: Sensitivity of climate change projections to uncertainties in the estimates of observed changes in deep-ocean heat content. *Climate Dynamics*, **34**, 735-745.
Sokolov, A. P., et al., 2009: Probabilistic Forecast for Twenty-First-Century Climate Based on Uncertainties in Emissions (Without Policy) and Climate Parameters. *Journal of Climate*, **22**, 5175-5204.
Solanki, S. K., I. G. Usoskin, B. Kromer, M. Schussler, and J. Beer, 2004: Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature*, **431**, 1084-1087.
Solomon, A., and D. Zhang, 2006: Pacific subtropical cell variability in coupled climate model simulations of the late 19th–20th century. *Ocean Modelling*, **15**, 236–249.
Solomon, S., R. Portmann, T. Sasaki, D. Hofmann, and D. Thompson, 2005: Four decades of ozonesonde measurements over Antarctica. *Journal of Geophysical Research-Atmospheres*, ARTN D21311, DOI 10.1029/2005JD005917. -.
Somot, S., F. Sevault, M. Deque, and M. Crepon, 2008: 21st century climate change scenario for the Mediterranean using a coupled atmosphere-ocean regional climate model. *Global and Planetary Change*, DOI 10.1016/j.gloplacha.2007.10.003. 112-126.
Son, S., et al., 2008: The impact of stratospheric ozone recovery on the Southern Hemisphere westerly jet. *Science*, DOI 10.1126/science.1155939. 1486-1489.
Son, S., et al., 2010: Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *Journal of Geophysical Research-Atmospheres*, ARTN D00M07, DOI 10.1029/2010JD014271. -.
SPARC-CCMVal, 2010: SPARC Report on the Evaluation of Chemistry-Climate Models.
Sperber, K. R., and H. Annamalai, 2008: Coupled model simulations of boreal summer intraseasonal (30-50 day) variability, Part 1: Systematic errors and caution on use of metrics. *Climate Dynamics*, **31**, 345-372.
Sperber, K. R., S. Gualdi, S. Legutke, and V. Gayler, 2005: The Madden-Julian oscillation in ECHAM4 coupled and uncoupled general circulation models. *Climate Dynamics*, **25**, 117-140.
Sperber, K. R., et al., 2010: Monsoon Fact Sheet: CLIVAR Asian-Australian Monsoon Panel.
Stainforth, D. A., et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403-406.
Stephens, G. L., and C. D. Kummerow, 2007: The remote sensing of clouds and precipitation from space: A review. *Journal of the Atmospheric Sciences*, **64**, 3742-3765.
Stephens, G. L., et al., 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211.
Stephenson, D., C. Coelho, F. Doblas-Reyes, and M. Balmaseda, 2005: Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus Series a-Dynamic Meteorology and Oceanography*. 253-264.
Stephenson, D. B., V. Pavan, M. Collins, M. M. Junge, R. Quadrelli, and C. M. G. Participating, 2006: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment. *Climate Dynamics*, **27**, 401-420.
Stevenson, S., B. Fox-Kemper, M. Jochum, R. Neale, C. Deser, and G. Meehl, 2011: Will there be a significant change to El Nino in the 21st century? . *J. Clim. submitted*.
Stocker, B. D., K. Strassmann, and F. Joos, 2011: Sensitivity of Holocene atmospheric CO(2) and the modern carbon budget to early human land use: analyses with a process-based model. *Biogeosciences*, **8**, 69-88.
Stoner, A. M. K., K. Hayhoe, and D. J. Wuebbles, 2009: Assessing General Circulation Model Simulations of Atmospheric Teleconnection Patterns. *Journal of Climate*, **22**, 4348-4372.
Stott, P. A., and C. E. Forest, 2007: Ensemble climate predictions using climate models and observational constraints. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, **365**, 2029-2052.
Strassmann, K. M., F. Joos, and G. Fischer, 2008: Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric CO2 increases and future commitments due to losses of terrestrial sink capacity. *Tellus Series B-Chemical and Physical Meteorology*, **60**, 583-603.
Stroeve, J., M. Holland, W. Meier, T. Scambos, and M. Serreze, 2007: Arctic sea ice decline: Faster than forecast. *Geophysical Research Letters*, **34**, -.
Su, H., D. E. Waliser, J. H. Jiang, J. L. Li, W. G. Read, J. W. Waters, and A. M. Tompkins, 2006: Relationships of upper tropospheric water vapor, clouds and SST: MLS observations, ECMWF analyses and GCM simulations. *Geophysical Research Letters*, **33**.
Sueyoshi, M., and T. Yasuda, 2009: Reproducibility and future projection of the ocean first baroclinic Rossby radius based on the CMIP3 multi-model dataset. *気象集誌*, **87**, 821-827.

Sun, D.-Z., Y. Yu, and T. Zhang, 2009: Tropical Water Vapor and Cloud Feedbacks in Climate Models: A Further Assessment Using Coupled Simulations. *Journal of Climate*, **22**, 1287-1304.
Sutton, R. T., B. W. Dong, and J. M. Gregory, 2007: Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations. *Geophysical Research Letters*, **34**.
Svensson, G., and A. Holtslag, 2009: Analysis of Model Results for the Turning of the Wind and Related Momentum Fluxes in the Stable Boundary Layer. *Boundary-Layer Meteorology*, **132**, 261-277-277.
Swart, N. C., and J. C. Fyfe, 2011: Ocean carbon uptake and storage influenced by wind bias in global climate models. *Nature Climate Change*.
Tagliabue, A., et al., 2009: Quantifying the roles of ocean circulation and biogeochemistry in governing ocean carbon-13 and atmospheric carbon dioxide at the last glacial maximum. *Climate of the Past*, **5**, 695-706.
Takahashi, M., 1999: Simulation of the stratospheric quasi-biennial oscillation in a general circulation model. *Geophysical Research Letters*, **26**, 1307-1310.
Takle, E. S., et al., 2007: Transferability intercomparison - An opportunity for new insight on the global water cycle and energy budget. *Bulletin of the American Meteorological Society*, **88**, 375-+.
Tao, W. K., and M. W. Moncrieff, 2009: MULTISCALE CLOUD SYSTEM MODELING. *Rev. Geophys.*, **47**.
Tao, W. K., et al., 2009: A MULTISCALE MODELING SYSTEM Developments, Applications, and Critical Issues. *Bulletin of the American Meteorological Society*, **90**, 515-+.
Taylor, K. E., 2011: A Summary of the CMIP5 Experiment Design.
Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2011: A Summary of the CMIP5 Experiment Design.
Tebaldi, C., and R. Knutti, 2007a: The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, DOI 10.1098/rsta.2007.2076. 2053-2075.
Tebaldi, C., and R. Knutti, 2007b: The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, **365**, 2053-2075.
Teixeira, J., et al., 2008: Parameterization of the atmospheric boundary layer. *Bulletin of the American Meteorological Society*, **89**, 453-458.
Teixeira, J., et al., 2011: Tropical and Subtropical Cloud Transitions in Weather and Climate Prediction Models: The GCSS/WGNE Pacific Cross-Section Intercomparison (GPCI). *Journal of Climate*, **24**, 5223-5256.
Teng, Q., J. Fyfe, and A. Monahan, 2007: Northern Hemisphere circulation regimes: observed, simulated and predicted. 867-879.
Teutschbein, C., F. Wetterhall, and J. Seibert, 2011: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. *Climate Dynamics*, **37**, 2087-2105.
Textor, C., et al., 2007: The effect of harmonized emissions on aerosol properties in global models - an AeroCom experiment. *Atmospheric Chemistry and Physics*, **7**, 4489-4501.
Thomas, H., et al., 2008: Changes in the North Atlantic Oscillation influence CO2 uptake in the North Atlantic over the past 2 decades. *Global Biogeochemical Cycles*, **22**.
Thompson, D. W. J., M. P. Baldwin, and J. M. Wallace, 2002: Stratospheric connection to Northern Hemisphere wintertime weather: Implications for prediction. *Journal of Climate*, **15**, 1421-1428.
Thorne, P. W., et al., 2011: A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. *J. Geophys. Res.*, **116**, D12116.
Thornton, P. E., J. F. Lamarque, N. A. Rosenbloom, and N. M. Mahowald, 2007: Influence of carbon-nitrogen cycle coupling on land model response to CO2 fertilization and climate variability. *Global Biogeochemical Cycles*, **21**.
Timbal, B., and D. Jones, 2008: Future projections of winter rainfall in southeast Australia using a statistical downscaling technique. *Clim. Change*, **86**, 165-187.
Timmermann, A., S. Lorenz, S.-I. An, A. Clement, and S.-P. Xie, 2007: The Effect of Orbital Forcing on the Mean Climate and Variability of the Tropical Pacific. *J. Climate*, **20**, 4147-4159.
Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and Internal Twentieth-Century SST Trends in the North Atlantic. *Journal of Climate*, **22**, 1469-1481.
Tjiputra, J. F., K. Assmann, M. Bentsen, I. Bethke, O. H. Ottera, C. Sturm, and C. Heinze, 2010: Bergen Earth system model (BCM-C): model description and regional climate-carbon cycle feedbacks assessment. *Geoscientific Model Development*, **3**, 123-141.

Tolika, K., C. Anagnostopoulo, P. Maheras, and M. Vafiadis, 2008: Simulation of future changes in extreme rainfall and temperature conditions over the Greek area: A comparison of two statistical downscaling approaches. *Global and Planetary Change*, **63**, 132-151.

Trenberth, K. E., and J. T. Fasullo, 2010: Simulation of Present-Day and Twenty-First-Century Energy Budgets of the Southern Oceans. *Journal of Climate*, **23**, 440-454.

Trenberth, K. E., D. P. Stepaniak, and J. M. Caron, 2000: The global monsoon as seen through the divergent atmospheric circulation. *Journal of Climate*, **13**, 3969-3993.

Tryhorn, L., and A. DeGaetano, 2011: A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. *International Journal of Climatology*, **31**, 1975-1989.

Tschumi, T., F. Joos, and P. Parekh, 2008: How important are Southern Hemisphere wind changes for low glacial carbon dioxide? A model study. *Paleoceanography*, **23**.

Tschumi, T., F. Joos, M. Gehlen, and C. Heinze, 2011: Deep ocean ventilation, carbon isotopes, marine sedimentation and the deglacial CO(2) rise. *Climate of the Past*, **7**, 771-800.

Tsujino, H., M. Hirabara, H. Nakano, T. Yasuda, T. Motoi, and G. Yamanaka, 2011: Simulating present climate of the global ocean-ice system using the Meteorological Research Institute Community Ocean Model (MRI.COM): simulation characteristics and variability in the Pacific sector. *Journal of Oceanography*, **67**, 449-479.

Tsushima, Y., et al., 2006: Importance of the mixed-phase cloud distribution in the control climate for assessing the response of clouds to carbon dioxide increase: a multi-model study. *Clim. Dyn.*, **27**, 113-126.

Uotila, P., A. Lynch, J. Cassano, and R. Cullather, 2007: Changes in Antarctic net precipitation in the 21st century based on Intergovernmental Panel on Climate Change (IPCC) model scenarios. *Journal of Geophysical Research-Atmospheres*, **112**, -.

Valcke, S., E. Guilyardi, and C. Larsson, 2006: PRISM and ENES: a European approach to Earth system modelling. *Concurrency and Computation-Practice & Experience*, **18**, 247-262.

van den Hurk, B., and E. van Meijgaard, 2010: Diagnosing Land-Atmosphere Interaction from a Regional Climate Model Simulation over West Africa. *Journal of Hydrometeorology*, DOI 10.1175/2009JHM1173.1. 467-481.

van Haren, R., G. J. van Oldenborgh, G. Lenderink, M. Collins, and W. Hazeleger, 2011 (submitted): SST and circulation trend biases cause an underestimation of European precipitation trends. *Climate Dynamics*.

van Oldenborgh, G. J., S. Y. Philip, and M. Collins, 2005: El Niño in a changing climate: a multi-model study. *Ocean Science*, **1**, 81–95.

van Roosmalen, L., J. H. Christensen, M. B. Butts, K. H. Jensen, and J. C. Refsgaard, 2010: An intercomparison of regional climate model data for hydrological impact studies in Denmark. *Journal of Hydrology*, **380**, 406-419.

van Ulden, A., and G. van Oldenborgh, 2006: Large-scale atmospheric circulation biases and changes in global climate model simulations and their importance for climate change in Central Europe. *Atmospheric Chemistry and Physics*. 863-881.

van Vliet, M., S. Blenkinsop, A. Burton, C. Harpham, H. Broers, and H. Fowler: A multi-model ensemble of downscaled spatial climate change scenarios for the Dommel catchment, Western Europe. *Clim. Change*, 10.1007/s10584-011-0131-8. 1-29.

——, 2011: A multi-model ensemble of downscaled spatial climate change scenarios for the Dommel catchment, Western Europe. *Clim. Change*, 10.1007/s10584-011-0131-8. 1-29.

van Vuuren, D. P., J. Feddema, J.-F. Lamarque, K. Riahi, S. Rose, S. Smith, and K. Hibbard: Work plan for data exchange between the Integrated Assessment and Climate Modeling community in support of Phase-0 of scenario analysis for climate change assessment (Representative Community Pathways).

Vancoppenolle, M., T. Fichefet, and H. Goosse, 2009a: Simulating the mass balance and salinity of arctic andantarctic sea ice. 2. importance of sea ice salinity variations. *Ocean Modelling*, **27**, 54-69.

Vancoppenolle, M., T. Fichefet, H. Goosse, S. Bouillon, G. Madec, and M. A. M. Maqueda, 2009b: Simulating the mass balance and salinity of arctic and antarctic sea ice. 1. model description and validation. *Ocean Modelling*, **27**, 33-53.

Vannière, B., E. Guilyardi, G. Madec, F. J. Doblas-Reyes, and S. Woolnough, 2011: Using seasonal hindcasts to understand the origin of the equatorial cold tongue bias in CGCMs and its impact on ENSO. submitted ed., Clim Dyn.

Vecchi, G. A., B. J. Soden, A. T. Wittenberg, I. M. Held, A. Leetmaa, and M. J. Harrison, 2006a: Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature*, **327**, 216–-219.

Vecchi, G. A., B. J. Soden, A. T. Wittenberg, I. M. Held1, A. Leetmaa, and M. J. Harrison, 2006b: Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature*, **327**, 216–-219.

Veljovic, K., B. Rajkovic, M. J. Fennessy, E. L. Altshuler, and F. Mesinger, 2010: Regional climate modeling: Should one attempt improving on the large scales? Lateral boundary condition scheme: Any impact? *Meteorologische Zeitschrift*, **19**, 237-246.

Vera, C., and G. Silvestri, 2009: Precipitation interannual variability in South America from the WCRP-CMIP3 multi-model dataset. *Climate Dynamics*, **32**, 1003-1014.

Verlinde, J., et al., 2007: The Mixed-Phase Arctic Cloud Experiment. *Bulletin of the American Meteorological Society*, **88**, 205-221.

Vincent, E. M., et al., 2011: Global study of ocean surface cooling induced by tropical cyclones: controlling mechanisms. submitted.

Vizcaino, M., U. Mikolajewicz, J. Jungclaus, and G. Schurgers, 2010: Climate modification by future ice sheet changes and consequences for ice sheet mass balance. *Climate Dynamics*, **34**, 301-324.

Vizcaino, M., U. Mikolajewicz, M. Groger, E. Maier-Reimer, G. Schurgers, and A. M. E. Winguth, 2008: Long-term ice sheet-climate interactions under anthropogenic greenhouse forcing simulated with a complex Earth System Model. *Climate Dynamics*, **31**, 665-690.

Volodin, E. M., 2008a: Methane cycle in the INM RAS climate model. *Izvestiya Atmospheric and Oceanic Physics*, **44**, 153-159.

——, 2008b: Relation between temperature sensitivity to doubled carbon dioxide and the distribution of clouds in current climate models. *Izvestiya Atmospheric and Oceanic Physics*, **44**, 288-299.

Volodin, E. M., N. A. Dianskii, and A. V. Gusev, 2010: Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya Atmospheric and Oceanic Physics*, **46**, 414-431.

Vrac, M., and P. Naveau, 2008: Stochastic downscaling of precipitation: From dry events to heavy rainfall (vol 43, art no W07402, 2007). *Water Resources Research*, **44**, -.

Waelbroeck, C., et al., 2009: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geoscience*, **2**, 127-132.

Waliser, D., K. W. Seo, S. Schubert, and E. Njoku, 2007: Global water cycle agreement in the climate models assessed in the IPCC AR4. *Geophysical Research Letters*, **34**.

Waliser, D., Y. Chao, J. Ci, H. Zhang, and K. Seo, 2011: How Well Do Global Climate Models Simulate Mean Sea Surface Salinity? . *Journal of Climate, submitted.*

Waliser, D. E., et al., 2003: AGCM simulations of intraseasonal variability associated with the Asian summer monsoon. *Climate Dynamics*, **21**, 423-446.

Waliser, D. E., et al., 2009: Cloud ice: A climate model challenge with signs and expectations of progress. *J. Geophys. Res.*, **114**, D00A21, doi 10.1029/2008jd010015.

Walther, A., J.-H. Jeong, G. Nikulin, C. Jones, and D. Chen, 2011: Evaluation of the warm season diurnal cycle of precipitation over Sweden simulated by the Rossby Centre regional climate model RCA3. *Atmospheric Research*, **Accepted**.

Wang, B., and Q. H. Ding, 2008: Global monsoon: Dominant mode of annual variation in the tropics. *Dynamics of Atmospheres and Oceans*, **44**, 165-183.

Wang, B., H. J. Kim, K. Kikuchi, and A. Kitoh, 2011: Diagnostic metrics for evaluation of annual and diurnal cycles. *Climate Dynamics*, **37**, 941-955.

Wang, C., and J. Picaut, 2004: Understanding ENSO Physics - A Review. *Earth's climate: the ocean-atmosphere interaction*, AGU, Washington D.C., C. Wang, S.-P. Xie and J.A. Carton Eds., 21-48.

Wang, J. F., and X. B. Zhang, 2008: Downscaling and projection of winter extreme daily precipitation over North America. *Journal of Climate*, **21**, 923-937.

Wang, S. S., A. P. Trishchenko, K. V. Khlopenkov, and A. Davidson, 2006: Comparison of International Panel on Climate Change Fourth Assessment Report climate model simulations of surface albedo with satellite products over northern latitudes. *Journal of Geophysical Research-Atmospheres*, **111**.

Wang, T., H. J. Wang, and D. B. Jiang, 2010: Mid-Holocene East Asian summer climate as simulated by the PMIP2 models. *Palaeogeography Palaeoclimatology Palaeoecology*, **288**, 93-102.

Wang, Y., M. Notaro, Z. Liu, R. Gallimore, S. Levis, and J. E. Kutzbach, 2008: Detecting vegetation-precipitation feedbacks in mid-Holocene North Africa from two climate models. *Climate of the Past*, **4**, 59-67.

Wang, Y., L. R. Leung, J. L. McGregor, D.-K. Lee, W.-C. Wang, Y. Ding, and F. Kimura, 2004: Regional Climate Modeling: Progress, Challenges, and Prospects. *Journal of the Meteorological Society of Japan*, **82**, 1599-1628.

Wang, Y. P., and B. Z. Houlton, 2009: Nitrogen constraints on terrestrial carbon uptake: Implications for the global carbon-climate feedback. *Geophysical Research Letters*, **36**.

Wania, R., I. Ross, and I. C. Prentice, 2009: Integrating peatlands and permafrost into a dynamic global vegetation model: 1. Evaluation and sensitivity of physical land surface processes. *Global Biogeochemical Cycles*, **23**.

Watanabe, M., S. Emori, M. Satoh, and H. Miura, 2009: A PDF-based hybrid prognostic cloud scheme for general circulation models. *Climate Dynamics*, **33**, 795-816.

Watanabe, M., M. Chikira, Y. Imada, and M. Kimoto, 2011: Convective control of ENSO simulated in MIROC. *J. Clim. submitted.*

Watanabe, M., et al., 2010: Improved Climate Simulation by MIROC5. Mean States, Variability, and Climate Sensitivity. *Journal of Climate*, **23**, 6312-6335.

Waugh, D., and V. Eyring, 2008: Quantitative performance metrics for stratospheric-resolving chemistry-climate models. *Atmospheric Chemistry and Physics*. 5699-5713.

Weaver, A. J., et al., 2001: The UVic Earth System Climate Model: Model Description, Climatology, and Applications to Past, Present, and Future Climates. *Atmos.-Ocean*, **39**, -428.

Webb, M., C. Senior, S. Bony, and J. J. Morcrette, 2001: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. *Climate Dynamics*, **17**, 905-922.

Webb, M., et al., 2006: On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Climate Dynamics*, **27**, 17-38-38.

Weber, S. L., et al., 2007: The modern and glacial overturning circulation in the Atlantic Ocean in PMIP coupled model simulations. *Climate of the Past*, **3**, 51-64.

Webster, P. J., A. M. Moore, J. P. Loschnigg, and R. R. Leben, 1999: Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98. *Nature*, **401**, 356-360.

Wehner, M. F., R. L. Smith, G. Bala, and P. Duffy, 2010: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate Dynamics*, **34**, 241-247.

Weigel, A. P., R. Knutti, M. A. Liniger, and C. Appenzeller, 2010: Risks of Model Weighting in Multimodel Climate Projections. *Journal of Climate*, **23**, 4175-4191.

Wetzel, P., E. Maier-Reimer, M. Botzet, J. H. Jungclaus, N. Keenlyside, and M. Latif, 2006: Effects of ocean Biology on the Penetrative Radiation in a Coupled Climate Model. *Journal of Climate*, **19**, 3973-3987.

Whetton, P., I. Macadam, J. Bathols, and J. O'Grady, 2007: Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophysical Research Letters*, ARTN L14701, DOI 10.1029/2007GL030025. -.

Wild, M., 2008: Short-wave and long-wave surface radiation budgets in GCMs: a review based on the IPCC-AR4/CMIP3 models. *Tellus Series a-Dynamic Meteorology and Oceanography*, **60**, 932-945.

Wild, M., C. N. Long, and A. Ohmura, 2006: Evaluation of clear-sky solar fluxes in GCMs participating in AMIP and IPCC-AR4 from a surface perspective. *Journal of Geophysical Research-Atmospheres*, **111**.

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences.* Academic Press, 467 pp.

Willett, K., P. Jones, P. Thorne, and N. Gillett, 2010: A comparison of large scale changes in surface humidity over land in observations and CMIP3 general circulation models. *Environmental Research Letters*, **5**, -.

Williams, K., and M. Webb, 2009: A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics*, DOI 10.1007/s00382-008-0443-1. 141-157.

Williams, K. D., and G. Tselioudis, 2007: GCM intercomparison of global cloud regimes: present-day evaluation and climate change response. *Climate Dynamics*, **29**, 231-250.

Williams, K. D., and M. E. Brooks, 2008: Initial tendencies of cloud regimes in the Met Office unified model. *Journal of Climate*, **21**, 833-840.

Williams, K. D., W. J. Ingram, and J. M. Gregory, 2008: Time variation of effective climate sensitivity in GCMs. *Journal of Climate*, **21**, 5076-5090.

Williamson, D. L., and J. G. Olson, 2007: A comparison of forecast errors in CAM2 and CAM3 at the ARM Southern Great Plains site. *Journal of Climate*, **20**, 4572-4585.

Williamson, D. L., et al., 2005: Moisture and temperature balances at the Atmospheric Radiation Measurement Southern Great Plains Site in forecasts with the Community Atmosphere Model (CAM2). *J. Geophys. Res.*, **110**, D15S16.
Willis, J. K., 2010: Can in situ floats and satellite altimeters detect long-term changes in Atlantic Ocean overturning? *Geophys. Res. Lett.*, **37**, L06602-.
Winterfeldt, J., and R. Weisse, 2009: Assessment of Value Added for Surface Marine Wind Speed Obtained from Two Regional Climate Models. *Monthly Weather Review*, DOI 10.1175/2009MWR2704.1. 2955-2965.
Winton, M., 2011: Do Climate Models Underestimate the Sensitivity of Northern Hemisphere Sea Ice Cover? *Journal of Climate*, **24**, 3924-3934.
Wise, M., et al., 2009: Implications of Limiting CO2 Concentrations for Land Use and Energy. *Science*, **324**, 1183-1186.
Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, **36**, -.
Wittenberg, A. T., A. Rosati, N. C. Lau, and J. J. Ploshay, 2006: GFDL's CM2 Global Coupled Climate Models. Part III: Tropical Pacific Climate and ENSO. *Jounal of Climate*, **19**, 698-722.
WMO, 2011a: (World Meteorological Organization), *Scientific Assessment of Ozone Depletion: 2010*, Global Ozone Research and Monitoring Project–Report, 516pp pp.
——, 2011b: *Scientific Assessment of Ozone Depletion: 2010*.
Wohlfahrt, J., et al., 2008: Evaluation of coupled ocean-atmosphere simulations of the mid-Holocene using palaeovegetation data from the northern hemisphere extratropics. *Climate Dynamics*, **31**, 871-890.
Woillez, M. N., M. Kageyama, G. Krinner, N. de Noblet-Ducoudré, N. Viovy, and M. Mancip, 2011: Impact of CO2 and climate on the Last Glacial Maximum vegetation. *Climate of the Past Discussions*, **7**, 1-46.
Woollings, T., 2008: Vertical structure of anthropogenic zonal-mean atmospheric circulation change. *Geophysical Research Letters*, **35**.
Woollings, t., and M. Blackburn, 2011: The North Atlantic jet stream under climate change, and its relation to the NAO and EA patterns. *Journal of Climate*.
Woollings, T., A. Hannachi, and B. Hoskins, 2010a: Variability of the North Atlantic eddy-driven jet stream. *Q.J.R. Meteorol. Soc.*, **136**, 856-868.
Woollings, T., A. Hannachi, B. Hoskins, and A. Turner, 2010b: A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing. *Journal of Climate*, **23**, 1291-1307.
Woollings, T., B. Hoskins, M. Blackburn, D. Hassell, and K. Hodges, 2010c: Storm track sensitivity to sea surface temperature resolution in a regional atmosphere model. *Climate Dynamics*, **35**, 341-353.
Woollings, T., A. Charlton-Perez, S. Ineson, A. G. Marshall, and G. Masato, 2010d: Associations between stratospheric variability and tropospheric blocking. *J. Geophys. Res.*, **115**, D06108-.
Wright, D. G., and T. F. Stocker, 1992: SENSITIVITIES OF A ZONALLY AVERAGED GLOBAL OCEAN CIRCULATION MODEL. *Journal of Geophysical Research-Oceans*, **97**, 12707-12730.
Wu, Q. G., D. J. Karoly, and G. R. North, 2008: Role of water vapor feedback on the amplitude of season cycle in the global mean surface air temperature. *Geophysical Research Letters*, **35**.
Wunsch, C., 2008: Mass and volume transport variability in an eddy-filled ocean. *Nature Geoscience*, **1**, 165-168.
Wyant, M. C., C. S. Bretherton, and P. N. Blossey, 2009: Subtropical Low Cloud Response to a Warmer Climate in a Superparameterized Climate Model. Part I: Regime Sorting and Physical Mechanisms. *Journal of Advances in Modeling Earth Systems*, **1**, doi:10.3894/JAMES.2009.3891.3897.
Xavier, P. K., J. P. Duvel, P. Braconnot, and F. J. Doblas-Reyes, 2010: An Evaluation Metric for Intraseasonal Variability and its Application to CMIP3 Twentieth-Century Simulations. *Journal of Climate*, **23**, 3497-3508.
Xiao, X., et al., 1998: Transient climate change and net ecosystem production of the terrestrial biosphere. *Global Biogeochemical Cycles*, **12**, 345-360.
Xie, L., T. Z. Yan, L. J. Pietrafesa, J. M. Morrison, and T. Karl, 2005: Climatology and interannual variability of North Atlantic hurricane tracks. *Journal of Climate*, **18**, 5370-5381.
Xie, S., J. Boyle, S. A. Klein, X. Liu, and S. Ghan, 2008: Simulations of Arctic mixed-phase clouds in forecasts with CAM3 and AM2 for M-PACE. *J. Geophys. Res.*, **113**, D04211.
Xie, S. P., H. Annamalai, F. A. Schott, and J. P. McCreary, 2002: Structure and mechanisms of South Indian Ocean climate variability. *Journal of Climate*, **15**, 864-878.

Xin, X.-G., T.-J. Zhou, and R.-C. Yu, 2008: The Arctic Oscillation in coupled climate models. *Chinese Journal of Geophysics-Chinese Edition*, **51**, 337-351.
Xu, Y., X. Gao, and F. Giorgi, 2010: Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Climate Research*, **41**, 61-81.
Xue, Y. K., R. Vasic, Z. Janjic, F. Mesinger, and K. E. Mitchell, 2007: Assessment of dynamic downscaling of the continental US regional climate using the Eta/SSiB regional climate model. *Journal of Climate*, **20**, 4172-4193.
Yeh, S.-W., J.-S. Kug, B. Dewitte, M.-H. Kwon, B. P. Kirtman, and F.-F. Jin, 2009: El Nino in a changing climate. *Nature*, **461**, 511-U570.
Yhang, Y. B., and S. Y. Hong, 2008: Improved physical processes in a regional climate model and their impact on the simulated summer monsoon circulations over East Asia. *Journal of Climate*, **21**, 963-979.
Yin, J., S. Griffies, and R. Stouffer, 2010: Spatial Variability of Sea Level Rise in Twenty-First Century Projections. *Journal of Climate*, DOI 10.1175/2010JCLI3533.1. 4585-4607.
Yokohata, T., M. J. Webb, M. Collins, K. D. Williams, M. Yoshimori, J. C. Hargreaves, and J. D. Annan, 2010: Structural Similarities and Differences in Climate Responses to CO2 Increase between Two Perturbed Physics Ensembles. *Journal of Climate*, **23**, 1392-1410.
Yokohata, T., et al., 2008: Comparison of equilibrium and transient responses to CO2 increase in eight state-of-the-art climate models. *Tellus Series a-Dynamic Meteorology and Oceanography*, **60**, 946-961.
Yokoi, S., Y. N. Takayabu, and J. C. L. Chan, 2009a: Tropical cyclone genesis frequency over the western North Pacific simulated in medium-resolution coupled general circulation models. *Climate Dynamics*, **33**, 665-683.
Yokoi, S., et al., 2011: Application of Cluster Analysis to Climate Model Performance Metrics. *Journal of Applied Meteorology and Climatology*, **50**, 1666-1675.
Yokoi, T., T. Tozuka, and T. Yamagata, 2009b: Seasonal Variations of the Seychelles Dome Simulated in the CMIP3 Models. *Journal of Physical Oceanography*, **39**, 449-457.
Yoshimori, M., T. Yokohata, and A. Abe-Ouchi, 2009: A Comparison of Climate Feedback Strength between CO2 Doubling and LGM Experiments. *Journal of Climate*, **22**, 3374-3395.
Yu, B., G. J. Boer, F. W. Zwiers, and W. J. Merryfield, 2011: Covariability of SST and surface heat fluxes in reanalyses and CMIP3 climate models. *Climate dynamics*, **36**, 589-605.
Yu, W., M. Doutriaux, G. Seze, H. LeTreut, and M. Desbois, 1996: A methodology study of the validation of clouds in GCMs using ISCCP satellite observations. *Climate Dynamics*, **12**, 389-401.
Zaehle, S., P. Friedlingstein, and A. D. Friend, 2010a: Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophysical Research Letters*, **37**.
Zaehle, S., A. D. Friend, P. Friedlingstein, F. Dentener, P. Peylin, and M. Schulz, 2010b: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance. *Global Biogeochemical Cycles*, **24**.
Zaliapin, I., and M. Ghil, 2010: Another look at climate sensitivity. *Nonlinear Processes in Geophysics*, **17**, 113-122.
Zhang, D., and M. J. McPhaden, 2006: Decadal variability of the shallow Pacific meridional overturning circulation: Relation to tropical sea surface temperatures in observations and climate change models. *Ocean Modelling*. 250-273.
Zhang, D., R. Blender, X. H. Zhu, and K. Fraedrich, 2011: Temperature variability in China in an ensemble simulation for the last 1,200 years. *Theoretical and Applied Climatology*, **103**, 387-399.
Zhang, F., and A. Georgakakos, 2011: Joint variable spatial downscaling. *Clim. Change*, 10.1007/s10584-011-0167-9. 1-28.
Zhang, Q., H. S. Sundqvist, A. Moberg, H. Kornich, J. Nilsson, and K. Holmgren, 2010a: Climate change between the mid and late Holocene in northern high latitudes - Part 2: Model-data comparisons. *Climate of the Past*, **6**, 609-626.
Zhang, X., 2010: Sensitivity of arctic summer sea ice coverage to global warming forcing: towards reducing uncertainty in arctic climate change projections. *Tellus*, **62A**, *220–227*.
Zhang, Y., and S. A. Klein, 2010: Mechanisms Affecting the Transition from Shallow to Deep Convection over Land: Inferences from Observations of the Diurnal Cycle Collected at the ARM Southern Great Plains Site. *Journal of the Atmospheric Sciences*, **67**, 2943-2959.
Zhang, Y., S. A. Klein, J. Boyle, and G. G. Mace, 2010b: Evaluation of tropical cloud and precipitation statistics of Community Atmosphere Model version 3 using CloudSat and CALIPSO data. *J. Geophys. Res.*, **115**, D12205.

Zhang, Y., et al., 2008: On the diurnal cycle of deep convection, high-level cloud, and upper troposphere water vapor in the Multiscale Modeling Framework. *J. Geophys. Res.*, **113**, D16105.

Zhao, M., I. M. Held, S. J. Lin, and G. A. Vecchi, 2009: Simulations of Global Hurricane Climatology, Interannual Variability, and Response to Global Warming Using a 50-km Resolution GCM. *Journal of Climate*, **22**, 6653-6678.

Zhao, T. L., et al., 2008: A three-dimensional model study on the production of BrO and Arctic boundary layer ozone depletion. *J. Geophys. Res.*, **113**, D24304.

Zhao, Y., and S. P. Harrison, 2011: Mid-Holocene monsoons: a multi-model analysis of the inter-hemispheric differences in the responses to orbital forcing and ocean feedbacks.

Zheng, X.-T., S.-P. Xie, and Q. Liu, 2011: Response of the Indian Ocean basin mode and its capacitor effect to global warming. *Jounal of Climate*, **in press**.

Zhu, H., H. Hendon, and C. Jakob, 2009: Convection in a Parameterized and Superparameterized Model and Its Role in the Representation of the MJO. *Journal of the Atmospheric Sciences*, **66**, 2796-2811.

Zhu, Y., and H. Wang, 2010: The Arctic and Antarctic Oscillations in the IPCC AR4 Coupled Models. *Acta Meteorologica Sinica*, **24**, 176-188.

Zwiers, F. W., X. Zhang, and Y. Feng, 2011: Anthropogenic Influence on Long Return Period Daily Temperature Extremes at Regional Scales. *Journal of Climate*, **24**, 881-892.

Tables

Table 9.1: Salient features of the AOGCMs and ESMs participating in CMIP5. Column 1: identification (Model ID) along with the calendar year ('vintage') of the first publication for each model; Column 2: sponsoring institution(s), main reference(s) and flux correction implementation (*not yet described*); Subsequent Columns for each of the 8 CMIP5 realms: component name, code independence and main component reference(s). Additionally, there are standard entries for the Atmosphere realm: horizontal grid resolution, number of vertical levels, grid top (low or high top); and for the Ocean realm: horizontal grid resolution, number of vertical levels, top level, vertical coordinate type, ocean free surface type ("Top BC"). This table information was automatically extracted from the CMIP5 online questionnaire (http://q.cmip5.ceda.ac.uk/) as of 12 November 2011.

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
ACCESS1.0 2011	Centre for Australian Weather and Climate Research	Aerosols XX%	Atmosphere XX%	Atmospheric Chemistry XX%	Land Ice XX%	Land Surface XX%	Ocean Biogeochemistry XX%	Ocean XX%	Sea Ice XX%
CanESM2 2010	Canadian Centre for Climate Modelling and Analysis	Aerosols XX%	Atmosphere T63L35 35 0.5 hPa XX%	Atmospheric Chemistry XX%	Land Ice XX%	Land Surface XX%	Ocean Biogeochemistry XX%	Ocean 256 X 192 40 5 m depth other XX%	Sea Ice XX%
BCC_AGCM2.1 2010	Beijing Climate Center, China Meteorological Administration	Aerosols XX%	Atmosphere XX%	Atmospheric Chemistry XX%	Land Ice XX%	Land Surface XX%	Ocean Biogeochemistry XX%	Ocean XX%	Sea Ice XX%

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
BCC_CSM1.1 2011	Beijing Climate Center, China Meteorological Administration	Not implemented XX% None	BCC_AGCM2.1 T42 T42L26 26 2.917hPa XX%	Not implemented	Not implemented	BCC-AVIM1.0 XX%	Ocean Biogeochemistry XX%	MOM4-L40 1° with enhanced resolution in the meridional direction in the tropics (1/3° meridional resolution at the equator) tripolar 40 20 Z-coordinate linear split-explicit XX%	SIS XX%
CMCC-CESM 2009	Centro Euro-Mediterraneo per I Cambiamenti Climatici Fogli et al., 2009; Vichi, et al., 2011	Aerosols XX%	Atmosphere XX%	Atmospheric Chemistry XX%	Land Ice XX%	Land Surface XX%	Ocean Biogeochemistry XX%	OPA8.2 2° zonal resolution, meridional resolution varying from 0.5° at the equator to 2° cos / south of 20°S Orca2_T 31 4,9999938 depth linear implicit XX% Madec et al., 1998	LIM2 XX% Fichefet and Morales-Maqueda, 1997; Fichefet and Morales Maqueda, 1999; Timmermann et al., 2005

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
CNRM-CM5 2010	Centre National de Recherches Meteorologiques - Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique Emile-Geay, and Madec, 2009; Voldoire, 2011	Not implemented	ARPEGE (Atmosphere) none tl127r 31 10 hPa XX% ARPEGE-Climat_V5 Version 5.2, 2011	Not implemented	Not implemented	SURFEX (Land and Ocean Surface) XX% Masson et al., 2003; Le Moigne et al., 2009	Not implemented	NEMO 0.7 degree on average ORCA1 42 5 m linear filtered XX%	Gelato5 (Sea Ice) XX% Salas-Mélia, 2002
EC-EARTH 2010	Europe	Not implemented	IFS c31r1 1.125 longitudinal spacing, Gaussian grid T159L62 62 1 XX%	Not implemented	Not implemented	HTESSEL XX%	Not implemented	NEMO_ecmwf The grid is a tripolar curvilinear grid with a 1 degree resolution. ORCA1 31 1 linear filtered XX%	LIM2 XX%
gfdl-esm2g 2011	Geophysical Fluid Dynamics Laboratory Delworth et al., 2006	Not implemented	Atmosphere XX% GFDL Global Atmospheric Model Development Team, 2004	Not implemented	Not implemented	Land Surface XX%	Ocean Biogeochemistry XX%	Ocean 1 degree tripolar360x210L50 63 0 other non-linear split-explicit XX%	SIS XX% Delworth et al., 2006; Winton, 2000

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
gfdl-cm2p1 2011	Geophysical Fluid Dynamics Laboratory Delworth et al., 2006	Aerosols XX%	Atmosphere 2.5 degree longitude, 2 degree latitude M45L24 24 midpoint of top box is 3.65 hPa XX% GFDL Global Atmospheric Model Development Team, 2004	Not implemented	Land Ice XX%	Land Surface XX%	Not implemented	Ocean 1 degree tripolar360x200L50 50 0 depth non-linear split-explicit XX%	SIS XX% Delworth et al., 2006; Winton, 2000
gfdl-hiram-c180 2011	Geophysical Fluid Dynamics Laboratory Delworth et al., 2006; Donner et al., 2011	Not implemented	Atmosphere Averaged cell size: approx. 50x50 km. C180L32 32 2,164 XX% Donner et al., 2011	Not implemented	Not implemented	Land Surface XX% Milly and Shmakin, 2002; Shevliakova et al., 2009	Not implemented	Not implemented Not implemented Not implemented Not implemented Not implemented Not implemented XX% None	Not implemented XX% None
gfdl-esm2m	Geophysical Fluid Dynamics Laboratory Delworth et al., 2006	Not implemented	Atmosphere XX% GFDL Global Atmospheric Model Development Team, 2004	Not implemented	Not implemented	Land Surface XX%	Ocean Biogeochemistry XX%	Ocean non-linear split-explicit XX%	SIS XX% Delworth et al., 2006; Winton, 2000

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
gfdl-cm3 2011	Geophysical Fluid Dynamics Laboratory Delworth et al., 2006; Donner et al., 2011	Aerosols XX%	Atmosphere ~200km C48L48 48 XX% Donner et al., 2011	Atmospheric Chemistry XX% Austin and Wilson, 2006 Horowitz et al., 2003 Sander et al., 2000	Not implemented	Land Surface XX% Milly and Shmakin, 2002; Shevliakova et al., 2009	Not implemented	Ocean 1 degree tripolar360x200L50 50 0 depth non-linear split-explicit XX%	SIS XX% Delworth et al., 2006; Winton, 2000
inmcm4 2009	Russian Institute for Numerical Mathematics Volodin et al., 2010	Not implemented	Atmosphere 2x1.5 degrees in longitude and latitude latitude-longitude 21 sigma=0.01 XX%	Not implemented	Land Ice XX%	Land Surface XX% Alekseev et al., 1998; Volodin and Lykossov, 1998	Ocean Biogeo Chemistry XX% Volodin, 2007	Ocean 1x0.5 degrees in longitude and latitude generalized spherical coordinates with poles displaced outside ocean 40 sigma=0.0010426 sigma linear implicit XX% Marchuk et al., 2010; Volodin et al., 2010.	Sea Ice XX% Iakovlev, 2008

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
IPSL-CM5A-LR 2010	Institut Pierre Simon Laplace	Not implemented	Atmosphere 96x95 equivalent to 1,9° x 3,75° LMDZ96x95 39 4Pa XX% Hourdin et al., 2006	Not implemented	Not implemented	Land Surface XX%	Ocean Biogeo Chemistry XX% Aumont et al., 2003; Aumont and Bopp, 2006	Ocean 2° ORCA2 31 0m depth linear filtered XX% Madec, 2008	Sea ice XX% Fichefet and Maqueda, 1997;Goosse and Fichefet, 1999. Timmermann et al., 2005
MIROC-ESM 2010	University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Aerosols XX%	Atmosphere 2.815x2.815degree T42 80 0.0036hPa or 85km XX%	Not implemented	Not implemented	Land Surface XX%	Ocean Biogeo Chemistry XX%	Ocean non-linear split-explicit XX% Hasumi and Emori, eds., 2004	Sea Ice XX%
MIROC4h 2009	University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Aerosols XX%	Atmosphere 0.5625x0.5625degree T213 56 about 0.9hPa XX%	Not implemented	Not implemented	Land Surface XX%	Not implemented	Ocean non-linear split-explicit XX% Hasumi and Emori, eds., 2004	Sea Ice XX%

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
MIROC5 2010	University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Aerosols XX%	Atmosphere XX%	Atmospheric Chemistry XX%	Land Ice XX%	Land Surface XX%	Ocean Biogeo Chemistry XX%	Ocean XX%	Sea Ice XX%
HadCM3 1998	UK Met Office Hadley Centre Collins et al., 2001; Gordon et al., 2000; Johns et al., 2003; Pope et al., 2000	Aerosols XX% Jones et al., 2001	Atmosphere HadAM3 (N48L19) 3.75 degrees in longitude by 2.5 degrees in latitude. N48 19 0,005 XX% Pope et al., 2000	Not implemented	Not implemented	Land Surface XX% Collatz et al., 1991; Collatz et al., 1992; Cox et al., 1999; Cox, 2001; Mercado et al., 2007	Not implemented	Ocean HadOM (lat: 1.25 lon: 1.25 L20) 1.25 deg in longitude by 1.25 deg in latitude N144 20 5 depth linear implicit XX% Cox, 1984; UNESCO, 1981	Sea Ice XX%

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
HadGEM2-CC 2010	UK Met Office Hadley Centre Bellouin et al., 2007; Collins et al., 2008	Aerosols XX% Bellouin et al., 2007	Atmosphere 1.875 deg in longitude by 1.25 deg in latitude N96 60 84132,439 XX% Davies et al., 2005	Atmospheric Chemistry XX% Bellouin et al., 2011; Jones et al., 2001	Land Ice XX% Johns et al., 2006	Land Surface XX% Cox et al., 1999; Essery et al., 2003	Ocean Biogeo Chemistry XX%	Ocean 1.875 deg in longitude by 1.25 deg in latitude N96 hybrid height linear implicit XX% Bryan, 1969; Cox, 1984; Johns et al., 2006	Sea Ice XX% McLaren et al., 2006; Thorndike et al., 1975
HadGEM2-ES 2009	UK Met Office Hadley Centre Bellouin et al., 2007; Collins et al. 2008	Not implemented	Atmosphere 1.875 degrees in longitude by 1.25 degrees in latitude N96 38 39254,8 XX% Davies et al. 2005	Atmospheric Chemistry XX% O'Connor et al. 2009 O'Connor et al., 2010	Land Ice XX% Johns et al., 2006	Land Surface XX% Cox et al., 1999; Essery et al., 2003	Ocean Biogeo Chemistry XX%	Ocean 1 deg by 1 deg between 30 N/S and the poles; meridional resolution increases to 1/3 deg at the equator N180 40 5 depth linear implicit XX% Bryan 1969; Cox 1984; Johns et al. 2006	Sea Ice XX% McLaren et al., 2006; Thorndike et al., 1975

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
MPI-ESM-MR 2009	Max Planck Institute for Meteorology Marsland et al., 2003.; Raddatz et al., 2007	Not implemented	ECHAM6 approx 1.8 deg T63 95 0.1 hPa XX%	Not implemented	Not implemented	JSBACH XX% Raddatz et al., 2007	HAMOCC XX% HAMOCC: Technical Documentation	MPIOM approx. 0.4 deg TP04 40 Z-coordinate linear implicit XX% Marsland et al., 2003	Sea Ice XX% Marsland et al., 2003
MPI-ESM-P 2009	Max Planck Institute for Meteorology Marsland et al., 2007; Raddatz et al., 2007	Not implemented	ECHAM6 approx 1.8 deg T63 47 0.1 hPa XX%	Not implemented	Not implemented	JSBACH XX% Raddatz et al., 2007	HAMOCC XX% HAMOCC: Technical Documentation	MPIOM linear implicit XX% Marsland et al., 2003	Sea Ice XX% Marsland et al., 2003
MPI-ESM-LR 2009	Max Planck Institute for Meteorology Marsland et al., 2003; Raddatz et al., 2007	Not implemented	ECHAM6 approx 1.8 deg T63 47 0.1 hPa XX%	Not implemented	Not implemented	JSBACH XX% Raddatz et al., 2007	HAMOCC XX% HAMOCC: Technical Documentation	MPIOM average 1.5 deg GR15 40 6 depth linear implicit XX% Marsland et al., 2003	Sea Ice XX% Marsland et al., 2003

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
GISS-E2-R 2011	NASA Goddard Institute for Space Studies USA Schmidt et al., 2006	Aerosols XX% Bauer et al., 2007; Koch et al., 2006; Menon et al. 2010; Tsigaridis and Kanakidou, 2007	Atmosphere 40 0.1 hPa XX%	G-PUCCINI XX% Shindell et al., 2006	Land Ice XX%	Land Surface XX%	Not implemented	Russell Ocean 32 0 m other XX%	Sea Ice XX%
GISS-E2-H	NASA Goddard Institute for Space Studies USA Schmidt et al., 2006.	Aerosols XX% Bauer et al., 2007; Koch et al., 2006; Menon et al., 2010; Tsigaridis and Kanakidou, 2007.	Atmosphere XX%	G-PUCCINI XX% Shindell et al., 2006	Land Ice XX%	Land Surface XX%	Not implemented	HYCOM Ocean hybrid Z-isopycnic non-linear split-explicit XX%	Sea Ice XX%
GISS-E2CS-H 2011	NASA Goddard Institute for Space Studies USA Schmidt et al., 2006	Aerosols XX% Bauer, et al., 2007; Koch et al., 2006; Menon et al., 2010; Tsigaridis and Kanakidou, 2007	Atmosphere 40 0.1 hPa XX%	G-PUCCINI XX% Shindell et al., 2006	Land Ice XX%	Land Surface XX%	Not implemented	HYCOM Ocean hybrid Z-isopycnic non-linear split-explicit XX%	Sea Ice XX%

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
GISS-E2CS-R 2011	NASA Goddard Institute for Space Studies USA Schmidt et al., 2006	Aerosols XX% Bauer et al., 2007; Koch, et al., 2006; Menon, et al., 2010; Tsigaridis and Kanakidou, 2007	Atmosphere 40 0.1 hPa XX%	G-PUCCINI XX% Shindell, et al., 2006	Land Ice XX%	Land Surface XX%	Not implemented	Russell Ocean 32 0 m other XX%	Sea Ice XX%
CCSM4 1° 2010	US National Centre for Atmospheric Research	Aerosols XX% Neale, et al., 2010; Oleson et al., 2010.	Atmosphere 0.9x1.25 f09 27 2.194067 hPa XX% Neale, et al., 2010; Lean et al., 1995	Not implemented	Not implemented	Land Surface XX%	Not implemented	Ocean 320x384 60 5 depth linear implicit XX%	Sea Ice XX%
CCSM4 2° 2010	US National Centre for Atmospheric Research	Aerosols XX% Neale et al., 2010; Oleson et al., 2010.	Atmosphere 1.9x2.5 f19 27 2.194067 hPa XX% Neale et al., 2010; Lean et al., 1995	Not implemented	Not implemented	Land Surface XX%	Not implemented	Ocean 320x384 60 5 depth linear implicit XX%	Sea Ice XX%

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
HadGEM2-ES 2009	UK National Centre for Atmospheric Science Bellouin et al., 2007; Collins et al., 2008	Aerosols XX% Bellouin et al., 2007	Atmosphere XX% Davies et al., 2005	Atmospheric Chemistry XX% O'Connor, et al., 2009; O'Connor et al., 2010	Land Ice XX% Johns et al., 2006	Land Surface XX% Cox et al., 1999; Essery et al., 2003	Ocean Biogeo Chemistry XX%	Ocean linear implicit XX% Bryan, 1969; Cox, 1984; Johns et al., 2006	Sea Ice XX% McLaren et al., 2006; Thorndike et al., 1975
HiGEM1-2 2009	UK National Centre for Atmospheric Science Bellouin et al., 2007; Collins et al., 2008	Aerosols XX% Bellouin et al., 2007.	Atmosphere XX% Davies et al., 2005	Atmospheric Chemistry XX% O'Connor, et al., 2009 O'Connor et al., 2010	Land Ice XX% Johns et al., 2006	Land Surface XX% Cox et al., 1999; Essery et al., 2003	Ocean Biogeo Chemistry XX%	Ocean linear implicit XX% Bryan, 1969; Cox, 1984; Johns et al., 2006	Sea Ice XX% McLaren et al., 2006; Thorndike et al., 1975
NorESM1-M 2011	Norwegian Climate Centre	Aerosols XX%	CAM4 1.9 degrees meridionally, 2.5 degrees zonally 26 3.54 hPa XX%	Atmospheric Chemistry XX%	Not implemented	CLM4 XX%	Ocean Biogeo Chemistry XX%	MICOM 1.125 degrees along the equator 53 1 hybrid Z-isopycnic XX%	CICE4 XX%
HadGEM2-AO 2008	Korean National Institute for Meteorological Research	Aerosols XX%	Atmosphere XX%	Atmospheric Chemistry XX%	Land Ice XX%	Land Surface XX%	Ocean Biogeo Chemistry XX%	Ocean XX%	Sea Ice XX%

Model ID Vintage	Institution Main Reference(s) Flux correction information	Aerosols Component Name Code Independence Reference	Atmosphere Name Horizontal Grid Number of Levels Grid Top Code Independence References	Atmospheric Chemistry Name Code Independence References	Land Ice Name Code Independence References	Land Surface Name Code Independence References	Ocean Biogeochemistry Name Code Independence References	Ocean Name Horizontal Grid Number of Levels Top Level Z Coordinate Top BC Code Independence References	Sea ice Name Code Independence References
CSIRO-Mk3-6-0 2007	Queensland Climate Change Centre of Excellence and Commonwealth Scientific and Industrial Research Organisation Chou and Lee, 2005; Gordon et al., 2010;Grant et al., 1999; Rotstayn et al., 2007; Rotstayn et al., 2010; Rotstayn and Lohmann, 2002; Sato et al., 1993	Aerosols XX% Rotstayn et al., 2007; Rotstayn et al., 2010; Rotstayn and Liu, 2009; Rotstayn and Lohmann, 2002.	Atmosphere ~1.875x1.875 mk36 18 18 XX% Gordon et al., 2002; Gordon et al., 2010; Rotstayn et al., 2010	Not implemented	Land Ice XX%	Land Surface XX% Gordon et al., 2002; Gordon et al., 2010	Not implemented	Ocean ~0.9x1.875 ocean horizontal 31 5 depth N/A XX% Gordon et al., 2002; Gordon et al., 2010; Rotstayn et al., 2010	Sea Ice XX%

Table 9.3: Features of Earth System Models of Intermediate Complexity (EMICs)

Model	Atmosphere	Ocean	Sea Ice	Coupling	Land Surface	Biosphere	Ice Sheets	Sediment and Weathering
Bern3D-LPJ (Ritz et al., 2011)	EMBM,2-D(φ, λ), NCL, 10° × (3 − 19)° (Ritz et al., 2011)	FG, 3-D, RL, ISO, MESO, 10° × (3 − 19)°, L32 (Müller and Roeckner, 2006)	0-LT, DOC, 2-LIT (Ritz et al., 2011)	PM, NH, RW (Ritz et al., 2011)	Bern3D component: 1-LST, NSM, RIV (Ritz et al., 2011); LPJ component: 8-LST, CSM (uncoupled hydrology) (Wania et al., 2009)	BO (Gangsto et al., 2011; Parekh et al., 2008; Tschumi et al., 2008), BT (Sitch et al., 2003; Stocker et al., 2011; Strassmann et al., 2008), BV (Sitch et al., 2003)	N/A	CS, SW, (Tschumi et al., 2011)

CLIMBER-2.4 (Petoukhov et al., 2000)	SD, 3-D, CRAD, ICL, 10° x 51°, L10 (Petoukhov et al., 2000)	FG with parameterized zonal pressure gradient, 2-D (φ, z), 3 basins, RL, 2.5°, L21 (Wright and Stocker, 1992)	1-LT, PD, 2-LIT (Petoukhov et al., 2000)	NM, NH, NW (Petoukhov et al., 2000)	1-LST, CSM, RIV (Petoukhov et al., 2000)	BO, BT and BV (Brovkin et al., 2002)	TM, 3-D, 0.75° x 1.5°, L20* (Calov et al., 2002)	N/A
CLIMBER-3a (Montoya et al., 2005)	SD, 3-D, CRAD, ICL, 7.5° x 22.5°, L10 (Petoukhov et al., 2000)	PE, 3-D, FS, ISO, MESO, TCS, DC*, 3.75° x 3.75°, L24 (Montoya et al., 2005)	2-LT, R, 2-LIT (Fichefet and Morales Maqueda, 1997)	AM, NH, RW (Montoya et al., 2005)	1-LST, CSM, RIV (Petoukhov et al., 2000)	BO(Six and MaierReimer, 1996), BT and BV (Brovkin et al., 2002)	N/A	N/A
DCESS (Shaffer et al., 2008)	EMBM, 2-box in φ with fit of mean surface temperature to a continuous Lagrange polynomial in φ, LRAD, CHEM* (Shaffer et al., 2008)	PCE (Parameterized circulation and exchange), 2-box in φ, MESO, L55 (Shaffer and Sarmiento, 1995; Shaffer et al., 2008)	PT (parameterized from mean surface temperature distribution in φ) (Shaffer et al., 2008)	NH, NW (Shaffer et al., 2008)	NST, NSM (Shaffer et al., 2008)	BO, BT (Shaffer et al., 2008)	N/A	CS, SW Additional Ocean sediment: VP (variable porosity), VB (variable bioturbation), OAR (oxic-anoxic remineralization), CC (carbonate carbon), OC (organic carbon) (Shaffer et al., 2008)
MESMO 1.0 (Matsumoto et al., 2008)	EMBM, 2-D (φ, λ), NCL, 10◦ × (3 − 19)◦ (Fanning and Weaver, 1996)	FG, 3-D, RL, ISO, MESO, 10◦ × (3 − 19)◦, L16 (Edwards and Marsh, 2005)	0-LT, DOC, 2-LIT (Edwards and Marsh, 2005)	PM, NH, RW (Edwards and Marsh, 2005)	NST, NSM, RIV (Edwards and Marsh, 2005)	BO (Matsumoto et al., 2008)	N/A	N/A
UVic 2.9 (Weaver et al., 2001)	DEMBM,2-D(φ,λ□□ NCL,1.8°x3.6° (Weaver et al., 2001)	PE,3-D,RG,ISO,MESO,1.8°x3.6° (Weaver et al., 2001)	0-LT,R,2-LIT(Weaver et al., 2001)	AM,NH,NW (Weaver et al., 2001)	1-LST,CSM,RIV (Meissner et al., 2003)	Biosphere: BO (Schmittner et al., 2005b), BT and BV (Cox, 2001)	TM, 3-D, 20 km x 20 km, L10 (Fyke et al., 2011)	CS, SW (Eby et al., 2009)

Notes:
(a) EMBM = energy-moisture balance model; DEMBM = energy-moisture balance model including some dynamics; SD = statistical-dynamical model; QG = quasi-geostrophic model; 1-D (φ) = zonally and vertically averaged; 2-D(φ, λ □□ =vertically averaged; 2-D(φ, z) = zonally averaged; 3-D = three-dimensional; LRAD = linearized radiation scheme;

CRAD = comprehensive radiation scheme; NCL = non-interactive cloudiness; ICL = interactive cloudiness; CHEM = chemistry module; horizontal and vertical resolutions: the horizontal resolution is expressed either as degrees latitude x longitude or as spectral truncation with a rough translation to degrees latitude x longitude; the vertical resolution is expressed as 'Lm', where m is the number of vertical levels.
(b) FG = frictional geostrophic model; PE = primitive equation model; 2-D (φ, z) = zonally averaged; 3-D = three-dimensional; RL = rigid lid; FS = free surface; ISO = isopycnal diffusion; MESO = parametrization of the effect of mesoscale eddies on tracer distribution; TCS = complex turbulence closure scheme; DC = parametrization of density-driven down-sloping currents; horizontal and vertical resolutions: the horizontal resolution is expressed as degrees latitude x longitude; the vertical resolution is expressed as 'Lm', where m is the number of vertical levels.
(c) n-LT = n-layer thermodynamic scheme; PD = prescribed drift; DOC = drift with oceanic currents; R = viscous-plastic or elastic-viscous-plastic rheology; 2-LIT = two-level ice thickness distribution (level ice and leads).
(d) PM = prescribed momentum flux; GM = global momentum flux adjustment; AM = momentum flux anomalies relative to the control run are computed and added to climatological data; NM = no momentum flux adjustment; GH = global heat flux adjustment; NH = no heat flux adjustment; GW = global freshwater flux adjustment; RW = regional freshwater flux adjustment; NW = no freshwater flux adjustment.
(e) NST = no explicit computation of soil temperature; n-LST = n-layer soil temperature scheme; NSM = no moisture storage in soil; BSM = bucket model for soil moisture; CSM = complex model for soil moisture; RIV = river routing scheme.
(f) BO = model of oceanic carbon dynamics; BT = model of terrestrial carbon dynamics; BV = dynamical vegetation model.
(g) TM = thermomechanical model; M = mechanical model (isothermal); 1-D (φ) = vertically averaged with east-west parabolic profile; 2-D (φ, λ □□ =vertically averaged; 3-D = three-dimensional; horizontal and vertical resolutions: the horizontal resolution is expressed either as degrees latitude x longitude or kilometres x kilometres; the vertical resolution is expressed as 'Lm', where m is the number of vertical levels.
(h) CS = complex ocean sediment model; CW = complex weathering model; SW = simple, specified or diagnostic weathering model.

Chapter 9: Evaluation of Climate Models

Coordinating Lead Authors: Gregory Flato (Canada), Jochem Marotzke (Germany)

Lead Authors: Babatunde Abiodun (South Africa), Pascale Braconnot (France), Sin Chan Chou (Brazil), William Collins (USA), Peter Cox (UK), Fatima Driouech (Morocco), Seita Emori (Japan), Veronika Eyring (Germany), Chris Forest (USA), Peter Gleckler (USA), Eric Guilyardi (France), Christian Jakob (Australia), Vladimir Kattsov (Russia), Chris Reason (South Africa), Markku Rummukainen (Sweden)

Contributing Authors: Alessandro Anav, Timothy Andrews, Johanna Baehr, Alejandro Bodas-Salcedo, Ping Chang, Aiguo Dai, Clara Deser, Bode Gbobaniyi, Fidel Gonzales-Rouco, Stephen Griffies, Alex Hall, Elizabeth Hunke, Tatiana Ilyina, Viatcheslav Kharin, Stephen A. Klein, Jeff Knight, Reto Knutti, Felix Landerer, Tatiana Pavlova, Florian Rauser, Mark Rodwell, Adam A. Scaife, John Scinocca, Hideo Shiogama, Jana Sillmann, Ken Sperber, David Stephenson, Bjorn Stevens, Mark Webb, Keith Williams, Tim Woollings, Shang-Ping Xie

Review Editors: Isaac Held (USA), Andy Pitman (Australia), Serge Planton (France), Zong-Ci Zhao (China)

Date of Draft: 16 December 2011

Notes: TSU Compiled Version

Figures

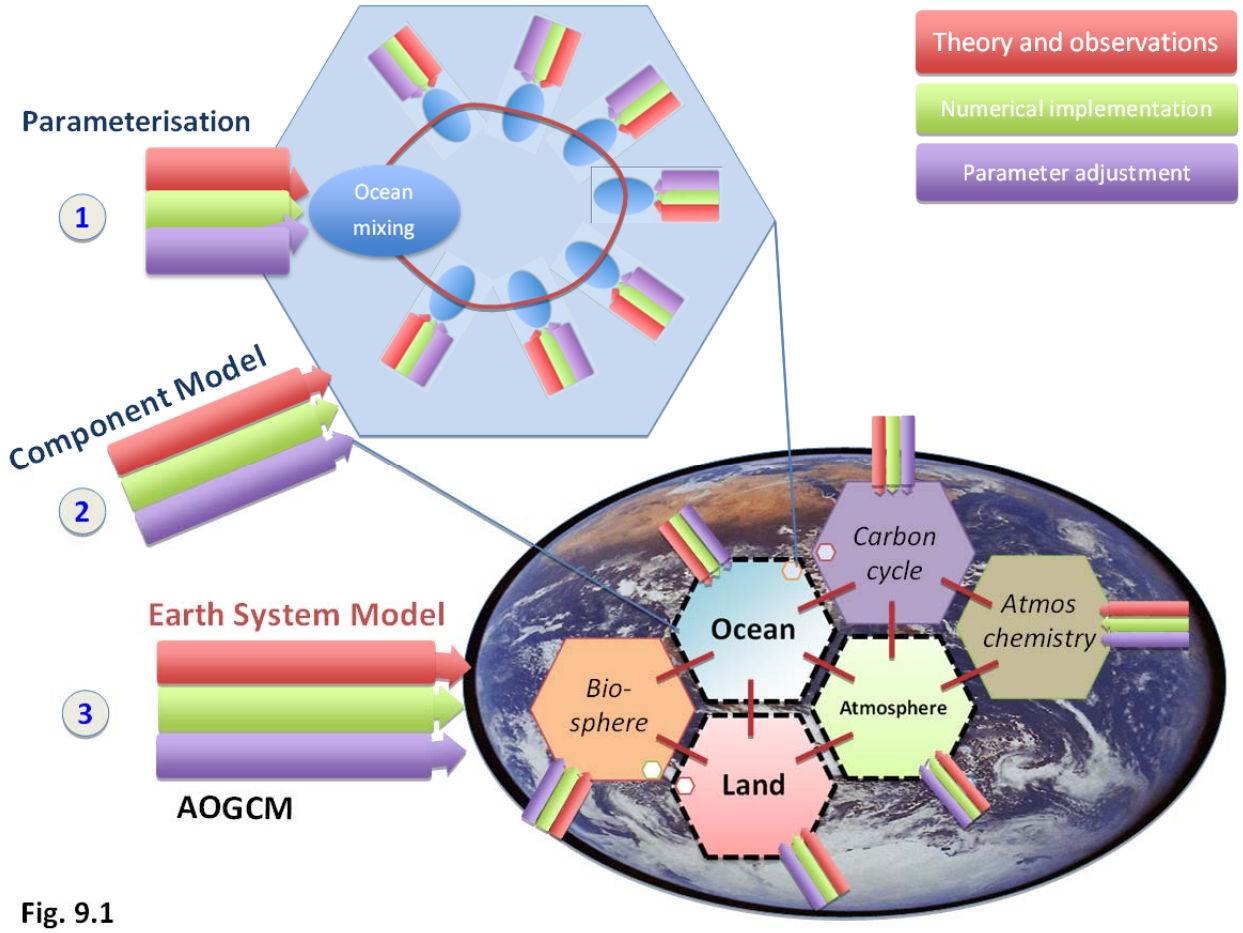


Figure 9.1: A sketch illustrating the interconnection of model components in Atmosphere-Ocean General Circulation Models (AOGCMs) and Earth system models (ESMs), and the way in which parameterisations are embedded in each component.

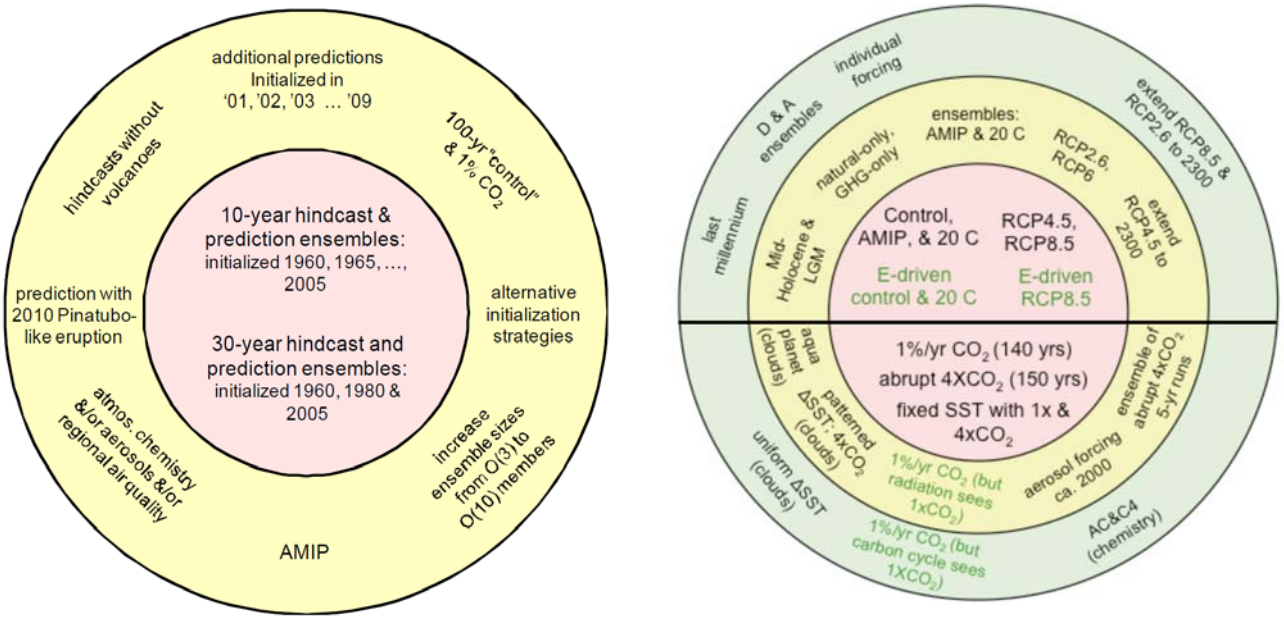


Figure 9.2: Left: Schematic summary of CMIP5 short-term experiments with tier 1 experiments (yellow background) organized around a central core (pink background). From (Taylor, Stouffer, & Meehl, 2011), their Figure 2. Right: Schematic summary of CMIP5 long-term experiments with tier 1 and tier 2 experiments organized around a central core. Green font indicates simulations to be performed only by models with carbon cycle representations, and "E-driven" means "emission-driven". Experiments in the upper hemisphere either are suitable for comparison with observations or provide projections, whereas those in the lower hemisphere are either idealized or diagnostic in nature, and aim to provide better understanding of the climate system and model behaviour. From (Taylor, et al., 2011), their Figure 3.

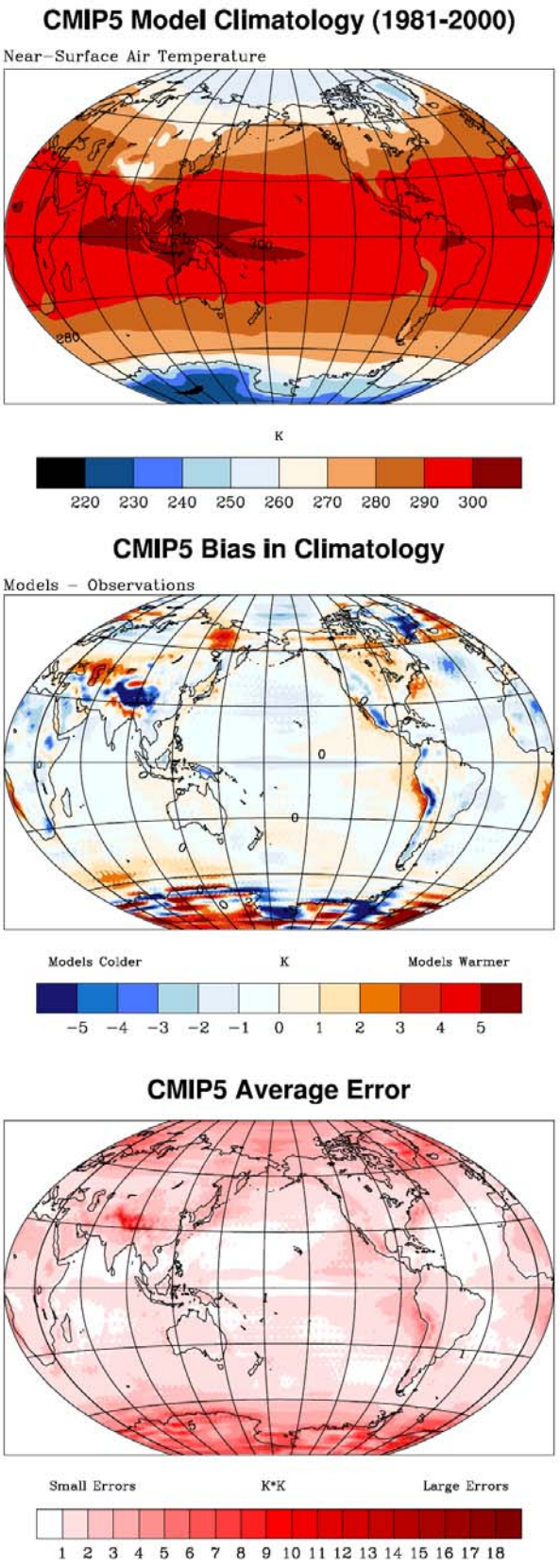


Figure 9.3: Annual mean surface (2 meter) air temperature (K) for the period (1985–2005). Top panel: Multi-model ensemble (MME) constructed with 11 available AOGCMs used in the CMIP5 historical experiment. Middle panel shows the MME bias compared to observations (Jones, 1999). Bottom panel shows the average of the individual model absolute biases.

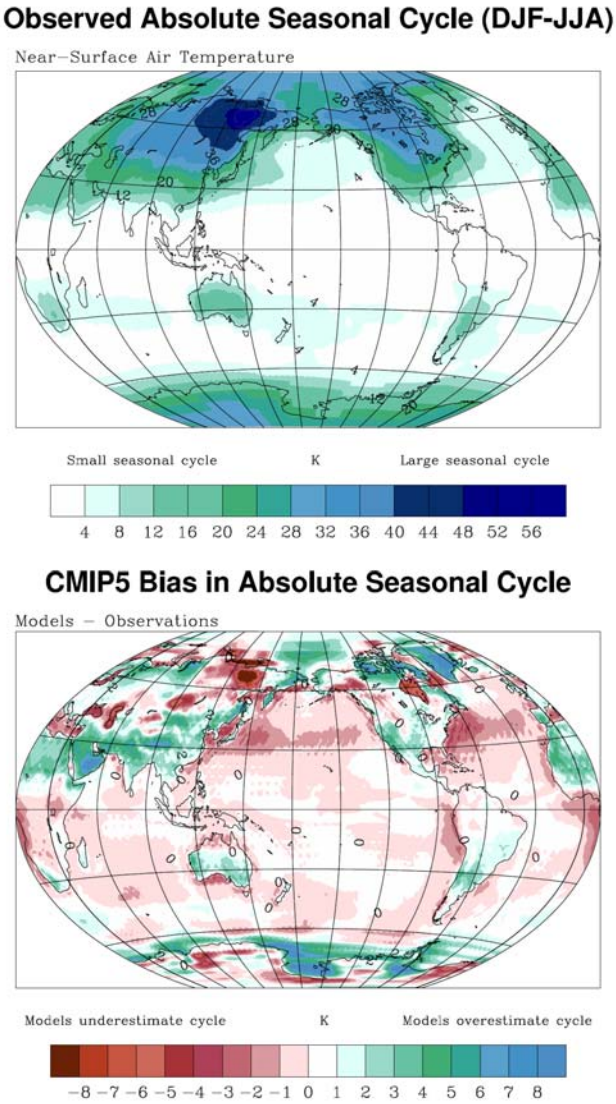


Figure 9.4: Annual surface (2 meter) air temperature (K) range (DJF-JJA) for the period (1985–2005). Top panel shows updated observations from Jones (1999). Bottom panel shows bias in the multi-model ensemble (MME) constructed with 11 available AOGCMs used in the CMIP5 historical experiment.

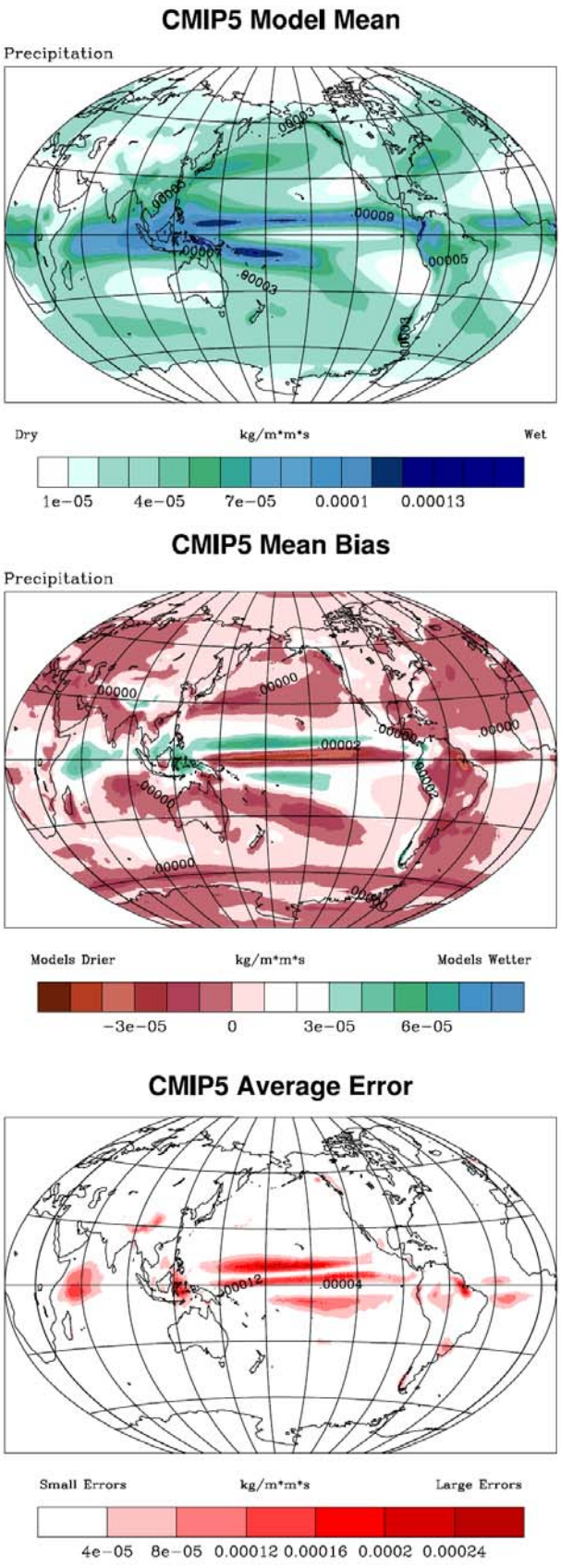


Figure 9.5: Annual mean precipitation for the period (1985–2005). Top panel: Multi-model ensemble (MME) constructed with 11 available AOGCMs used in the CMIP5 historical experiment. Middle panel shows the MME bias with updated observations (Adler et al., 2003). Bottom panel shows the average of the individual model absolute biases.

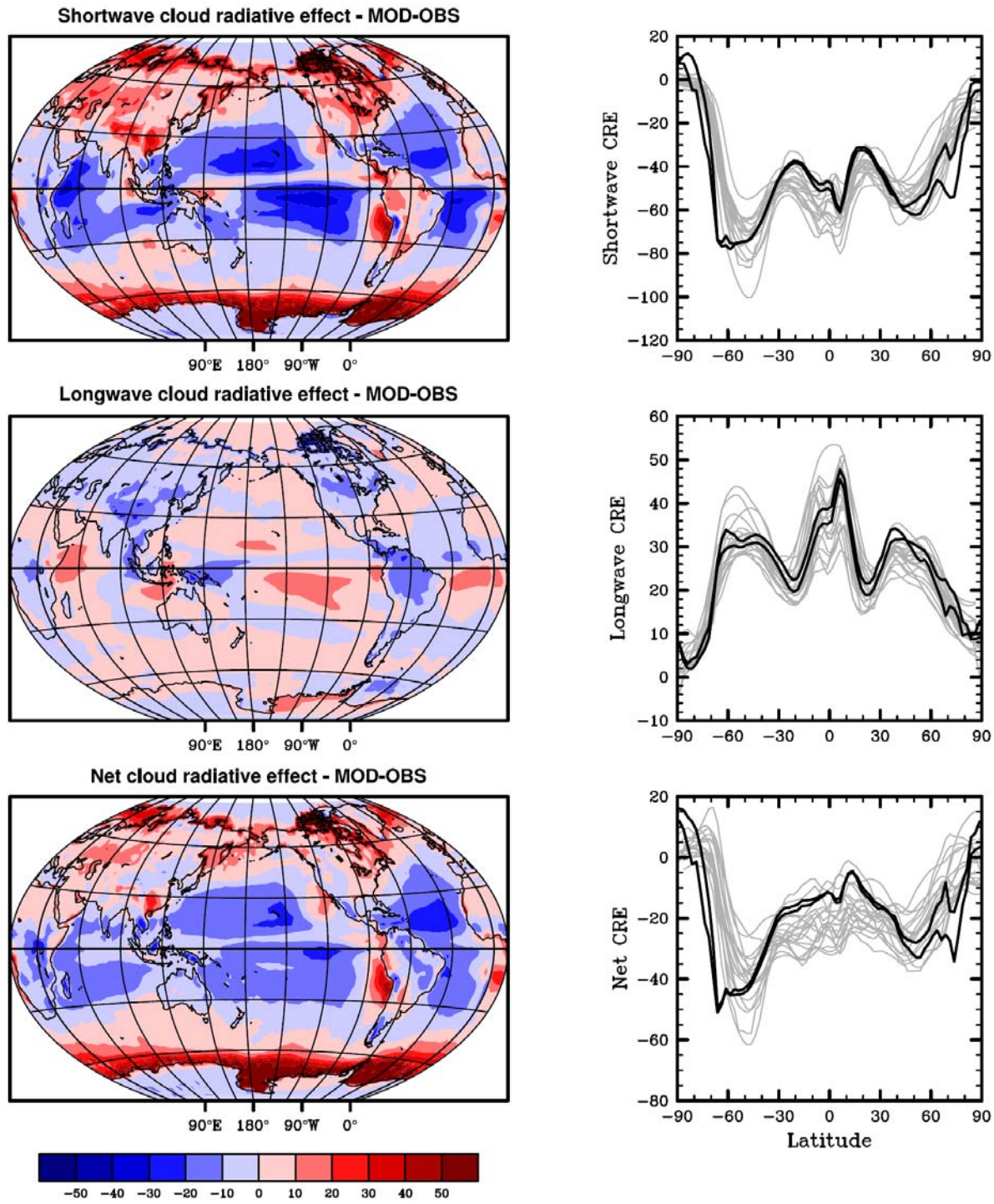


Figure 9.6: Annual mean errors in shortwave (top left), longwave (middle left) and net (bottom left) cloud radiative effect of the CMIP3 multi-model mean. Also shown are zonal averages of the absolute values of the same quantities from observations (CERES ES-4 and ERBE S-4G, thick black lines) and individual models (thin grey lines). For a definition of cloud radiative effect and maps of its absolute values, see Chapter 7.

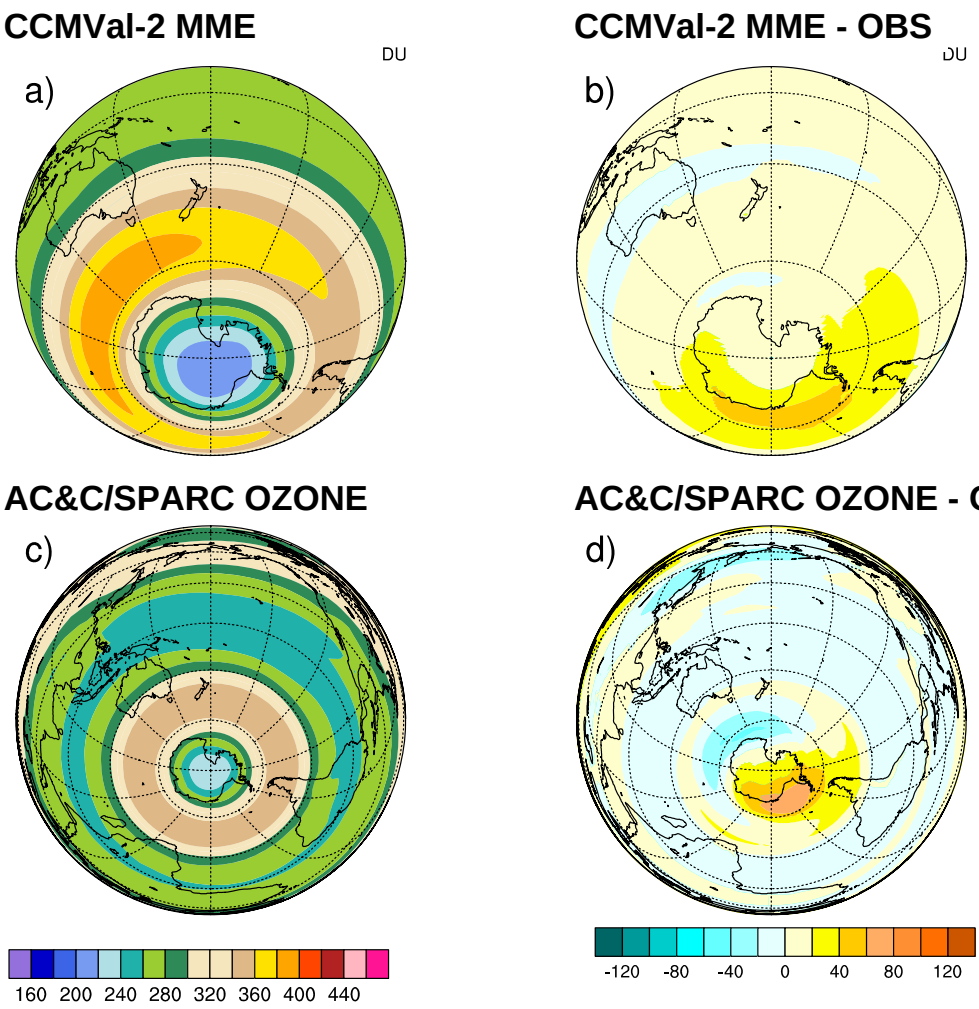


Figure 9.7: September to November total column ozone climatology (1980–1999) from the CCMVal-2 multi-model mean (a) and the bias of it from the NIWA database (b). (c,d) same as (a,b), but for the AC&C / SPARC ozone database that was used as forcing in a subset of the CMIP5 model simulations. Ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere. Modified from Cionni et al. (2011).

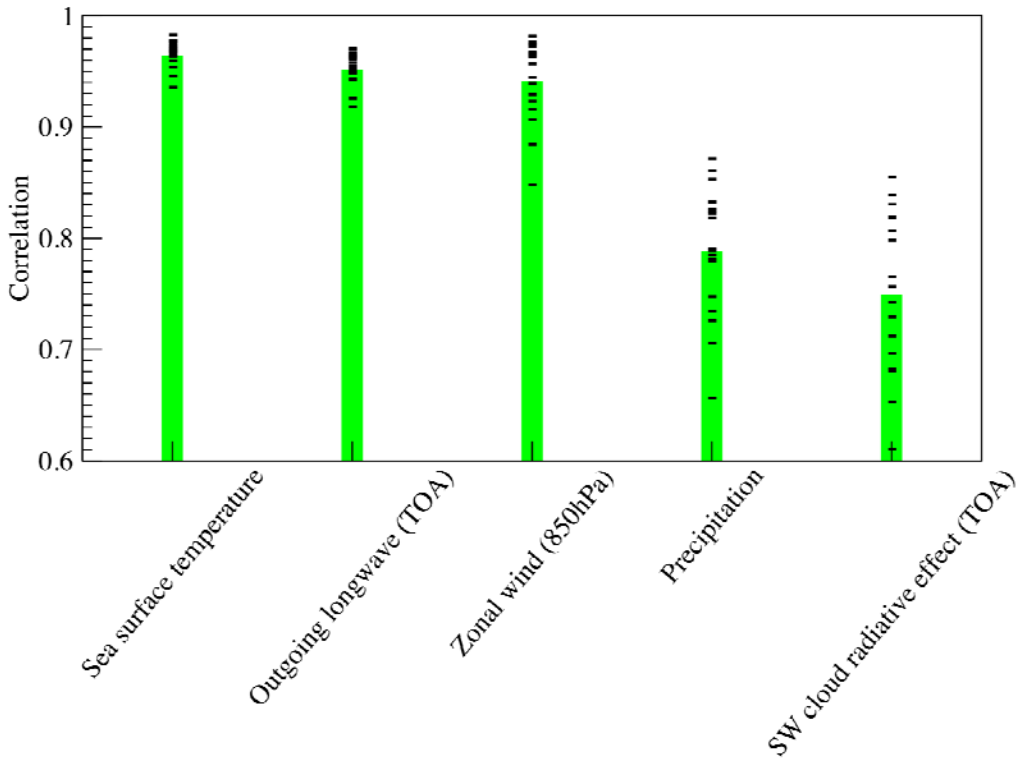


Figure 9.8: Global annual mean climatology (1980–1999) pattern correlations between CMIP3 simulations and corresponding observations (see Table [9.x] for the default references for each field). Results for sea surface temperature and SW cloud radiative effects exclude data pole-ward of 50 degrees in both hemispheres. Individual model results are identified as dash marks. The green bars represent the average result for each variable.

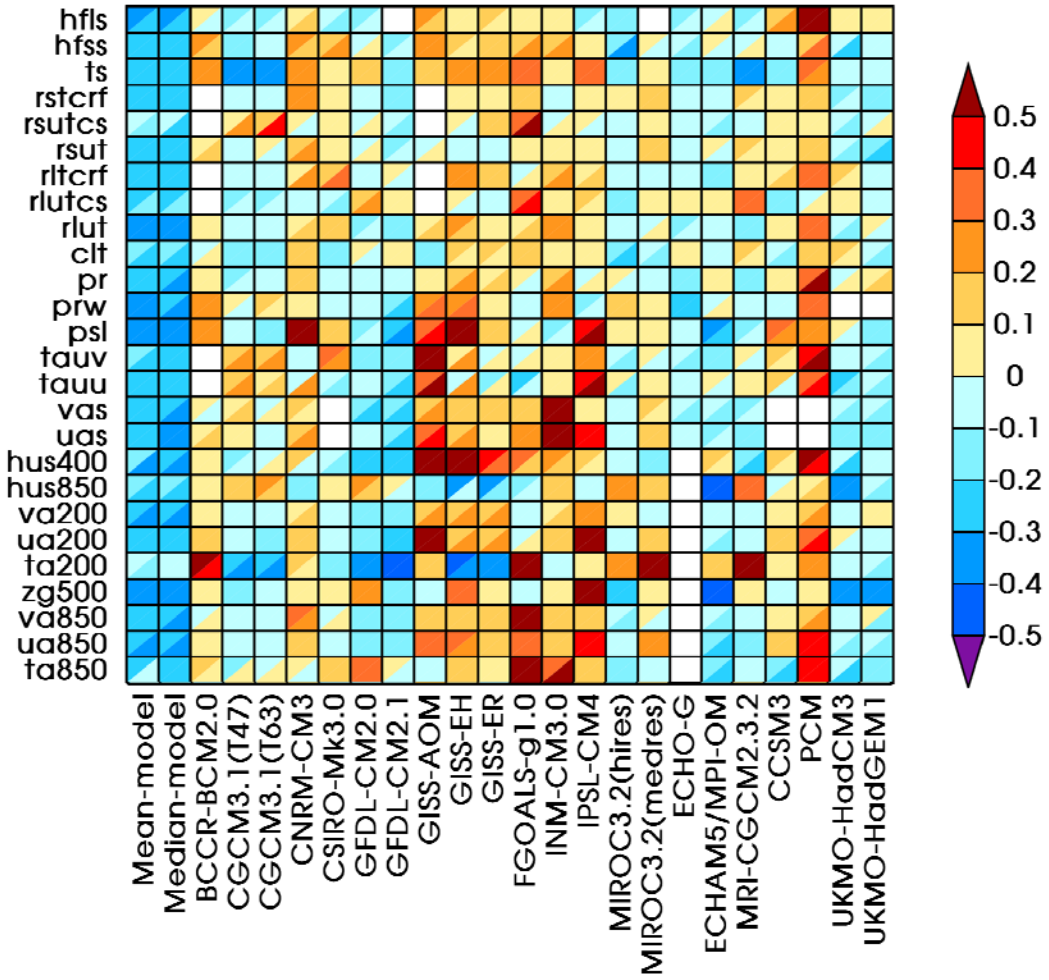


Figure 9.9: Relative error measures for 20th century CMIP3 models, based on the global annual cycle climatology (1980–1999) in the historical (20c3m) experiments. Treating each variable independently, the space-time RMSE is normalized by the median result across all models. A value of 0.3 indicates an error 30% larger than the median error, whereas –0.3 is 30% smaller than the median error. A diagonal splits each grid square showing the relative error with respect to both the primary (upper left triangle) and the alternate (lower right triangle) reference data sets. Taken from Gleckler et al. (2008). The two left hand columns depict the relative error for both the multi-model mean and median (which is distinct from the normalization by the median of individual models).

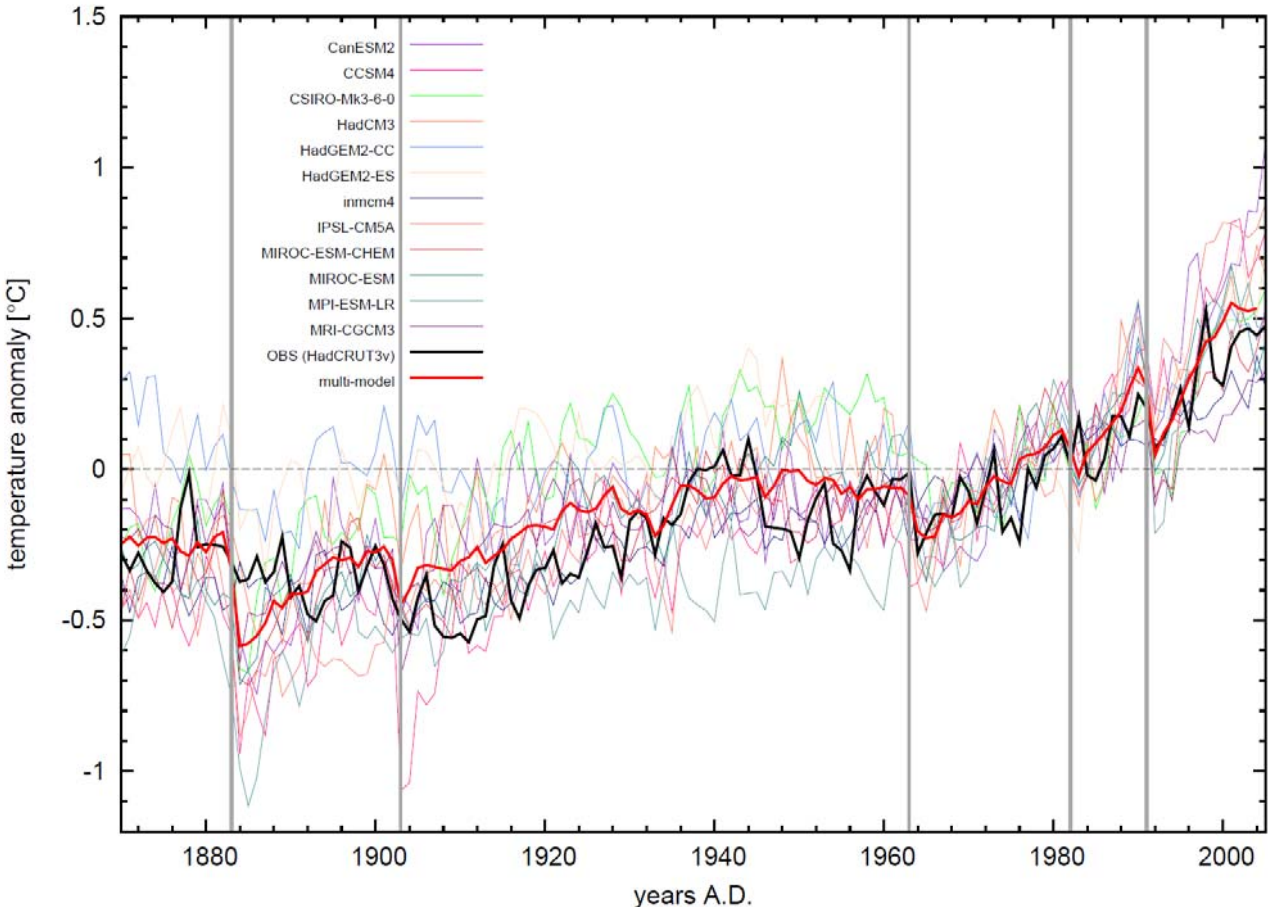


Figure 9.10: Observed and simulated annual mean global average anomaly time series of surface air temperature. Lines (thin) show results from single simulations currently available for CMIP5. Thick black and red lines represent the observations and the multi-model mean respectively. Vertical grey bars represent times of major volcanic eruptions. Observational data are the HadCRUTT3v merged surface temperature, 2 meter of land and surface over the ocean. The current plot shows 2 meter temperature over the land and ocean for model simulations (to be updated with merged surface temperature). All anomalies are with respect to a 1961–1990 climatology.

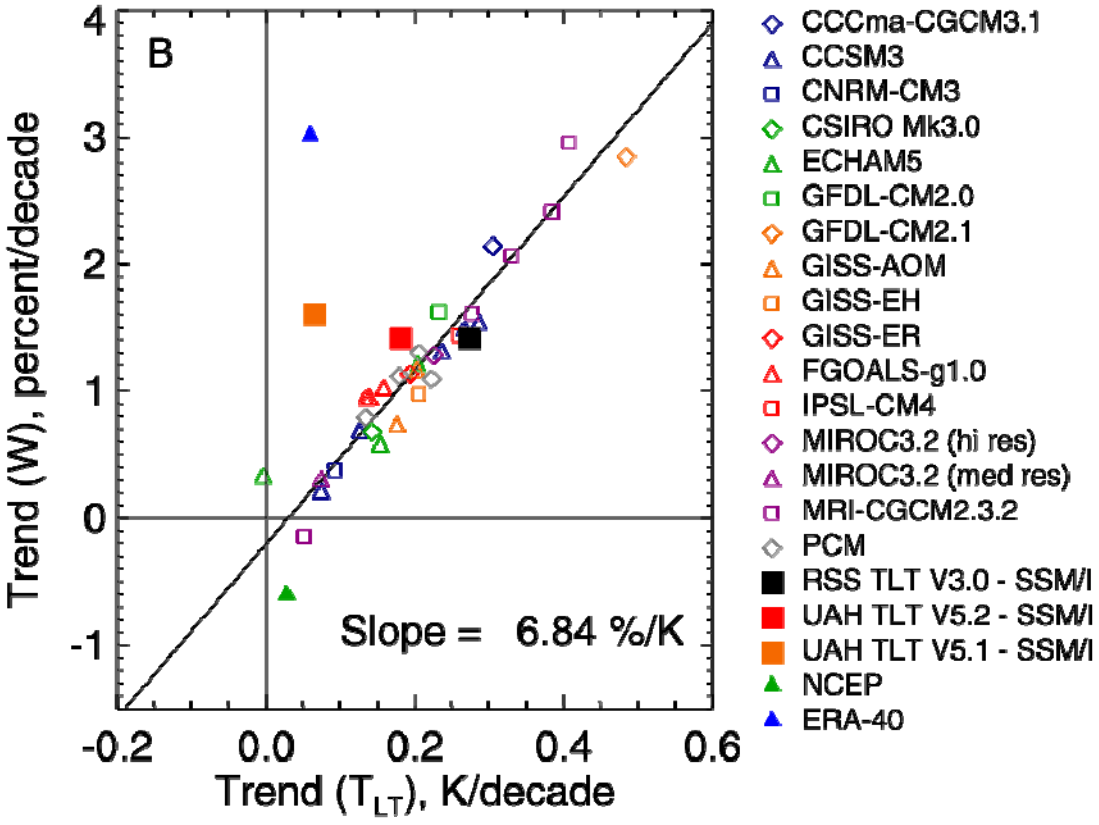


Figure 9.11: Scatter plot of the variability of W as a function of the trend in W as a function of the TLT trend for the tropical oceans. Trends are calculated over the periods given in Table 1. In Figure 3a, UAH V5.2 and UAH V5.1 yield nearly identical results, so the UAH V5.2 data point is hidden. The lines shown bisect the two different linear fits obtained with first W, then TLT assumed to be the dependent variable. Isobe et al. (1990) show that this is a good method for finding an estimate of an underlying relationship in the presence of unknown measurement errors and or scatter that is not strictly related to measurement error, as is the case here. The climate model and reanalysis results are for the 1981–1999 period, while the satellite results are for the 1988–2006 period, so the trend results from the satellite data and the models and reanalysis cannot be directly compared. Also, for UAH V5.1, the calculations are performed over the 1988–2005 period when both SSM/I and UAH 5.1 data are available.

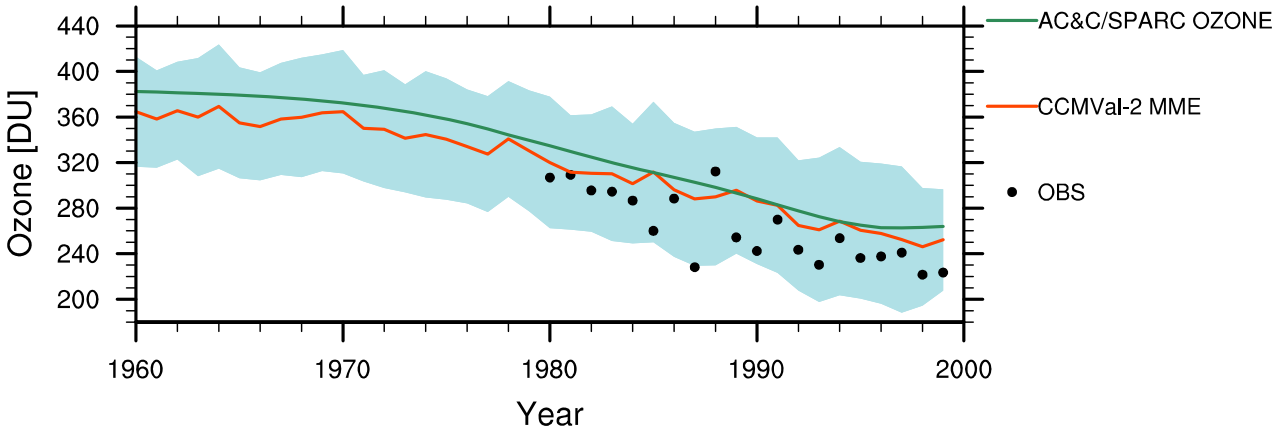


Figure 9.12: Time series of total column ozone over Antarctica (averaged from 60–90°S) from 1960 to 2000 for the CCMVal-2 multi-model mean (red line) and standard deviation (blue shaded area) in comparison to the AC&C / SPARC ozone database (green line) and observations from the NIWA database (black dots). Ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere. Modified from Cionni et al. (2011).

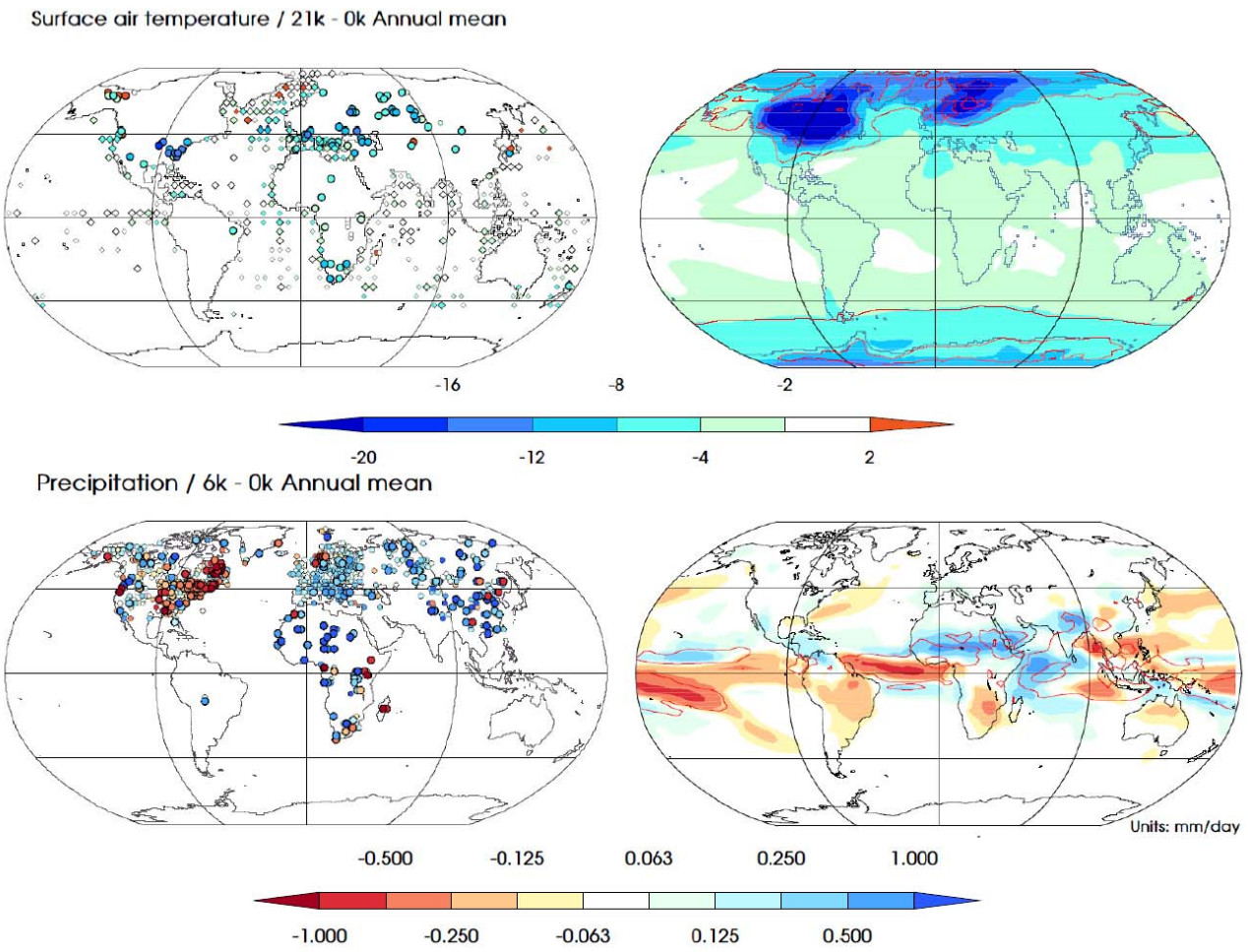


Figure 9.13: Surface air temperature difference between the Last Glacial Maximum (21 ka) and today (top), and precipitation difference between the mid-Holocene (6 ka) and today (bottom), as shown by palaeo-environmental data (left) and the PMIP2 ensemble model simulations (right). The top uses reconstructed and simulated mean temperature of the coldest month (K) for the LGM, the bottom uses reconstructed and simulated mean annual precipitation (mm/day) for the MH. The land reconstructions are from (Bartlein et al., 2010) and the ocean reconstruction are from (Waelbroeck et al., 2009). On the right figures the red line highlights the root-mean square of the inter-model differences.

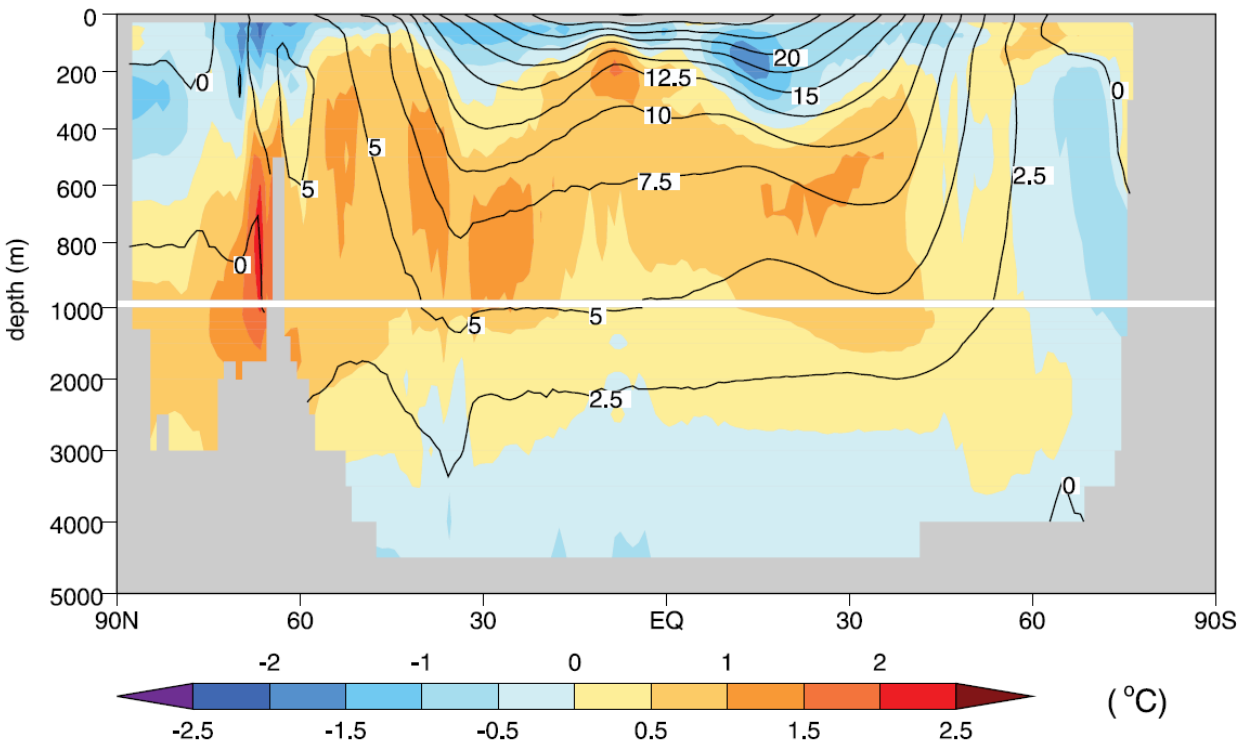


Figure 9.14: [PLACEHOLDER FOR SECOND ORDER DRAFT: Figure from AR4; to be redone from CMIP5 when results available.] Time-mean observed potential temperature (°C), zonally averaged over all ocean basins (labelled contours) and multi-model mean error in this field, simulated minus observed (colour-filled contours). The observations are from the 2004 World Ocean Atlas compiled by (Levitus, Antonov, & Boyer, 2005) for the period 1957 to 1990, and the model results are for the same period in the 20th-century simulations in the CMIP3 ensemble.

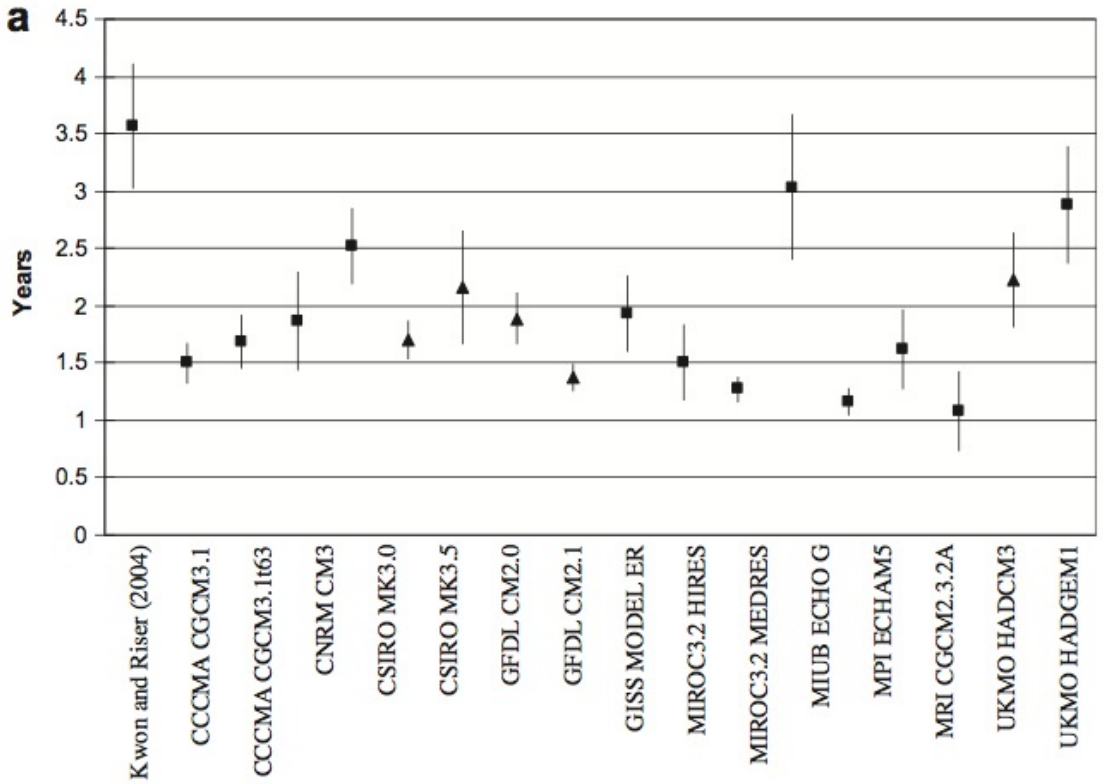


Figure 9.15: Sub-Tropical Mode Water (STMW) turnover time for various models compared with (Kwon & Riser, 2004); time is calculated by annual maximum volume divided by annual production. Values are means; error bars give ranges of one standard deviation. Square data symbols indicate those models with a distinct (if small) secondary water mass transformation rate peak corresponding to STMW formation. Triangular data symbols indicate those models with broad, diffuse, or indiscernible STMW formation peak (from (McClean & Carman, 2011).

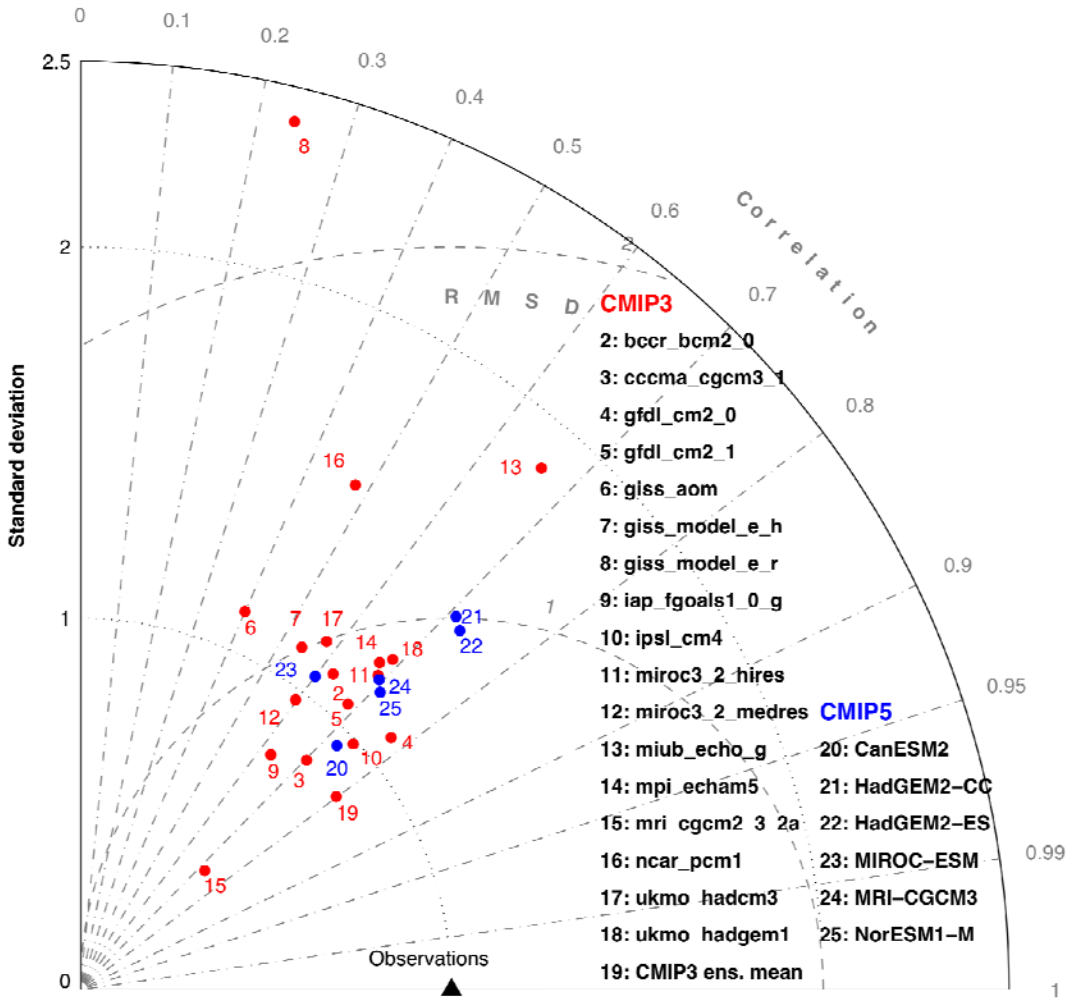


Figure 9.16: Taylor diagram of the dynamic sea-level height seasonal cycle climatology (1987–2000). The radial coordinate shows the standard deviation of the spatial pattern, normalised by the observed standard deviation. The azimuthal variable shows the correlation of the modelled spatial pattern with the observed spatial pattern. The root-mean square error is indicated by the dashed grey circles about the observational point. Analysis is for the global ocean, 50°S–50°N. The reference dataset is AVISO, a merged satellite product (Ducet, Le Traon, & Reverdin, 2000), which is described in Chapter 3. Figure currently shows results for the CMIP3 models and the CMIP5 data currently available.

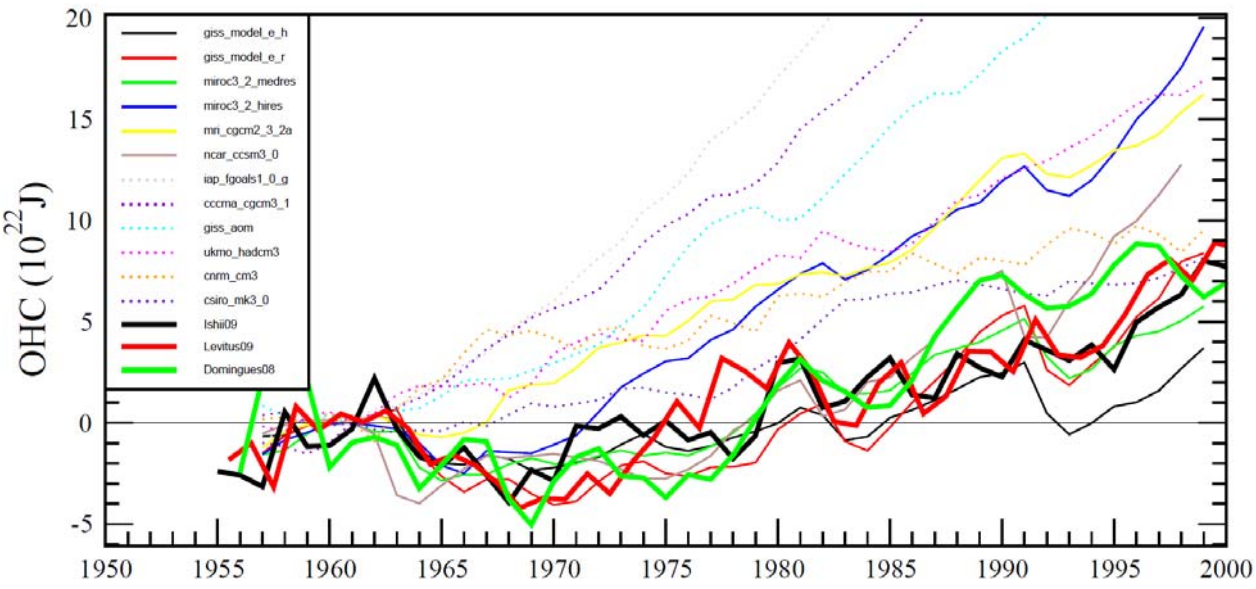


Figure 9.17: Time series of observed and simulated (CMIP3) global ocean heat content (0–700 m) anomalies during the second half of the 20th Century. The three observational estimates (thick lines) are discussed in Chapter 3. Individual (one per model) simulations are shown, with solid lines for models that included volcanic forcings and dashed lines for those that did not. When updated with CMIP5 results, this figure may evolve into multiple panels, e.g., to depict averaging across multiple realizations to better capture trends and include results from historically forced ESMs.

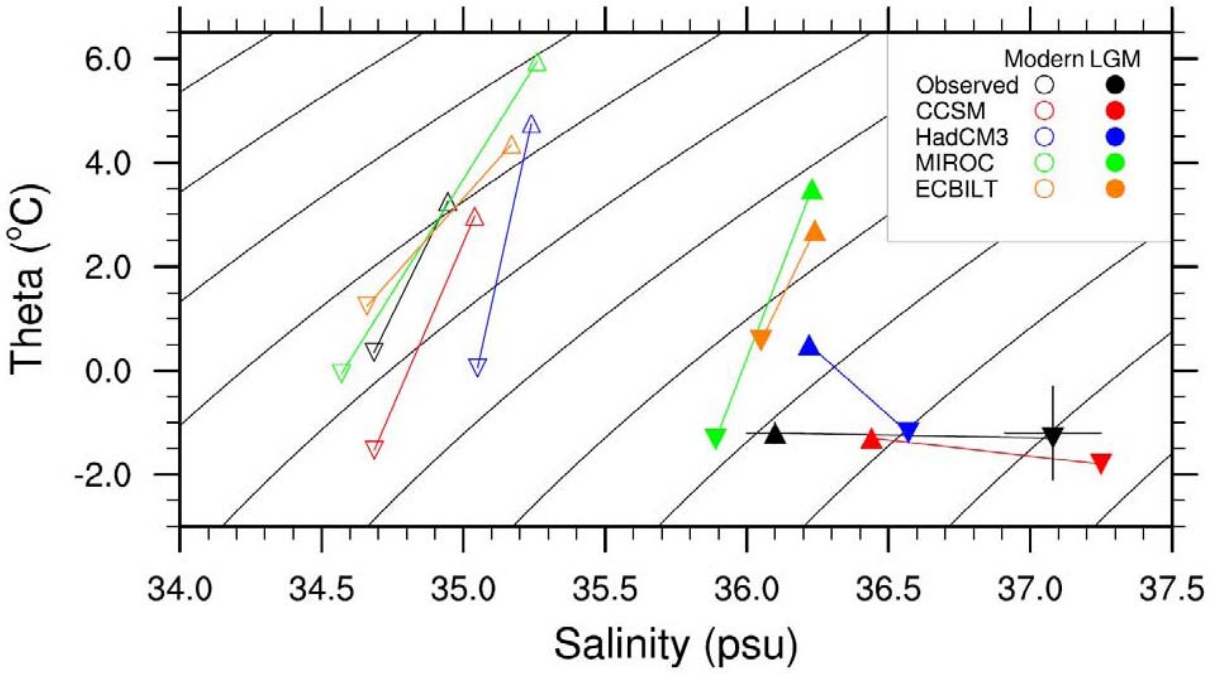


Figure 9.18: Temperature and salinity for modern (open symbols) and LGM (filled symbols) as estimated from data (with error bars) at ODP sites (Adkins et al., 2002) and predicted by the PMIP2 models. Site 981 triangles) is located in the North Atlantic (Feni Drift, 55_N, 15_W, 2184 m). Site 1093 (upside down triangles) is located in the South Atlantic (Shona Rise, 50_S, 6_E, 3626 m). OnlyCCSMincluded a 1 psu adjustment of ocean salinity at initialization to account for fresh water frozen into LGMice sheets; HadCM, MIROC, and ECBilt LGM predicted salinities have been adjusted to allow comparison. Show quantitatively how deep-ocean properties can be evaluated for both modern and palaeoclimate. From (Otto-Bliesner et al., 2007).

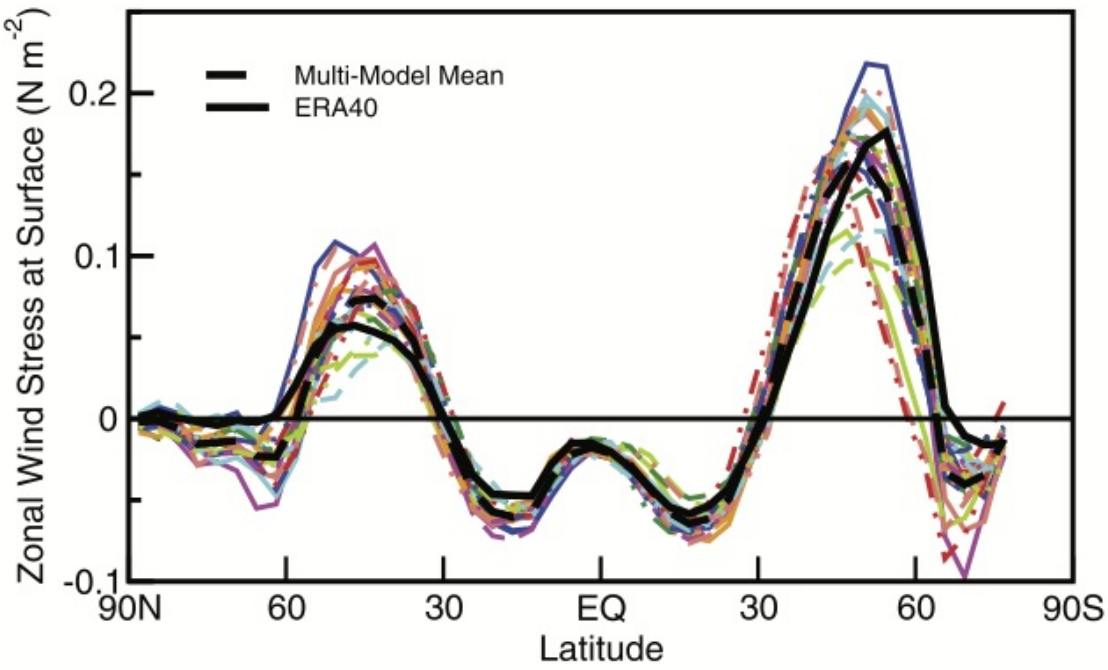


Figure 9.19: Zonal mean zonal wind stress over the oceans in CMIP3 20th century simulations. [PLACEHOLDER FOR SECOND ORDER DRAFT: Final figure will include CMIP5 results and more observational estimates.]

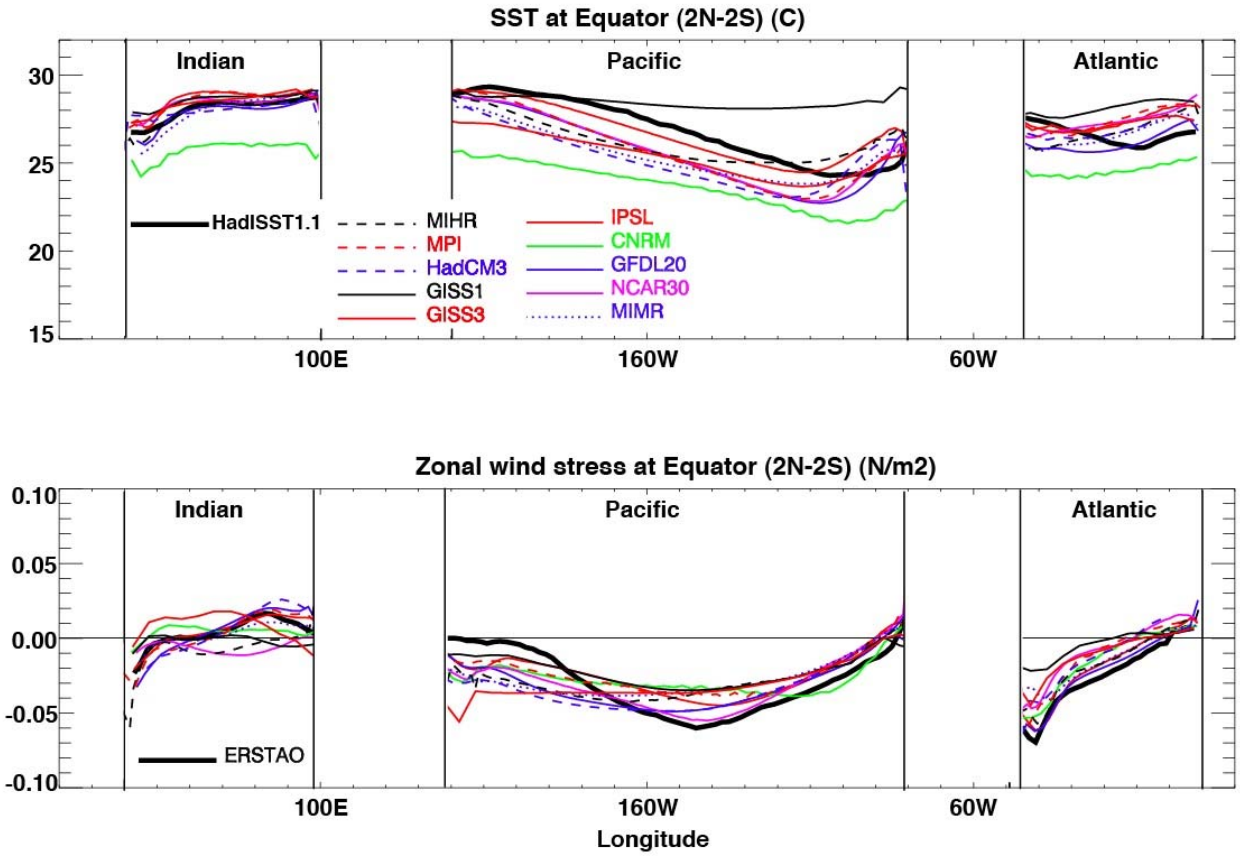


Figure 9.20: (a) SST and (b) zonal wind stress along equator in the Indian, Pacific, and Atlantic Oceans for the CMIP3 and CMIP5 20th-century simulations. Observations are from HadISST1.1 for SST (Rayner et al., 2003) and ERSTAO for wind stress (Menkes et al., 1998).

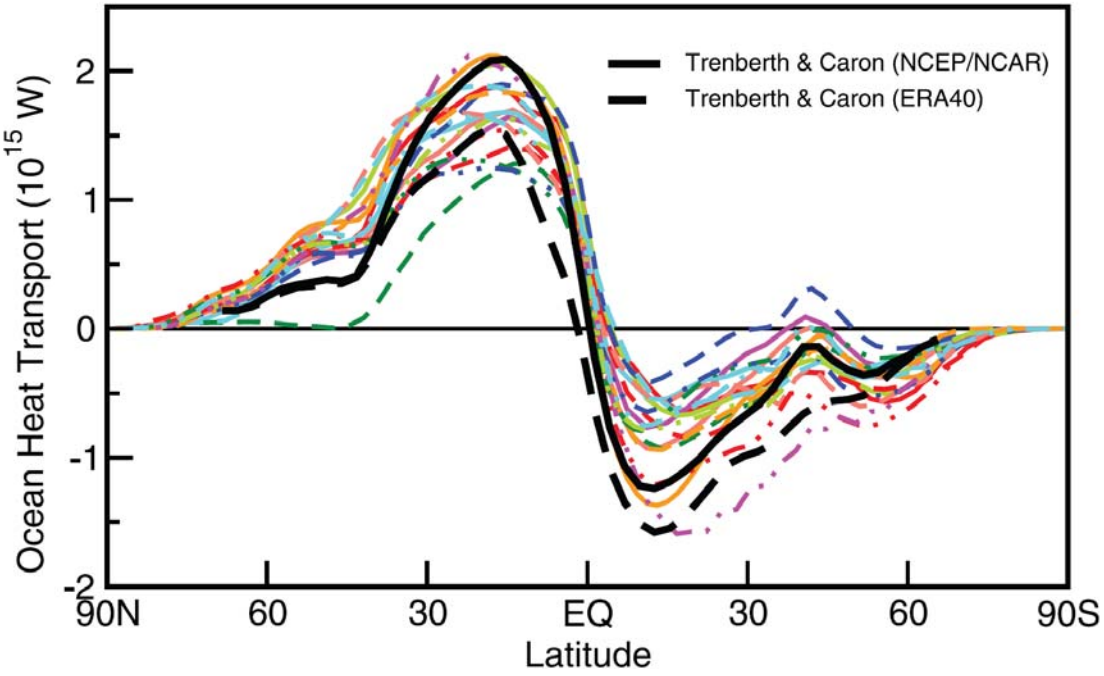


Figure 9.21: [PLACEHOLDER FOR SECOND ORDER DRAFT: From AR4, to be redone using CMIP5 and more observational estimates.] Annual mean, zonally averaged oceanic heat transport implied by net heat flux imbalances at the sea surface, under an assumption of negligible changes in oceanic heat content. The observationally based estimate, taken from (Trenberth & Caron, 2001) for the period February 1985 to April 1989, derives from reanalysis products from the National Centers for Environmental Prediction (NCEP)/NCAR (Kalnay et al., 1996) and European Centre for Medium Range Weather Forecasts 40-year reanalysis (ERA40); (Uppala et al., 2005). The model climatologies are derived from the years 1980 to 1999 in the 20th-century simulations in the CMIP3.

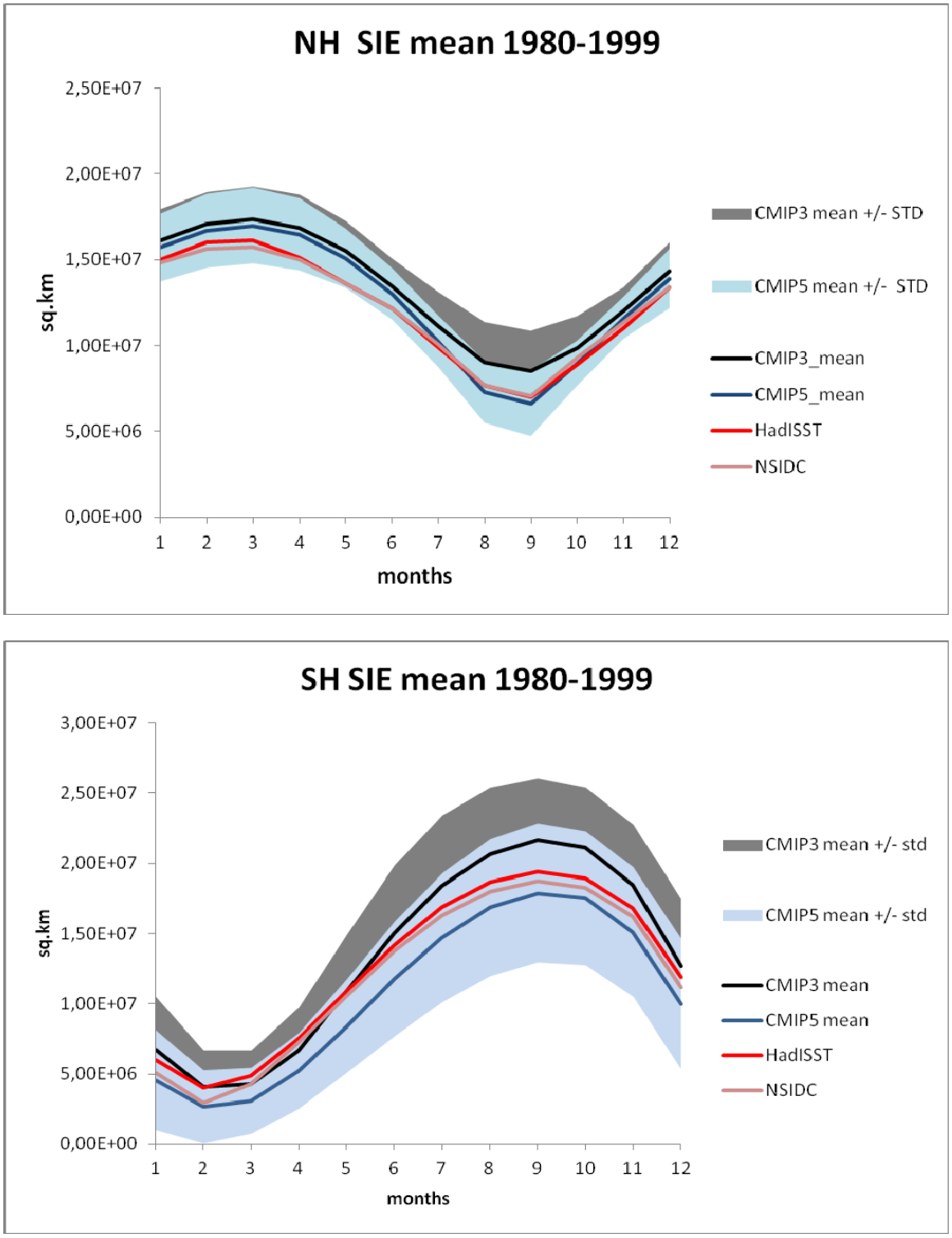


Figure 9.22: Mean sea ice extent (the ocean area within 15% sea ice concentration) seasonal cycle in the Northern (upper panel) and Southern (lower panel) hemispheres as simulated by the CMIP5 (blue line) and CMIP3 (black line) ensembles. The observed sea-ice extent cycles (1980–1999) are based on the Hadley Centre Sea Ice and Sea Surface Temperature – HadISST (Rayner, et al., 2003) (red line) and the National Snow and Ice Data Center – NSIDC (Fetterer, Knowles, Meier, & Savoie, 2002) (brown line) data sets. The shaded areas show the inter-model standard deviation for each ensemble.

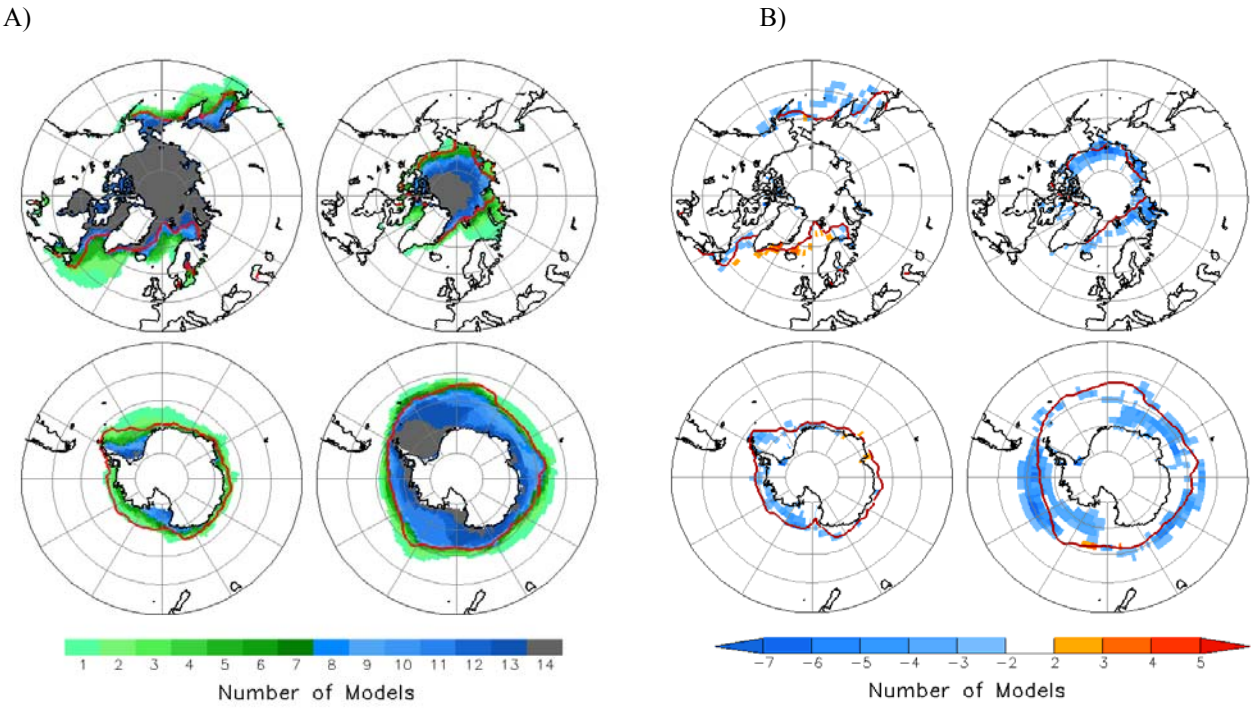


Figure 9.23: Sea ice distribution in the Northern Hemisphere (upper panels) and the Southern Hemisphere (lower panels) for March (left) and September (right). A) AR5 baseline climate (1986–2005) simulated by 14 of CMIP5 AOGCMs. For each 1° × 1° longitude-latitude grid cell, the figure indicates the number of models that simulate at least 15% of the area covered by sea ice. B) AR4 baseline climate (1980–1999) differences between 14 CMIP5 and 14 CMIP3 (AR4 (Randall et al., 2007) Figure 8.10) AOGCMs. For each 2.5° × 2.5° longitude-latitude grid cell, the figure indicates the difference in the number of CMIP5 and CMIP3 models that simulate at least 15% of the area covered by sea ice. The observed 15% concentration boundaries (red line) are based on the Hadley Centre Sea Ice and Sea Surface Temperature – HadISST data set (Rayner, et al., 2003).

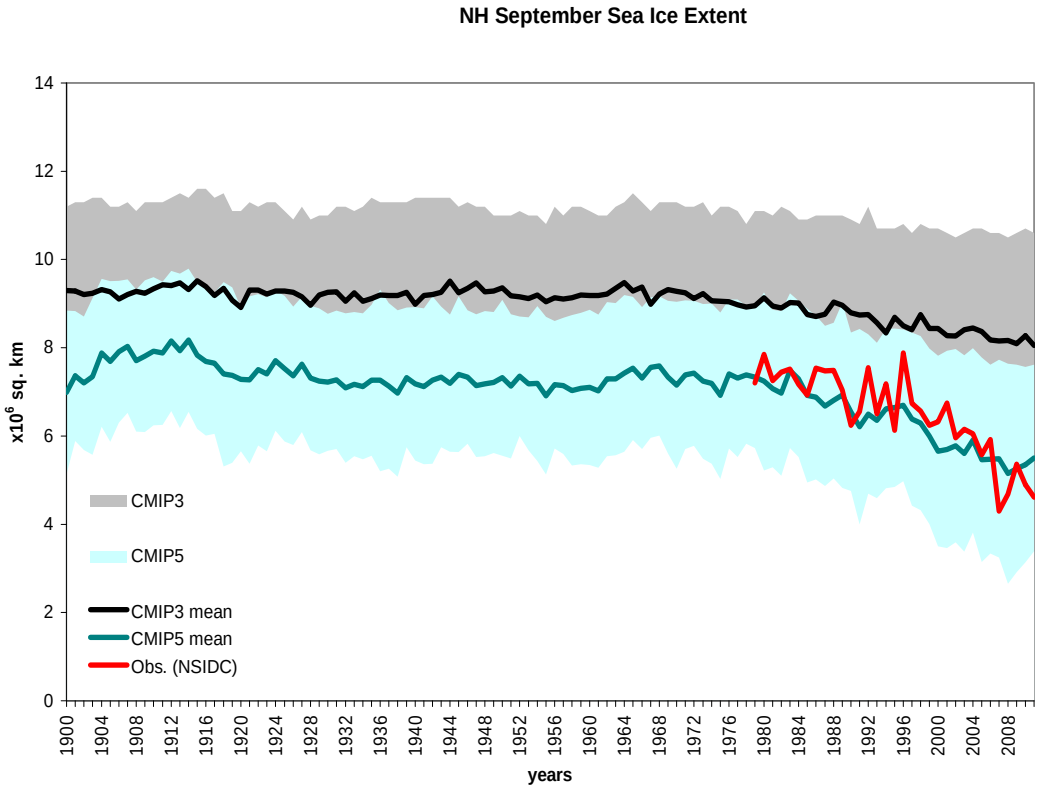


Figure 9.24: Arctic September sea-ice extent from observations (NSIDC, red line) (Fetterer, et al., 2002), the 12 CMIP5 model multi-model ensemble mean (dark greenish line), the 12 CMIP3 model multi-model ensemble mean (black line), and one standard deviation ranges of the model estimates (bluish and grey shadings, correspondingly). Note that these are September means, not yearly minima.

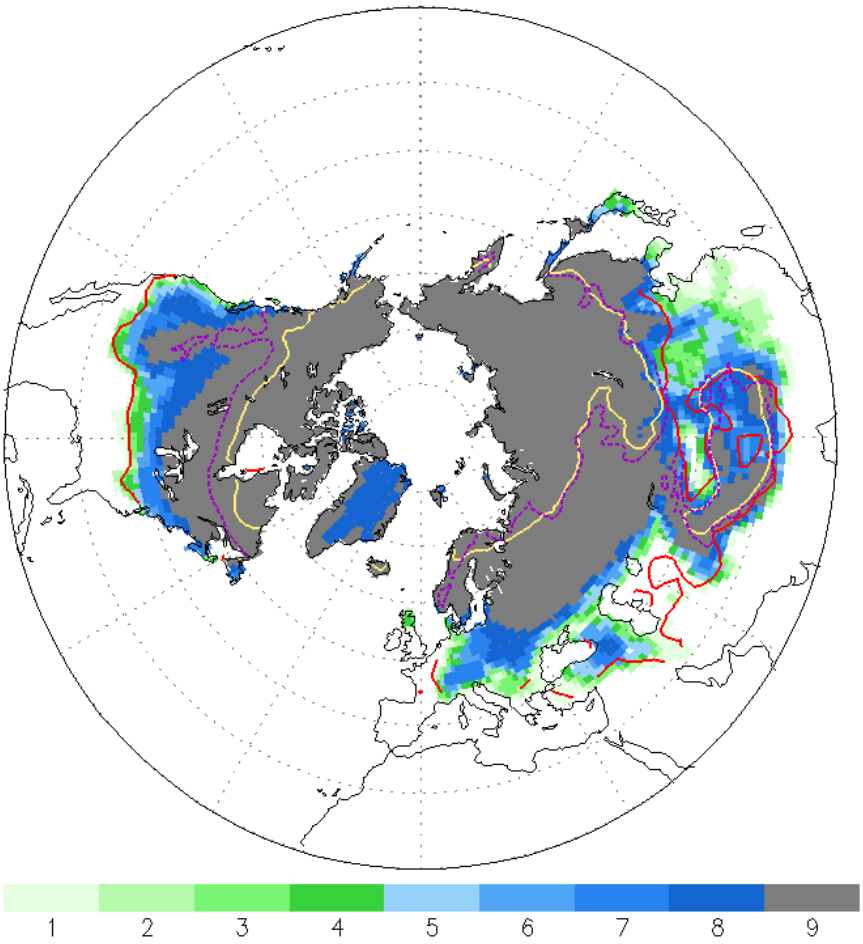


Figure 9.25: Terrestrial snow-cover distribution in the Northern Hemisphere simulated by 9 CMIP5 AOGCMs for February. For each 1° × 1° longitude-latitude grid cell, the figure indicates the number of models that simulate at least 5 kg m^{-2} of snow water equivalent. The observed 20% concentration boundaries (red line) are based on the (Robinson & Frei, 2000) and cover the period 1986–2005. The annual mean 0°C isotherm at 3.3 m depth averaged across the 9 AOGCMs (yellow line) is a proxy for the permafrost boundary. Observed permafrost zonation in the Northern hemisphere (magenta dashed line) is based on (Nelson, Anisimov, & Shiklomanov, 2002).

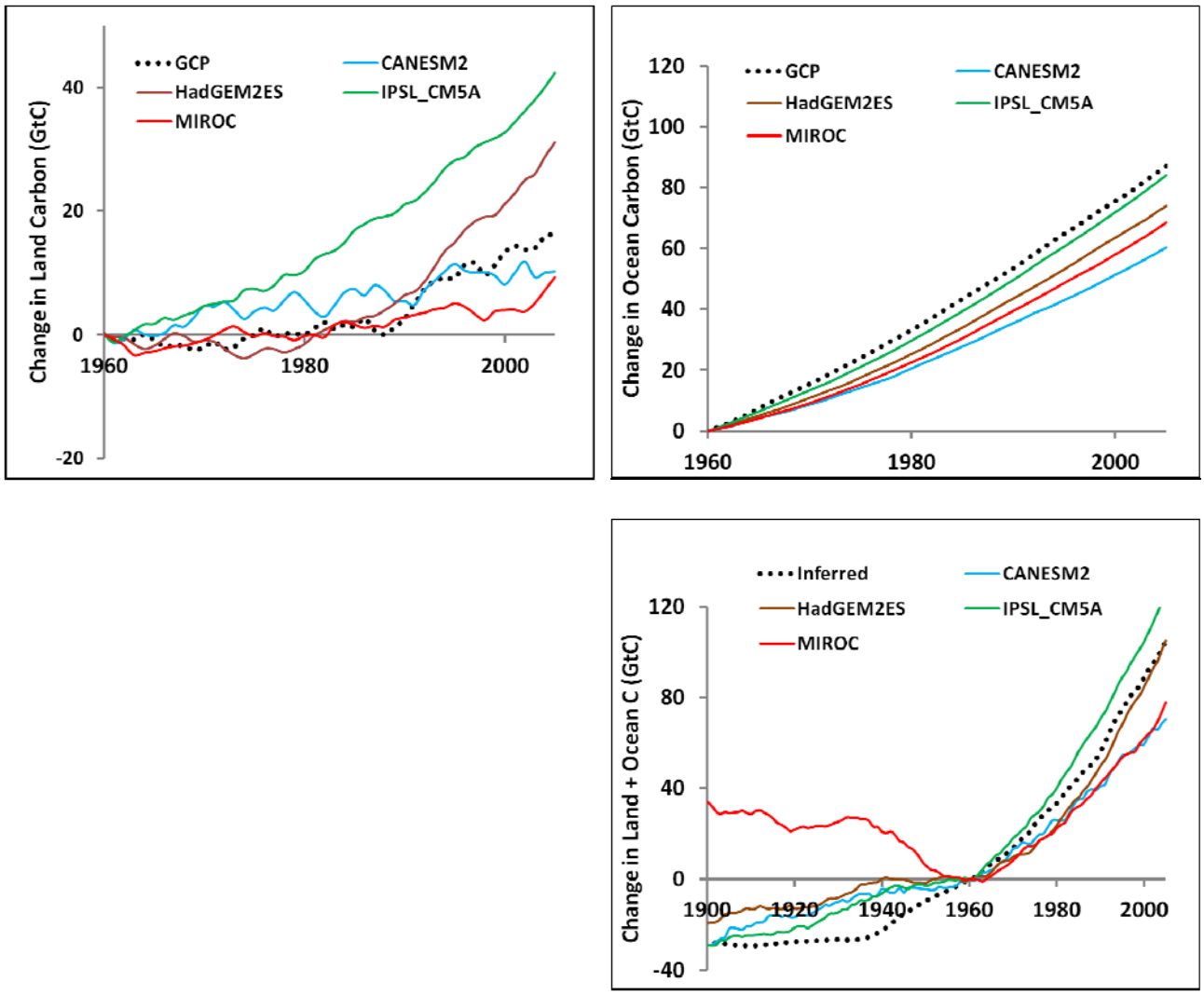


Figure 9.26: Simulation of land carbon uptake (top left) and ocean carbon uptake (top right) in the CMIP5 Earth System Models (ESMs), for the period 1960–2005, relative to 1960. All of these models include the impact of land-use changes on land carbon storage. For comparison, the observation-based estimates provided by the Global Carbon Project ("GCP", (Le Quere et al., 2009) are also shown as the dotted line. The bottom right panel shows the sum of the land and ocean uptake from 1900 to 2005, again relative to 1960.

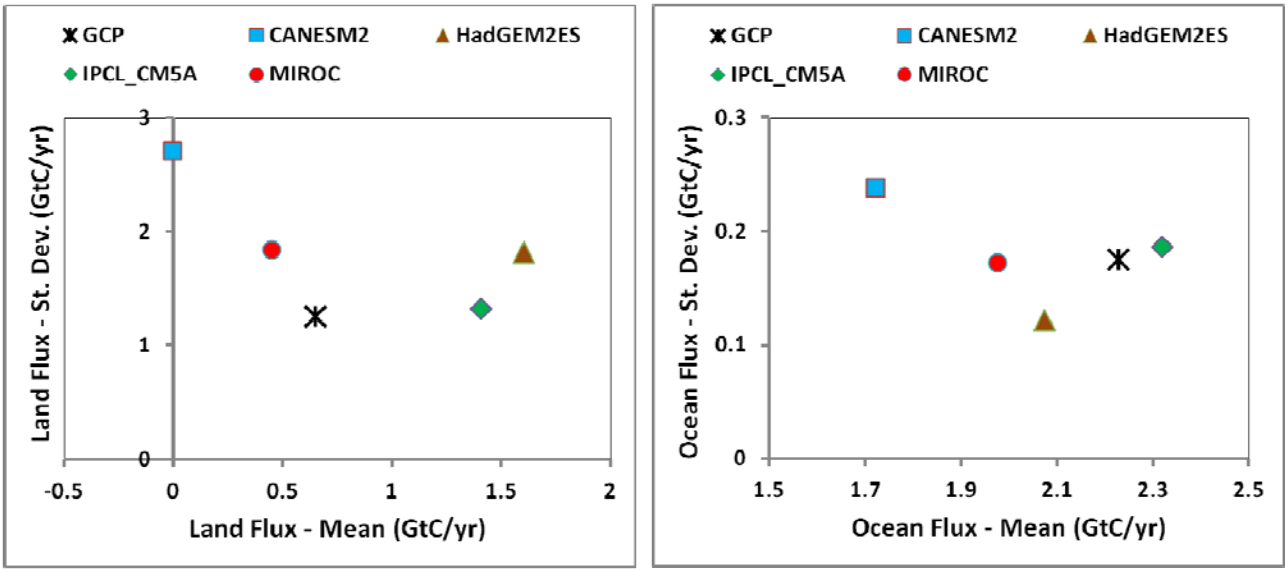


Figure 9.27: Simulation of net land CO_2 flux (top left) and ocean CO_2 flux (top right) in the CMIP5 Earth system models (ESMs), for the period 1995–2005. In each panel the mean flux over the period is plotted on the x-axis, while the standard deviation of the annual fluxes is plotted on the y-axis. For comparison, the observation-based estimates provided by the Global Carbon Project ("GCP", (Le Quere, et al., 2009) are also shown as the dotted line.

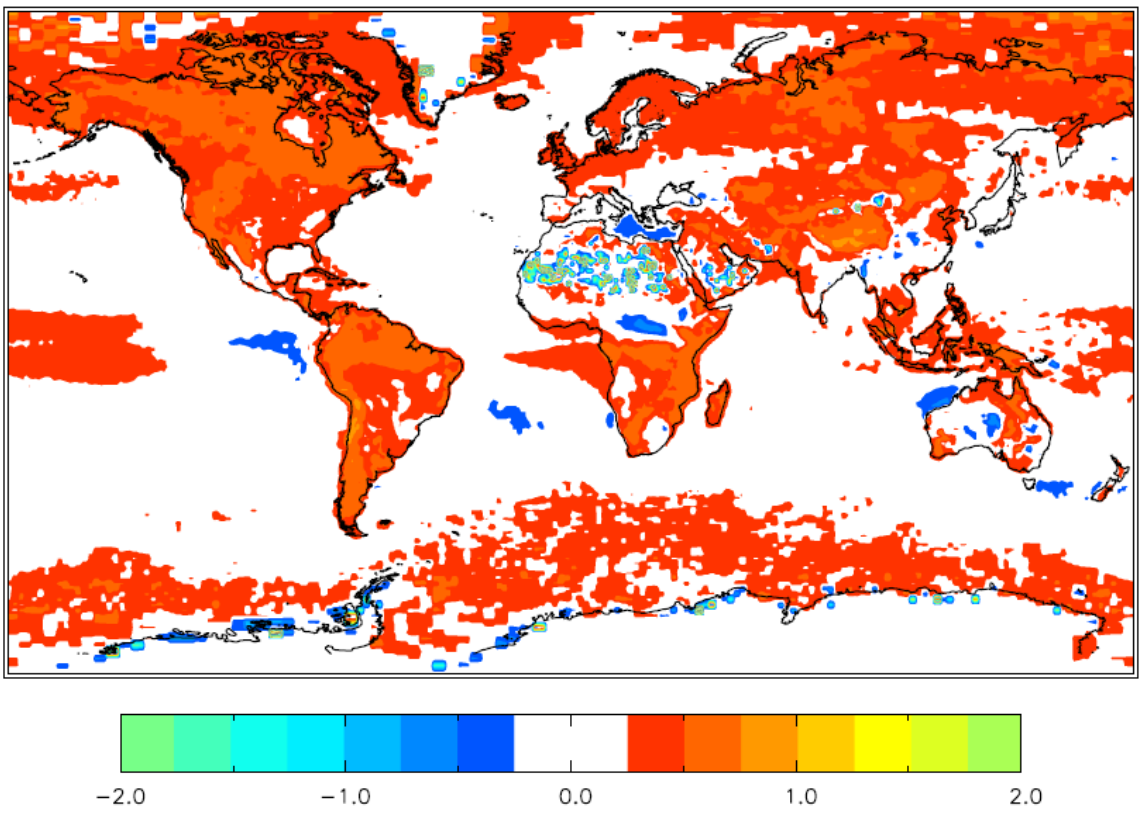


Figure 9.28: The relative error in visible aerosol optical thickness (AOT) from the median of a subset of CMIP5 models' historical simulations, relative to satellite retrievals of AOT. The figure was constructed following (Kinne et al., 2006). The satellite AOT is from the MODIS instrument on the NASA Terra satellite from 2001 through 2005. The data version is MODIS 4; the model output is from CSIRO Mk3-6-0, GISS ER-2, HadGEM2-ES, IPSL CM5A-LR, and NorESM1-M.

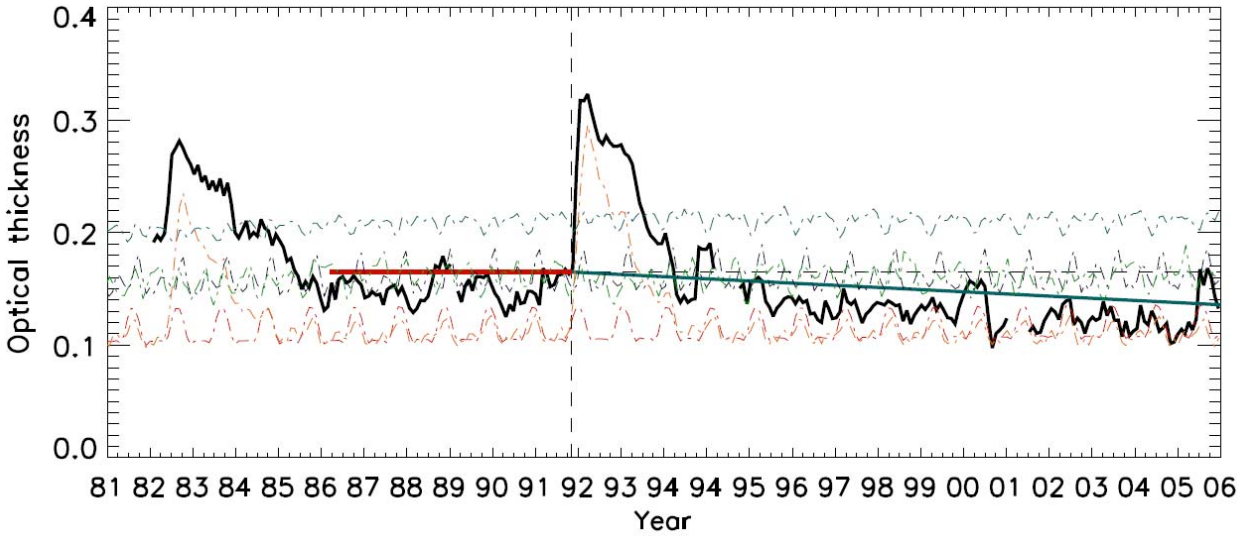


Figure 9.29: Time series of the global oceanic-mean AOT from individual CMIP5 models' historical simulations against the time series of global oceanic-mean AOT from the Global Aerosol Climatology Project (GACP). The figure is constructed following (Mishchenko et al., 2007). The "brightening" trend shown with the straight green line is discussed in Chapter 7. The model output is from CSIRO Mk3-6-0, GISS ER-2, HadGEM2-ES, IPSL CM5A-LR, and NorESM1-M.

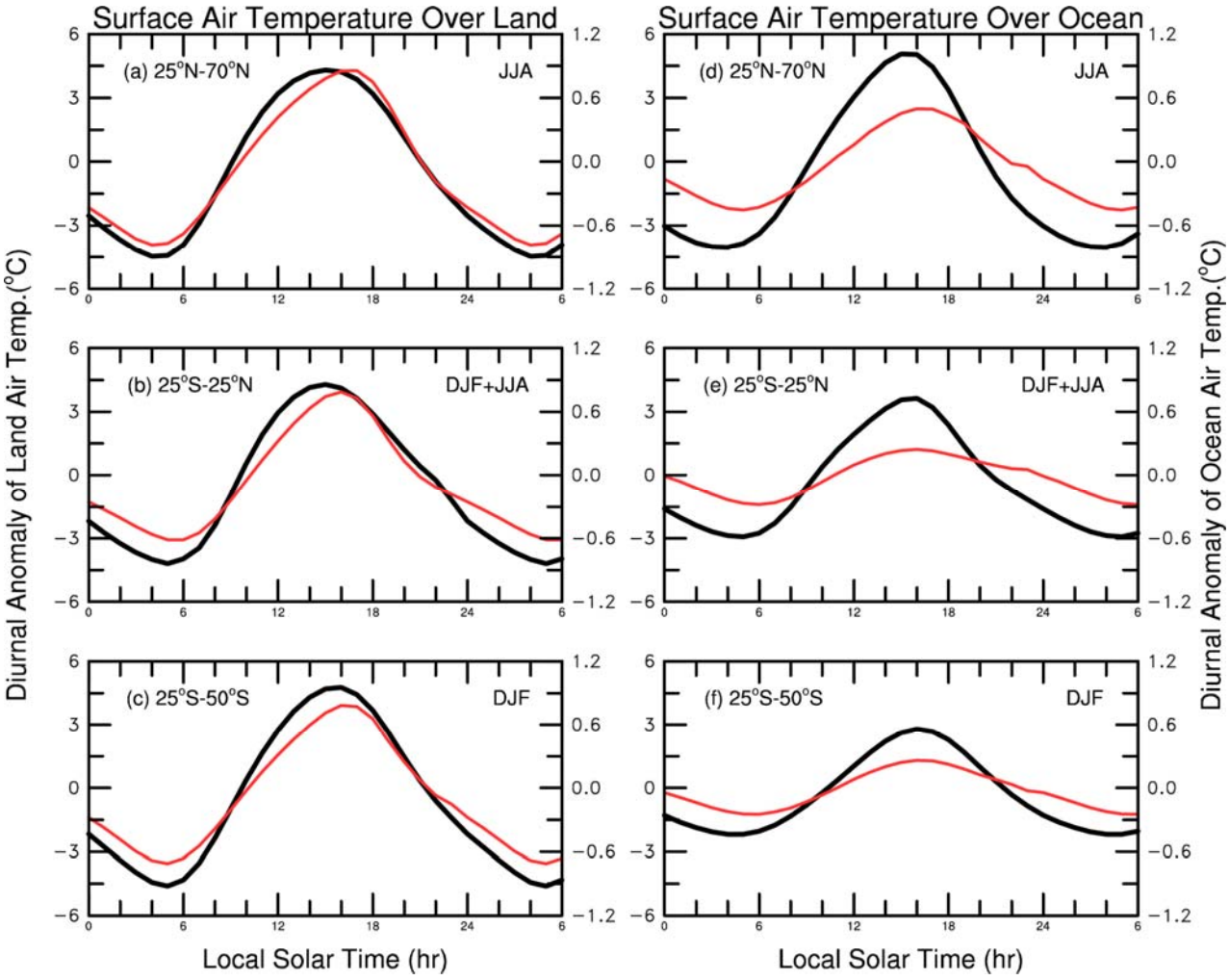


Figure 9.30: Composite diurnal cycle of surface air temperature from observations (black line) and CMIP3 models (coloured lines) averaged over land (left) and ocean (right) areas for three different zones. Adapted from (Dai & Trenberth, 2004).

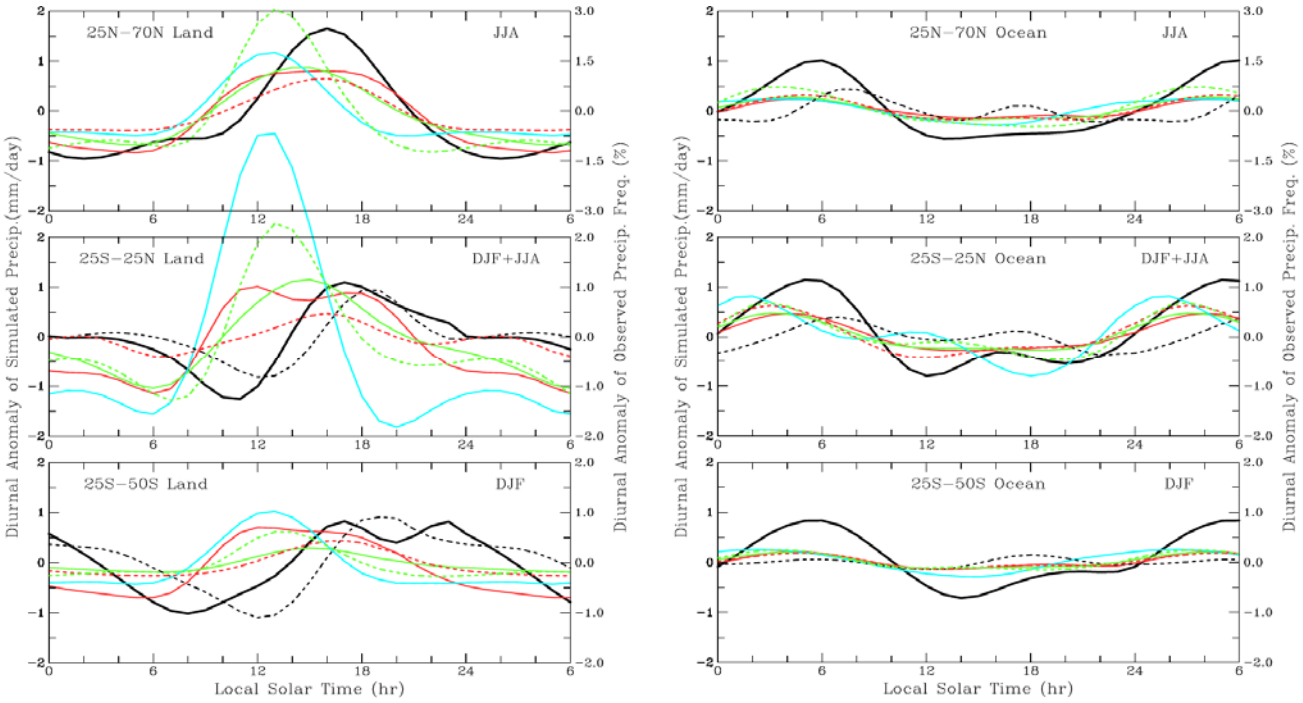


Figure 9.31: Composite diurnal cycle precipitation from observations (black) and a subset of CMIP3 models (coloured lines) averaged over land (left) and ocean (right) areas for three different zones. Adapted from (Dai, 2006).

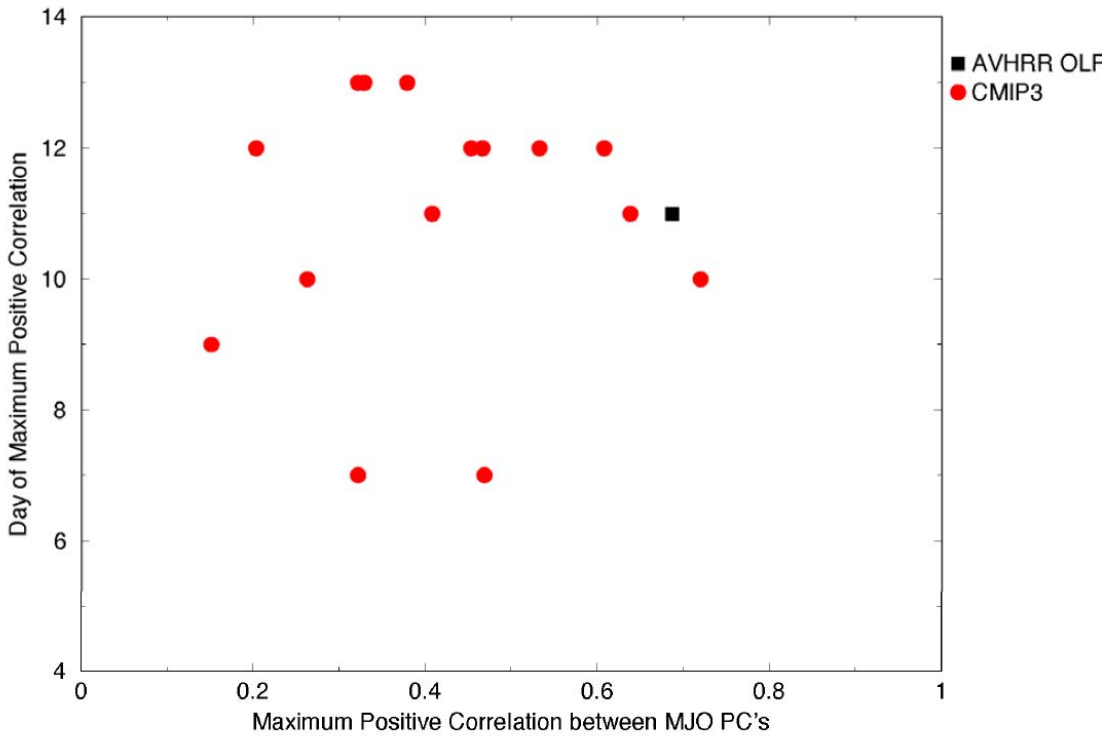


Figure 9.32: Outgoing Longwave Radiation (OLR), 20–100 day filtered, from observations and each of the CMIP3 models' simulations of 20th-century climate is projected on the two leading Empiricol Orthogonal Functions (EOF's) of OLR that constitute the Madden-Julian Oscillation (MJO). Shown is the maximum positive correlation between the resulting MJO Principal Components (PC's) and the time lag at which it occurred for all winters (November-March). The maximum positive correlation is an indication of the coherence with which the MJO convection propagates from the Indian Ocean to the Maritime Continent/western Pacific and the time lag is approximately 1/4 of the period of the MJO. Most models have weaker coherence in the MJO propagation (smaller maximum positive correlation), and some have periods that are too short compared to observations. One CMIP3 model is not shown as its day of maximum positive correlation was –16, indicating that this model is incorrectly dominated by westward propagation. Constructed following (Sperber, Gualdi, Legutke, & Gayler, 2005).

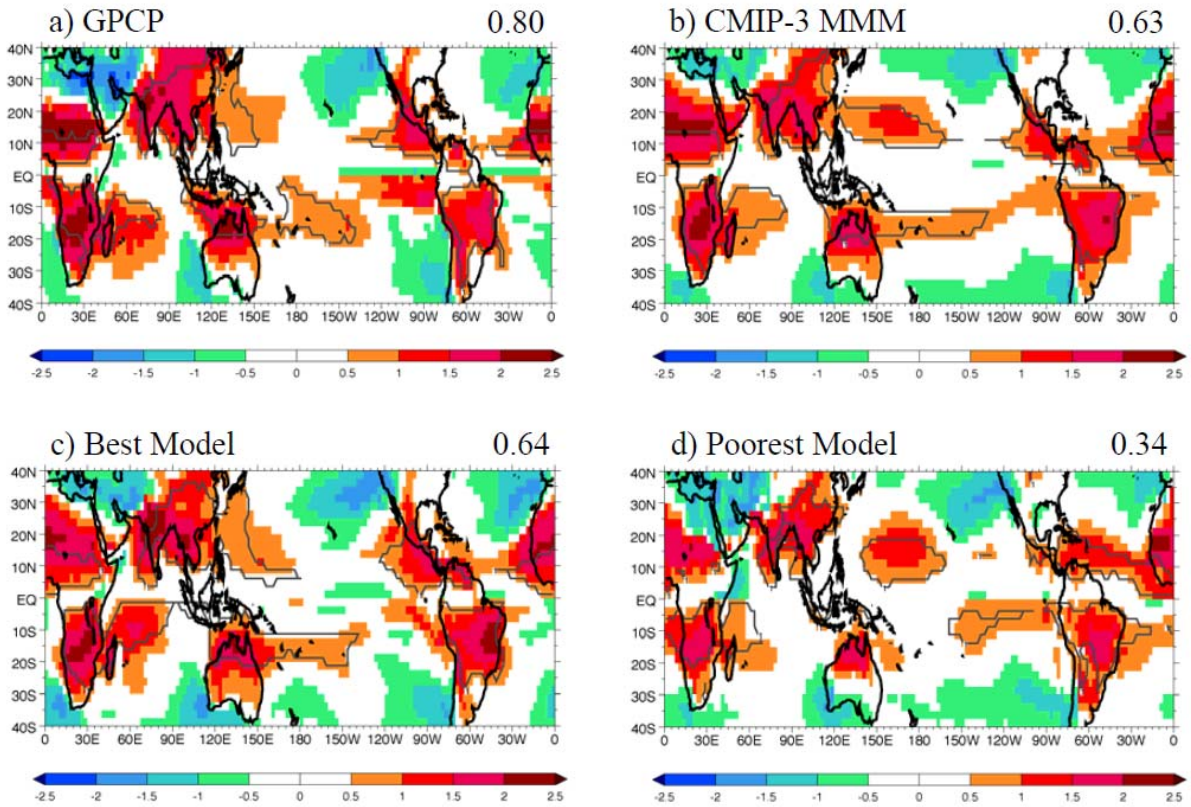


Figure 9.33: Monsoon precipitation intensity (shading, mm/day) and monsoon precipitation domain (lines) are shown for (a) observations from GPCP, (b) the CMIP3 multi-model mean, (c) the best model, and (d) the worst model in terms of the threat score for this diagnostic. The threat scores indicate how well the models represent the monsoon precipitation domain compared to the GPCP data. The threat score in panel (a) is between GPCP and CMAP rainfall to indicate observational uncertainty. A threat score of 1.0 would indicate perfect agreement between the two datasets. See (Wang & Ding, 2008; Wang, Kim, Kikuchi, & Kitoh, 2011); and (Kim et al., 2011) for details of the calculations..

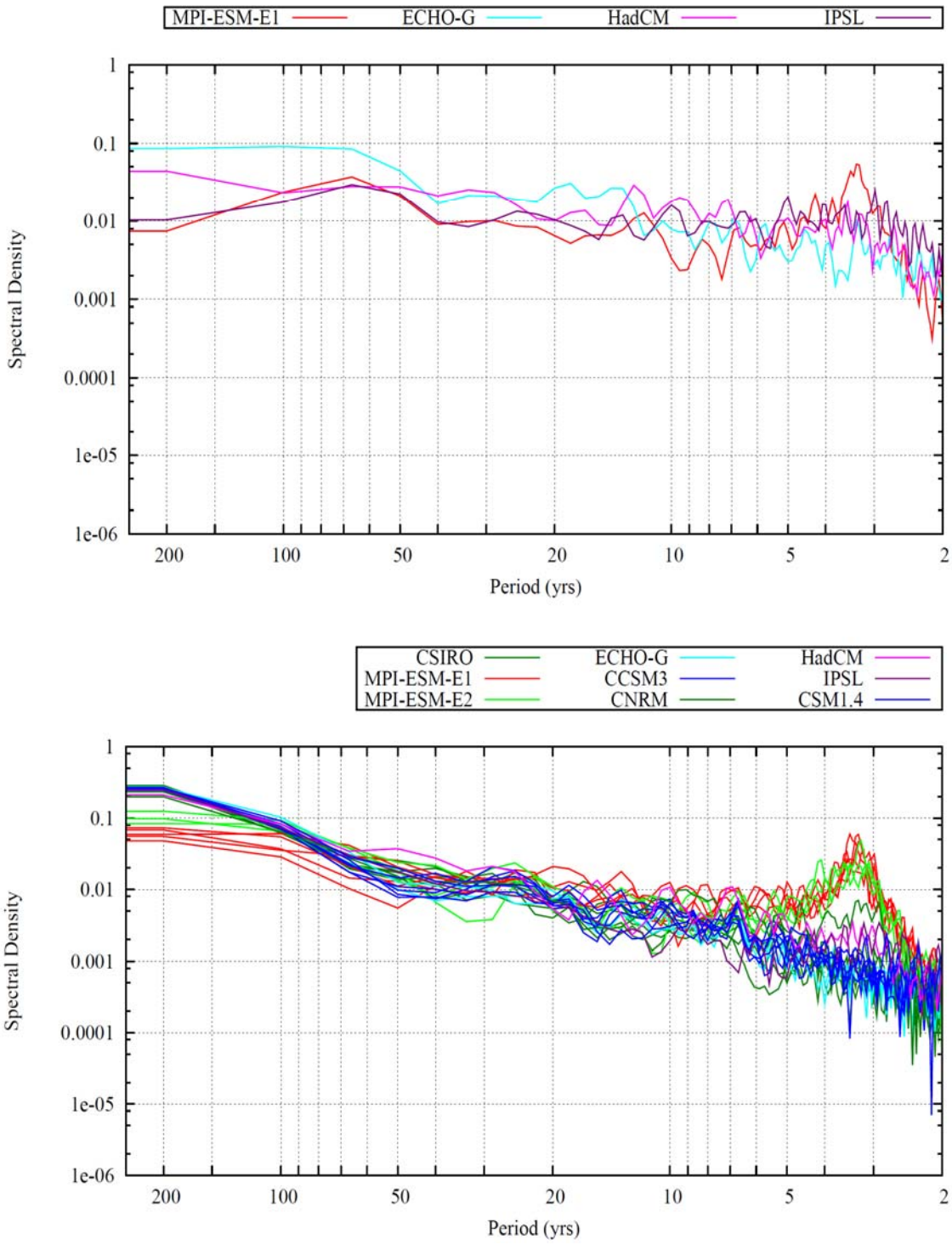


Figure 9.34: Power spectral density of NH temperature for a) several simulations of the last millennium performed with CMIP3-generation models (see Chapter 5) b) long pre-industrial simulations for a subset of the same models. In a) the model were all forced by the long-term evolution of the atmospheric trace gases, tropospheric aerosols (except ECHOG and CCSM3), solar irradiance, volcanism eruption (except for IPSL), even though from different reconstructions. The two MPI-ESM simulations differ by the magnitude of the change in the solar irradiance between the Little Ice Age and the present (0.1% in E1 instead of 0.25% in E2 and the other simulations) to better reflect the recent revised estimate by (Solanki, Usoskin, Kromer, Schussler, & Beer, 2004). A subset of simulations also includes the volcanic forcing or the evolution of land use (MPI-ESM and CNRM). In the MPI-ESM simulations the carbon cycle is interactive.

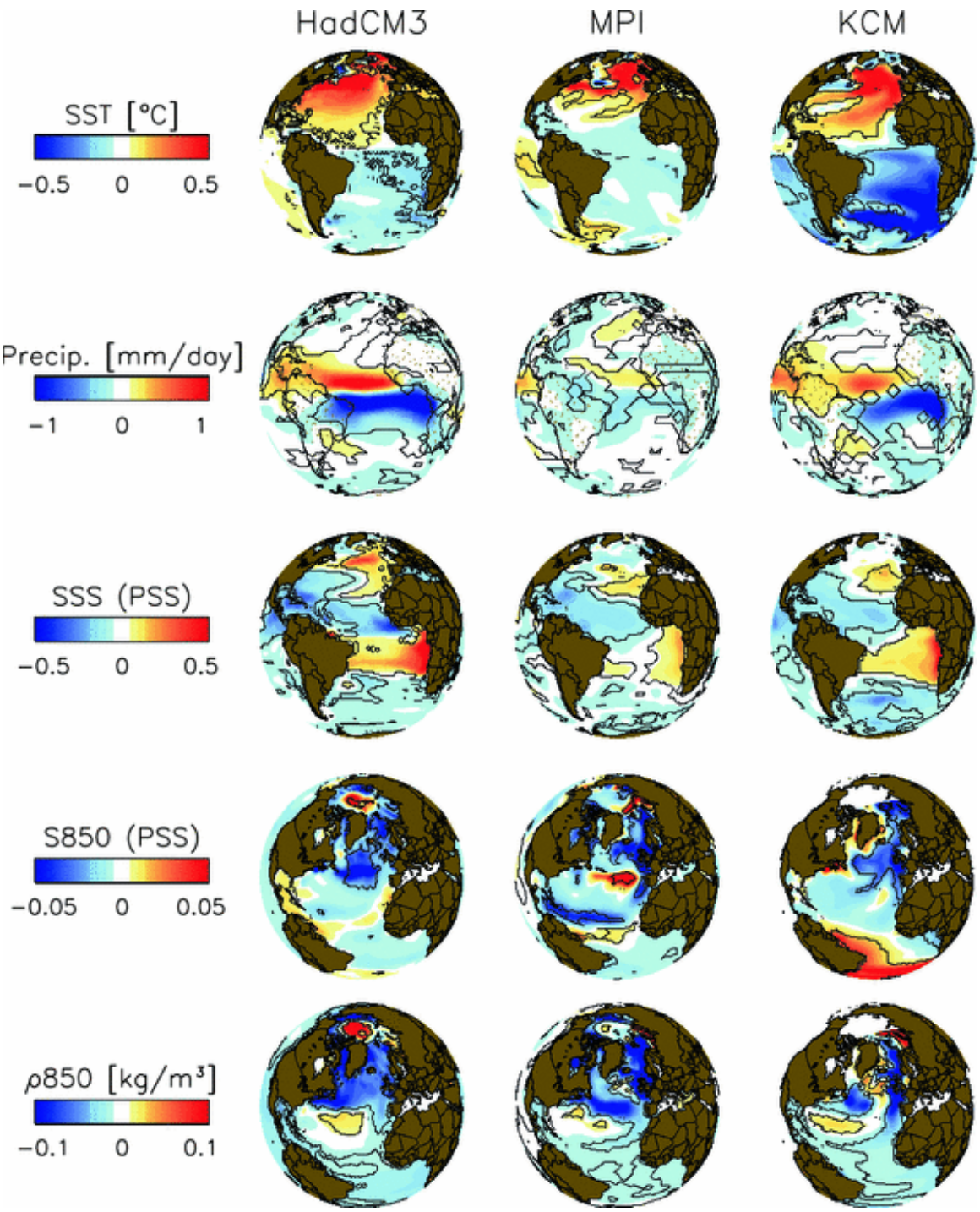


Figure 9.35: From top to bottom: SST composites using AMOC time series; precipitation composites using cross-equatorial SST difference time series; equatorial salinity composites using ITCZ-strength time series; subpolar-gyre depth-averaged salinity (top 800–1,000 m) using equatorial salinity time series; subpolar gyre depth averaged density using subpolar gyre depth averaged salinity time series. From left to right: HadCM3, MPI-ESM, and KCM. Black outlining signifies areas statistically significant at the 5% level for a two-tailed t test using the moving-blocks bootstrapping technique (Wilks, 1995) (Figure 3 from Menary et al. (2011)).

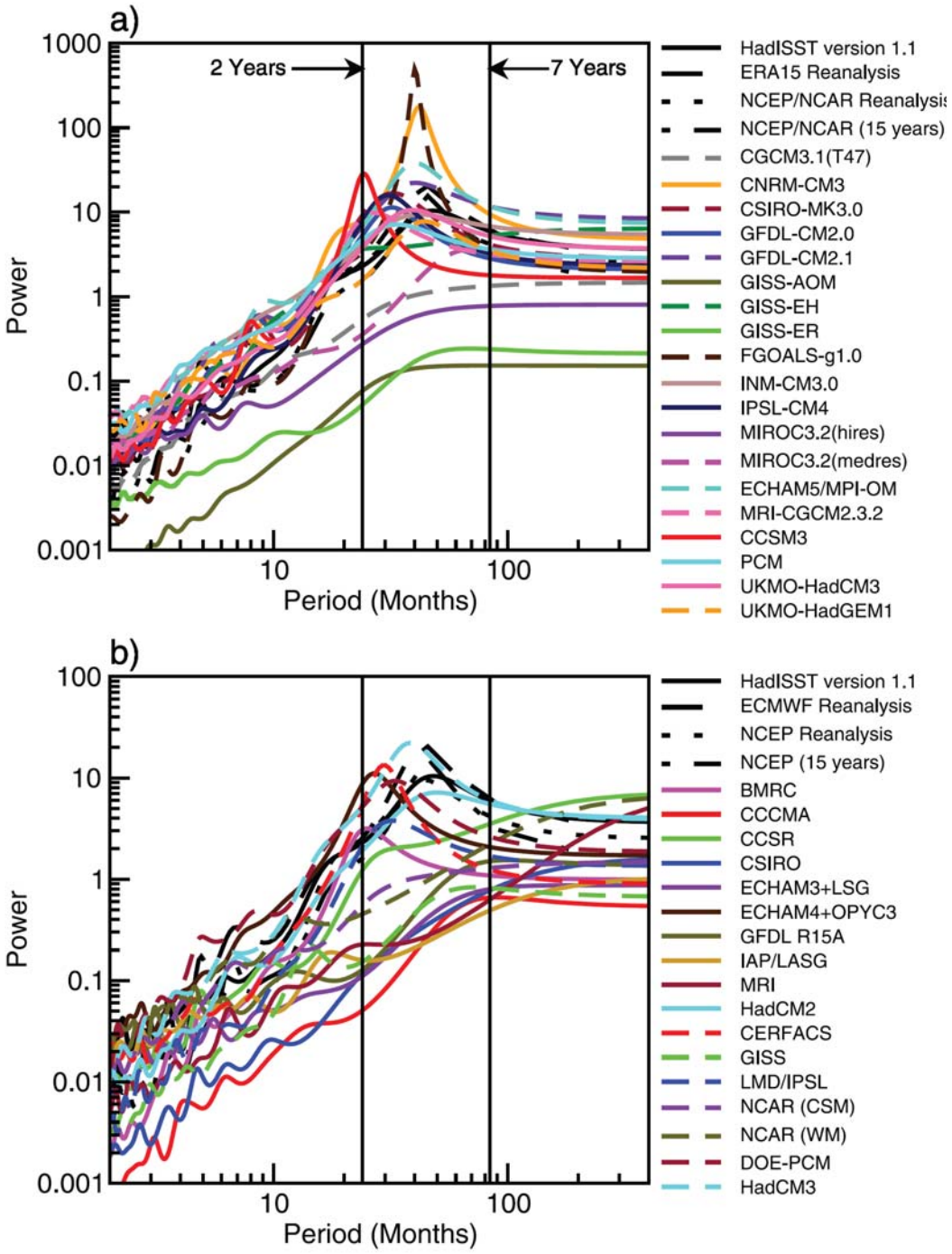


Figure 9.36: [PLACEHOLDER FOR SECOND ORDER DRAFT: Figure from AR4; to be updated.] Maximum entropy power spectra of surface air temperature averaged over the NINO3 region (i.e., 5°N to 5°S, 150°W to 90°W) for (a) the CMIP3 models and (b) the CMIP2 models. Note the differing scales on the vertical axes and that ECMWF reanalysis in (b) refers to the European Centre for Medium Range Weather Forecasts (ECMWF) 15-year reanalysis (ERA15) as in (a). The vertical lines correspond to periods of two and seven years. The power spectra from the reanalyses and for SST from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) version 1.1 data set are given by the series of solid, dashed and dotted black curves. Adapted from (AchutaRao & Sperber, 2006) and Sperber.

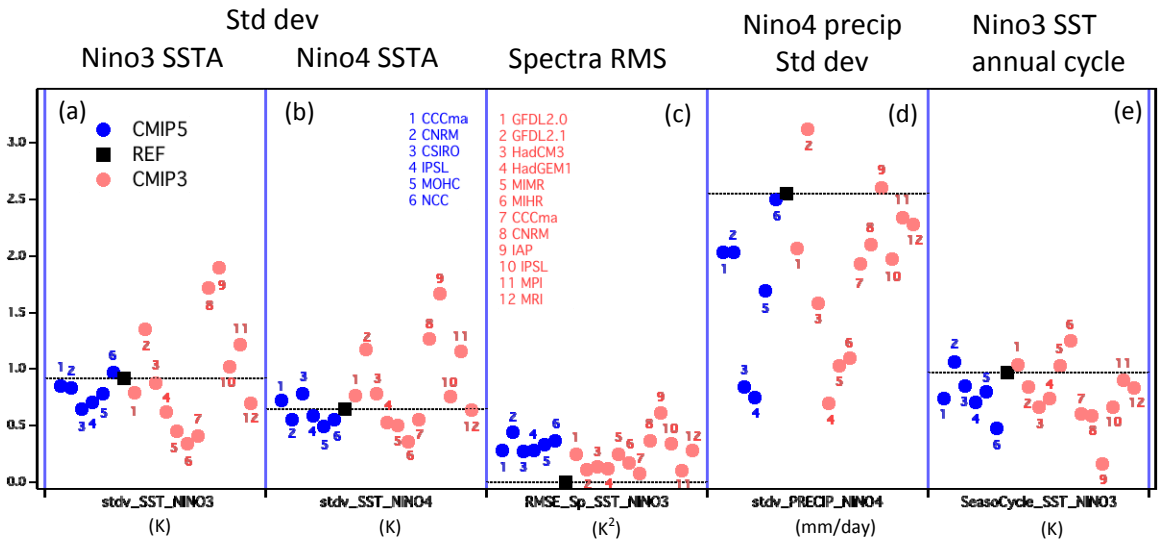


Figure 9.37: ENSO metrics comparing CMIP3 and CMIP5 [PLACHOLDER FOR SECOND ORDER DRAFT: To be updated with more CMIP5 results.]

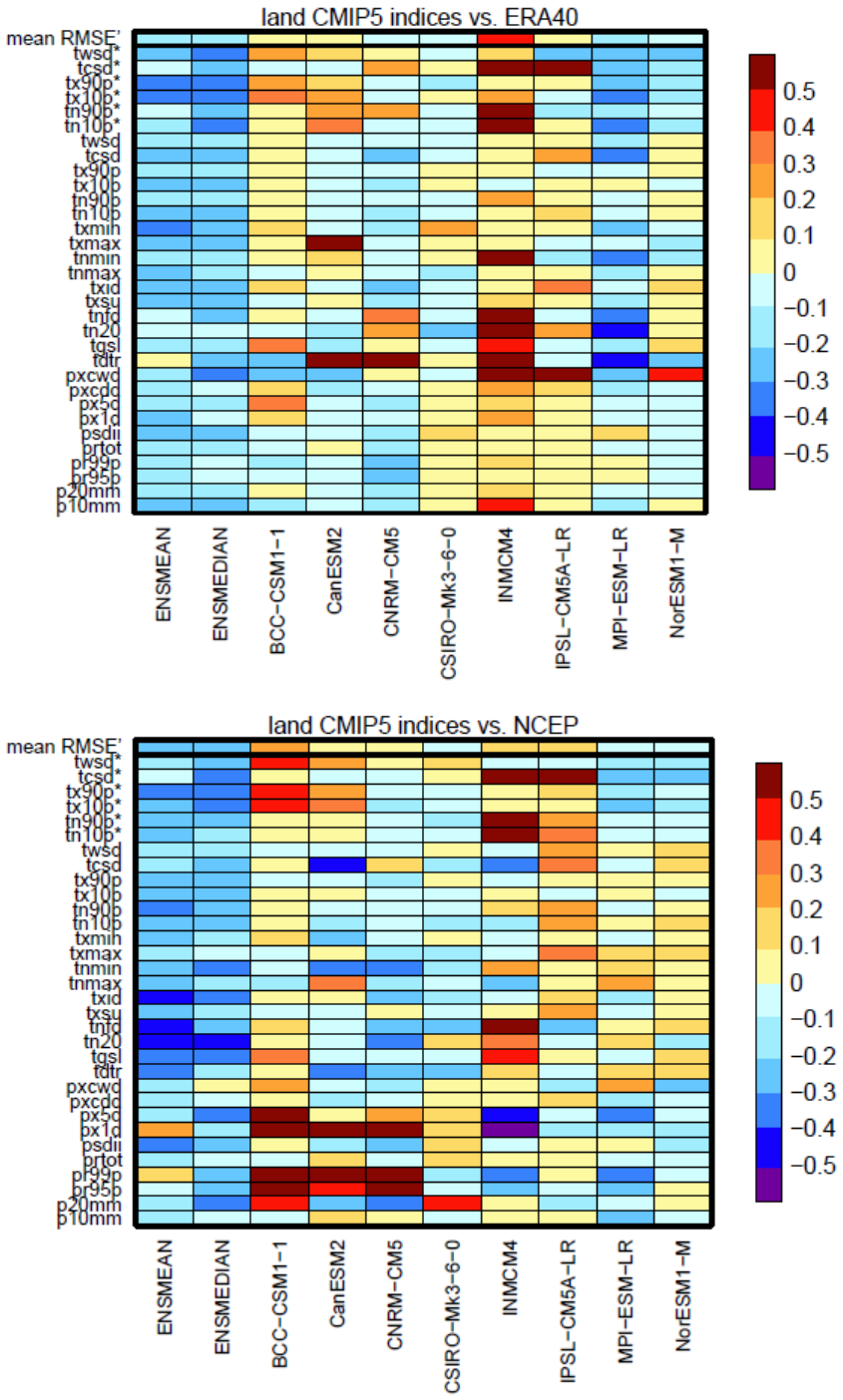


Figure 9.38: Portrait diagram display (Gleckler, Taylor, & Doutriaux, 2008) of relative error metrics for the CMIP5 temperature and precipitation indices compared to ERA40 (top) and NCEP (bottom) re-analyses for the base period 1961–1990. Only land areas are considered. Top row in each diagram indicates the mean RMSE across all indices for a particular model. The indices marked with a * are bias-corrected (indices are calculated using the bias-corrected minimum and maximum temperature time series but using the temperature thresholds estimated from the re-analysis, so that the bias in simulated temperature variability can be assessed). Indices shown are 'Warm/Cold spell duration' (twsd/tcsd), 'Warm/Cool days' (tx90p/tx10p), 'Warm/Cool nights' (tn90p/tn10p), 'Min/Max T2MAX' (txmin/txmax), 'Min/Max T2MIN' (tnmin/tnmax), 'Ice/Summer/Frost days' (txid/txsu/tnfd), 'Tropical nights' (tn20), 'Growing season length' (tgsl), 'Diurnal temperature range' (tdtr), 'Consecutive wet/dry days' (pxcwd/pxcdd), 'Max 5-day/1-day precipitation amount' (px5d/px1d), 'Simple daily intensity index' (psdii), 'Annual total wet-day precipitation' (prtot), 'Extremely/Very wet days' (pr99p/pr95p) and 'Number of very heavy/heavy precipitation days' (p20mm/p10mm) from Klein Tank et al. (2009). (Haylock et al., 2008).

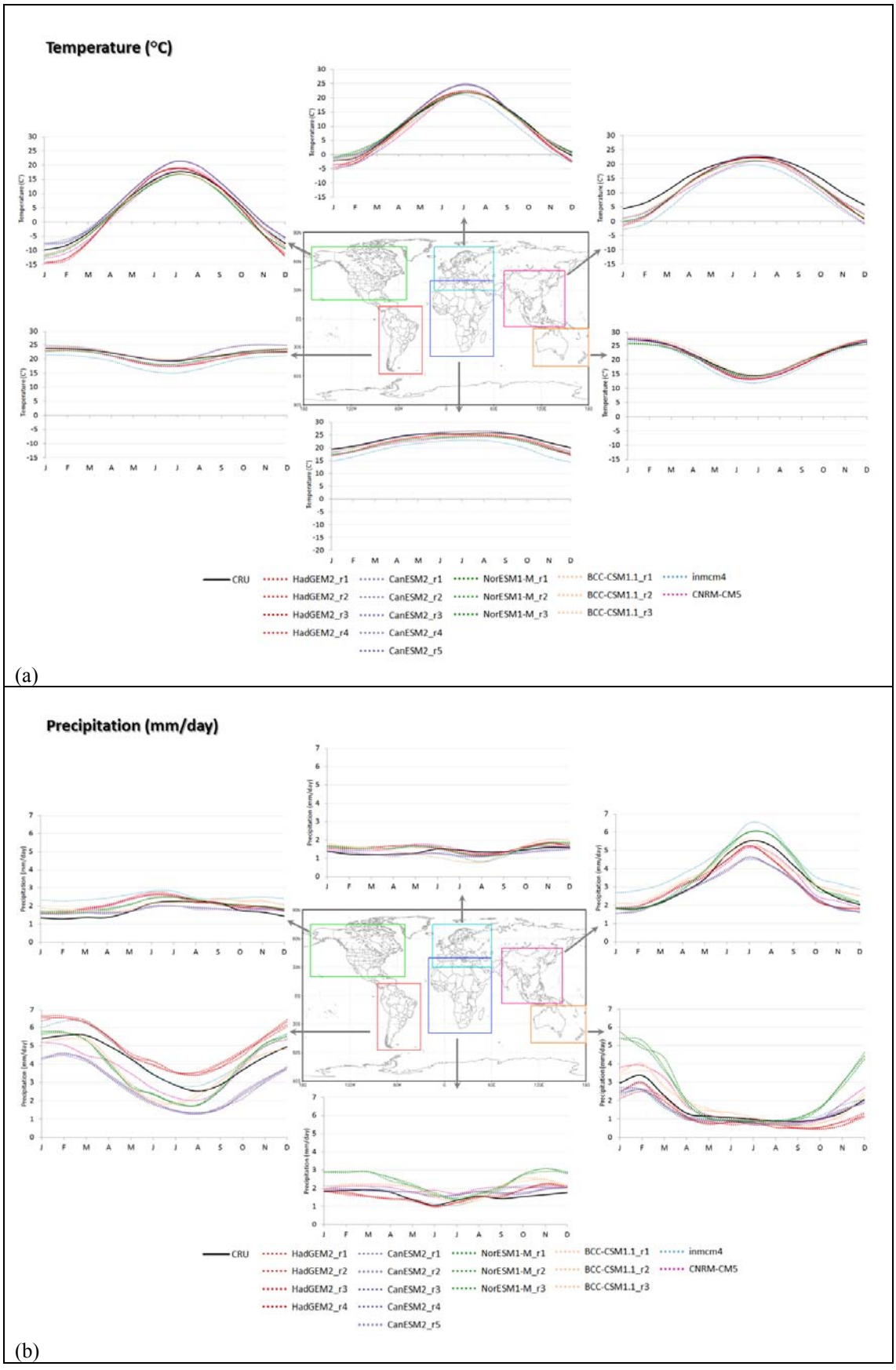


Figure 9.39: Mean annual cycle of temperature (a) and precipitation (b) from CMIP5 GCM (dotted lines) historical runs and CRU (solid lines) data for the indicated areas. Average is taken over land points over the period 1979 to 2005. Units are mm/day for precipitation and °C for temperature.

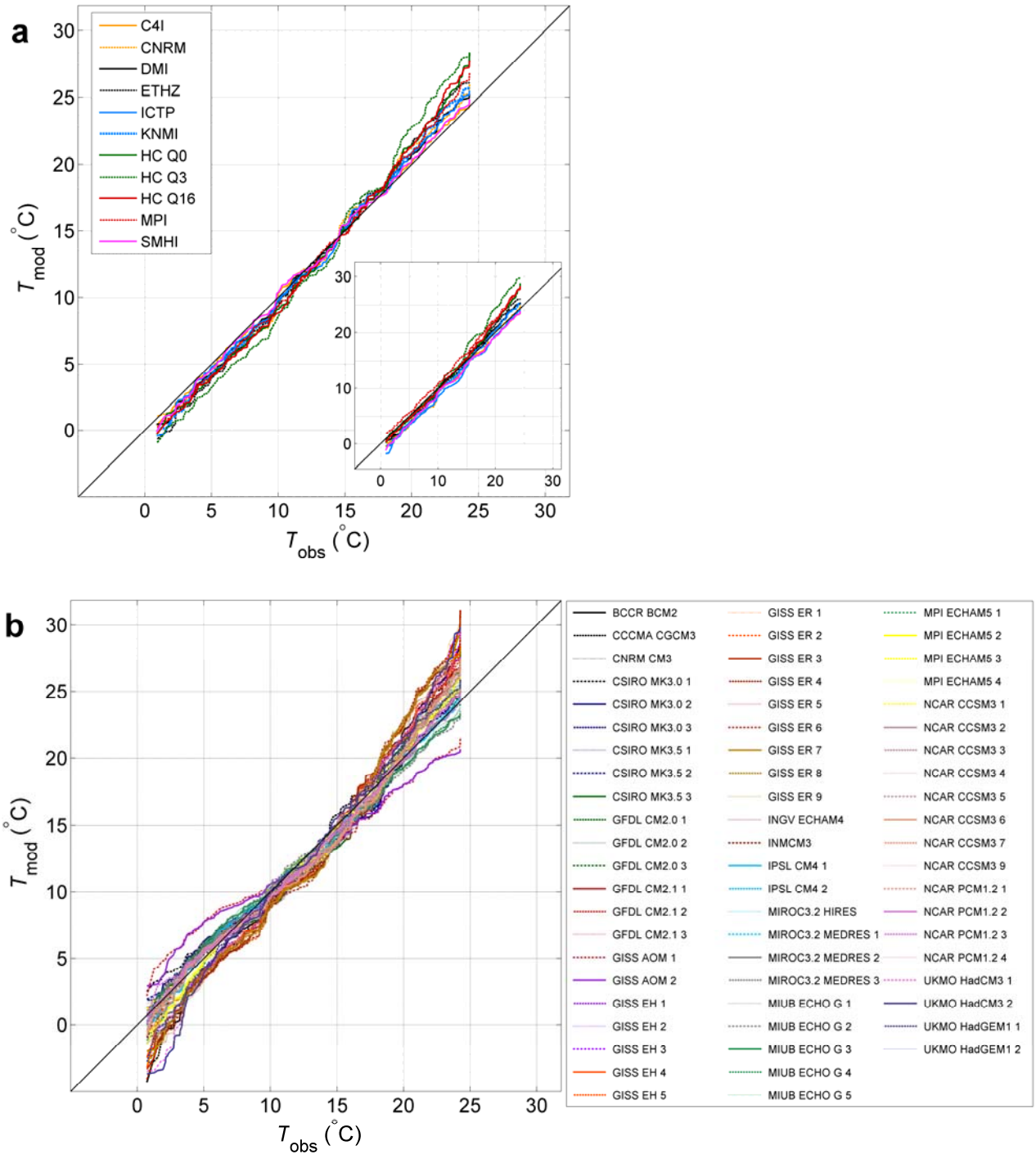


Figure 9.40: Ranked modelled versus observed monthly mean temperature for a Mediterranean subregion, for the 1961–2000 period. The RCM data (panel a) are from Christensen et al. (2008) and are adjusted to get a zero mean in model temperature with respect to the diagonal. The GCM data in panel b are from CMIP3 and adjusted to get a zero mean in model temperature with respect to the diagonal. Figure after Boberg and Christensen (2011).

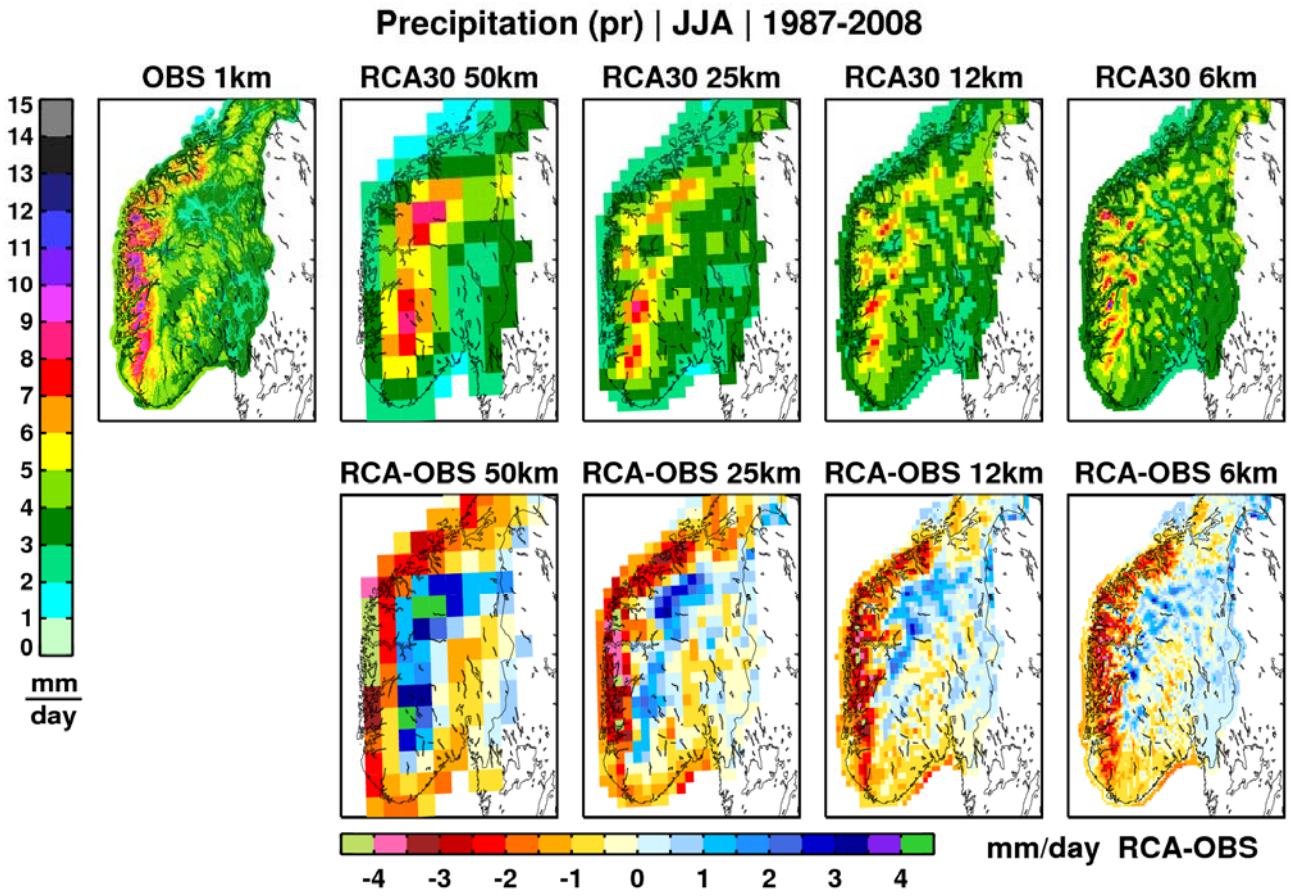


Figure 9.41: Summer seasonal mean (JJA, 1987–2008) in Southern Norway gridded observational precipitation with 1 km resolution from Met.no and RCM-simulated precipitation with boundary conditions from the ERA40 reanalysis and ECMWF operational analysis (top row). The RCM has been run at four different resolutions ranging from 50 to 6 km. Differences between the simulated precipitation and the gridded observations aggregated from 1 km to respectively 50, 25, 12 and 6 km grids are shown in the bottom row. The model runs are from Walther et al. (2011).

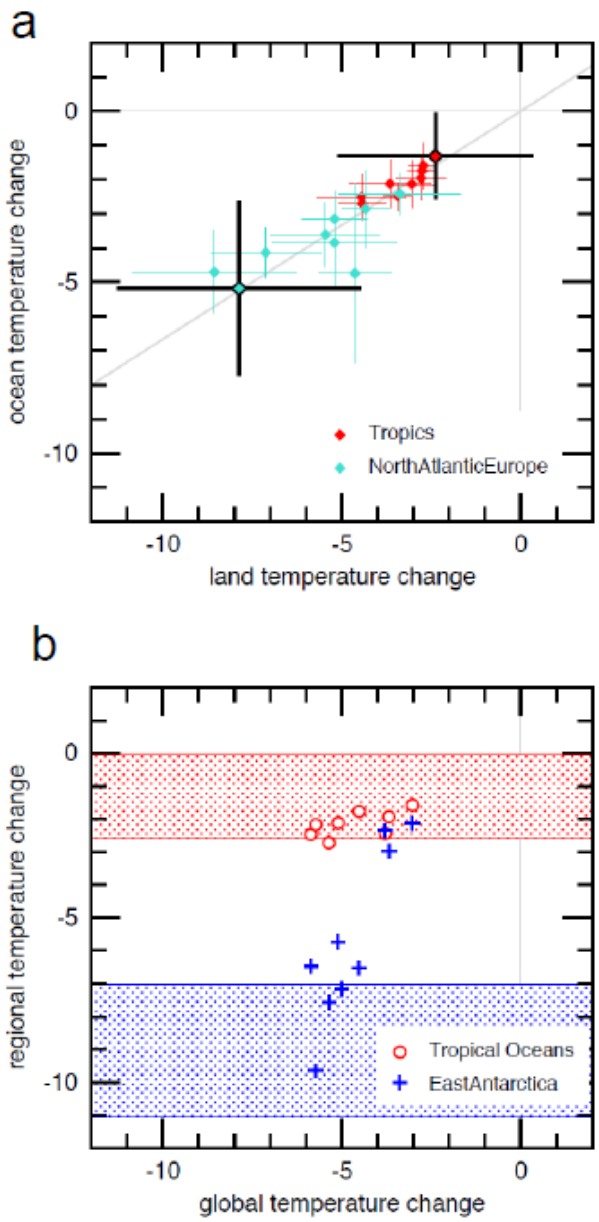


Figure 9.42: (a) Comparison of simulated and observed changes in annual mean temperature, LGM compared to modern and ocean compared to land, (b) simulated relationship between regional cooling in the tropics and over eastern Antarctica and global cooling. This figure is adapted from (Crucifix, 2006; Cunningham et al., 2009; Kageyama et al., 2006; Masson-Delmotte et al., 2006; Otto-Bliesner et al., 2009) using the PMIP2 dataset (Braconnot et al., 2007). In a) the colour dots represent the different model results for the two regions, and the large crosses the estimates for the ocean and land surface data from (Waelbroeck, et al., 2009) and (Bartlein, et al., 2010) respectively. In b) the different points represent different model results and the hatched bars the estimates with error bars for the (Waelbroeck, et al., 2009) SST reconstruction and East Antarctica air temperature reconstruction from ice cores (Masson-Delmotte, et al., 2006).

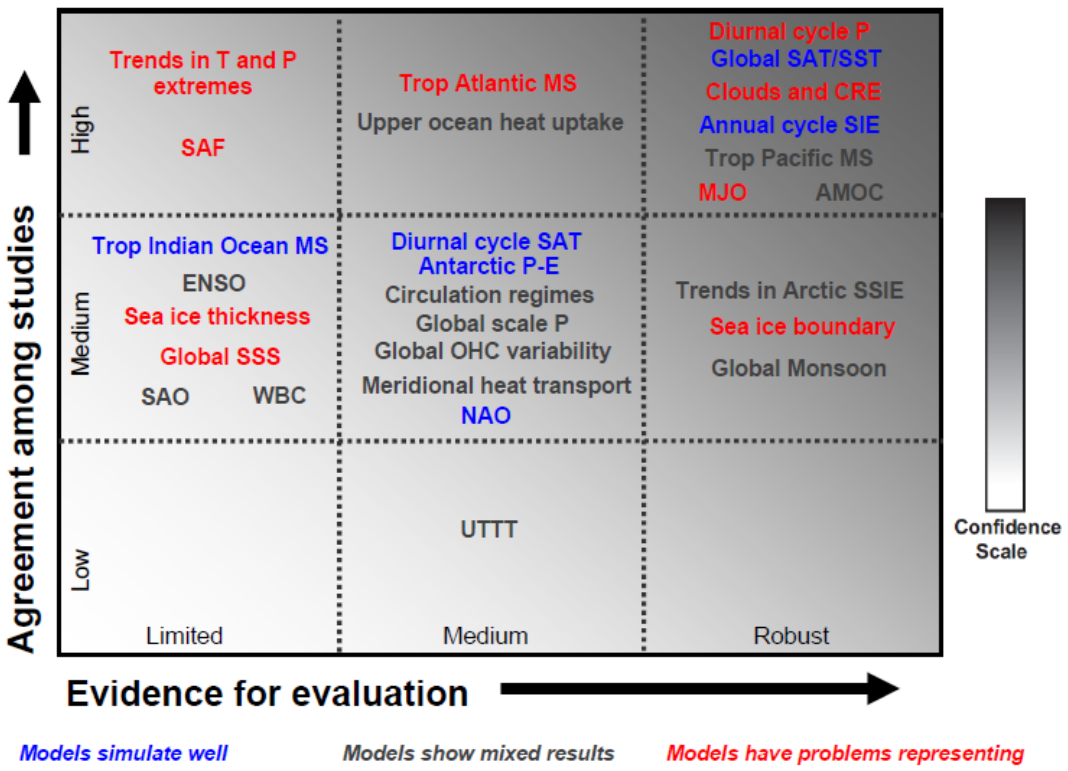


Figure 9.43: Summary of the findings of Chapter 9 with respect to how well the CMIP3 models simulate important features of the climate of the 20th century [PLACEHOLDER FOR SECOND ORDER DRAFT: Will be updated with CMIP5 models]. Confidence in the assessment increases towards the top-right corner as suggested by the increasing strength of shading. Features that current state-of-the-art AOGCMS and ESMs simulate well, show mixed results, or have problems representing are shown in blue, grey, and red, respectively. The figure highlights the following key features (subject to revisions), with the sections that back up the assessment added in brackets:

Annual cycle SIE: Annual cycle Arctic and Antarctic Sea Ice Extent (Section 9.4.3)
AMOC: Atlantic Meridional Overturning Circulation (Section 9.4.2.6)
Antarctic P-E: Antarctic Precipitation minus Evaporation (Section 9.6.2)
Circulation regimes: Blocking events and others circulation regimes (Section 9.5.2.2)
Clouds and CRE: Clouds and Cloud Radiative Effects (Section 9.4.1)
ENSO: El Niño Southern Oscillation (Section 9.5.3.7)
Global Monsoon: see Section 9.4.2
Global Scale P: Global scale precipitation (Section 9.4.1)
Meridional heat transport: see Section 9.4.2.6
MJO: Madden Julian Oscillation (Section 9.5.2.2)
NAO: Northern Annual Mode (Section 9.5.3.7)
OHC: Ocean Heat Content (Section 9.4.2)
SAF: Snow albedo feedback (Sections 9.8.3)
SAO: Southern Annual Mode (Section 9.5.3.7)
SAT: Surface Air Temperature (Section 9.4.1)
SIE: Sea Ice Extent (Sections 9.4.3 and 9.8.3)
SSIE: September SIE (Sections 9.4.3 and 9.8.3)
SSS: Sea Surface Salinity (Section 9.4.2.6)
SST: Sea Surface Temperature (Sections 9.4.1 and 9.4.2.6)
Trends in T and P Extremes: Trends in temperature and precipitation extremes (Section 9.5.4)
Trop Atlantic / Pacific MS: Tropical Atlantic / Pacific Mean State (Section 9.4.2)
Trop Indian Ocean MS: Tropical Indian Ocean Mean State (Section 9.4.2.6)
Upper ocean heat uptake: see Section 9.4.2
UTTT: Upper tropospheric temperature trends (Section 9.4.1)
WBC: Western Boundary Current (Section 9.4.2.6)

[PLACEHOLDER FOR SECOND ORDER DRAFT: This figure is a preliminary version subject to revisions; it will be updated with more CMIP5 results and will possibly be separated into different panels that show: Panel 1: Climatologies and trends; Panel 2: Variability including extremes; Panel 3: Regional performance.]

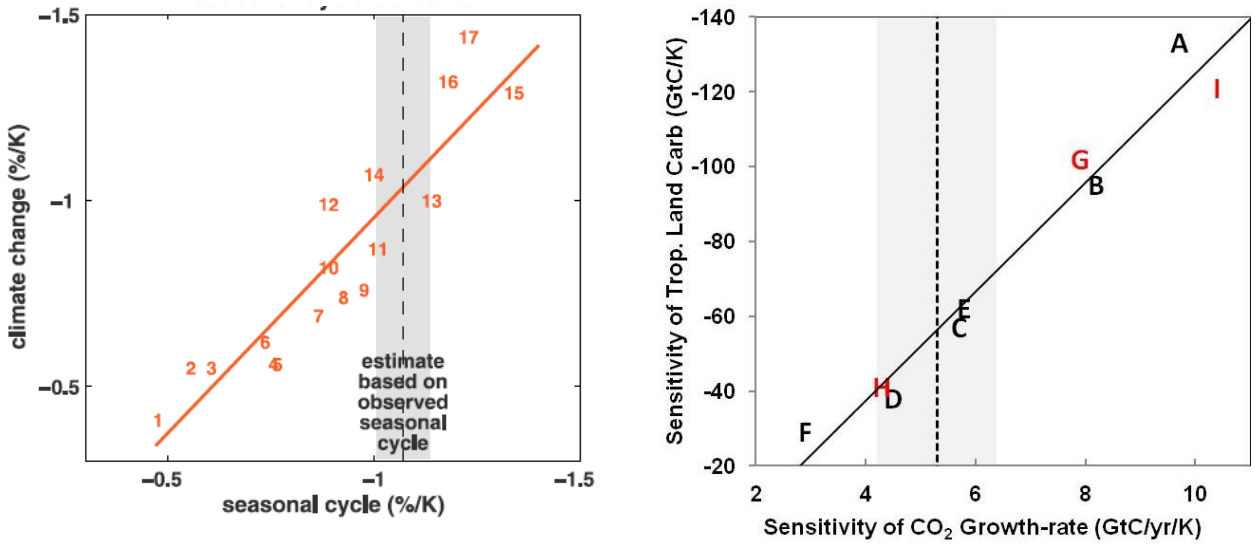


Figure 9.44: *Left:* Scatter plot of simulated springtime snow albedo feedback ($\Delta\alpha_s/\Delta T_s$) values in climate change (ordinate) versus simulated springtime $\Delta\alpha_s/\Delta T_s$ values in the seasonal cycle (abscissa) in transient climate change experiments with 17 AOGCMs from CMIP3 (α_s and T_s are surface albedo and surface air temperature, respectively). From Hall and Qu (2006); [update with CMIP5 data; show CMIP3 in different colour in addition]. *Right:* Constraint on the climate sensitivity of land carbon in the Tropics (30°N-30°S) from interannual variability in the growth-rate of global atmospheric CO_2. This version is based on C^4MIP GCMs (black labels), and three land carbon "physics ensembles" with HadCM3 (red labels). The y-axis is calculated over the period 1960-2099 inclusive, and the y-axis is calculated over the period 1960-2010 inclusive. In both cases the temperature used is the mean (land+ocean) temperature over 30°N-30°S. The vertical grey band shows the estimated sensitivity of the observed global CO_2 growth-rate to the observed tropical mean temperature.

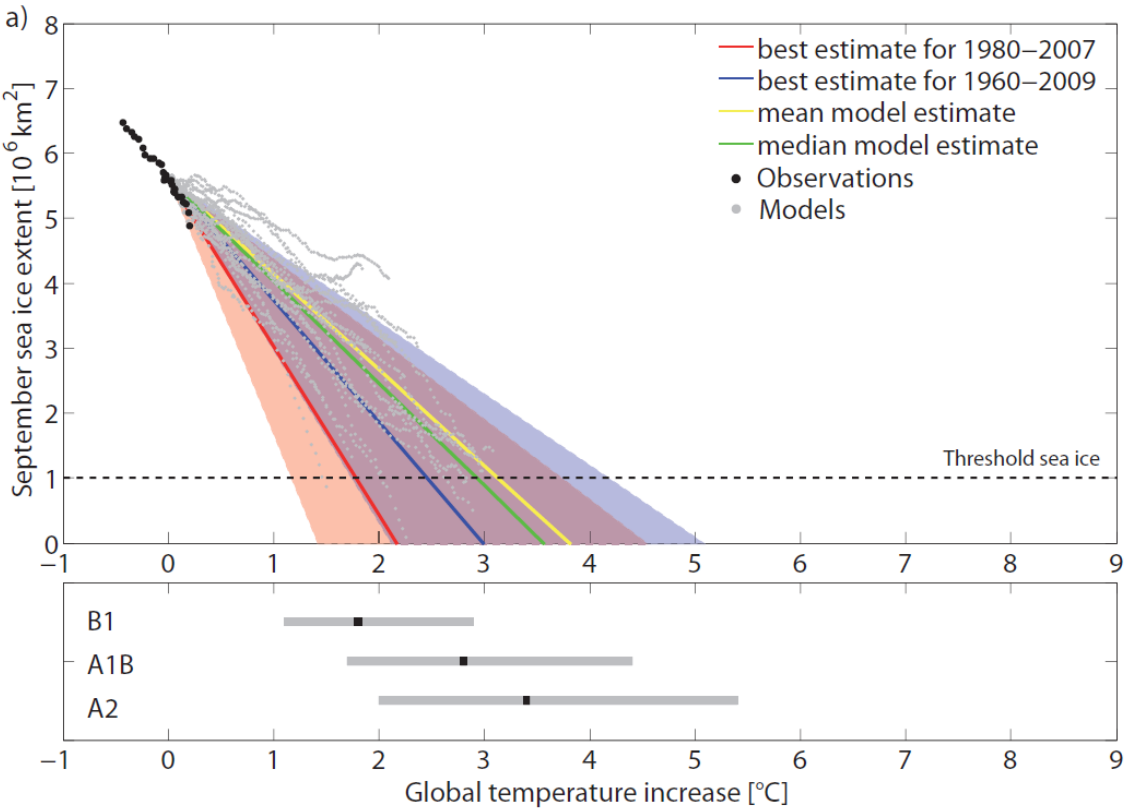
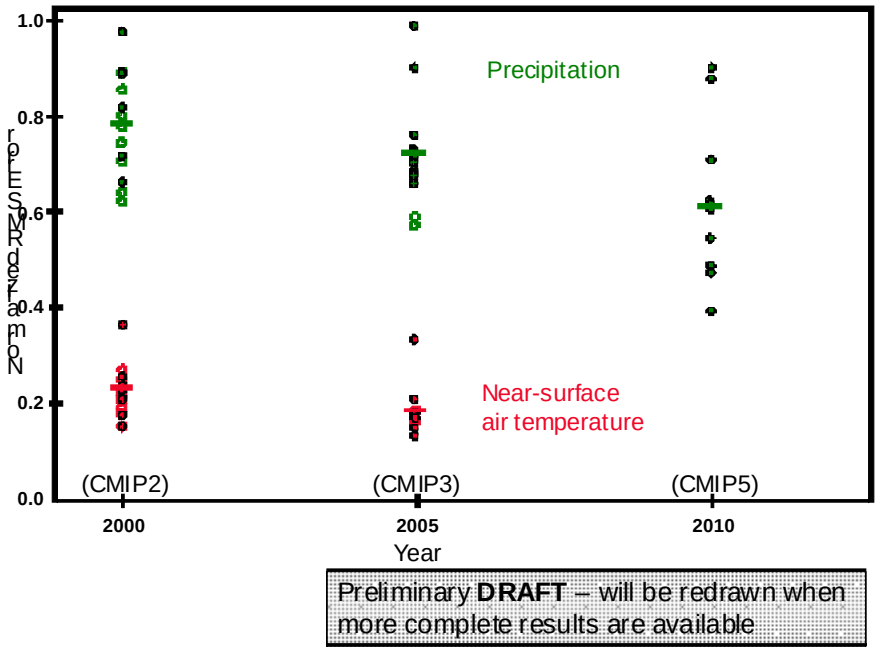


Figure 9.45: Projected decline of Arctic sea ice area with increasing global temperature. Shaded areas depict the uncertainty range (red based on observations from 1980-2007 and blue from 1960-2010). The time period in the legend indicates the time window that is used to estimate polar amplification. The models are calibrated to start at the current observational point (1980-2007) and show points for sea ice larger than 1.0 million km^2. Warming in 2090-2099 and associated uncertainties for three SRES non-intervention emission scenarios from (IPCC, 2007) are indicated at the bottom. From (Mahlstein & Knutti, 2011).



FAQ 9.1, Figure 1: Quantitative examination of model skill as measured in the three recent phases of CMIP (CMIP2, CMIP3 and CMIP5). The RMSE is normalized in each case by the observational standard deviation to facilitate comparison across variables. Results are shown for global precipitation and surface air temperature. [PLACEHOLDER FOR SECOND ORDER DRAFT: This figure very preliminary; will be updated as additional CMIP5 simulations become available. Additional fields may also be included in future renditions.] Redrafted from (P. Gleckler, K. Taylor, and C. Doutriaux, 2008) and updated with CMIP5 results.