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: Johanna Baehr, Alejandro Bodas-Salcedo, Bode Gbobaniyi, Stephen Griffies, Elizabeth Hunke, Tatiana Ilyina, Stephen A. Klein, Reto Knutti, Felix Landerer, Florian Rauser, Mark Rodwell, Adam A. Scaife, John Scinocca, Hideo Shiogama, Ken Sperber, Bjorn Stevens, Keith Williams, Tim Woollings

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

Date of Draft: 15 April 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, with more traditional presentation of 'error maps' included in the supplementary material.

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 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 with particular characteristics or process parameterisations in a model. This necessarily requires careful and extensive documentation of each model, something that has also improved since the AR4.

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

- Specific bullet points will be included here based on analysis that will be presented in the Chapter. At this point, we include only a placeholder, as results from the CMIP5 models are not yet available. The bullet points will note particular areas in which modelling capability has improved since the AR4, and of course areas in which improvements have stagnated, or areas in which significant errors/biases remain.

The Executive Summary will conclude with a few paragraphs assessing the overall state/capability of climate models, a summary of findings related to the sources of model errors and uncertainties, and an assessment of the 'credibility' of climate models for the applications that follow in Chapters 10-14. This will be written once the CMIP5 model results are available and evaluated.

9.1 Climate Models and their Characteristics

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.1.1 Introduction

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 quantitative 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 here will be particularly on the models whose results will be used in 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., 2007); (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) will also be 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 methods, in Section 9.2, we will evaluate recent and longer-term records as simulated by global models in Section 9.3, variability and extremes in Section 9.4, downscaling and regional scale climate simulation in Section 9.5 and conclude with a discussion of the sources of model errors in Section 9.6 and the relation between model performance and the credibility of future climate projections in Section 9.7.

9.1.2 Overview of Models to be Evaluated

The models used in climate research range from simple energy balance models to full complexity 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., 2006c; Held, 2005). Applications include simulating historical climate, predicting near-term climate change on seasonal to decadal time scales, making projections of future climate change over the coming 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 Earth System Models (ESMs)

These are the current state of the art climate models in the AR5, in terms of the extent to which the overall climate 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, stratospheric ozone, etc. 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 about one month for a 100-year simulation.

9.1.2.2 Atmosphere-Ocean General Circulation Models (AOGCMs)

These models were the "standard" climate models 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 make projections from future pathways of concentrations. These models are still extensively used, and in particular are run (often at higher resolution) for seasonal to decadal climate prediction applications in which biogeochemical feedbacks are less important than they are in century-scale projections.

9.1.2.3 Earth-System Models of Intermediate Complexity (EMICs)

These models 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 climate system components not yet included in 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)

These models focus on a particular geographical region, usually employing a limited area grid driven at the boundaries by output from a global climate model or reanalysis system (Rummukainen, 2010; Wang et al., 2004). Many RCMs include an atmospheric component and a land surface component, with SST and sea ice as boundary conditions, although some include interactive oceans, sea-ice, and/or vegetation components (e.g., Dorn et al., 2009; Doscher et al., 2010). Typically, the representations of climate processes are comparable those in AOGCMs and ESMs (and in many cases share much of the same model code). The typical application of an RCM is to 'downscale' global model projections providing higher-resolution regional-scale information for impact studies.

[PLACEHOLDER FOR FIRST ORDER DRAFT: Refer to model complexity figure in Chapter 1]

9.1.3 The Path to Model Improvement

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.1.3.1 The Model Development Process

The models evaluated in this chapter consist of many components (atmosphere, ocean, ice, land surface, etc.) and, as highlighted earlier, some of them include descriptions of biogeochemical processes. All comprehensive climate models are founded on well-known physical laws and basic principles (e.g., energy and momentum conservation). The two principal steps in developing a climate model are:

i) Find the mathematical expressions for the physical laws as they apply to the system to be modelled. This often requires theoretical work for instance in deriving and simplifying the mathematical expressions that best describe the system;

ii) Implement the mathematical expressions derived in i) on a computer. This requires the development of numerical methods that allow the efficient solution of the mathematical expressions for each subsystem. Most models implement these numerical methods on some form of grid (both horizontal and vertical), the best known of which is the latitude-longitude-height grid.

The climate system is a very complex natural systems and modelling it using the approach above requires significant supercomputing resources. Hence, additional constraints due to limitations in computing power further influence the two basic steps above. Despite great progress in supercomputing even the most powerful computers today (and in the foreseeable future) require compromises to be made when designing a climate model. These important and competing constraints cover three main areas:

i) The numerical implementation described above allows for a choice of grid-spacing in the model components, often referred to as "model resolution" (see Section 9.1.3.4). Higher model resolution, i.e., smaller grid spacing, generally leads to more accurate models but also leads to higher computational cost. Model resolutions affordable today imply that certain processes important to climate, such as cloud processes (see Chapter 7), are excluded from the numerical solutions and have to be represented through simple conceptual models usually referred to as parameterisations.

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 in a model must be made, with each increase in complexity usually accompanied by an increase in computational cost.

iii) The climate system is highly non-linear and as a consequence often shows chaotic behaviour. A single model simulation therefore only represents one of the possible pathways the climate system might follow and it is necessary to carry out a number of simulations either with several models or by performing an ensemble of simulations with an individual model (each ensemble member having been perturbed in some way) to allow some evaluation of uncertainties. It is self-evident that increasing the number of simulations through an ensemble approach increases the computational cost.

As a consequence of the compromises required in modelling the climate system, trade-offs are introduced depending on the aim of the study. This naturally leads to the several classes of models as introduced in Section 9.1.2.

As the discussion above highlights, the development of climate models takes place at three distinct but strongly connected levels that are illustrated in Figure 9.1. Having developed the numerical description of one of the model components (e.g., atmosphere, ocean, etc.), it is necessary to develop descriptions of all the important processes that are excluded from the model by the necessary choice of model resolution (Level 1 in Figure 9.1). This process of parameterisation has become very complex (Jakob, 2010) and is often carried out by isolating the process of interest and developing the best possible conceptual model of it, constrained by available observations, computational resources and current knowledge (e.g., (Randall et al., 2007)). The values of any adjustable parameters contained in the conceptual models are usually chosen, within the bounds of our knowledge, to ensure the best possible process representation - a process often referred to as model tuning (See "Tuning" box – Box 9.1).

[START BOX 9.1 HERE]

Box 9.1: Potential Box on Model Tuning [joint with Chapter 7]

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 physical principles, and a non-fundamental component, whereby important processes are described as best one can (McWilliams, 2007). This mix of fundamental and non-fundamental components in the description of the climate system makes climate models interesting from an epistemological perspective (Oreskes et al., 1994; Petersen, 2000; Stevens and Lenschow, 2001; Mueller, 2010), and underlies much of their imprecision (McWilliams, 2007).

As an example, the primitive equations used by most comprehensive climate models provide a fundamental description of the fluid mechanics of the atmosphere and ocean, at least for scales of motions that most climate models choose to represent. The primitive equations are a well-defined limit of the Navier-Stokes equations, a limit whose mathematical properties (integrability) are actually better understood than the Navier-Stokes equations themselves (Cao and Titi, 2007). In contrast, smaller scale or non fluid-dynamical processes, either because their degrees of freedom are too numerous to be individually represented or because their physics is incompletely known, comprise much of the non-fundamental, and by definition uncertain, component of a comprehensive climate model. However, even for the fundamental components of the system, uncertainty arises through discretisation of the equations (e.g., Oden and Prudhomme, 2002).

The construction of a comprehensive climate model involves making choices concerning the discretisation;

which component processes to incorporate; the basic form of the equations describing these processes; and the parameter values that give these descriptions their specific form. Usually model tuning is identified only with the final step of model development, that being the choice of adjustable parameters (Randall and Wielicki, 1997).

Tuning is thus inexorably connected to judgments as to what constitutes a skillful 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. Similarly, global observational constraints on quantities like the amount of water in the atmosphere, its distribution among its different phases, total rainfall, and the partitioning of the top of the atmosphere energy budget into its shortwave and longwave components, or clear and cloudy contributions, also guide the development of simulations and their presentation, as plausible representations of the climate system (Roeckner et al., 2006; Donner et al., 2011 (in press); Collins et al., 2006b). Such considerations also apply on other components of Earth System models such as atmospheric chemistry or the carbon cycle.

As our knowledge of the past and current state of the climate system increases, this knowledge is incorporated into the process of model tuning. For instance, given an equally good representation of the mean state, one's choice of convection scheme may be guided by the ability of different schemes to support realistic variability in tropical circulations associated with phenomena such as El Niño, the Madden Julian Oscillation, or the seasonal migration of the inter-tropical convergence zones. Momentum transport by gravity waves may be adjusted, either through changes to parameterizations of unresolved gravity waves, or one's choice of model resolution or even representations of topography, to better represent circulations in the stratosphere, or the position of the storm tracks, or standing waves in the mid-latitudes. Ocean mixing parameters may be adjusted to better represent the thermal structure of the equatorial ocean or the strength of the meridional overturning circulation in the Atlantic. The degree to which one incorporates specific processes may be guided by the ability of a model system to represent the observed warming of the 20th century (e.g., Kiehl, 2007), or preconceptions about the potency of particular effects (e.g., Hoose et al., 2009). Model development and tuning are hence guided by an awareness of deficiencies in the simulation of current and past climate states, an awareness that fundamentally complicates the *a posteriori* evaluation of models.

Our lack of process understanding means that much model tuning is to some extent *ad hoc*. While tuning parameters are usually associated with processes that regulate the flow of energy, moisture or momentum through the climate system, the exact parameters differ among models and, with a few notable exceptions, do not generally correspond to observable quantities. Even in the cases when tuning parameters are associated with observational parameters, they are often undetermined by these quantities, in part because they often arise from effective descriptions of processes. As example consider a common tuning parameter for establishing the amount of condensed water in the atmosphere, namely the rate at which a distribution of stratiform clouds produce precipitation, as a function of the low order moments of the distribution of condensed cloud water. While the rate at which raindrops are formed from spatially homogeneous distributions of cloud droplets is reasonably well known, the effective rate for inhomogeneous distributions, particularly when they interact, is not well constrained experimentally or observationally. Concerning ice microphysical processes, large gaps in our understanding remain even for the case of homogeneous distributions of condensate. This lack of understanding on the level of specific processes means that parameterizations are developed, or parameters are chosen, from a family of plausible solutions largely through trial and error, with the success of particular trials determined based on the overall performance of the simulation. This fact increasingly motivates research at the process level, whereby fine-scale models are used to help bridge the gap between process level understanding, and the effective descriptions of processes that one incorporates in climate models (e.g., Randall et al., 2003a; Eden and Greatbatch, 2008), thereby limiting at least the family of plausible choices.

The introduction of a new parameterization or a new choice of parameters does not necessarily improve all important aspects of a simulation. It is quite common that certain aspects of a simulation are improved at the costs of others. For instance, the representation of the meridional overturning circulation in the Atlantic may be improved by one parameter choice, but the fidelity of the equatorial sea surface temperatures may be degraded (Mauritsen, 2011 (in preparation)). Similarly models that are tuned to well represent inter-seasonal

variability in tropical convection often do so by degrading the representation of the mean state (Kim et al., 2011 (submitted)). Counterexample: simulation of blockings and mean model biases (Scaife et al., 2010).

Likewise deficiencies in the representation of some processes may be tolerated because they compensate deficiencies in the representation of other processes. For instance unrealistically low ocean albedos at high latitudes have been tolerated in the past because ice-melt ponds were not incorporated in the description of sea ice, or excessive absorption of UV radiation has been tolerated because it was known to compensate for deficiencies in ozone climatologies (Mauritsen, 2011 (in preparation)). When compensating effects are understood, it is only a matter of time and resources before they are addressed. When they are not understood, however, it is difficult to implement improvements to component processes, because in such cases the improvement of one process can lead to an overall degradation in the quality of the simulation.

In summary, model tuning is inseparable from model development and arises naturally from the fact that climate simulations depend to a significant degree on the uncertain representation of poorly understood processes. Model tuning can be seen as the process by which our understanding of the current and present state of the climate system is incorporated into the development of climate models. These facts influence the use and evaluation of climate models and have led some to question whether climate models are a reliable source of knowledge for those aspects of the system toward which they were not tuned, such as anthropogenic climate change. But through the course of climate model development it has proven impossible to develop and tune equally plausible models that do not produce significant warming under increasing greenhouse gas concentrations, a fact that underlines the broad consensus behind the climate projections presented in all IPCC reports.

[END BOX 9.1 HERE]

Once descriptions for all the important processes have been found, single model components are assembled and evaluated individually (Level 2). For instance, the atmospheric component is evaluated by prescribing sea surface temperature (Gates et al., 1999) or the ocean and land components are evaluated by prescribing the atmospheric input (Barnier et al., 2006; Griffies et al., 2009a). At this stage of model development, it might be necessary to alter some of the parameters in the process representation within their uncertainty bounds to fulfil well-known global constraints in the individual components, a second phase of model tuning (See “Tuning” box, Box 9.1). Usually the parameter adjustments made at this stage are much smaller than those made initially when designing the process description.

The final step of model development is the integration of all components into the full climate model and the evaluation of that model as a whole (Level 3 in Figure 9.1). A faithful representation (which is application- and metrics-dependent) of the climate system in these coupled models is the ultimate goal of climate modelling. As the coupling of the various components may lead to imbalances in the now significantly more complex model, a final small parameter adjustment may become necessary and is carried out at the time of coupling (See “Tuning” box, Box 9.1). The choice of which components to include derives from the purpose of the model simulations as well as the state of our scientific knowledge. The number of model components has increased steadily from the FAR to this report as scientific knowledge about the climate system and the maturity of component models has increased.

[INSERT FIGURE 9.1 HERE]
Figure 9.1: A simple figure that illustrates the three levels of model components and that highlights the different components of Atmosphere Ocean General Circulation Models (AOGCMs) and Earth System Models (ESMs).

The process of building a comprehensive climate model is very complex and time-consuming and typically requires several years and dozens of scientists and software engineers. The increased demand on such highly skilled human resources has led to the development of community shared software infrastructures to combine model components or standardise model outputs and archival (Collins et al., 2005; Valcke et al., 2006). The increasing modular nature of process-descriptions and model components leads to a sharing and exchange of model components in the climate modelling community. The result is that several of the models evaluated here contain similar or even identical components. For instance [X] of the [Y] models used here

employ the same ocean component [ZZZ] albeit with slightly different settings. This sharing of components has consequences for evaluating the ensemble of the CMIP models, as not all of them can be considered entirely independent (see Section 9.7). Table 9.1 provides an overview over the models used here, including an indication where model components are similar between models.

[INSERT TABLE 9.1a OR TABLE 9.1b HERE]
[Two possibilities are presented as options]
Table 9.1a: Prototype Table 1 - List of coupled AOGCMs that participated at CMIP5 /MMD PCMDI. Selected model features of participating Earth System Models and important references. The models are listed by IPCC identification (Model ID) along with the calendar year of the first publication of results from each model. Also listed are respective sponsoring institutions, resolution information for atmosphere and ocean, features of the Carbon Cycle in ocean and land, details of the sea ice implementation, details of the terrestrial ecosystem implementation, and details on the coupled system. [This table would be aligned to fill a page in landscape mode, similar to AR4]
Table 9.1b: Prototype Table 2 – Significant properties of the AOGCMs used for CMIP5.
Selected model features of participating Earth System Models and important references. The models are listed by IPCC identification (Model ID) along with the calendar year of the first publication of results from each model and the respective sponsoring institutions. Also listed are references on the implementation of atmospheric dynamics, physics and chemistry, information on the implementation of oceanic dynamics, biogeochemistry and sea ice, information on the respective terrestrial ecosystem model, and details on the coupled system. [This table is a more detailed table that would go in portrait mode over two pages]

9.1.3.2 Parameterisations

As highlighted in Section 9.1.3.1, parameterisations are included in all model components to represent the processes that cannot be explicitly resolved in the numerical representations of the physical laws embedded in climate models. 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 (Section 9.1.3.1). The purpose of this section is to highlight major new developments in the parameterisations employed in each model component.

9.1.3.2.1 Atmosphere

Due to the limitations in their resolution (Section 9.1.3.4) atmospheric models need to 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 AR4, and has resulted in improved simulations of tropical variability (Bechtold et al., 2008; Chikira, 2010; Chikira and Sugiyama, 2010; Neale et al., 2008). Another focus has been on improving of the transport of momentum in convection parameterisations (Richter et al., 2008). The work on an alternative approach to convection in climate models by using so-called super-parameterisations has progressed since AR4 and continues to yield promising results albeit at much increased computational cost, making the approach very useful in model evaluation (Demott et al., 2007, 2010; Khairoutdinov et al., 2008; Tao and Moncrieff, 2009; Tao et al., 2009; Zhu et al., 2009) but preventing its use for the climate scenario simulations used in this report.

Previous assessment reports 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 in GCMs, typically based on the use of probability density functions of thermodynamic variables in specifying the model cloud structures (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 (Morrison and Gettelman, 2008). The increasing use of sophisticated treatments of aerosols (see Chapter 7) in climate models 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 the representation of the atmospheric boundary layer since AR4 were focussed on several aspects, most notably on the basic representation of boundary-layer processes, the representation of the stable boundary layer and the representation of boundary layer clouds (Teixeira et al., 2008). A major new approach to treat cloud-topped boundary layers has emerged with the mixed Eddy-Diffusivity-Mass-Flux (EDMF) approach (Neggers, 2009; Neggers et al., 2009; Siebesma et al., 2007) and has been implemented in at least [X] models used in this report [to be checked; IPSL (tbc)]. Realistic treatment of the stable boundary layer in GCMs remains difficult and major research efforts are dedicated to resolving many of the outstanding issues (Beare et al., 2006; Cuxart et al., 2006; Svensson and Holtslag, 2009). The treatment of boundary layer clouds and improvements since AR4 are discussed in Chapter 7 [tbc, or included here].

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) in global general-circulation models (GCMs) (e.g., Palmer et al., 1986; McFarlane, 1987). These initial parameterisations of orographic GWD 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 (dynamical motions of the troposphere such as 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 come under intense scrutiny as a consequence of coordinated efforts such as the core project “Stratospheric Processes And their Role in the Climate” (SPARC) of the World Climate Research Programme.

[PLACEHOLDER FOR FIRST ORDER DRAFT: Description of aerosol treatment in models in Chapter 7 (tbc)]

[PLACEHOLDER FOR FIRST ORDER DRAFT: Include some information on cloud changes; cross-chapter coordination with Chapter 7 needed.]

[PLACEHOLDER FOR FIRST ORDER DRAFT: Include more relevant references and areas of improvements for all aspects as model papers become available.]

9.1.3.2.2 Ocean

Mesoscale and submesoscale eddy parameterisations

Ocean components in contemporary climate and earth system models generally employ grids that are too coarse to explicitly represent mesoscale eddies, even though eddy transport is critical for setting the observed water mass properties and tracer distributions. To address the resolution gap, ocean models implement a version of the (Gent and McWilliams, 1990) eddy parameterisation that computes an eddy-induced tracer advection in addition to that obtained from the model's resolved velocity field (Gent et al., 1995; McDougall and McIntosh, 2001). This parameterisation, along with the neutrally oriented downgradient diffusion proposed by Solomon (1971), Redi (1982), and McDougall and Church (1986), provide the framework for nearly all mesoscale eddy parameterisations used in current climate models. Within that framework, research efforts have focused on how parameterised eddy fluxes in the ocean interior interact with parameterised non-hydrostatic processes dominating the planetary boundary layer, (e.g., Gnanadesikan et al., 2007; Ferrari et

al., 2008; Ferrari et al., 2010; Danabasoglu et al., 2008). Another focus concerns specification of the eddy diffusivity setting the scale for parameterised eddy fluxes, (e.g., Held and Larichev, 1996; Visbeck et al., 1997; Danabasoglu and Marshall, 2007; Eden and Greatbatch, 2008; Eden et al., 2009; Marshall and Adcroft, 2010). Such refined details of the eddy parameterisation are important for both 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). For ocean climate models of higher resolution, a question of growing importance concerns the need to partially parameterise unresolved mesoscale eddies without damping eddies explicitly resolved, with Roberts and Marshall (1998) and Smith and Gent (2004) proposing methods to address this issue.

In addition to mesoscale eddies, there is a growing awareness 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 ocean climate models.

Parameterisations of dianeutral transformation

There is a growing focus in the physical oceanography community on climate model parameterisations of dianeutral mixing associated with breaking gravity waves, with this topic the subject of a NSF/NOAA Climate Process Team in the USA. Basically, this effort aims to provide physical rigor to the prototype abyssal tidal mixing parameterisation of (Simmons et al., 2004) that is now used in several climate models (e.g., (Jayne, 2009)), and which aims to remove the ad hoc nature of the earlier (Bryan and Lewis, 1979) prescription. It also aims to capture more of the mixing processes associated with internal tides and other processes (see (MacKinnon et al., 2009) for an overview). In addition to traditional mixing processes, (Klocker and McDougall, 2010) identify the importance for water mass transformation, especially in the Southern Ocean, of processes arising from the complex equilibrium thermodynamics of seawater. These processes are present in models that utilize a realistic equation of state (e.g., (IOC et al., 2010)) to define the neutral directions used to orient lateral tracer diffusion. 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 effort, with associated developments making their way into some of the current generation of climate models (Danabasoglu et al., 2010; Jackson et al., 2008b; Legg et al., 2009).

Ocean model formulation and experimental protocols

As reviewed by (Griffies et al., 2000; Griffies et al., 2009b), a fundamental choice affecting ocean model architecture and parameterisations involves the vertical coordinate. There are broadly two classes of vertical coordinate used for AR5 ocean climate models: geopotential level models and isopycnal or hybrid layer models. The layered models are used in a minority of climate models, though advances have been made to develop this approach beyond that in AR4 (Megann et al., 2010). Work on the complementary issues of horizontal gridding, such as finite element approaches, remain untested in current coupled climate applications, although some work is underway using ice-ocean component models (Danilov, 2011).

The Boussinesq approximation is commonly used by ocean models, in which the fluid kinematics are based on volume rather than mass conservation. The calculation of global mean sea level in Boussinesq models requires an adjustment to account for missing non-Boussinesq steric effects (Greatbatch, 1994; Griffies, 2011). In some ocean models, the liquid ocean is closed to the exchange of matter with other climate components, thus necessitating the use of unphysical virtual tracer fluxes (e.g., virtual salt fluxes) to partially account for this transfer. This approach, which has been criticized for its unphysical nature and potential to expose simulations to spurious feedbacks (Griffies et al., 2001; Griffies et al., 2005; Huang, 1993; Yin et al., 2010b), is increasingly being abandoned in favour of natural boundary conditions that allow water to cross the ocean surface. This transport can become especially important with advances in river models allowing for varying concentrations of biogeochemical tracers. The mass exchange across the ocean surface also facilitates the depression of the liquid ocean surface under the influence of sea ice and atmospheric loading. Many ocean models, however, ignore the impact from such loading, which means they assume a zero mass sea ice and atmosphere.

Benchmarking global ocean-ice climate models is being coordinated (Griffies et al., 2009a) with the specification of Common Ocean-ice Reference Experiments (COREs). Many groups are making use of this protocol for developing new model configurations (Danilov, 2011).

Ocean biogeochemical (OBGC) models

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 function of the saturation state of seawater with respect to calcite (Gehlen et al., 2007; Ilyina et al., 2009; Ridgwell et al., 2007) or pCO_2 (Heinze, 2004). In addition to calcite, also aragonite – another mineral form of $CaCO_3$ with different solubility has been included in a OBGC model (Gangsto et al., 2008). 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 sediments representation in the models (Heinze et al., 2009).

CMIP5 OBGC models are based on so called NPZD-type models which part marine ecosystems into nutrients, plankton, zooplankton, and detritus. These models provide us with most likely response of oceanic CO_2 uptake to climate, but are limited by lack of understanding 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 yet uncertain implications for Earth system models. Oceanic uptake of CO_2 is variable and is determined by the interplay between the biogeochemical and physical processes in the ocean. Most CMIP5 models are z-coordinate models; introducing an isopycnic coordinate system has resulted in large variability in the interior ocean biogeochemical processes (Assmann et al., 2010).

9.1.3.2.3 Land

Earth's climate is strongly affected by the nature of the land-surface, including the vegetation and soil type and the amount of water stored on the land as soil moisture, snow and groundwater. For example, it has been suggested that changes in the state of the land-surface, which in turn changed the energy and water fluxes to the atmosphere, played an important part in the severity and length of the 2003 European drought (Fischer et al., 2007a). Vegetation and soils affect the surface albedo, which determines the amount of sunlight absorbed by the land. The land surface also affects the partitioning of rainfall into evapotranspiration, which cools the surface and moistens the atmosphere, and runoff, which provides much of our freshwater.

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 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). In addition, more land-surface schemes now couple to a Dynamic Global Vegetation Model (DGVM), so that vegetation cover can be changed interactively in response to changes in climate, and atmospheric CO_2 (Cramer et al., 2001; Sitch et al., 2008), and in some cases also changes in the nitrogen cycle (Thornton et al., 2009; Zaehle et al., 2010a), near-surface ozone (Sitch et al., 2007) and diffuse light (Mercado et al., 2009a).

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, physical sea ice processes are well understood and well represented in models; for example, the basic thermodynamic description has been available for 40 years (Maykut and Untersteiner, 1971). Likewise, a relatively straightforward representation of sea ice dynamics is nearly 35 years old (Hibler, 1979) and something like this is now included in most modern climate models. These thermodynamic and dynamic models capture the first-order behaviour of sea ice in the climate system. Since the AR4, progress in improving sea-ice components in climate models has apparently slowed down. Sea ice model development now follows two paths, both arguably addressing higher-order effects: (1) more precise descriptions of physical processes and characteristics such as microstructure evolution and anisotropy, and (2) extensions of the model for "Earth system" simulations, for example by including biological and chemical species.

The Arctic and Antarctic sea ice packs are mixtures of open water, thin first-year ice, thicker multiyear ice, and thick 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 does 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 within a model grid cell, and an energy based description of mechanical redistribution that converts thinner ice to thicker ice under convergence and shear.

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.

Sea ice dynamics

Many climate models currently employ some version of the nonlinear viscous-plastic (VP) constitutive law proposed by Hibler (1979) to represent the internal ice stresses that arise from deformation. A variant of this scheme developed by (Hunke and Dukowicz, 1997) adds elastic behaviour (so called EVP model), thus permitting a fully explicit implementation with an acceptably long time step. This version has been implemented in many climate models. The recent Jacobian-Free Newton-Krylov approach (Lemieux et al., 2010) promises computationally efficient VP solutions for AOGCMs.

In most models, ice area fraction, volume, energy, and snow volume and energy are advected horizontally. In addition, there may be equations describing the transport of tracers such as biogeochemical inclusions. 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). The second-order-accurate incremental remapping scheme (Lipscomb and Hunke, 2004) projects model grid cells backward in time along Lagrangian trajectories, preserving monotonicity when the field gradients are limited appropriately, although this may reduce the accuracy locally.

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 a multiple-scattering radiative-transfer scheme (Briegleb and Light, 2007), in which absorptive effects of melt ponds and inclusions such as dust and algae can be simulated.

In the sea-ice component of many climate models, the sea ice salinity is assumed fixed and it is used in the calculation of fresh water and salt exchanges at the ice-ocean interface. Some models allow the value to vary in time (Schramm et al., 1997), while others assume a salinity profile that is constant in time, 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 biogeochemistry, which in sea ice depends critically 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. These inclusions interact with sea ice through radiative transfer and other physical processes such as flushing by melt water.

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, which affect the albedo, conductivity, and biogeochemical processes and thereby play a role in climate change. There are several schemes for modelling the ponds that form on the surface of melting sea ice. The simplest and most widely used method does not involve tracking melt water, but rather decreases the ice surface albedo under warm, melting conditions, e.g., (Vancoppenolle et al., 2009b). A second method tracks melt-pond area and volume for each ice thickness category, but applies those quantities only for radiative calculations using the Delta-Eddington multiple-scattering scheme (CCSM4, reference needed). These methods simulate the critical radiative effect of melt ponds, but not their hydrological influence (e.g., the delay of internal ice cooling as ponds refreeze in the fall). More realistic melt pond schemes are under development.

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

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.1.3.3.1 Widespread development of Earth system models

In order to project the coupled evolution of the Earth's physical climate, chemistry, and biogeochemistry, many traditional AOGCMs employed in the AR4 have been enhanced and transformed into Earth System Models (ESMs). Since AOGCMs generally do not incorporate the requisite chemical and biogeochemical cycles, the concentrations of long-lived greenhouse gases (LLGHGs) and other radiatively active species are prescribed from offline sources. The conceptual issue with prescribing concentrations is that the physical climate controls the natural sources and sinks of CO_2 and CH_4, the two most important LLGHGs. The omission of internally-consistent feedbacks between the physical, chemical, and biogeochemical processes in the climate system is an inherent feature of AOGCMs forced with prescribed concentrations. The enhancements in ESMs make 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, or alternatively, to diagnose these sources (with feedbacks included) in simulations with specified concentrations. 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. 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).

9.1.3.3.3 Methane cycle and permafrost

In addition to prognostic carbon cycles, an increasing number of ESMs are also incorporating methane cycles to quantify the feedbacks from changes in these cycles due to a warmer climate. The possibility that human-induced warming might cause permafrost to melt and release some of the stored carbon stocks in the form of methane has prompted further extensions to ESMs to treat these processes. Several models now include representations of permafrost distributions and processes (Khvorostyanov et al., 2008a) that facilitate simulations of the evolution of the permafrost carbon stock in warmer climates (Khvorostyanov et al., 2008b). In several of the current generation of ESMs (Volodin et al., 2010), the emissions of methane from permafrost and from other natural and anthropogenic sources are integrated with representations of the terrestrial and oceanic methane cycles (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). The 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 (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 (Smith et al., 2010b; Smith et al., 2010c), 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, and the strength of the Brewer-Dobson circulation (Butchart et al., 2010; CCMVal, 2010; Son et al., 2008; Son et al., 2010; WMO, 2011). 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 circulation 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 a coupled interactive ocean 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 chemistry-climate models evaluated in CCMVal (e.g., Richter et al., 2008) have been incorporated into ESMs (Scinocca et al., 2008). 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 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 (Butchart et al., 2010; Eyring et al., 2010b; SPARC-CCMVal, 2010), stratospheric dynamics can in turn influence the tropospheric circulation (Arblaster and Meehl, 2006; Baldwin and Dunkerton, 2001; Bell et al., 2009; Cagnazzo and Manzini, 2009; Charlton et al., 2004; Ineson and Scaife, 2009; Shaw and Perlwitz, 2010). To reduce uncertainties in the representation of climate variability on seasonal to multi-decadal timescales and to improve the representation of the upper troposphere, many of the climate models participating in CMIP3 have been further developed to include a fully resolved stratosphere with a model top above the stratopause. The subset of *X* models that performed the CMIP5 models with these so-called high-top models allows a multi-model comparison to the standard set of low-top models with a model top in the middle stratosphere. The differences in climate variability and impact on surface climate between these two classes have been explored in a variety of studies [references needed to synthesis papers on high top / low top comparison, including Charlton-Perez et al. and Manzini et al., in preparation] and are assessed in the subsequent Sections and in Chapter 10.

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 climatic 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 include 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 groups 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 to include the diurnal cycle, which was shown to improve SST bias in the tropical Pacific (Bernie et al., 2008; Ham et al., 2010a). 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 [examples of CMIP5 models to be included here], or by computing the chlorophyll distribution with an ocean biogeochemical model [examples of CMIP5 models to be included]. 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

[PLACEHOLDER FOR FIRST ORDER DRAFT]

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 bases an increased portion of simulated processes on known equations, which is expected to lead to better representation of finer scale structures, such as atmospheric and oceanic eddies and vertical stratifications, 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. Recent findings on the impacts of refined resolution on model performance are assessed in Section 9.6.3.3.

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. 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.

Advancement in computational capacity enables models to be run at ever higher resolution. This may lead to a stepwise, rather than incremental, improvement in model performance. (e.g., Roberts et al., 2004; Shaffrey et al., 2009; [references to be added]). 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) (Smith et al., 2000; McWilliams et al., 2008). Such mesoscale eddy permitting ocean models better capture the large amount of energy contained in the ocean mesoscale, including fronts, boundary currents, and time dependent eddy features (e.g., McClean et al. 2006). 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, accepted in Ocean Modelling; Delworth et al., 2011, in preparation).

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 with a dynamic ocean model, given current computational capacity. Moreover, a cloud-system-permitting model still depends on parameterisations for cloud microphysics and moist turbulence. Thus, a practical application of this kind of simulation for climate timescales requires further computational and scientific advancement.

9.1.3.4.2 Downscaling methods

Regional climate models share many of the same model resolution issues with global climate models. As a rule, regional climate models typically employ higher resolution than global climate models. Resolution increases tend to occur in parallel in regional and global models, so this relationship persists. Since the AR4, typical regional climate model resolution has increased from around 50 km to around 25 km (e.g., Christensen et al., 2010). Also, resolutions around 10 km are being increasingly considered (e.g., van Roosmalen et al., 2010), and some RCMs are run on yet higher resolution (e.g., Kanada et al., 2008).

Higher RCM resolution facilitates the representation of spatial and temporal detail. The representation of larger-scale features, however, may not be affected, when these are already skilfully represented in the global models (Sanchez-Gomez et al., 2009). However, increasing resolution in regional downscaling may cause some degradation of area means owing, for example, to errors in sub-grid-scale parameterisations. Benefits of higher resolution may be expected for regions characterised by variable topography, as well as for extremes in general. Indeed, positive effects of higher RCM resolution have been noted for, e.g., convective precipitation (Rauscher et al., 2010), near-surface temperature and wind (Kanamaru and Kanamitsu, 2007). There may be diminishing returns, however, as Rojas (2006) noted more improvement when increasing RCM resolution from 135 km to 45 km that when going from 45 km to 15 km, despite the very variable regional topography in the domain. (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. Indeed, the quality of boundary conditions is a fundamental aspect in evaluation of downscaling methods. With regards to climate projections, improvements in AOGCMs that provide boundary conditions are fundamental for the resulting downscaling results.

As in the case of AOGCMs, care needs to be taken to develop physical parameterisations alongside increasing model resolution. Another shared feature with AOGCMs is the unavoidable trade-off between higher resolution, model complexity and the extent of simulations to be made, due to computing resource constraints. A further difficulty is that data limitations can impact the development and evaluation of RCMs due to the relative sparseness of observational networks in many regions. This has consequences for evaluating high resolution model results on the daily scale as well as for extremes (Hofstra et al., 2010).

9.2 Techniques for Assessing Model Performance

[PLACEHOLDER FOR FIRST ORDER DRAFT]

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. Such evaluation also encompasses the identification and interpretation of the spread in state-of-the-art model simulations. Uncertainties in climate projections arises from internal variability, model uncertainty, and scenario uncertainty (Hawkins and Sutton, 2009). 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 (see Section 9.7).

In the AR4, the evaluation of climate models was mainly done qualitatively by applying different kinds of diagnostics (e.g., time series or spatial maps) to compare simulated and observed fields. Since the AR4, performance metrics, which are statistical measures of agreement between a simulated and observed quantity (or covariability between quantities), have been extensively used. Performance metrics derived from a variety of observationally-based diagnostics can provide an objective synthesis and visualization of model performance (Cadule et al., 2010; Gleckler et al., 2008; Pincus et al., 2008b; Waugh and Eyring, 2008) and enable a 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, 2008a)). 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 also been emphasized since the AR4 (Knutti et al., 2010b).

Despite these develoments, 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 is more difficult because evaluation is required 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, [reference needed]), which has increased the potential for evaluating Earth system models. Observational uncertainty can be included in model evaluation either by accounting for 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 still do not supply formal error estimates, but where multiple estimates exist they often incorporate the same underlying measurments 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) is impeding progress.

9.2.2 New Developments in Model Evaluation Aapproaches

[PLACEHOLDER FOR FIRST ORDER DRAFT]

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 a quantitative comparison, several studies have applied statistical measures (e.g., root-mean square error, centred and uncentred pattern correlations, M statistic, averages of some statistics) of overall model performance (overall metrics) compared with observed climatology, trend or variability (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, 2008b; Whetton et al., 2007). Many of these studies have also tested, through statistical analyses of metrics across models in an ensemble, whether the performance of a model in reproducing current or past climate is related to the behaviour of the model in projecting future climate. They have found that such relations between metrics of overall performance and projections are generally weak in the global and sub-continental scales (Abe et al., 2009; Giorgi and Coppola, 2010; Knutti et al., 2010a; Raisanen et al., 2010; Whetton et al., 2007). Furthermore, different overall metrics produce different rankings of models (Gleckler et al., 2008). These findings suggest the need for more process-oriented, rather than overall, metrics for evaluating the credibility of modelled future projections.

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 converges 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 modern climate models heavily rely on conceptual models of the processes represented in them. It is therefore necessary to evaluate the representation of those 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), the results are averaged within categories that describe physically important regimes of the system under study. Different types of regime-identification have been developed and applied to the evaluation of processes in climate models since AR4, such as 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 which processes might be responsible for errors identified in the more classical approaches described in Section 9.2.2.1, and for assisting in the selection of the key processes for further analysis using even more in-depth process studies (Jakob, 2010).

As the representation of processes underpins climate models (see Section 9.1.3.1) a wide range of in-depth process studies is regularly carried out to evaluate and improve the process representations in all model components. These studies are facilitated through the removal of model components or process descriptions from the host-model and the use of off-line simulations, or through the use of initial value techniques (see Sections 9.2.2.5 and 9.6.3.1). The results of such simulations are compared to process measurements from detailed field studies or to results from more sophisticated process models (Randall et al., 2003b). Numerous new important process-related data sets to support such evaluations have been collected since the AR4 (Illingworth et al., 2007; May et al., 2008; Moncrieff et al., 2011; 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 types of studies are of utmost importance to test the realism of the process formulations that ultimately underpin the global model simulations evaluated in this chapter.

[PLACEHOLDER FOR FIRST ORDER DRAFT: More YOTC results, CINDY/DYNAMO, others; more examples for land and ocean to be added.]

9.2.2.3 Instrument Simulators

Satellites provide nearly global coverage, sampling all meteorological conditions. This makes them powerful tools for model evaluation. The geophysical properties retrieved from inverse modelling applied to the raw observations (radiances) are called retrievals (Stephens and Kummerow, 2007). Satellite retrievals 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., 2008b) but the quantitative analysis of these comparisons is difficult. Modelled and retrieved variables are not consistently defined (e.g., total cloudiness), satellite sensors have limitations (e.g., finite sensitivity, fixed viewing geometry), and the retrievals involve various assumptions to make the inversion problem tractable. All these factors introduce inconsistencies between the retrievals and the model variables.

Satellite instrument simulators are pieces of software applied to model outputs that mimic the observational process with the aim of bringing models and observations onto the same ground to facilitate more direct comparisons. Until recently, the simulator approach has focused on the simulation of ISCCP products (Klein and Jakob, 1999; Webb et al., 2001; Yu et al., 1996). Since AR4, simulators for other instruments have been developed and applied to models: passive imagers such as MISR (Marchand and Ackerman, 2010) or MODIS (Pincus et al., 2011), lidars like CALIOP/CALIPSO (Chepfer et al., 2008) and ICEsat (Chepfer et al., 2007; Wilkinson et al., 2008), the CloudSat radar (Bodas-Salcedo et al., 2008; Haynes et al., 2007; Masunaga et al., 2008), and the TRMM precipitation radar (Masunaga et al., 2008). The Cloud Feedback Model Intercomparison Project (CFMIP) community has brought together several of these simulators under the same interface: the CFMIP Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011). COSP includes the following simulators: ISCCP, MISR, MODIS, PARASOL, CALIPSO, and CloudSat. RTTOV (Saunders et al., 1999) is a fast radiative transfer code that can also be linked to COSP to produce clear-sky brightness temperatures for many different channels of past and current infrared and passive microwave radiometers. In addition to model evaluation, satellite simulators are used for other purposes like algorithm development (Masunaga et al., 2010). [Acronyms to be defined.]

The simulator approach has been used not only for the evaluation of clouds and precipitation, but also for other variables like upper tropospheric humidity by using the 6.7 micron infrared channel (Allan et al., 2003; Brogniez et al., 2005; Iacono et al., 2003; Ringer et al., 2003; Zhang et al., 2008) and the 183.31+/-1GHz microwave channel (Bodas-Salcedo et al., 2011; Brogniez and Pierrehumbert, 2007).

9.2.2.4 Paleoclimate Studies

Past climates offer a wide range of climate fluctuations that can be used to test the model 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 during the past 25 years for the Last Glacial Maximum (21000 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 of the Biome 6000 project (Prentice et al., 1998) over land.

Paleo proxies, such as pollen or $\delta^{18}O$ in ice cores, are indirect measurements of climatic conditions (see Chapter 5). Two types of approach are found in the literature to test model results. The indirect method consists of transforming the proxy records into climate variables allowing for direct comparison with model outputs. It is important in this case to consider climate variables that best characterize 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. in press). Recent work on marine proxies suggests that, depending on the region, the same proxy is not necessarily dominated by the same aspects 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). Specific filters can be developed to facilitate comparison of model results with data.

The direct method or forward modelling approach consists of simulating the proxy indicators. This can be done by using either a specific off-line model, or the inclusion of specific processes in the new generation of ESMs. For example, in the 1990s, biome models were used to simulated biome distribution in equilibrium with the simulated past climate (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 [references needed]. Solutions have been proposed to consider mega biomes and to transfer properly the vegetation types that differ from one model to another into a common set of vegetation types (Harrison and Prentice, 2003). New data sets are also being proposed to test the biogeochemical components of ESMs looking at dust or fires [references needed]. Also, some models can be run with a representation of water isotopes which allows direct comparison of the model output with isotopic measurements in different proxy records (LeGrande et al., 2006). On the marine side there is a growing interest in simulations of ocean tracers such as $\delta^{13}C$ of $\delta^{14}C$, and of marine biochemistry that can provide more direct comparison with ocean proxies (Tagliabue et al., 2009).

The evolution in the estimation of the goodness of fit between model and data follows closely what is done for present day climate with the use of ensemble simulations, the development of diagnoses that can be easily applied to a wide range of model outputs, and the use of more complex statistical approaches that consider the uncertainties in both the climate simulations and the proxy records (Brewer et al., 2007; Hargreaves, in preparation).

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. Because of this, weather prediction is termed an "initial value" problem. Weather forecasts can be readily "verified" against subsequent observations a few days later. In contrast, climate predictions and projections (see Chapter 11) simulate the statistics of weather seasons to centuries in advance, and the opportunities to "verify" climate predictions and, particularly, multi-decade projections are limited. But despite their clear differences, both weather predictions and simulations of future climate are performed with essentially the same atmospheric model components. In the AR4, it was mentioned that climate models could now be integrated as weather prediction models if they are initialised appropriately (Phillips et al., 2004). An advantage of testing a model's ability to predict weather is that some of the sub-grid scale physical processes that are parameterised in models (e.g., cloud formation, convection) can be evaluated on time scales characteristic of those processes, without the complication of feedbacks from these processes altering the underlying state of the atmosphere (Boyle et al., 2005; Martin et al., 2010; Pope and Stratton, 2002; Williamson et al., 2005). Model error remains a major source of uncertainty in climate predictions, particularly at the regional scale (Hawkins and Sutton, 2009). Importantly, perturbed parameter ensembles show that climate forecasts are highly sensitive to uncertainties in "fast" weather-related model physics (Stainforth et al., 2005b). Such studies suggest that there is much to be gained from a unified weather-climate modelling framework where the routine developments in weather prediction lead implicitly to better climate models (Martin et al., 2010).

Prior to each operational forecast, typically several million observations of the atmosphere and surface are assimilated to produce the initial conditions (or "analysis") with which the forecast is initialised. The comparison of model and observations implicit within the data assimilation process provides an important opportunity to assess the model's physics. For example, in the absence of observation bias, mean differences between the background (or "first guess") forecast and the analysis indicate systematic model error. Such errors are manifested as erroneous initial tendencies in the forecast (Klinker and Sardeshmukh, 1992). More recently, data assimilation experiments with a full operational weather forecast model, but at climate model resolution, have shown that the initial tendency methodology casts doubt on some parameter perturbations associated with high climate sensitivities (Rodwell and Palmer, 2007). Such approaches avoid many of the uncertainties in the assessment of a model's long-term climate associated with interactions, feedbacks, compensating errors and the reduction in degrees of freedom. Importantly, since these methodologies assess the realism of model physics, they can be used to produce 'process metrics' that complement the 'circulation metrics' (Gleckler et al., 2008) more commonly used within the climate community.

Ensemble data assimilation provides the opportunity for a more systematic assessment of model error, including the need for a stochastic physics component. Approaches are being developed and applied to

simplified models that permit the estimation of optimal model parameters (Koyama and Watanabe, 2010). Such approaches could provide a less ad-hoc methodology for the construction of a perturbed-parameter climate model ensemble in future.

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. The initialisation of the ocean and other slowly-varying features of the climate system has been shown to increase forecast skill (Smith et al., 2007). However coupled atmosphere-ocean models show great spread in ocean circulation features such as the strength of the thermohaline circulation. Such model uncertainty has the potential to increase uncertainty in regional climate predictions. As with atmospheric data assimilation, it is evident that ocean data assimilation (Balmaseda et al., 2008; Bell et al., 2004) will provide a useful opportunity for the assessment of ocean processes at their characteristic timescales.

The assessment of all processes at their characteristic timescales, using data assimilation and initial value techniques, provides a logical framework for climate model evaluation and development. This approach is process-oriented and thus complementary to the more traditional circulation-metric assessment of models. Processes are assessed where observational data is available, but developments in a given process should lead to global improvements, wherever that particular process is relevant.

9.2.2.6 Evaluation Techniques for RCMs

The standard approach to evaluating an RCM is essentially the same as for an AOGCM in that the model results and observations are directly compared. Owing to their higher resolution, RCM evaluation can take advantage of the detail available in some high-resolution regional reanalysis product such as the North American Regional Reanalysis (Mesinger et al., 2006). In addition, direct comparisons to meteorological and experimental station data can often be made.

Transferability experiments involve applying an RCM to different region from that for which it was initially developed. This provides a way of exploring RCM biases which are often the consequence of region-specific tuning in model parameterisations (Takle et al., 2007). The most common conclusion from RCM intercomparison studies is that no single model outperforms the others (e.g., (Takle et al., 2007)). While a model might exhibit good accuracy in one variable for some specific region, it may appear more unreliable for other variables in the same region or for the same variable in another region. The source of the biases varies. In some cases, the regional model may simply not have schemes for describing some processes, which are not important in its home region but very important in other regions. For example, an RCM that does not have sea ice process may perform well in the tropics but not in the Arctic. In other cases, the biases may come from tuning some parameters in the model parameterisation schemes to obtain more accurate simulation of a desired variable over the domain of interest, whereas such tuning may give poor results over different domains. Transferability experiments can help in exposing model biases and the factors responsible for them, which in turn serves to increase understanding and lead to better implementations of important parameterisations in models (Takle et al., 2007; Gbobaniyi, 2010) typical transferability experiment, RCMs are applied to multiple domains with the model options kept fixed for all domains simulated. A transferability intercomparison often involves several models that are run in a coordinated set of transferability experiments where, as much as possible, the same configuration is used (lateral boundary conditions, domain size and resolution) on a prescribed suite of domains for periods where high temporal resolution observations are available. Simulations for the first transferability intercomparison project (The Inter-CSE Transferability study, ICTS) was conducted in 2005 by the Transferability Working Group (TWG).

The transferability of an RCM can be assessed in different ways, for example by a direct comparison of the simulated variables with observations, or by comparing the simulated atmospheric processes (e.g., for the boundary layer) with observations. Takle et al. (2007) formulated four transferability hypotheses: (i) models show no superior performance on their domains of origin as evaluated by their accuracy in reproducing the diurnal cycles of key surface hydrometeorological variables; (ii) for all climatic regions and periods having convective precipitation during both day and night, alternative parameter settings in convective schemes at a specific resolution result in changes of intensity and diurnal phasing of precipitation that are correlated; (iii)

no single domain provides climatic conditions for developing and tuning a regional climate model that result in measurably better regional climate model performance on all climate domains in the transferability domain ensemble; and (iv) for all non-monsoon climatic regions experiencing weak large scale forcing, daytime surface fluxes are correlated with the height of cloud base. Results of RCM evaluation and transferability studies will be summarized in Section 9.6.3.3].

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 major climate modelling centres. The PPE is developed within the context of a single climate model. The merits of each are distinct and both can contribute to the model evaluation process.

Although the emphasis of this assessment 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 perturbed 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 (e.g., simple box models of the energy balance and carbon cycle and Earth-system Models of Intermediate Complexity (EMICs), (Forest et al., 2006, 2008; Forest et al., 2002b; 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 of the atmosphere or the ocean with higher computational requirements (Brierley et al., 2010a; Collins et al., 2007; Collins et al., 2006a; Sanderson et al., 2008a; Stainforth et al., 2005a). As a model evaluation tool, ensemble methods are used in a two-step process involving EMICs and climate change detection statistics to establish probability distributions for model simulations consistent with historical observations. Model diagnostics (e.g., transient climate response (TCR)) are first assigned with probabilities (or weights) via a PPE from an EMIC and then the values of the diagnostics from the multi-model ensemble are assessed in this context. These results are discussed in Section 9.6.3.4.

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., 2008a; Reichler and Kim, 2008b), the development of ensemble techniques for climate modelling is an active area of research (see special issue Phil.Trans.A v.365(n.1857) [reference(s) needed]) that addressed 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 (e.g., MOM ocean models) and so certain lineages exist at most major modelling centres [examples TBD from Table 9.1]. This suggests that groups of models share biases, and the assumption of model independence is not correct (Annan, 2011; Masson and Knutti, 2011). The effective number of independent models is therefore likely to be smaller than the actual number of models in the MME (Annan, 2011; Jun et al., 2008; Knutti, 2010; Knutti et al., 2010a; Pennell and Reichler, 2011; Tebaldi and Knutti, 2007a). Consequently, the use of *ad hoc* multi-model ensembles is now generally recognized as not providing a rigorous assessment of uncertainty. By exploring the likelihoods of each model in reproducing historical events, a likelihood-based ensemble can be derived to avoid this simplistic use of the MME (Sokolov et al., 2010; Tebaldi and Knutti, 2007a). These methods also in principle permit generating ensembles with targeted uses. For example, projecting future temperatures would require a different distribution than one intended to produce sea-level rise projections, but it remains a challenge to define the metrics of model performance that might improve climate projections with a weighted ensemble (see Section 9.7).

9.2.3 The Role of Model Intercomparisons

The evaluation of climate models is a challenging undertaking. Gauging the extent to which climate models realistically capture fundamental processes of nature 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. MIP experiments, which are usually designed within a community-based framework (e.g., via working groups with representation from many modelling centres), 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 that their results can be compared 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 exercises after additional model developments, it becomes possible to determine how models might be improving over time, as more realistic processes are incorporated into the models.

All models suffer from systematic errors, and model evaluation is a necessary step towards their identification. Benchmark MIP experiments offer a pathway to distinguish between the errors found in an individual model from certain errors which might be found in all models. Organized MIPs often define more than one experiment, some of which might be designed to help isolate the causes of an error identified in all models, or simply to focus on a particular scientific problem that many participants are interested in.

A multi-model perspective has been extremely useful in evaluating to what extent climate models agree with observations, examining future projections of climate change, or exploring the relationship between the two. Identification of features that are consistent across models (robustness) is one example how a multi-model context has helped to characterize model results. The collective result, often expressed as a multi-model average, remains a useful description of model results, at least partly because in some settings it compares more favourably with observations than individual models.

Organized intercomparisons offer a mechanism for the climate research community to work more closely together. They have fostered a research environment that requires model results to be more open and available for independent assessment. Perhaps mundanely, the collaborative environment of MIPs has enabled motivated scientists to develop conventions and protocols that enable an unprecedented level of model evaluation via the "economies of scale" offered by well-organized and systematic experimentation.

9.2.4 Overall Summary of Approach that Will be Taken in this Chapter

The model evaluation assessed 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, 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. 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. This path of increasing focus takes us to more regional evaluation of model performance, including process-oriented evaluation, the importance of model resolution, and approaches to augment regional information with downscaling techniques.

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 our evaluation focuses on increasing detail, a sampling of well-established advanced 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; examine 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 become a comparable challenge. Efforts to formally describe model uncertainties are discussed in Section 9.6.3. 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). Our evaluation focuses on a multi-model perspective (e.g., CMIP3, CMIP5) and the inter-model spread between the individual models. The discrepancy between the individual models is a lower bound estimate of model uncertainty. In principal it is possible to identify a selection of models that agree most closely with available observations, and at certain stages of our model evaluation we do so. 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 be a useful gauge of the reliability of projections are addressed in Section 9.7.

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

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.3.1 Introduction – Basic Characterization of Climate Model Experiments

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.3.1.1 Structure of the Historical Experiments

In contrast to the CMIP3 ensemble of centennial-length simulations using AOGCMs assembled for AR4 (Meehl et al., 2007), the CMIP5 collection assembled for the AR5 also includes initialized decadal-length projections and long-term experiments using ESMs (Taylor, 2011). The observable properties of the basic mean states from these experiments are evaluated against the historical data record in this Section. These observational evaluations address two principal requirements for climate models to 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 for future climate (Senior and Mitchell, 2000). Second, many relationships among climatic forcings, feedbacks, and responses manifested in projections of future climate change can be tested using the observational record (Soden and Held, 2006). However, fidelity to 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 (Sections 9.3.1.3 and 9.3.1.4).

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 9.3.2 through 9.3.5, respectively. The integration of chemical and biogeochemical cycles with the physical climate system is a general property of the ESMs included in the AR5 multi-model ensemble. The formulations and observational evaluations of the carbon and sulfur cycles in the ESMs are presented in Sections 9.3.6 and 9.3.7, respectively.

9.3.1.2 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 present-day conditions corresponding to 2005 [van Vuuren et al., xxxx]. The second phase covers the future spanning 2005 to 2100 CE and is described by a collection of Reference 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 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. A comprehensive set of historical anthropogenic emissions and LULCC

records have been assembled for the AR5 experiments in order to produce a 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). The VOCs are further resolved into emissions for individual species including alcohols, hydrocarbons, aromatic hydrocarbons, ethers, esters, formaldehyde, ketones, acids, and other 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 transitions in 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., 2010a); 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.

9.3.1.3 Relationship of Observational Initialization and Decadal Predictive Uncertainty

The CMIP5 archive includes a new class of decadal-prediction experiments (Meehl et al., 2009b). The design of these experiments accounts for both for secular trends in external forcing and for the effects of internal variability in the climate system. The goal of the experiments is to understand the relative roles of forced changes and internal variability in historical and near-term climate variables. These experiments are comprised of two sets of hindcast and prediction ensembles initialized from 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. Initialized experiments have shown considerable skill in reproducing the observed temperature record (Keenlyside et al., 2008; Mochizuki et al., 2010; Pohlmann et al., 2009; Smith et al., 2007) and the interannual variations in Atlantic hurricane frequency (Smith et al., 2010a). The distinguishing feature of the decadal-length simulations relative to the centennial-length experiments is the use of realistic oceanic initial conditions to capture the state and trends of the upper ocean (Mochizuki et al., 2010). The limits of skilful predictability are dictated by the characteristic timescales for the onset of chaos in the climate system for the dominance of the ocean state by forced response rather than initial conditions. In current models, the time horizon for skilful predictability is approximately 10 years (Branstator and Teng, 2010). Results from these experiments will be described in detail in Chapter 11; here we focus on evaluation of the models used in such predictions.

9.3.2 Atmosphere

Many aspects of atmospheric models have been more extensively evaluated than other component models of the climate system. One reason for this is the diversity of observationally-based data available to test atmospheric model performance. Near global data sets exist for energy fluxes at the top of the atmosphere, cloud cover, the vertical structure of temperature and winds, ozone, and other important properties simulated by atmospheric models. Moreover, promising new data sets are helping advance the diagnosis of models in new ways, e.g., the three dimensional structure of clouds e.g., (Bodas-Salcedo et al., 2008; Chepfer et al., 2008). Our ability to evaluate atmospheric model performance and better understand the climate system is severely limited by deficiencies in observational systems – a concern that cannot be under-emphasized - but available data for the atmosphere is much more comprehensive than it is for the oceans, land and cryosphere.

Several studies have used basic but well-understood performance metrics to gauge the overall agreement between the CMIP3 models and the observed mean state (Gleckler et al., 2008; Reichler and Kim, 2008b) or

have focused on a particular subset of ECVs (John and Soden, 2007; Pincus et al., 2008b; van Ulden and van Oldenborgh, 2006). As many key features of the simulated atmosphere have been qualitatively described in previous assessments, we rely more heavily on ECV performance metrics to summarize the extent to which climate models agree with observations. In doing so, we are better positioned to quantify a variety of useful information including: 1) model improvements, 2) the relative strengths and weaknesses of the different models, 3) what models simulate well and where they are still lacking. Model evaluation of this kind serves to assess the overall agreement between observed and simulated ECVs, but identification of the root cause of simulation errors requires a process-oriented evaluation (see Section 9.3.2.5 and 9.6.2).

As discussed in Section 9.1, all component models of the climate system are built upon fundamental principles such as the conservation of energy, momentum, and mass. For atmospheric models, realistic simulation of the energy and water cycles are particularly important, and of obvious relevance to society. Our selection of ECVs used for model evaluation reflects this. The observations we will use in this Section are summarized in Table 9.2.

[INSERT TABLE 9.2 HERE]
Table 9.2: [PLACEHOLDER FOR FIRST ORDER DRAFT]. Observationally-based estimates of atmospheric Essential Climate Variables evaluated in Chapter 9; alternates are shown in parentheses. [to be updated]

9.3.2.1 Basic State Simulation of Selected Atmospheric Essential Climate Variables

Temperature is among the most basic ECVs. Insolation at the top of the atmosphere is prescribed in climate models, but all other factors that determine the distribution of temperature are simulated, including the radiative effects of clouds, energy transport, and storage of heat in the upper-ocean and land surface. Figure 9.2 depicts the error in climatological (1981–2005) temperature for a composite of 2-meter air temperature over land and SST over the ocean, and in particular illustrates (at least qualitatively) the improvements in this basic climate variable between the CMIP3 and CMIP5 collection of models.

[PLACEHOLDER FOR FIRST ORDER DRAFT: Climatological surface temperature and error changes from CMIP3 to CMIP5.]

[INSERT FIGURE 9.2 HERE]
Figure 9.2: [PLACEHOLDER FOR FIRST ORDER DRAFT] [AR4 Figure 8.2.] (a) Observed climatological annual mean SST and, over land, surfaceair temperature (labelled contours) and the multi-model mean error in these temperatures, simulated minus observed (colour-shaded contours). (b) Size of the typical model error, as gauged by the root-mean-square error in this temperature, computed over all AOGCM simulations available in the MMD at PCMDI. The Hadley Centre Sea Ice and Sea Surface Temperature (HadISST; Rayner et al., 2003) climatology of SST for 1980 to 1999 and the Climatic Research Unit (CRU; Jones et al., 1999) climatology of surface air temperature over land for 1961 to 1990 are shown here. The model results are for the same period in the 20th-century simulations. In the presence of sea ice, the SST is assumed to be at the approximate freezing point of seawater (–1.8°C).

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 (Section 9.1). Figure 9.3 compares observationally-based estimates of precipitation with the CMIP5 multi-model average. Precipitation generally decreases with latitude due to reduced evaporation and saturation vapour pressure at cooler latitudes, and qualitatively models capture this large-scale feature reasonably well. A well-known and persistent bias is evident in the structure of the tropical convergence zones (see Section 9.3.3.3) and has been extensively analyzed in the CMIP3 simulations (e.g., Lin, 2007a). [Zonal mean precipitation rather than a map being considered. Doing so would enable CMIP3 and CMIP5 multi-model averages to be shown.]

[INSERT FIGURE 9.3 HERE]
Figure 9.3: An extension of Figure 8.5 of the AR4 (above), showing 4 panels: (a) observations (with differences between 2 estimates shown in contours), (b) the multi-model means of the CMIP3 models, (c) the CMIP5 models, […] or presented as zonal averages.

The global annual-mean precipitable water is the measure of the total moisture content of the Earth's atmosphere. For AOGCMs evaluated in CMIP3 and AR4, the values of precipitable water agree 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 AR4 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 20C3M 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 climate system, and current models reproduce these patterns with considerable fidelity relative to the NASA CERES data sets (Pincus et al., 2008b). With respect to this metric, the performance of the individual models in the AR4 ensemble is not readily distinguishable from that of the average of all the models. 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 AR4 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 models used in the ECMWF reanalysis (Pincus et al., 2008b).

Since these observations are readily available to the climate community and the models are adjusted to improve the agreements against the satellite record, comparisons against surface fluxes represent more discriminating measures of the emergent physical fidelity of the climate simulations. On average, the AR4 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}. In the longwave, the variance in downwelling fluxes is dominated by the inter-model variation in these fluxes under clear-sky conditions. In the shortwave, 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 AR4 ensemble. Wild (2006) demonstrates that several AR4 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 vapor, 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 AR4 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).

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. Studies such as this are helpful in demonstrating how models improve over time. 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. However, construction of such an overall index is rather arbitrary, and it is unclear to what extent it should be used to make quantitative judgements about the relative performance of different models. We are now in a better position to gauge model improvement of individual ECVs because two generations of models (CMIP3 and CMIP5) can be evaluated with a majority of groups not using flux adjustments.

In what follows, we use the root-mean-square error (RMSE) as a statistical measure to compare models with observations. This basic error measure was used in the TAR and AR4 model evaluation chapters, and

continues to be used in routine model evaluation. Like all other statistical measures, the RMSE has limitations that need to be considered when interpreting results, e.g., areas where errors are large have a disproportionate influence on the overall measure. The RMSE continues to be used however, because its limitations are well understood and easy to interpret. Importantly, the RMSE also has several advantageous properties such as a simple relation to the standard deviation and correlation, and the possibility to decompose errors into orthogonal components (overall bias, departures from the zonal mean, etc.). [PLACEHOLDER FOR FIRST ORDER DRAFT: Inclusion of a brief comparison of RMSE results with alternate statistical measures to be considered following publication of the results from CMIP5.]

Figure 9.4 shows annual cycle space-time RMS errors for a suite of well-observed ECVs as simulated in CMIP3 and CMIP5. Results are normalized by the observed standard deviation of individual variables so that they can be collectively compared. With individual models identified, this figure demonstrates how models continue to improve [tbc] and how some models agree with observations better than others at simulating the mean state of certain ECVs.

[INSERT FIGURE 9.4 HERE]

Figure 9.4: Normalised RMS error in simulation of climatological patterns of monthly precipitation, mean sea level pressure and surface air temperature. Recent AOGCMs (circa 2005) are compared to their predecessors (circa 2000 and earlier). Models are categorised based on whether or not any flux adjustments were applied. The models are gauged against the following observation-based datasets: Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997) for precipitation (1980–1999), European Centre for Medium Range Weather Forecasts 40-year reanalysis (ERA40; Uppala et al., 2005) for sea level pressure (1980–1999) and Climatic Research Unit (CRU; Jones et al., 1999) for surface temperature (1961–1990). Before computing the errors, both the observed and simulated fields were mapped to a uniform 4° x 5° latitude-longitude grid. For the earlier generation of models, results are based on the archived output from control runs (specifically, the first 30 years, in the case of temperature, and the first 20 years for the other fields), and for the recent generation models, results are based on the 20th-century simulations with climatological periods selected to correspond with observations. (In both groups of models, results are insensitive to the period selected.) [Taken from Figure 8.11 of Chapter 8 of the AR4, This figure would be substantially modified, comparing CMIP3 and CMIP5 (rather than CMIP2 and CMIP3). All individual models would be color coded and identified. Results would be shown for a number of ECVs (Ts, pr, clt, OLR, water vapour, etc.) Perhaps only models without flux correction would be shown (i.e., most). This is an alternate to a "portrait plot" of relative model performance.]

[This figure would be an overall summary of mean climate performance with metrics. PG can prepare a test version (CMIP3-only) to be available by ZOD deadline. This figure is an alternate to the blue-red portrait plot (Figure 3 of Gleckler et al., 2008). It shows much of the same information, but visually it emphasizes more how the distribution of errors differ across variables than it does highlight which models are better or lesser performers. Another advantage it has over a portrait plot is that it will also enable visualization of how well models have improved since CMIP3. With all the models identified on a per variable basis however, all the information about relative model performance would be there (all one has to do is 'connect the dots').]

Figure 9.5 is a Taylor diagram (Taylor, 2001; Chapter 8 of the TAR) showing model performance for a selection of ECVs. This summary figure shows the standard deviation, correlation and RMSE for the space-time (1981-2005 climatological annual cycle with global mean removed) agreement of CMIP5 models with selected observational estimates. The results are normalized by the observed standard deviation to enable including multiple fields on the same diagram. A wide range of simulation skill is clearly evident, with average pattern correlations as high as 0.99 for the lower tropospheric temperature, and well below 0.8 for simulated total cloud cover and precipitation. This range of skill reflects a distinction between fields that are explicitly resolved and those that are parameterised. [to be expanded]. There is also some correspondence between the skill of the observed fields to the quality of observations available.

[INSERT FIGURE 9.5 HERE]

Figure 9.5: Multivariable Taylor diagrams of the 20th century CMIP3 annual cycle climatology (1980 – 1999) for NHEX (20°N–90°N). Color identified for each variable, with dots representing individual models and triangles denote multi-model mean. Standard deviation is normalized by the observed value for each

variable. Taken from Gleckler et al. 2008 [Models to be updated with CMIP5 as well as all observations, and results to be shown for the global annual cycle. Individual models to be identified.]

[A similar figure was shown in the TAR, however there was no discussion of the reasons for the range of skill for different variables. There is a slight redundancy between the information in Figures 9.4 and 9.5. Both show relative model performance on a per-variable basis. If there is a need to select one over the other, Figure 9.4 is better suited to summarize model improvement. This can be done with a Taylor diagram (using arrows), but the figure gets complicated fast with multiple variables and models. What the Taylor diagram offers here is a quantitative spread in the skill of different variables, which is perhaps less of a priority to document than model improvement. One possibility that has been discussed is to have several of the complementary figures in the Supplementary Material, i.e., not just plots but metrics results also.]

The broad scope performance metrics discussed above are a typical first-step toward quantifying model agreement with ECV observations. It is important to recognize, however, that when these metrics are applied the results can be sensitive to such factors 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 (Gleckler et al, 2008). More confidence in these quantitative measures can be gained when their robustness is backed up with an examination of the impact of various choices made in their construction.

[PLACEHOLDER FOR FIRST ORDER DRAFT: above to be modified with results from CMIP5]

9.3.2.2 Global-Scale Changes in Selected Atmospheric Essential Climate Variables

Before we examine the ability of the CMIP5 models to reproduce well-known modes of climate variability (Section 9.4), we first assess how well the models capture the broad features of observed changes in a global context. Evaluation of recent trends in climate models is complicated by the fact that the range of results is not just due to structural errors in the models, but may be influenced by uncertainties in external forcings as well [references needed]. De-convolving the importance of model and forcing uncertainties in the historical simulations is an important topic addressed in Chapter 10. Here we focus on quantifying the level agreement between the model simulations and observations.

Several studies have focused on the ability of the CMIP3 models to simulate observed trends (e.g., Fyfe et al., 2010). One topic that has received considerable attention is the extent to which models adequately capture the warming of the tropical troposphere during the satellite era. (Douglass et al., 2008) concluded that the tropospheric temperature trends in models disagree with observations to a statistically significant extent. This apparent inconsistency between models and observations was contradicted in (Santer et al., 2008), with a demonstration that Douglass et al. (2008) applied an inappropriate statistical test and had neglected observational trend uncertainties associated with natural variability.

[The amount of text needed for assessing work that compares observed and simulated trends, which is often done in D&A studies, requires clarification. Discussion being expanded to cover more CMIP3 results, e.g., (Zhang et al., 2007) for precipitation and (Santer et al., 2007) for water vapour to be considered.]

We examine how well the CMIP5 models capture the recent trends and variability of two well-observed ECVs, surface temperature and column-integrated water vapor (over the global oceans, 50°N–50°S).

Figure 9.6 shows the observed and simulated (multi-model) linear trends of temperature (1979–2005) and water vapour (1988–2005), the periods of which are constrained by the availability of near-global coverage offered by satellite measurements. [This (temperature only) component previously was shown in the detection and attribution chapter – clarification needed if this information be assessed in the model evaluation chapter and as a part of this section. Qualitative similarity between the multi-model and observed trends to be discussed, along inter-model spread.]

[INSERT FIGURE 9.6 HERE]

Figure 9.6: Taken from the AR4 Figure 9.6 (D&A Chapter), the above figure shows local linear trends of surface temperature from 1979–2005. The proposed figure would add similar panels for water vapour (1988–2005). Inter-model s.d. to be shown with contours.

[INSERT FIGURE 9.7 HERE]
Figure 9.7: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure to be constructed consists of two scatter plots, one for surface temperature, and one for water vapour. On each plot, the y-axis corresponds to the amplitude of the global average linear trend, and on the x-axis, the spatial pattern correlation of the simulated and observed linear trends.

Based on the trend maps of individual simulations used to construct Figure 9.6, Figure 9.7 consists of two scatter plots, one for surface temperature, and one for water vapour. On each plot, the y-axis corresponds to the amplitude of the global average linear trend, and on the x-axis, the spatial pattern correlation of the simulated and observed linear trends.

[This scatter plot would show results for observations and individual models (identified). Inclusion of multiple realizations of individual models would help demonstrate the sensitivity of the results to internal variability and the short record length. The metrics shown on these scatter plots would compare observed and individual simulation linear trends and their spatial structure. There would be lots to discuss here. If global scale variability is to be consider in addition to trends …]

[INSERT FIGURE 9.8 HERE]
Figure 9.8: Taken from Santer et al. (2009). [PLACEHOLDER FOR FIRST ORDER DRAFT: Similar to the above except replacing the y-axis with the spatial pattern correlation of the observed and simulated monthly variability. One figure for SST, and one for water vapour.]

[Figure 9.8 similar to Figure 9.7 except highlighting the amplitude and pattern similarity of monthly variability rather than trend similarity. Variability would be computed with linear trends removed. One axis of the scatter plots would be the amplitude of monthly variability, and the other would be the correlation between observed and simulated pattern of variability. This would be based on the performance metrics of (Santer et al., 2009) that have previously been used to evaluate the CMIP3 models.]

[The last two figures would complement the mean state ECVs assessment with metrics of global scale trends (Figure 9.7) and variability (Figure 9.8).]

[Several works-in-progress are addressing model agreement with observations (at global scale) as a function of timescale. This advancement would likely be folded into later drafts.]

9.3.2.3 Ozone

Ozone, another important ECV, 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. Since the AR4, there is increasing observational and modelling evidence that Antarctic stratospheric ozone loss has contributed to changes in southern high-latitude climate (WMO, 2011). 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 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 through a process-oriented approach (CCMVal, 2010; Eyring et al., 2005; Waugh and Eyring, 2008), see Section 9.3.2.5. They are also captured in CMIP3 models that prescribe time-varying ozone (Son et al., 2010). However, in AR4 a subset of models prescribed ozone as a climatological zonal mean rather than 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), see also Chapter 10.

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 main

stratospherically induced changes on high latitude surface climate over the past decades [to be confirmed], although Waugh et al. (2009) demonstrated in a study with a single CCM that trends in the location of the 850 hPa jet and the southern annular mode (SAM) are underestimated if ozone is prescribed as time-varying zonal mean field rather than calculated interactively. In addition, Chapter 10 of SPARC-CCMVal (2010) showed that CCMs participating in the second round of the Chemistry-Climate Model Validation (CCMVal-2) Activity model intercomparison have a mean stratospheric climate and variability that is much closer to the observations than CMIP3 models and that interannual variability in the troposphere also tends to be better simulated. As an example for ozone, Figure 9.9 shows a comparison of mean total column ozone (1980–1999) from the CCMVal-2 models and the AC&C /SPARC ozone database to the NIWA assimilated database [this figure to be updated with CMIP5 models that have interactive stratospheric chemistry]. The CCMVal-2 multi-model mean agrees well with observations (panels c, d) although large discrepancies exist for the individual models (panels a, b). As noted in Cionni et al. (2011), total column ozone over Antarctica in the AC&C/ SPARC is higher than in the NIWA assimilated 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). [PLACEHOLDER FOR FIRST ORDER DRAFT: evaluation of stratospheric ozone in CMIP5 models with interactive chemistry and comparison to CCMVal and CMIP5 models that are forced with prescribed zonal mean ozone fields]. Other possible impacts of the ozone hole on surface climate that have been more recently investigated but not fully quantified in model simulations yet include a southward shift of the Southern Hemisphere stormtrack (Lynch et al., 2006), decreases of carbon uptake over the Southern Ocean (Le Quere et al., 2007; Lenton et al., 2009) and increases in sea-ice area averaged around Antarctica (Goosse et al., 2009; Sigmond and Fyfe, 2010).

[INSERT FIGURE 9.9 HERE]

Figure 9.9: Simulated present-day (1980–1999 average) absolute September to November total column ozone (a) and the bias of it from TOMS data (b). The four models that deviate most from observations (two low-biased and two high-biased) are shown in (a, b) to illustrate the spread. The average of 18 CCMVal-2 models in the large panels is shown in (d,e), and the AC&C / SPARC ozone database in (f, g). Changes in total column ozone over Antarctica (averaged from 60°S–90°S) from 1960 to 2000 for the above datasets are shown in panel (h). Ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere. Modified from Eyring et al., J. Clim, (2011), in preparation. [Update with CMIP5 models that have interactive stratospheric chemistry.]

Tropospheric ozone in the historical period has mainly 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 lead to some differences in tropospheric ozone burden and radiative forcing estimates compared to previous studies, mainly due to a smaller increase in 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 6-year (2005-2009) climatology from MLS/OMI satellite measurements, although tropospheric column ozone is generally 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). [PLACEHOLDER FOR FIRST ORDER DRAFT: evaluation of tropospheric ozone in CMIP5 models with interactive chemistry.]

9.3.2.4 Aerosols

In the RCP protocol 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 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 harmonized (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).

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

The LGM and mid-Holocene are now considered as benchmarking periods to test the ability of climate models to represent a climate different from the modern one. We consider here results obtained for AOGCMs or EMICs in CMIP5, comparing the 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 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 a wide interest in testing the ability of climate models to reproduce observed trends over more recent periods. In this line, the transition from the Medieval warm period and the little ice-age offers 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 (except for land use, Jungclaus et al., 2010; Pongratz et al., 2009).

Figure 9.10 makes an overall assessment of the model performance, considering a suite of key diagnoses reported in various studies [inclusion of figures = map to be decided]. They consider large scale patterns, at the hemispheric or continental scale, consistent changes between land and ocean, and more regional aspects in ocean and continental regions for which several in depth analyses are available. The models reported here have different level of complexity and resolution than those used for future climate projections (see Table 9.1). The diagnoses use the most recent update of continental dataset described in (Bartlein et al., 2010) and ocean datasets from MARGO reported in (Waelbroeck et al., 2009) as well as [inclusion of other datasets depending on new papers]. Different methods were used to measure the goodness of fit between model and data. The different sources of uncertainties in the data or model outputs are treated with different level of complexity depending on the analyses (see supplementary material). Some of the simple diagnoses were already reported in the TAR (McAvaney et al., 2001) or in the AR4 (Jansen et al., 2007).

[INSERT FIGURE 9.10 HERE]
Figure 9.10: [PLACEHOLDER FOR FIRST ORDER DRAFT: Figure to be constructed summarizing results from paleoclimate comparisons.]

[The diagnostics (i.e., maps) from which metrics above are calculated to be provided in the Supplementary Material. Coordination with Annex I needed.]

9.3.3 Simulation of Recent and Longer Term Trends – 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 for coupled modes of variability such as ENSO. In this Section we focus on the evaluation of model performance in simulating the mean state of the thermodynamic and dynamic ocean properties, the surface fluxes, and their impact on the simulation of ocean heat content and sea level as well as the 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.

9.3.3.1 Simulation of Mean Temperature and Salinity Structure

The zonal distribution of potential temperature and salinity offers a first evaluation of the performance of climate models in simulating the different regions of the ocean (upper ocean, thermocline, deep ocean). [Changes from CMIP3 to CMIP5, simpler models, to be discussed.]

[INSERT FIGURE 9.11 HERE]
Figure 9.11: [PLACEHOLDER FOR FIRST ORDER DRAFT] From AR4, to be redone from CMIP5: Figure 8.9. 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 MMD at PCMDI.

[Figure 9.11 yet to be done: zonal T (and S) + errors + model discrepancy, consider multiple observations Levitus, Ishii and Domingues].

More 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 by, respectively, (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^{o}C. The biases arose because the convective regime was restricted to the upper 500 dbar; thus, intermediate water that in reality is formed by convection is, in the model, 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).

The capacity of the ocean to take up heat and other atmospheric quantities has a marked seasonality, especially at higher latitudes where the summer stratified upper ocean acts as a barrier and the winter deepening of the mixed layer offers a window to the deeper ocean. The capacity of the models to represent this seasonal feature is assessed in Figure 9.12.

[Further assessment when CMIP5 evaluations available.]

[Assessment of water masses (NADW, ABBW, AAIW, …) to be included if/when available.]

[INSERT FIGURE 9.12 HERE]
Figure 9.12: [PLACEHOLDER FOR FIRST ORDER DRAFT] [From (de Jong et al., 2009) to show ability of models to capture the mixed-layer depth in a crucial region for ocean uptake of heat or carbon. Another region could be shown.]

[Possible other Figures for Supplementary Material.: Zonal mean Mixed Layer Depth, maps of normalised error vs. standard deviation.]

9.3.3.2 Simulation of Sea Level and Ocean Heat Content

Steric and dynamical components of sea level (SL) changes are simulated by the current generation of climate models and can be evaluated with available observations. Ice-sheet models are actively under development, but are generally not included in the CMIP5 simulations. Modifications of the geoid by the gravitational effect and glacial isostatic adjustment influence relative regional sea level are not included in CMIP5 models. However, because these effects have negligible influence on steric and dynamic changes they can be computed diagnostically.

Some of the differences in simulated dynamic sea level changes between models may have to do with configuration choices such as model resolution and such as the rigid-lid and Boussinesq approximations. In principle, evaluation of simulated ocean heat content (OHC) is more straightforward than sea level because OHC depends only on ocean temperature, however, as a result of limitations in the observations there are challenges in the model evaluation of both OHC and SL changes. For model evaluation, it is convenient to

consider them collectively because of the near-linear correspondence between large-scale variations of OHC and the thermosteric contribution to SL changes.

About half of the historical simulations in CMIP3 included the effects of volcanic eruptions. (Gleckler et al., 2006) showed that inclusion of volcanic aerosols led to significantly smaller trends of the 20th century OHC and SL compared to CMIP3 models that did not include volcanic aerosol forcing. A recent hypothesis (Gregory, 2010) suggests, however, that in order to properly estimate simulated heat uptake and sea-level changes associated with volcanic eruptions, an additional set of control runs (i.e., not part of CMIP5) may be necessary to correct for a prolonged ocean cooling associated with episodic volcanic forcing. Thus, our model evaluation of simulated OHC/SL will need to address not only uncertainties in external forcing, but possible limitations in experimental design as well.

AchutaRao et al. (2007) showed that the variability structure of simulated OHC agreed much better with observations when the model data was sampled according to when and where actual measurements exist. Making use of a reduced space-time reconstruction, including corrections for identified biases in the historical measurements, Domingues et al. (2008) further demonstrated that the trends and overall global variability structure of the CMIP3 models agrees better with observations when volcanic aerosols are included. These features are further explored with a regional evaluation of these simulations (Domingues et al, in preparation; [reference needed / Domingues, xxxx]). Yin et al. (2010a) examine the time-mean patterns in the CMIP3 simulations induced by ocean dynamics, and find that the models generally capture the broad scale inter- and intra-basin characteristics seen in the observations, albeit with varying degrees of accuracy among models. Projections of 21st century regional dynamic sea level changes are very uncertain due to inter-model spread [reference to AR4 needed], but at least for some regions Yin et al. (2010) found this spread can be reduced by selecting only a subset of CMIP3 models based on (rather crude) performance skill.

[Several papers in preparation use metrics to evaluate the simulated global annual cycle of ocean heat content and SSH in the CMIP3 simulations [reference needed / Landerer, xxxx]. These are much in the spirit of the broad scale metrics used to evaluate the atmosphere in Section 9.3.2. A proposed figure (Figure 9.13) for CMIP5 will show results in the form of a scatter plot. One axis will show near-global ocean annual cycle RMSE errors of SSH and the other axis will be the same for OHC. These will be constructed from independent observations, namely altimetry and in-situ temperature measurements.]

[Several detection and attribution studies [reference needed / Gleckler, xxxx) are showing time series that compare the CMIP3 OHC trends and variability with a new suite of observational estimates that correct for identified biases. A figure such as the one included here can be re-worked to focus more on the model-obs comparison with less emphasis on observational uncertainty (Figure 9.14). Of practical concern – because CMIP5 ocean data will be on the native model grids - OHC figures for CMIP5 may be difficult to complete in time for the AR5. Another possibility would be to show maps of observed and multi-model OHC trends, just like that proposed with surface temperature in Section 9.2.2. In the same vein, metrics (on a scatter plot) for these trends could easily be constructed.]

[INSERT FIGURE 9.13 HERE]

Figure 9.13: [PLACEHOLDER FOR FIRST ORDER DRAFT: For CMIP5 we will show results in the form of a scatter plot. One axis will show near-global ocean annual cycle RMSE errors of SSH and the other axis will be the same for OHC. Note these will be constructed from independent observations, namely altimetry and in-situ temperature measurements.]

[INSERT FIGURE 9.14 HERE]

Figure 9.14: [Taken from Gleckler et al., D&A paper, in preparation] Global average evolution of upper ocean (0–700 m) volume average temperature. a) New and old estimates of Ishii et al. (2003, 2009) and Levitus et al. (2005, 2009), b) the impact of using different XBT bias correction methods demonstrated within a common analysis framework c) Ishii et al. (2009) and Levitus et al. (2009) spatially complete (SC) and subsampled (SS) as described in text, d) Recent observed estimates and the CMIP3 ensemble common signal (ECS) for the subset of models including volcanic eruptions, those without, and all models. All times series are departures from 1957–1990 climatologies (°C), and those in a) and c) are computed from spatially complete data. For visual clarity, all observational data are 5-year running averages. The 20CEN simulations

in CMIP3 end in 1999 or 2000. [The emphasis of this figure is on observational uncertainty, with multi-model results only shown in panel (d). This information can be re-worked to focus more directly on model-obs comparison.]

9.3.3.3 Simulation of Circulation Features Important for Climate Response

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.3.3.3.1 Simulation of recent ocean circulation

Atlantic Meridional Overturning Circulation

The Atlantic Meridional Overturning Circulation (AMOC) plays a key role in present-days climate. The Atlantic overturning circulation consists of northward transport of shallow warm water overlying a southward transport of deep cold water and is thought to be 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. Therefore, AMOC estimates have had to be inferred from hydrographic measurements; 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, coupled 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., 2005; Schneider et al., 2007). However, AMOC 'time-series' based on synoptic measurements represent sparse sampling (once every few years or decades) with serious aliasing problems, and are not capable of representing AMOC variability or long-term trends. Continuous AMOC monitoring was started at 26°N about half a decade ago (Cunningham et al., 2007); the four-year mean has been determined as 18.7 Sv with a standard 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 [CMIP5 results expected and text to be updated and referenced]. Model ability might directly lead to an assessment of 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 a comparison of observed and simulated AMOC variability (see Section 9.4.3.2).

[No figure in main text, but a metric (MOC strength in Sv vs. recent estimates) to integrate into a larger Glecker et al. style table to be considered.]

[Figure on latitude/depth MOC in Atlantic for models to be included in Supplementary Material.]

Western Boundary Currents

The relatively low horizontal resolution of coupled models leads to western boundary currents that are too weak and diffuse, and hence biases in heat transport, SST, SSS and subtropical mode water formation (Kwon et al., 2010). In addition, the models tend to underestimate the magnitudes of the covariance between the SST and the heat fluxes compared to the re-analyses (Yu et al., 2011). Errors in the simulated time-mean state of the ocean then lead to errors in the models capturing the 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 CGCMs 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.

AchutaRao et al. (2007) found no evidence to support the previously held view that CMIP3 models systematically underestimate the true variability of the ocean temperature and heat content.

[Figure on barotropic stream function for models to be included in Supplementary Material.]

Southern Ocean Circulation

The Southern Ocean is an important driver for the meridional overturning circulation which 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.

Tropical Circulation

Yokoi et al. (2009) showed that the CMIP3 models had significant problems in accurately representing the annual cycle of the Seychelles Chagos thermocline ridge in the South West Indian Ocean, a feature important for the Indian monsoon and tropical cyclone activity in this basin (Xie et al., 2002). These problems arise because the models have difficulty in capturing the asymmetric nature of the monsoonal winds over the basin resulting in a semi-annual harmonic in the local Ekman pumping over the ridge region being too weak compared to observations. In the tropical eastern Pacific, CMIP3 biases in meridional wind lead to model errors in equatorial SST and simulation of ENSO (De Szoeke and Xie, 2008). See also Section 9.3.3.5 for equatorial evaluation (with relevance for ENSO).

9.3.3.3.2 Simulation of glacial ocean circulation

The simulations of the last glacial maximum (LGM -- 21000 years ago) provide additional information on the ability of climate model to simulate the thermohaline circulation and the characteristics of water masses. 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 and the boundary between North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW) was substantially shallower than today and that the NADW formation was less intense (Curry and Oppo, 2005; Duplessy et al., 1988; McManus et al., 2004) . The comparison of PMIP2 simulations, including GCMs and model of intermediate complexity, show a wide range of model responses of the AMOC to the LGM forcing (Weber et al., 2007), 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. (Otto-Bliesner et al., 2007) compared the results of 4 PMIP2 simulations with the deep ocean data from Adkins et al. (Adkins et al., 2002) and [Figure 9.15 to be done] provides an update with the recent CMIP5 simulations. 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 increase 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 (large brine rejection in CCSM). [To be updated when results are available.]

[INSERT FIGURE 9.15 HERE]

Figure 9.15: 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).

9.3.3.4 Simulation of Surface Fluxes and Meridional Transports

Additional heat and GHGs can be sequestered in the deeper ocean only where the ocean is ventilated, which mostly happens in the mid and high latitudes where the mixed layer deepens and allows a direct connection

between the surface and the deeper ocean. An accurate simulation of the mixed layer depth and structure is therefore important; both are closely connected to the surface fluxes of heat, freshwater, and momentum.

9.3.3.4.1 Heat flux and meridional heat transport

[PLACEHOLDER FOR FIRST ORDER DRAFT: Discuss time mean zonal mean total surface heat flux over the oceans (Figure to be included in Supplementary Material) as compared to reanalysis (ERA40, ERAint, OAFLUX, other. Change between CMIP3 and CMIP5. Describe relative contribution of radiative and turbulent HF components to errors in models. The ocean meridional heat transport is tightly connected to the surface heat flux. Discuss change between CMIP3 and CMIP5 + performance of simpler models + role of ocean model resolution. Contribution from advection (gyre + overturning). Meridional heat transport in the ocean (Figure to be included in Supplementary Material.).]

9.3.3.4.2 Fresh water flux and meridional salt transport

[PLACEHOLDER FOR FIRST ORDER DRAFT: The zonal mean fresh water flux (Evaporation minus Precipitation plus continental Runoff) is evaluated in models (Figure to be included in the Supplementary Material [global or just North Atlantic to be decided]). This flux has several strong regional circulation controls (Arctic, AMOC, ACC, Bay of Bengal, …). Surface salinity integrates the surface fresh water flux and is hence a good indicator of model performance of this flux (Figure to be included in Supplementary Material: SSS). The meridional salt transport (Figure to be included in the Supplementary Material) gives an overview of the capacity of the models to connect sources and sinks of E-P+R (caveat: almost no observations to constrain this – re-analysis quality poor and diverse).]

9.3.3.4.3 Momentum flux

The main surface currents are wind driven. Hence a large fraction of the performance of ocean models in reproducing the former depend on the latter.

The zonal mean zonal wind stress is shown in [Figure to be included in the Supplementary Material if no major change since AR4].

[Structure and link with model performance above to be discussed; map of wind stress curl could be added in Supplementary Material; as well as a diagnostic of meridional wind stress.]

9.3.3.5 Simulation of Tropical Pacific Mean State and Annual Cycle

[Other possible subsection on tropical/equatorial assessment (cold tongue error, double ITCZ…) to be considered; location to be clarified; includes: ocean, atmosphere, coupling and links between variability and mean state); considered for inclusion in Section 9.6, less ENSO-centric, and/or distributed in several subsections.]

Simulating the time-mean properties in the tropics has continually been a challenge for coupled GCMs. Though most models can internally generate the fundamental mechanisms that drive El Niño properties, most models simulate a mean zonal equatorial wind stress that is too strong and that has an annual amplitude that is also too strong (Guilyardi et al., 2009b; Lin, 2007b; Lin, 2007a). This has profound effects on ENSO behaviour in that it limits the regimes in which interannual anomalies can develop. Indeed, several studies have shown that a large amplitude of the seasonal cycle usually implies a weak El Niño and vice versa (Fedorov and Philander, 2001; Guilyardi, 2006). Similarly, the meridional extent of the wind variability, of importance for ENSO phase change, is too confined near the equator (Capotondi et al., 2006; Zelle et al., 2005). The “double Intertropical Convergence Zone (ITCZ)” problem, in which a symmetrization of the circulation across the equator leads to a spurious Southern Hemisphere ITCZ and is associated with excessive precipitation over much of the tropics, remains a major source of model error in simulating the annual cycle in the tropics (Lin, 2007a), and can ultimately impact the fidelity of the simulated El Niño (Guilyardi et al., 2003).

Similarly, there are still large differences in how the models reproduce the mean state of the tropical ocean, including the mean thermocline depth and slope along the equator and the structure of the equatorial currents (Brown et al., 2010a). Along the equator in the Pacific, the models have difficulty capturing the correct intensity and spatial structure of the East Pacific cold tongue. Often, the simulated cold tongue is too

equatorially confined, extends too far to the west and is too cold (see Figure 4 of Reichler and Kim, 2008b). These recurrent biases, already present in CMIP1 fifteen years ago, arise from numerous factors including overly strong trade winds, leading to increased cooling via oceanic upwelling, mixing, and latent heat flux to the atmosphere; a diffuse thermocline structure, leading to improper sensitivity of SST to anomalous upwelling and vertical mixing; insufficient surface and penetrating solar radiation, and weak ocean vertical mixing in the subtropics, leading to subsurface temperature errors along the equator; and weak tropical instability waves, resulting in too little meridional spreading of SST anomalies during cold events (Lin, 2007a; Meehl et al., 2001; Wittenberg et al., 2006).

There are also errors in the tropical Pacific seasonal cycle, both in SST and wind: many models exhibit an overly strong seasonal cycle in the east Pacific and/or a spurious semi-annual cycle, possibly tied to the lack of sufficient meridional asymmetry in the background state (Guilyardi, 2006; Li and Philander, 1996; Timmermann et al., 2007) and/or errors in the water vapor feedbacks feedbacks (Wu et al., 2008). The lack of marine stratocumulus clouds in the eastern part of the tropical Pacific is still a major issue in CGCMs (Lin, 2007a), and, associated with a too weak coastal upwelling along the coast of Peru and Chili, leads to a warm bias in these regions. Nevertheless, the CMIP3 models show a clear improvement over previous generation models, as shown in AchutaRao and Sperber (2006) and Reichler and Kim (2008b). [to be updated with CMIP5 results]

[Possible figures to illustrate this subsection:
- SST, taux and depth of 20°C along equator (all longitudes, 1 curve per model)
- Equatorial zonal current in longitude/depth Section for all models (in Supplementary Material)
- Update of Figure 1 Guilyardi et al., BAMS, 2009
- lon/time hovmoeller at equator (3 basins) of seasonal cycle of SST
- lat/time precip in east Pac or other diags by Lin (double ITCZ)

As space is an issue, we could have the figures in the Supplementary Material and just use some of the metrics developed in a table (given the number of models, not sure this would take less space).]

9.3.3.6 Summary Ocean

[PLACEHOLDER FOR FIRST ORDER DRAFT: To be done once CMIP5 results are fully analyzed.]

9.3.4 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 kilometers. Most differences between PMW products tend to offset and thus trends and anomalies are generally in better agreement, than the absolute extent. Insufficiency of sea ice thickness observations remains an impediment for model evaluation; however, (Koldunov et al., 2010) have shown that the ECHAM5/MPI-OM AR4 simulation has significant ice thickness errors, which originate from the incorrectly simulated atmospheric state.

[PLACEHOLDER FOR FIRST ORDER DRAFT: The following text is a placeholder until CMIP5 models are available for analysis.]

Despite the significant differences between models, the CMIP5 multimodel 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 throughout the year (Figure 9.16). This difference is of the same order as differences between various sea ice extent products. In many models, however, the regional distribution of sea ice is poorly simulated, even if the hemispheric extent is approximately correct.

[INSERT FIGURE 9.16 HERE]

Figure 9.16: [PLACEHOLDER FOR FIRST ORDER DRAFT] Baseline climate (1980–1999) model mean sea ice extent seasonal cycle in the Northern (left) and Southern (right) hemispheres as simulated by CMIP3 models (1 and solid) and observed (2 and dashed). The shaded are shows the intermodel standard deviation. 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). [To be replaced by a similar figure for CMIP5 (for the AR5 baseline period 1986–2005).]

There has been no dramatic increase in sophistication of sea ice treatment in CMIP5 AOGCMs compared to CMIP3. The improvement in simulating sea ice in the former, as a group, is not evident (compare Figure 9.17 with AR4 Figure 8.10). In some models, however, the geographical distribution and seasonality of sea ice is now better reproduced.

[INSERT FIGURE 9.17 HERE]
Figure 9.17: [PLACEHOLDER FOR FIRST ORDER DRAFT] Baseline climate (1980–1999) sea ice distribution in the Northern Hemisphere (upper panels) and the Southern Hemisphere (lower panels) simulated by 17 of CMIP3 AOGCMc for March (left) and September (right). For each 2.5° × 2.5° longitude-latitude grid cell, the figure indicates the number of 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). [To be replaced by a similar figure for CMIP5 (same period as above – thus comparable to that in AR4 Figure 8.10).]

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, both positive and negative, are still poorly quantified. As noted in 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. Also important are errors in surface energy fluxes, which may result from inadequate parameterisations of the atmospheric boundary layer in high latitudes and from generally poor simulation of polar cloudiness which is evident from the large inter-model scatter. Potentially important small scale processes, such as convection in brine pockets or in melt ponds, are not included in the sea ice components of current AOGCMs. Possible impacts of black carbon aerosols that induce atmospheric warming and black carbon on snow and ice that decreases the surface albedo have not yet been examined in AOGCMs.

The CMIP5 models appear to have limited abilities to generate sufficient unforced atmospheric variability (as found for CMIP3 ECHAM5/MPI-OM by Koldunov et al., 2010). The major Arctic warming event from the 1920s through the 1940s [is/is not] simulated better in CMIP5 compared to CMIP3. On the other hand, at least some of the CMIP5 models do simulate rapid changes in the Arctic sea ice due mainly to natural variability (Figure 9.18).

[INSERT FIGURE 9.18 HERE]
Figure 9.18: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure showing CMIP5 model ability to generate natural (unforced) variability of sea ice extent. [Placeholder Figure from Holland et al., 2008.]

9.3.5 Land Surface, Fluxes, and Hydrology

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.3.5.1 Surface Fluxes

The primary role of land surface schemes within climate models has been to calculate surface-to-atmosphere fluxes of heat, water and momentum. Offline intercomparison studies such as PILPS (the "Project for Intercomparison of Land-Surface Parameterisation Schemes"; Henderson-Sellers et al., 1995) have historically been very useful in highlighting the differences between the surface energy and water partitioning simulated by different land models, even when they are driven by identical observed meteorological variables. At the time of the IPCC AR4 the impacts of these land model differences were also being detected in atmosphere-only climate simulations. Comparison of the surface water balance simulated by 20 atmospheric global circulation models, against observations over Global Energy and Water Cycle

Experiment Coordinated Enhanced Observing Period basins, identified differences arising from both the simulation of precipitation and also differences in the partitioning of that precipitation into evaporation and runoff (Irannejad and Henderson-Sellers, 2007). The key uncertainty remains the simulation of soil moisture in these models, which controls both the water and energy balance at the land-surface.

The increasing availability of eddy-covariance flux measurements, such as FLUXNET (Baldocchi et al., 2001), has made it possible for land-surface modellers to calibrate and validate their models against direct observations of surface energy and water fluxes. This has undoubtedly led to improvements in the performance of individual land models (Friend et al., 2007; Stockli et al., 2008; Blyth et al., 2010), but there is as yet no evidence of a reduction in the overall range of land model responses, partly because a necessary increase in the complexity of land models for Earth System modelling applications.

9.3.5.2 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 amplifying in most regions, such that a dry anomaly in soil moisture reduces subsequent precipitation which tends to maintain the anomaly. Prior to the IPCC 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 agreed upon (Koster et al., 2004). Soil moisture is 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., 2011a). 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., 2011a). Climate change is expected to increase the extent of semi-arid areas on the globe, and so a better understanding and quantification of the relevant processes would significantly help to reduce uncertainties in future-climate scenarios (Seneviratne et al., 2010).

9.3.5.3 Dynamic Global Vegetation and Nitrogen Cycling

At the time of the IPCC 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 of the climate modelling centres now also incorporate dynamic vegetation in at least a subset of the runs that they have submitted to the IPCC AR5.

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 AR5 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 AR5 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., 2010b). By contrast, tropical ecosystems are thought to be phosphorus rather than nitrogen limited.

9.3.5.4 Land-Use Change

Another major innovation in the land component of the AR5 climate models is the inclusion of land-use change. Changes in land-use associated with the spread of agriculture and 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. Despite the key role of land cover change at regional scales, climate model projections from IPCC AR4 excluded anthropogenic land-cover change.

There has been significant progress on modeling the role of land cover change since the IPCC 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).

9.3.5.5 The Impact of Different Atmospheric Pollutants

Vegetation is affected differently by different atmospheric pollutants, and this means that the effects of changes in atmospheric composition on vegetation, and feedbacks to climate, cannot be understood purely in terms of radiative forcing (Huntingford et al., 2011). There have been advances in representing the relevant processes in land models since the IPCC AR4.

CO_2 increases affect the land through climate change, but also directly through CO_2-fertilization of photosynthesis, and 'CO_2-induced stomatal closure' which tends to increase plant water-use efficiency. Observational studies have shown a direct impact of CO_2 on the stomatal pores of plants, which regulate the fluxes of water vapor and CO_2 at the leaf surface. In a higher CO_2 environment, stomata typically reduce their opening since they are able to take up CO_2 more efficiently. By transpiring less, plants increase their water-use efficiency, which consequently affects the surface energy and water balance. If transpiration is suppressed via higher CO_2, the lower evaporative cooling may also lead to higher temperatures (Cruz et al., 2010). There is also the potential for significant positive impacts on freshwater resources, but this is still an area of active debate (Gedney et al., 2006; Piao et al., 2007; Betts et al., 2007).

By contrast, increases in near surface ozone have strong negative impacts on vegetation by damaging leaves and their photosynthetic capacity. As a result historical increases in near surface ozone have probably suppressed land carbon uptake and therefore increased the rate of growth of CO_2 in the 20th century. Sitch et al. (2007) estimate that this indirect forcing of climate change almost doubles the contribution that near-surface ozone made to 20th century climate change.

Atmospheric aerosol pollution also has a direct impact on plant physiology by changing the quantity and nature of the sunlight reaching the land-surface. Increasing aerosol loadings from around 1950 to 1980, associated predominantly with the burning of sulphurous coal, reduced the amount of sunlight at the surface, which has been coined 'global dimming' (Wild et al., 2006). Since plants need sunlight for photosynthesis, we might have expected to see a slow-down of the land carbon sink during the global dimming period, but we didn't. Mercado et al. (2009b) offer an explanation for this based on the fact that plants are more light-efficient if the sunlight is 'diffuse'. Aerosol pollution would certainly have scattered the sunlight, making it more diffuse, as well as reducing the overall quantity of sunlight reaching the surface. It has been estimated that 'diffuse radiation fertilization' won this battle, enhancing the global land-carbon sink by about a quarter from 1960 to 2000 (Mercado et al., 2009b). This would also imply a potential decline in the land carbon sink as atmospheric aerosol loadings reduce

9.3.6 Carbon cycle

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.3.6.1 The Terrestrial Carbon Cycle

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 underestimate the seasonal amplitude of CO_2 in the northern hemisphere by factors of 2 to 3 (Randerson et al., 2009). This conclusion is quite 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).

The accurate simulation of the Amazon is important for representing its buffering of atmospheric CO2 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 within the context of CCSM reproduced the gross primary productivity (GPP) of the Amazonian forests to within 14% (Lewis et al., 2011), the models overestimate the above-ground live biomass by between 130 and 190% and underestimate the carbon stored in soils by between –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.

The accurate reproduction of Amazonian GPP is qualitatively consistent with results from an intercomparison of CO2 exchange at 44 sites in North America, which shows that the simulated net ecosystem exchange (NEE) in forested areas has normalized mean absolute errors of between –0.7 and –1.1 (Schwalm et al., 2010). Errors in other biomes with significant seasonal cycles in leaf area index are much larger, and in general the mean squared distance between models and observations of NEE is ~10 times the observational uncertainty. Model skill is only weakly correlated with elevated levels of drought (Schwalm et al., 2010), which is a necessary but not sufficient condition for maintenance of skill under systematically drier conditions projected for 2100 over the tropics and subtropics.

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).

9.3.6.2 The Oceanic Carbon Cycle

Recent advances in the observational evaluation of ecosystem-biogeochemical ocean 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. Current AOGCMS produce quite skilful simulations of present-day temperatures with pattern correlations to the observations approaching 0.98 and temporal normalized standard deviations very close to 1 (Schneider et al., 2008). The simulated MLDs at low latitudes also correspond reasonably well to observations, but the models overestimate the MLDs in the intermediate and northern extratropics by up to a factor of two.

Evaluation of OBGC models has been performed within C4M studies with a focus on regional implications of climate change for oceanic uptake of CO_2 (Roy et al., 2011); a similar framework will be applied in the near future for evaluating CMIP5 OBGC models. Declining rates of net ocean CO2 uptake that have been observed in the temperate North Atlantic are broadly reproduced by historical hindcast simulations with

oceanic ecology-biogeochemistry models (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 in the hindcasts. However, models that have been evaluated against estimates 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 reproduction of 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).

9.3.7 Sulfur Cycle

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.3.7.1 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 spanning 1980 to present. 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. In Europe, land-based emissions decreased by 70% between 1980 and 2000 resulting in a drop from 55 $TgSO_2$ emitted in 1980 to 15 $TgSO_2$ released in 2004. In the United States, model-estimated emissions declined by 26.3% between 1985 and 2000. Over that same time period, Asian emissions of SO_2 increased by 31.5%, although the growth was reversed between 1995 to 2000 when emissions declined from 38.5 $TgSO_2$ to 34.4 $TgSO_2$ due to a variety of industrial and societal factors including the introduction of emission-abatement policies in China. Subsequent economic growth has resulted in an increase in Chinese emissions of 53% from 21.7 to 38.5 $TgSO_2$ between 2000 and 2006. 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 chemical transport models (CTMs) forced with meteorological reanalyses. The results from these CTMs can be used to evaluate the historical simulations of sulphate burdens from ESMs. 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. 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 AR4 climate 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 in the data sets input to the AR4 models.

9.3.7.2 Principal Sources of Uncertainty in Projections of Sulphate Burdens

In contrast to the AR4 multi-model simulation ensemble, the CMIP5 ensemble is based upon a single internally consistent set of SO_4 concentrations and SO_2. 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 to sulphur-cycle processes, uncertainties in emissions accounted for 53.3% the ensemble variance (Ackerley et al., 2009). The next largest source of uncertainty is associated with the wet scavenging of sulphate, which

accounts 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 intermodal 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 primarily due to differences in the treatment of chemical production, transport, and removal from the Earth's 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.

9.4 Simulation of Variability and Extremes

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.4.1 Introduction

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 a 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 the response 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 9.3) 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.

9.4.2 Diurnal and Seasonal Cycles

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.4.2.1 Diurnal and Seasonal Cycle of Physical Climate Variables

The diurnally varying input of energy by the sun is one of the most strongly forced external modes of the climate system. Through complex interactions of solar radiation with the land surface and vegetation as well as the upper-most layers of the ocean and the transfer of energy from the surface to the lower layers of the

atmosphere, the varying energy supply leads to strong and 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 landscapes respond to the strong diurnal forcing at different timescales, most prominently land and ocean, even more complex diurnal patterns can emerge near topographic features, in particular coastlines. As the diurnal cycle of many of the ECVs depends on complex interactions of many different physical processes, an evaluation of the diurnal cycle in climate models is an attractive way of evaluating process interactions in the models. The main focus of this Section is the evaluation of the diurnal cycle of temperature and precipitation, although some examples for other variables, such as low-level winds and surface exchanges will also be given.

This paragraph should describe the evaluation of the diurnal cycle of temperature both over land and over the ocean, similar to the work of (Dai and Trenberth, 2004) but extended to all models. Models tend to capture the overall amplitude and phase of the diurnal cycle well over land, but tend to not have much of a diurnal cycle over the ocean. The importance of the diurnal variations of sea surface temperature has only recently been recognized and coupled models have begun to include it (Bernie et al., 2008; Danabasoglu et al., 2006; Ham et al., 2010b). In some models the inclusion of diurnally varying SST has led to a significant reduction in some long-standing model biases, such as cold biases in the tropical Pacific (Bernie et al., 2008; Ham et al., 2010b) as well as to improved simulations of the MJO (Bernie et al., 2008).

Due to the complex interactions of many physical processes the simulation of the diurnal cycle of precipitation over land remains one of the biggest challenges in climate modelling. GCMs have been shown in the past to have difficulties to correctly simulate the diurnal cycle of precipitation over land, in particular when convective processes are involved, as is frequently the case of tropical land and extratropical land areas in summer. Many of the CMIP3 models showed the well-known model error of a rainfall peak too early in the day (Dai, 2006) and of rainfall occurring too frequently at too low an intensity in many models (Dai, 2006; Stephens et al., 2010). In the CMIP5 models this situation has [improved/not changed/deteriorated – awaiting results]. 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) [to be updated with more recent examples]. Increased atmospheric model resolution (Ploshay and Lau, 2010) as well as the inclusion 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 to the simulation of the diurnal cycle of precipitation, although the physical reasons for these improvements remain poorly understood.

While much of the focus of diurnal cycle studies is naturally on temperature and precipitation, several other important diurnal variations have been evaluated in GCMs, such as the simulation of the diurnal cycle of low –level winds [reference to upcoming Gabls publications needed] as well as the simulation of diurnally driven low-level jets [to be confirmed/references needed]. The diurnal cycle of surface energy and water fluxes has also been extensively evaluated using a variety of observations and process modelling approaches. [Text referencing figures related to diurnal cycle, examples of which are provided in Figures 9.19 and 9.20, to be included.]

[INSERT FIGURE 9.19 HERE]
Figure 9.19: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure for diurnal cycle of temperature (from Dai and Trenberth, 2004). This could be replaced by a harmonic analysis of phase and amplitude, which would reduce it to two panels, or by selected line graphs, which would allow for all models to be shown.

[INSERT FIGURE 9.20 HERE]
Figure 9.20: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure for the evaluation of diurnal cycle of precipitation (example figure from Dai, 2006). Line plots are proposed for main report, maps for supplementary material. Those figures could be replaced by a harmonic analysis of phase and amplitude.

[PLACEHOLDER FOR FIRST ORDER DRAFT: Text to be added on seasonal cycle.] Figure 9.21 showing model errors in simulating aspects of seasonal cycle.

[INSERT FIGURE 9.21 HERE]

Figure 9.21: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure illustrating CMIP5 models' ability to simulate general features of the seasonal cycle. Placeholder figure is from the AR4.]

9.4.2.2 Diurnal and Seasonal Cycle of Vegetation and Carbon Cycle

[PLACEHOLDER FOR FIRST ORDER DRAFT: to be completed with CMIP5 results]

9.4.2.3 Large Scale Monsoon 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) Assessment diagnostics and metrics of a models ability to simulate the monsoon domain and its intensity were introduced by Wang and Ding (Wang and Ding, 2008) As shown in Figure 9.22 these measures are based on the hemispheric summer minus winter values, providing a large-scale view of the Earth's monsoon systems in terms or precipitation and circulation (Kim et al., submitted). The CMIP3 multi-model ensemble (MME) generally reproduces the observed spatial patterns but somewhat underestimates the extent and intensity, especially over Asia and for the North American Monsoon. Metrics for the performance of the individual models are given in Figure 9.23, which shows the circulation vs. precipitation plots for (a) the pattern correlation for intensity and (b) the threat scores for representing the monsoon domain. Relative to observations, Figure 9.23a shows that the pattern correlation of intensity for the circulation is better simulated than the pattern correlation of precipitation intensity. Similarly, the Figure 9.23b threat scores (a categorical metric, (Wilks, 1995) indicate that the circulation domain is better represented than is the precipitation domain. Consistent with previous generations of models (Sperber and Palmer, 1996; Sperber and Grp, 1999), the circulation is better represented than the precipitation. Importantly, the results indicate that models with the greater skill in simulating the precipitation also have better skill at representing the circulation. [to be updated with CMIP5 results and new reference]

[INSERT FIGURE 9.22 HERE]
Figure 9.22: [PLACEHOLDER FOR FIRST ORDER DRAFT: Proposition with CMIP5 / CMIP3 results] The approximate extent of the global monsoon domain (solid line) and monsoon intensity (shading) are shown for precipitation (a and b) and 850hPa wind speed (c and d). The monsoon domain is defined where the local summer-minus-winter precipitation rate (850hPa windspeed) exceeds 2.5 mm/day (2.5 m/s). (a) and (c) are based on GPCP precipitation and NCEP/DOE Reanalysis-2, respectively. (c) and (d) show the multi-model mean from the CMIP3 20c3m simulations. After Kim et al. (2011)

[INSERT FIGURE 9.23 HERE]
Figure 9.23: [PLACEHOLDER FOR FIRST ORDER DRAFT: Proposition this kind of figure with CMIP5 results] Evaluation of the CGCMs' performance on the climatological global monsoon intensity (left) and domain (right). The regression coefficient is shown in lower-right corner of each panel. The domain used is 0°–360°E, 40°S–45°N. The threat score has a range of 0–1, with 1 indicating perfect agreement with observations. After Kim et al. (2011)

Large variations of the monsoon systems have been recorded in paleo proxy records (see chapter5). They show for example that the boreal summer monsoon was stronger and penetrated further inland during the mid-Holocene in response to a larger seasonal cycle of insolation in the northern Hemisphere (and smaller seasonal cycle in the Southern Hemisphere). The representation of the northward shift of the rainbelt in the Sahel region has improved (or not) in the last generation of climate models [to be updated from CMIP3/CMIP5], even though most models still underestimated the amount of precipitation north of 18°N (Braconnot et al., 2007b). Comparison with data syntheses over East Asia show that the PMIP2 simulations reproduce well the precipitation in China except for in the central parts of China, but that the model spread is large (Wang et al., 2010).

9.4.2.4 Monsoon and Intraseasonal Variability

Intraseasonal (ISO, 30–70 day) variations are a key component of the monsoon, modulating convection across the tropics. There is some evidence that models that are better able to simulate the seasonal mean

climate tend to better simulate intraseasonal variability, particularly in the monsoon region (Kim et al., 2008). During the boreal winter the eastward propagating ISO (know as the Madden-Julian Oscillation) predominantly impacts the deep tropics, while during the boreal summer there is also northward propagation over much of southern Asia (Annamalai and Sperber, 2005). The simulation of the Madden-Julian Oscillation is still a challenge for climate models (Kim et al., 2009; Lin et al., 2006; Sperber et al., in press; Xavier et al., 2010). However, contrary to the previous generation of models (Waliser et al., 2003). Sperber and Annamalai (Sperber and Annamalai, 2008) showed that the CMIP3 models were able to simulate eastward propagating intraseasonal convection over the Indian Ocean, though only two of seventeen were able simulate the observed northward propagation during boreal summer.

Coordinated efforts have been undertaken to improve our understanding of monsoon intraseasonal variability (Hendon et al., submitted; Sperber and Annamalai, 2008) and to design diagnostics and metrics for model evaluation (Kim et al., 2009) (CLIVAR MJOWG 2009). For example, 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 provides information about how well models simulate eastward propagation and its associated time scale. The maximum positive correlations presented in Figure 9.4.2.6 indicate that all of the CMIP3 models have less coherent eastward propagation compared to observations

[INSERT FIGURE 9.24 HERE]
Figure 9.24: [PLACEHOLDER FOR FIRST ORDER DRAFT] (From Kim et al., 2009) illustrating model performance in simulating MJO.

[To be updated. Process-oriented diagnostics and metrics are in development, and they should provide insight in to the reasons why simulation of the MJO is challenging.]

9.4.3 Simulation of Variability Around the Mean State

The mean climate is by definition the time average of the numerous time 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 the modes of variability whose timescales ranges from a few weeks (e.g., blocking regimes) to multi-decadal features that can modulate the centennial 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 emphasize recent research on the interactions between modes of variability via teleconnections, the processes involved, and model improvements since the AR4.

9.4.3.1 High Latitudes Modes

The Northern and Southern Annular modes (respectively, NAM and SAM) are the dominant modes in the extra-tropical circulation in both hemisphere on time scales from the intra-seasonal to the interdecadal (Thompson and Wallace, 2000). Gerber et al. (2008) showed that the CMIP3 coupled models captured the broad temporal features of both modes as well as their main inter-hemispheric differences. However, models substantially over-estimated the 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 the re-analyses. There are also considerable biases in the Southern Hemisphere eddy-driven jet stream in the CMIP3 models (Kidston and Gerber, 2010). In terms of the spatial patterns, (Raphael and Holland) 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.

[INSERT FIGURE 9.25 HERE]
Figure 9.25: [PLACEHOLDER FOR FIRST ORDER DRAFT] [tbd]

9.4.3.1.1 Simulation of the NAO/NAM

There is still considerable debate over the difference between the hemispheric Northern Annular Mode (NAM) and the more regional North Atlantic Oscillation (NAO). (Feldstein and Franzke, 2006) showed that in practice both approaches identify the same events in the analysis of the observed record, so in this context the distinction is less important. However, as described in AR4, climate models have a tendency to overestimate the teleconnection between the Atlantic and Pacific basins, so that patterns of model variability tend to resemble the NAM more than the NAO.

While climate models successfully simulate the basic features of the NAM, they tend to overestimate its decorrelation timescale (Gerber et al., 2008), a bias which is linked to the climatological bias in the latitude of the jet streams (Barnes and Hartmann, 2010; Kidston and Gerber, 2010). The asymmetry in persistence between positive and negative phases of the NAO was also not well simulated in the climate model analysed by (Woollings et al., 2010b). Some recent studies have assessed the skill of climate models in representing the shape of the distribution of NAO variability. The distribution of the NAO in reanalyses has pronounced negative skewness, and analyses of individual climate models have shown that they do not represent this asymmetry correctly (Coppola et al., 2005; Woollings et al., 2010b). Similarly, the CMIP3 models tend to misrepresent the distribution of variability 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 correlated with the climatological equatorward jet bias in the models (Barnes and Hartmann, 2010). As described in AR4, climate models in the past have been unable to simulate the recent observed level of multi-decadal variability in the NAO/NAM without imposed forcing. However, at least some AOGCMs have now been shown to exhibit multi-decadal variability of this magnitude in unforced control simulations (Selten et al., 2004; Semenov et al., 2008). Several potential influences have been suggested for the observed variations, so it is unclear to what extent the underestimation of late 20th century trends reflects real problems with the variability in models. Further evidence has emerged of the coupling of NAM variability between the troposphere and the stratosphere, and even climate models with improved stratospheric resolution appear to underestimate the vertical coupling (Morgenstern et al., 2010). 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) also 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.

9.4.3.2 Atlantic Modes

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.4.3.2.1 AMOC variability

On short timescales of less than 5 years, some AMOC simulations have shown a comparable magnitude of variability as the RAPID array observations at 26°N, although with a tendency to underestimate the variability (Baehr et al., 2009; Balan Sarojini et al., 2011; Marsh et al., 2009). This underestimate results from an underestimate of the density-driven contribution (Balan Sarojini et al., 2011), and might point to deficiencies in the simulation of temperature and salinity in the models (Baehr et al., 2009). Some of these deficiencies in the simulation of the AMOC 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).

On a broader perspective than 26°N, it has yet to be established how reliably models represent the meridional coherence of the AMOC. Variability of the AMOC observed outside the sub-tropical gyre further north has not been found to be coherent with AMOC variability observed at 26°N (Bingham and Hughes, 2009; Willis, 2010). And only recently, gyre-specific AMOC changes previously found in models (Baehr et al., 2009; Biastoch et al., 2008; Bingham et al., 2007) have also been identified in historical hydrographic data (Lozier et al., 2010), questioning for the variability the canonical picture of a single, basin-scale North Atlantic overturning.

9.4.3.2.1 Atlantic multi-decadal variability

[PLACEHOLDER FOR FIRST ORDER DRAFT] (Kravtsov and Spannagle, 2008)

[INSERT FIGURE 9.26 HERE]

Figure 9.26: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure illustrating models' ability to simulate Atlantic multidecadal variability. The placeholder figure merely shows this for one particular model (CCCma).

9.4.3.2.2 Tropical zonal and meridional modes

[PLACEHOLDER FOR FIRST ORDER DRAFT: Atlantic Niño, WES; given the likely focus on MOC predictability in Chapter 11, part of this could fit in this other chapter.]

9.4.3.3 Pacific Modes

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.4.3.3.1 El Niño-Southern Oscillation

The El Niño-Southern Oscillation (ENSO) phenomenon is well recognized as the dominant mode of natural climate variability in the tropical Pacific on seasonal to interannual time scales. During the last decades, there has been steady progress in the simulation and seasonal prediction of ENSO and its global impacts using AOGCMs (AchutaRao and Sperber, 2002; Guilyardi et al., 2009b; Randall et al., 2007). Improvements in model formulation have led to a better representation of the spatial pattern of the sea surface temperature (SST) anomalies in the eastern Pacific and of ENSO's periodicity (AchutaRao and Sperber, 2006). Compared to previous generation models, some of CMIP3 models can now not only simulate the mean state and the annual cycle with some degree of fidelity but also the tropical interannual variability, without the use of the flux corrections. Indeed, many AOGCMs now exhibit a behaviour that is qualitatively similar to that of the real-world ENSO – a considerable achievement given the complexity of the interactions involved. Most of the CMIP3 models can now produce a reasonable climatology and ENSO without flux adjustments, enhancing their physical credibility for simulating ENSO and its response to climate changes. [to be updated with CMIP5 results]

Despite this progress, multi-model analyses show that serious systematic errors remain in the simulated background climate (time mean and annual cycle, see Section 9.3.3) as well as in the simulated natural variability (Capotondi et al., 2006; Guilyardi, 2006; van Oldenborgh et al., 2005; Wittenberg et al., 2006) and updates. Several studies have pointed out that these coupled models errors can often be traced back to the atmosphere component and more specifically 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). [to be updated with CMIP5 results]

AOGCMs produce a variety of El Niño variability time scales (Figure 9.27: Niño 3 spectra): model spectra range from very regular near-biennial oscillations to spectra that are close to the observed 2 to 7 years (DISCUSS CHANGE CMIP3 to CMIP5, if any). 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 some tendency to have ENSO peaking in boreal winter (refs). All these biases combine to generate errors in ENSO amplitude, period, irregularity, skewness or spatial patterns (Guilyardi et al., 2009b; Leloup et al., 2008; and other CMIP5 references). [ENSO diversity (CP vs. EP) to be discussed.]

[INSERT FIGURE 9.27 HERE]

Figure 9.27: [PLACEHOLDER FOR FIRST ORDER DRAFT] ENSO power spectra for different models. [Placeholder figure taken from AR4 – to be updated]

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) (see metrics). The complex interactions, the main biases, and the variability described above (and with a number of likely others as discussed below) together with model structural diversity make it difficult to

clearly identify the origin of deficiencies in simulated ENSO. Nevertheless, it is likely that progress can be made. AOGCMs do appear now to exhibit many of the key processes and interactions thought to control the ENSO cycle in the real world (as detailed in Section 9.6.2).

Observations and climate models alike indicate large multi-decadal changes in ENSO properties (Li et al., 2011; Wittenberg, 2009). Individual ENSO proxies each record a local ENSO teleconnection. Hence, the non-stationarity of ENSO teleconnections together with the diversity of events (e.g., east Pacific El Niño versus central Pacific El Niño, or strong La Niña versus weak La Niña) makes it very difficult to infer robust changes about ENSO itself from any single proxy. Understanding the relative roles of external forcing (orbital, solar cycle, GHG) vs. internal dynamics in shaping this variability requires ambitious proxy synthesis, to validate the models (REFS TO COME). (Lin, 2007a) shows that CMIP3 models display a wide range of skill in simulating the interdecadal variability of ENSO. 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 many 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 displays 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. [to be updated with CMIP5 results]

9.4.3.3.2 Pacific decadal variability (PDO)

[PLACEHOLDER FOR FIRST ORDER DRAFT]
What is new since AR4 (research and CMIP analysis). Some potential references:

Parker, D., C. Folland, A. Scaife, J. Knight, A. Colman, P. Baines, and B. Dong (2007), Decadal to multidecadal variability and the climate change background, J. Geophys. Res., 112, D18115, doi:10.1029/2007JD008411.

Newman, Matthew, 2007: Interannual to Decadal Predictability of Tropical and North Pacific Sea Surface Temperatures. J. Climate, 20, 2333–2356. doi:10.1175/JCLI4165.1.

Kwon, Young-Oh, Michael A. Alexander, Nicholas A. Bond, Claude Frankignoul, Hisashi Nakamura, Bo Qiu, Lu Anne Thompson, 2010: Role of the Gulf Stream and Kuroshio–Oyashio Systems in Large-Scale Atmosphere–Ocean Interaction: A Review. J. Climate, 23, 3249–3281, doi:10.1175/2010JCLI3343.1.

9.4.3.3.3 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 (inccrease) in tropical Pacific SST is significantly correlated with a spin-up (spin-down) of the STCs. 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.) 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.

9.4.3.4 Indian Ocean Variability

In the Indian Ocean region, variability is dominated by the Indian Ocean zonal dipole mode (IOD) (Saji et al., 1999; Webster et al., 1999) and, at higher latitudes, by the subtropical SST dipole mode (Behera and Yamagata, 2001) which appears part of a hemispheric response to tropical atmospheric forcing (Fauchereau et al., 2003; Hermes and Reason, 2005). The latter are discussed in Section 9.4.3.8. The ability of CMIP3 AOGCMs to represent the general features of the IOD appears related to their ability to simulate the mean state of the equatorial Indian Ocean (Saji et al., 2006). These models show a large spread in the modelled depth of the 20^{o}C isotherm in the eastern equatorial Indian Ocean. It is unclear why the CMIP3 models show

such a large spread but it may be related to differences in the various model parameterisations of vertical mixing (Schott et al., 2009). Another source of error in the ability of models to represent the IOD arises because one of the triggers for this mode is ENSO and the models have difficulty in representing the ocean teleconnection of ENSO through the Indonesian archipelago. Models that simulate a deeper thermocline off Sumatra also tend to show a larger correlation value between indices of ENSO and the IOD than do models with a shallower thermocline (Saji et al., 2006). Ihara et al. (Ihara et al., 2009) showed that, under global warming scenarios, the western equatorial Indian Ocean warms more than the east, i.e., in the same direction as happens in a positive IOD event. Both the interannual and lower frequency variability of the tropical dipole mode index appear to be well captured by the SINTEX-F1 CGCM (Tozuka et al., 2007).

9.4.3.5 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, xxxx]) as well as ozone (Butchart et al., 2003; Shibata and Deushi) 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).

9.4.3.6 Intraseasonal Variability

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.4.3.6.1 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 GCMs 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 in the events identified as blocking and their climatologies (Barriopedro et al., 2010a). The diagnosed blocking frequency can be very sensitive to details of the method used, 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 potentially 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 upwards 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). In terms of more general analyses of circulation regimes, there is further evidence that climate models can simulate the broad features of observed regimes (Teng et al., 2007). There is also further 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 level from reasonable to poor 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.

9.4.3.6.2 Wind events and ENSO

In addition to the Madden Julian Oscillation (MJO), a large part of the intraseasonal wind variability over the Pacific warm pool occurs as Westerly wind Events (WWEs) (Harrison and Vecchi, 1997). Both observational and modelling studies demonstrated that WWEs are associated with the onset and maintenance of warm SST conditions (Gebbie and Tziperman, 2009; Lengaigne et al., 2002; Lengaigne et al., 2003; Marshall et al., 2009). Air-sea interactions involving WWEs, oceanic Kelvin waves and western Pacific warm pool were therefore suggested to be instrumental to the fast growth of El Niño events and to the modulation of ENSO characteristics (see Lengaigne et al., 2004a for a review). In addition to WWEs, ISO wind variability in the eastern part of the basin was also on some occasion suggested to influence ENSO events both in observations (Takayabu et al., 1999) and models (Lengaigne et al., 2004b). [to be updated with equivalent studies in CMIP3/CMIP5 models: e.g., Neale et al., 2008]

9.4.3.6.3 ENSO forcing of the NAO

ENSO has remote influences throughout the tropics and extratropical effects are well established over North America. Although Atlantic links have been much less clear (e.g., Trenberth and Caron, 2000), recent observational evidence points to a clear relationship between ENSO and the North Atlantic with a negative NAO response in late winter during El Nino and the opposite during La Nina (Bronnimann et al., 2004; Moron and Gouirand, 2003). During El Nino there is a southward shift in the Atlantic storm track and cold conditions over Northern Europe and Eastern United States and warm anomalies over Canada and the Mediterranean region, all of which have been reproduced in model studies (Bulic and Brankovic, 2007; Cagnazzo and Manzini, 2009; Fraedrich and Müller, 1992; Ineson and Scaife, 2009). There are tropospheric pathways by which the tropical Pacific affects Atlantic-European winter climate, for example via a Rossby wave emanating from the tropics to the North Atlantic (Bell et al.; Toniazzo and Scaife) or directly across the North American continent due to downstream effects of the Pacific North American pattern. A second type of pathway occurs through modulation of the Aleutian Low in the North Pacific and a change in planetary wave driving of the stratosphere (Hamilton; Manzini et al.; Taguchi and Hartmann; van Loon and Labitzke) and its subsequent tropospheric response which may also explain the intraseasonal transition and the strong late Winter response to ENSO (Bell et al.; Cagnazzo and Manzini; Ineson and Scaife). Other aspects are still debated such as the robustness of the modeled response to La Nina (Manzini et al., 2006) and the apparent non-linearity or possibly non-stationarity of the ENSO response (Greatbatch et al., 2004; Mathieu et al., 2004; Toniazzo and Scaife, 2006). Recent evidence however points to a stationary response in time (Bronniman 200?). The connection between ENSO and the NAO also provides a potential role for ENSO in the future climate response over the North Atlantic (e.g., Müller and Roeckner, 2006). [to be updated with related studies in CMIP5 models]

9.4.3.6.4 MJO and North Atlantic weather regimes

[PLACEHOLDER FOR FIRST ORDER DRAFT: Cassou (2008) and Lin et al. (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. Studies analysing this in models to be confirmed]

[Other interactions: NPO/ENSO (Seasonal footprinting mechanism) e.g., Alexander et al., 2010]

9.4.3.7 Teleconnections

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.4.3.7.1 Pacific North American pattern

[PLACEHOLDER FOR FIRST ORDER DRAFT] Some references:

Stoner, Anne Marie K., Katharine Hayhoe, Donald J. Wuebbles, 2009: Assessing General Circulation Model Simulations of Atmospheric Teleconnection Patterns. *J. Climate*, **22**, 4348–4372.

Seager, Richard, Robert Burgman, Yochanan Kushnir, Amy Clement, Ed Cook, Naomi Naik, Jennifer Miller, 2008: Tropical Pacific Forcing of North American Medieval Megadroughts: Testing the Concept with an Atmosphere Model Forced by Coral-Reconstructed SSTs. *J. Climate*, **21**, 6175–6190.

Linkin, Megan E., Sumant Nigam, 2008: The North Pacific Oscillation–West Pacific Teleconnection Pattern: Mature-Phase Structure and Winter Impacts. *J. Climate*, **21**, 1979–1997.

9.4.3.7.2 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 (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.

9.4.3.7.3 Southern Hemisphere subtropical basin modes

Despite their association with the rainfall variability of large areas in the Southern Hemisphere (Behera and Yamagata; Nicholls; Reason), there has been little attempt to assess the ability of coupled models to represent the generation and evolution of these modes. Suzuki et al. (Suzuki et al., 2004) used the Japanese NIED CGCM to study the annual cycle of the subtropical dipole mode in the South Indian Ocean, finding that it is mainly driven by atmospheric anomalies and that the SST changes only impact back on the atmosphere late in the austral summer phase of this mode.

Other references found: Watterson, I. G. (2009), Components of precipitation and temperature anomalies and change associated with modes of the Southern Hemisphere. International Journal of Climatology, 29: 809–826. doi:10.1002/joc.1772.

9.4.3.7.4 ENSO – West African Monsoon

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., 2007). 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.

9.4.3.7.5 Other teleconnections

[PLACEHOLDER FOR FIRST ORDER DRAFT]

[Link between SST and fluxes (Shin, S.-I., P. D. Sardeshmukh, and K. Pegion (2010), Realism of local and remote feedbacks on tropical sea surface temperatures in climate models, J. Geophys. Res., 115, D21110, doi:10.1029/2010JD013927.) Coordination with Chapter 11 (Section 11.2.1 on internal mechanisms of variability) and Chapter 14 (Section 14.3) needed. Power spectra to be included. Cold ocean and warm land pattern to be included if new studies.]

9.4.3.9 Interannual Variability in Terrestrial Sources and Sinks

Both coupled biogeochemistry/land 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., 2010c).

When these components are linked into 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.

9.4.3.10 Internal and Interannual Variability in Ocean Sources and Sinks

Most of the biogeochemical ocean 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.

9.4.4 Extreme Events

Extreme events, which are of great importance for the society as they bring about natural disasters, are realizations of a tail of probability distribution functions of natural variability of climate or weather. They are higher-order statistics than climatological mean states and, thus, generally expected to be more difficult to realistically represent in climate models. Extreme events take place in various time scales, such as daily, intra-seasonal, inter-annual and inter-decadal, corresponding to the different time scales of variability. 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).

9.4.4.1 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 performance 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. Meehl et al. (2009a) found that a 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 due to more uniform warming across the USA. Christidis et al. (2007) found that the observed lengthening of the growing season (often considered as one of extremes indices) in the second half of 20th century was consistently simulated by climate models in a context of detection and attribution. See also Chapter 10.

[Placeholder for CMIP5 results. CMIP3 results might be replaced by CMIP5 ones, or improvement seen in CMIP5 over CMIP3 might be described.]

9.4.4.2 Extreme Precipitation

(Kharin et al., 2007) concluded that precipitation extremes simulated by CMIP3 models were plausible in the extratropics, but uncertainties in the Tropics are very large, both in the models and the available observationally based datasets. Simulated precipitation extremes are known to be highly resolution

dependent. Growing evidence has shown that high-resolution models that are recently available (0.5 degree or higher atmospheric horizontal resolution) can capture the magnitude of extreme precipitation realistically (e.g., Wehner et al., 2010). See also Section 9.6.3.3.

Since AR4, attention to a higher temporal resolution for extreme precipitation has emerged, i.e., hourly precipitation extremes, rather than daily. Lenderink and Meijgaard (2008) have analysed 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. This feature was partly (only for very high extremes) reproduced by a 25 km resolution regional climate model.

Similar to the temperature extremes, studies that compared modelled and observed historical trends of precipitation extremes have emerged. (Min et al., 2011) 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, albeit with a smaller amplitude than observed [See also Chapter 10].

[Placeholder for CMIP5 results. CMIP3 results might be replaced by CMIP5 ones, or improvement seen in CMIP5 over CMIP3 might be described.]

9.4.4.3 Other Extremes

One of the remarkable findings on other extremes 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 notably encouraging in the context of model evaluation.

[PLACEHOLDER FOR FIRST ORDER DRAFT: More literature on various kinds of extremes either from CMIP3 or CMIP5 analyses.]

9.4.4.4 Summary

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.5 Downscaling and Simulation of Regional Scale Climate

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.5.1 Introduction

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.5.1.1 Downscaling

Downscaling is a collection of methods that are used to add detail to AOGCM results. Regional scale climate downscaling and simulation is pursued with various statistical methods , regional climate models (RCM, a.k.a dynamical downscaling), and variable-resolution-GCMs. RCMs are comparable to AOGCMs in their physical representation of climate processes, except that *coupled* regional climate models are still rather few (Dorn et al., 2009; Doscher et al., 2010; Smith et al., 2011a; Somot et al., 2008). As in the case of AOGCMs, coupled RCMs can provide a fuller representation of the regional climate system.

The typical resolution of regional climate models is higher than that of typical AOGCMs. This implies more detailed representation and simulation of processes that are affected by variable topography, as well as of other spatial detail including many types of extremes. A basic tenet is that RCMs' provide added value compared to AOGCMs. However, this is not self-evident (Laprise et al., 2008). Studies do, however, indicate that added value indeed arises.

Downscaling facilitates climate change impact research, but it is also a tool for testing and developing process descriptions. AOGCMs exhibit various systematic biases in their regional circulation regimes. RCMs, on the other hand, can be driven with boundary conditions from global reanalyses, in so-called "perfect boundary condition" mode. This allows some separation of biases in the representation of regional processes due to large-scale biases and arising from inadequacies in the process descriptions themselves. Such simulations are also more straight-forward to compare to observations, as the large-scale variability corresponds to the observed one.

9.5.1.2 Recent Developments of the Downscaling Techniques

Recent developments of regional climate downscaling concern longer perfect boundary condition simulations, increased resolution, coordinated experiments and ensemble analyses as well as extension of model evaluation to additional processes (Rummukainen, 2010). Studies of RCM internal variability has led to new insights on the applicability of deterministic and statistical evaluation, respectively. The use of RCMs has also spread to include most land regions of the world.

Perfect boundary condition simulations have benefited from the continued development of global reanalyses at the ECMWF, NCEP and JMA. For example, simulations that cover several decades allow for extended evaluation of variability and extremes compared to runs that address shorter periods. The latter are still useful, however, in probing the representation of physical processes. For example, Yhang and Hong (2008) test improvements of land surface, boundary layer and cumulus parameterisation schemes in a specific RCM and arrive at a better simulation of the East Asian summer monsoon. Dorn et al. (2009) show that in a coupled RCM for the Arctic, substantial improvements on sea ice simulation can be gained by improved ice growth, ice albedo and snow cover descriptions.

RCM resolution has in many cases increased to 10–25 km, although use is still made of 50 km resolution runs as well. Higher resolution facilitates evaluation of extremes, but also challenges the availability of observations for model evaluation (e.g., Nikulin et al., 2011; Driouech et al., 2009). If observations are available gridded at a lower resolution than the resolution in an RCM, the observational estimates can be expected to be more attenuated compared to reality, and the RCM. This can be even more of an issue in case of extremes in an RCM with a subgridscale land cover separating forested and open land surface conditions, while gridded observations come from measurements in open land locations. On the other hand, station observations may significantly differ from what is simulated in a grid box of an RCM over complex terrain. Influences like these need to be accounted for in assessing downscaling skill.

Specific studies with particular RCMs can provide useful insights. However, as is the case for AOGCMs (see Section 9.2.3), more comprehensive assessment of models' uncertainty and skill can be facilitated with coordinated experiments. Such regional downscaling studies have recently become much more 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).

[PLACEHOLDER FOR FIRST ORDER DRAFT: CORDEX]

[INSERT FIGURE 9.28 HERE]
Figure 9.28: [PLACEHOLDER FOR FIRST ORDER DRAFT] Map of recent coordinated downscaling study regions. From Giorgi et al., WMO Bulletin 2009, 58:3, 175-183. [NB: Need copyright or redrawing.]

9.5.2 Fidelity of Downscaling Methods and Value Added

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.5.2.1 Assessment of Skill

Assessing the skill of downscaling methods uses many of the same approaches as those that are used for AOGCMs. However, there are also fundamental differences. Not least, the skill of regional climate downscaling is conditional on the quality of the driving data, e.g., lateral boundary conditions, as well as

SSTs if these are used as forcing for RCMs. The latter play a role also for variable-resolution AGCMs. The underlying RCM skill can be better assessed in perfect boundary condition simulations, as they allow for both deterministic and statistical assessment.

Models that are constrained by boundary conditions still exhibit internal variability which varies with the synoptic situation, season, as well as model domain and other choices in model set-up (Alexandru et al., 2007; Leduc and Laprise, 2009; Nikiema and Laprise, 2010; Rapaić et al., 2010). Internal variability limits deterministic RCM skill for specific situations, short periods and small scales, but these effects are small in climate applications (Laprise et al., 2008; Separovic et al., 2008). Internal variability can also be constrained by spectral nudging or other techniques (Misra, 2007); however, their application may also lead to deterioration of some features, e.g., precipitation extremes (Alexandru et al., 2009).

Comparison of modelled and observed climatological seasonal means has a long history as an evaluation method of RCM skill. Process-oriented skill assessment has,however, gained ground, e.g., whether RCMs can simulate observed relationships. Sasaki and Kurihara (2008) find, for a specific RCM running at high resolution, that it reproduces the observed relationships between precipitation and elevation. Driouech et al. (2010) show that a specific variable-resolution AGCM reproduces rather well the observed link between north Atlantic weather regimes and local precipitation. Hirschi et al. (2011b) find that a number of RCMs reproduce observed relationships between soil moisture and extreme temperature, but with some overestimation. This may indicate that changes in warm summertime conditions including heat waves, as well as precipitation features may suffer from systematic biases in climate change simulations (Christensen et al., 2008; Kostopoulou et al., 2009). Döscher et al. (2010) find that a coupled RCM reproduces empirical relationships between Arctic sea ice extent and areal sea ice thickness characteristics and NAO variations. Another area that has been under considerable scrutiny is the West African monsoon region (e.g., Boone et al., 2010; Druyan et al., 2010). Some general findings for the latter region are that RCMs display biases in their simulated latent heat fluxes, positive precipitation biases and a general tendency to place the monsoon too far to the north and to have it mistimed. This emphasises the need to pursue targeted evaluation of atmosphere-land surface coupling in the region (see also van den Hurk and van Meijgaard, 2010).

In addition to evaluating RCM results themselves, there is a need for quantitative measures of the value added by an RCM to the AOGCM results used to drive it. The field remains rather unexplored, but some attempts at performance-based ranking of RCMs have been made. There is, of course, a fair deal of subjectivity in designing such methods. Christensen et al. (2010) note that important aspects to cover are, first, those that RCMs can provide added value on, e.g., extremes and mesoscale signals and, second, basic climate data such as consistency with the driving boundary conditions. Some of the suggested metrics lead to striking differentiation among RCMs (Lenderink, 2010), whereas other choices do not. The latter may imply general skilfulness of models, or a shortcoming of the metric. Nevertheless, Coppola et al. (2010), Kjellström et al.(2010) demonstrate that weighted sets of RCMs outperform sets without weighting, in terms of bias and RMSE of temperature and precipitation. The degree to which this applies tends to vary, however, with the choice of the particular metric, region and season. An option is to combine several metrics, which, however, brings in an additional choice to be made – how to do the combining (Christensen et al., 2010).

In general, a fundamental issue is how evaluation findings on the performance of a downscaling method translate into skill in downscaling climate change projections. Buser et al. (2009) look at how assumptions on whether underlying biases in RCMs vary with changing conditions and whether this affects the projection outcomes. They find that the projected summertime warming in the European Alpine region, , obtained from a combination of several RCMs, was 3.4°C and 5.4°C for assumptions of constant bias and constant relationship, respectively. Changes in winter were not sensitive in this sense. Buser et al. (2010) find that considering these two bias assumptions together, leads to overall lower summer and autumn warming, and a larger winter warming in different parts of Europe, according to several RCMs. The reason for this contrast can be traced back to deficiencies in how the RCMs' reproduce interannual temperature variability in the different seasons.

9.5.2.2 Value Added

As has been discussed above, downscaling is expected to add value to AOGCM results, in particular in regions with variable topography (land-sea distribution, large lakes, orographic detail). The representation of

extremes is also expected to improve, given the higher resolution of regional climate models. Improved representation of extremens when resolution is regionally enhanced by means of interactive nesting – of an RCM in an AOGCM – has also been explored (Inatsu and Kimoto, 2009). On the other hand, a regional model appliced to East Asia was found to suppress storm activity that was excessive in the particular AOGCM under to provide the boundary conditions. Fox-Rabinovitz et al. (2008) found that stretched grid model configurations reduced errors compared to when they were run on a uniform, but lower, resolution. However, Déqué et al. (2010) found that while a collection of 9 AGCMs with variable resolution requires a computationally lower effort than an equivalent high resolution AGCM with a regular grid, the simulated mean climate is similar.

Some examples of specific value added are given by Feser (2006), who documented added value for near-surface temperature, and by Winterfeldt and Weisse (2009) who found that some coastal wind characteristics were improved by dynamical downscaling. Donat et al. (2010) found in turn that downscaling improved European storm damage estimates.

[Text to be added referring to Figures 9.29, 9.30 and 9.31]

[INSERT FIGURE 9.29 HERE]
Figure 9.29: Monthly mean model temperature bias vs. observed monthly mean temperature for Region MD (Mediterranean) for the period 1961–2000. Raw data shown for MPI model, while full curves represents polynomial fit to underlying data. Dashed curve segments based on fit excluding the 25% warmest of all months are added to the right. Taken from (Christensen et al., 2008)

[INSERT FIGURE 9.30 HERE]
Figure 9.30: Monthly mean model precipitation bias vs. observed monthly mean precipitation for Region SC (Scandinavia) for the period 1961–2000. Raw data shown for MPI model, while full curves represent polynomial fit to underlying data. Dashed curve segments based on fit excluding the 25% wettest of all months are added to the right. No model precipitation data available for the UCLM PROMES experiment due to a temporary error in the data archive. Taken from (Christensen et al., 2008)

[INSERT FIGURE 9.31 HERE]
Figure 9.31: [PLACEHOLDER FOR FIRST ORDER DRAFT] Global display of regional model skill, as a synthesis figure. Preferably [tbc] for the same regions as those displayed in [Figure 9.28]. Generic results for various regions (sample, appropriately, results from ENSEMBLES (Europe), NARCCAP (North America), RMIP (Asia), CLARIS (South America), CORDEX (multiple regions)). [So far, ENSEMBLES accommodates the largest set of models for a specific region and features the longest hindcasts; Figure to be created, no suitable placeholder available.]

9.5.3 Transferability Experiments

A more comprehensive approach to RCM evaluation can be obtained from results of experiments with coordinated simulation set-up (see Section 9.1.5.2). Some studies argue, however, that a coordinated model setup may favour or disfavour specific RCMs, in terms of assessing inter-RCM skill (Farda et al., 2010). This is because many RCMs are developed and subsequently applied primarily for a specific region. This may limit the range of climate processes, as well as evaluation, that are considered. Evaluation of RCMs run for different regions – in transferability experiments – may help to expose shortcomings that are not apparent in standard experiments

Some transferability studies have shown that RCMs exhibit home biases. For example, (Takle et al., 2007) evaluated the diurnal cycles of surface energy fluxes using five models and showed that models do have home domain advantages. Gbobaniyi et al. (2011, submitted) used data from five RCMs and showed that the European models demonstrate a clear home advantage over the North American models in simulating seasonal and diurnal cycles at a site in Cabauw, the Netherlands.

Studies have also shown that RCMs generally perform poorly over some specific geographical regions. For instance, in the Arctic Regional Climate Model Intercomparison Project (ArcMIP), RCMs show large-errors in simulating the 2-m temperature over land have, surface radiation fluxes, and cloud cover (Rinke et al.,

2006; Koltzow et al., 2003). In the international North American Monsoon Experiment (NAME), the simulation of intense convective precipitation over complex terrain provided a unique challenge to the RCMs (Gutzler et al., 2005). In the Baltic Sea Experiment (BALTEX), the models underestimated planetary boundary layer height, and as a result overestimated cloud amount below 900mb and underestimated it at 800mb (van Meijgaard et al., 2001). Roads et al. (2003) reported that the models used in the International Research Institute (IRI) project underestimated the total rainfall over South America; the models had excessive variability in simulated rainfall near the Andes, and performed better in winter and summer than in the transition seasons. Gbobaniyi et al. (2011, in press) found the RCMs exhibit some difficulty in reproducing the diurnal and seasonal temperature over the tropical stations in South America.

9.6 Sources of Model Errors and Uncertainty

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.6.1 Introduction

The previous Sections 9.3, 9.4, and 9.5 have dealt with climate models' capability to simulate recent and longer-term records, variability and extremes, and regional-scale climate, respectively. 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 9.6 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 conceptually be 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 AR4 significant progress has been made in understanding climate model error and spread. Section 9.6.2 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 9.6.3 considers targeted numerical experiments, devoted in turn to the application of climate models in weather forecasting ("Transpose AMIP", Section 9.6.3.1), useful because some important model errors manifest themselves within simulation days; the simulation of key periods in the past, useful because models are applied in configurations for which they have not been tuned (Section 9.6.3.2); the effect of high spatial resolution, important because a greater portion of processes is shifted from where we are uncertain about representation (unresolved) to where we know the underlying equations (resolved), thus reducing "modelling error" (Section 9.6.3.3); and perturbed-physics ensembles in which uncertain model parameters are varied systematically (Section 9.6.3.4). Section 9.6.4 assesses the rich literature since AR4 devoted to understanding the feedbacks causing spread in climate sensitivity, which we expect still to be the prime source of spread in projected climate warming (together with uncertainty in greenhouse gas emissions and ocean heat uptake, see Chapter 12). Section 9.6 ends with a synthesis (Section 9.6.5), linking the lessons learned from the process-based analysis assessed here to the model errors assessed in Sections 9.3, 9.4, and 9.5.

9.6.2 Process Oriented Evaluation

Our focus in previous Sections has been on the comparison of observed and simulated mean state of ECVs at the global scale. These routine tests provide a valuable summary of overall model performance but their broad scope does not isolate selected processes or feedbacks believed to be of importance for realistic simulation of the climate change. Process-oriented evaluation is often applied over limited areas to focus on particular processes or phenomena (e.g., monsoons, deep convection).

What is called process oriented evaluation in this Section is to measure agreement between a model simulation and observations 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 in the phenomena, hence ensuring the phenomena is correctly represented for the right reasons, and not via error compensation. Thus, it is often possible to relate the performance 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. This feature of process oriented evaluation is crucial for understanding model spread in future projections and constraining future projections by observed data.

Since AR4, following the pioneering work by Hall and Qu (2006), who have successfully related the model performance on current seasonal cycle to the credibility of snow albedo feedback in projected future climate, work in the same direction has been extended. Boe et al. (2009b) have demonstrated that most CMIP3 models underestimated significantly the observed trend in Arctic sea-ice decline. This, together with the relationship they found between historical and projected trends, leads to an observational constraint on future projection of sea-ice change (in this case, underestimation of sea-ice decline in most models was implied). Similarly, Boe et al. (2009a) have shown that many CMIP3 models might overestimate the efficiency of polar oceanic mixing. Together with the relationship between deep ocean heat uptake and transient global surface warming, this induces an observational constraint implying that many models may underestimate future transient surface warming.

Examples for evaluation in terms of the mechanisms involved in phenomena include that of ENSO. Starting from the linearized SST equation, Jin et al. (2006) and Kim and Jin (2010a) derived a coupled stability index (referred to the Bjerknes-index or BJ index) that details the ocean-atmosphere feedbacks involved in ENSO. They identified five different feedbacks: the mean advection and upwelling feedback (always negative), the thermal damping rate (due to surface heat fluxes and also negative), the zonal advection feedback (positive), the Ekman pumping feedback (positive) and the thermocline feedback (positive). Kim and Jin (2010b) applied this process-based analysis of ENSO to the CMIP3 multi-model ensemble and demonstrated a significant positive correlation between ENSO amplitude and ENSO stability as measured by the BJ index. When respective components of the BJ index obtained from the coupled models are compared with those from observations, it is revealed that most coupled models underestimate the thermodynamic damping effect and the positive effect of the zonal advective and thermocline feedback.

Process oriented evaluation is also useful for evaluating ability of a model to project regional phenomena. 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, which is a regional phenomenon. They evaluated the ability of each model in the CMIP3 ensemble to simulate the particular seasonal march of the storm-track activity and argued that the first spring storm over Japan is likely to occur earlier in a future warmer climate, based on the models that show the highest reproducibility of the storm-track activity measured with a particular metric.

Evaluation of chemical 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, which in this case was stratospheric ozone. This is a key aspect of the CCMVal model evaluation concept which is applied to more accurately identify the sources of model errors and to circumvent the case 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 have been assessed for the first time in the SPARC CCMVal report, and the upper troposphere / lower stratosphere (UTLS) has been explicitly examined (Gettelman et al., 2010; Hegglin et al., 2010). The model performance in the UTLS was found to be better than might have been expected based on the spatial resolution of the models, although the lack of observations is still a major limitation for model evaluation in this region. Radiation schemes have been shown to be sufficient for representing the major causes of observed stratospheric temperature changes and the main radiative drivers of surface climate. In addition, chemistry schemes have also been shown to agree generally well with benchmark schemes. The identification of model deficiencies through a process-oriented evaluation has overall 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 to identify remaining common systematic errors which include 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). An extension of the CCMVal concept of process-oriented evaluation to Earth system models was recently suggested by Eyring et al., J. Clim, in preparation, 2011. [to be updated with CMIP5 studies]

[Process-oriented cloud metrics to be discussed here; (Williams and Webb, 2009)]

[More examples from CMIP5 analyses to be added if available.]

9.6.3 Targeted experiments

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.6.3.1 Transpose AMIP

It is well understood that sources of numerous 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 highly valuable to test their response when initialized with an observed large-scale state, before errors in the processes can substantially alter or feedback to the large-scale. This test is 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 involving several of the atmosphere models used for IPCC climate simulations. 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; [Senior et al., 2009 book reference needed], 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. Recent examples following the Transpose-AMIP methodology include the identification that overestimates in winter-hemisphere tropical precipitation during solstice seasons is present in the first few days of a forecast of some models (Moncrieff et al., 2011) whereas 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 the simulated forecast precipitation have predictive power as 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 appear to 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., 2010). To the extent that errors are present from early in the forecast, this guides scientists to focus on improving the parameterisations directly rather than to focus on other processes such as a land-surface model or the interactions with the large-scale state of the sub-grid atmospheric processes. The Transpose-AMIP methodology also allows scientists involved with the improvement of climate models to test new parameterisations against advanced process observations that 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). [Results from Transpose-AMIP to be assessed once they become available.]

9.6.3.2 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 intermodel differences and the reasons of model mismatches with observations. Several groups have also run additional sensitivity experiments to test the role of model biases that have been identified.

The vegetation feedback, resulting from the replacement of one type of vegetation by another has received considerable attention because it was shown to produce a positive feedback on the initial climate perturbation (see AR4). New results for the LGM confirm the initial estimates and show that the vegetation induces a cooling ranging from [xx] to [yy] °C [reference needed]. [regional information to be added if available]. However models tend to produce [tbc based on Woilliez et al. (in preparation) and O'Ishi et al. (in preparation]. It also confirms that the role of CO_2 on vegetation account for about [xx%] of the vegetation change stressing that its variations need to be considered in the atmosphere coupling in model with dynamic vegetation (Harrison and Prentice, 2003). For the mid Holocene the vegetation feedback was shown to translate the seasonal insolation forcing in the northern hemisphere into an annual mean warming, but model tends to produce too large mid continental aridity that is not seen in data (Wohlfahrt et al., 2008). These feedbacks have been revisited by (Otto et al., 2009). Using an experimental set up that allow to separate the response into the pure feedbacks from ocean and vegetation and their mutual interactions (Dallmeyer et al., 2010). They show that that the largest feedback in high latitude was due to the ocean-ice albedo feedback that produce warmer condition around the Arctic and that vegetation was responding to this warming and not directly causing it. [additional studies to be included if available.] In the tropical regions, simulations of the mid-Holocene with ESM including dynamic vegetation confirm [tbc] that the vegetation feedback is not as large as previously inferred from asynchronous coupling and that changes in soil moisture has an impact that can be as large as that of vegetation in some models [tbc] [include new results if available]. The new simulations confirms the role of vegetation temperature, but show that model biases in the simulation of the present day simulation prevent the simulation of the right vegetation changes. O'Ishi et al. (in preparation).

Model biases in paleoclimate simulations have been investigated in different ways. For example, using the MIROC3.2 model coupled to a slab ocean model (Ohgaito and Abe-Ouchi, 2009) and Abe-Ouchi (in preparation) tested the impact of the SST biases in the control simulation. They show that the *pattern* of the preindustrial SST has larger impact than the *magnitude* 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. [Link with convection (Zheng, in preparation)].

The sensitivity of the simulated change in AMOC at the LGM to different aspect of the fresh water forcing, including river runoff, precipitation minus evaporation or ice calving has been reported in several publications (Kageyama et al., xxxx; Oki et al., in preparation; Xu et al., in preparation). 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)

Model resolution is another aspect that has been explored.

[If available, include millennium then discuss vertical resolution and the impact on better representation of the polar vortex and improved regional agreement of model results with paleo proxy data (Koerper et al., in preparation).]

[Overall conclusion to be added]

9.6.3.3 Sensitivity to Resolution

[PLACEHOLDER FOR FIRST ORDER DRAFT]

9.6.3.3.1 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 0.5 degree horizontal resolution realistically reproduce observed intensity of extreme daily precipitation over the continental United States, while lower resolution runs severely underestimated the intensity.

As for oceanic components, although in principle solving the same equations, laminar ocean models simulate an ocean that is quite distinct from that simulated by turbulent ocean models. 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) (Barnier et al. 2011). Results from the Drakkar hierarchy suggest that some large-scale climate indices, such as variability in the meridional overturning and poleward heat transport, may not be very sensitive to grid resolution.

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 10km or finer) coupled to fine scale (order 1/2 degree 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, in preparation) point to the dual importance of atmosphere and ocean resolution for ameliorating biases associated with the double ITCZ occuring in the tropical Pacific (Lin, 2007a).

Overall, recent evidence confirms that improvement of resolution is a promising way to improve various aspects of model performance. However, it should be noted that improved resolution sometimes causes a degradation of performance by breaking a balance of compensating errors in a well-tuned lower resolution model, as in the case of ENSO amplitude of Marti et al. (2010).

9.6.3.3.2 RCMs

Although AOGCMs have increased their horizontal resolution, many impact studies require still higher resolution as input to hydrological or crop productivity modelling. Regional climate models, either singly or doubly nested, can be used to downscale to resolutions as fine as 4 km (e.g., Sasaki et al, 2008; Kusaka et al 2010). This increase in horizontal resolution generally causes an increase in precipitation especially around mountainous areas where topography becomes more accurately represented and peaks can be higher. RCM runs over North America (Leung and Qian, 2003) and over Europe using an ensemble of 9 different models (Rauscher et al, 2010) have illustrated this, and it has been shown that precipitation statistics from an RCM at high resolution tend to be more like observations than the driving AGCM (Kanada et al. (2010)).

The representation of surface processes in an RCM can have an important impact on the simulated climate. For example, Samuelsson et al (2010) show that by including a lake model in an RCM, the collective presence of the relative shallow lakes in Europe can contribute to an increase in air temperature over adjacent land. Improvements in the surface representation can generate finer atmospheric structures (Giorgi and Marinucci, 1996), and Castro et al (2005) show that as the resolution gets finer, the kinetic energy and its variance improve.

Regional Climate Models (RCMs) are strongly constrained by conditions prescribed on their lateral boundaries. In addition to sampling errors arising from natural internal variability (Laprise et al. 2008), errors in the RCM solution reflect both errors attributable to the RCM itself and those attributable to

boundary conditions or the manner in which they are prescribed. Typically, boundary conditions are applied ether by adopting a 'buffer zone' of 5 to 10 grid points around the RCM domain over which the boundary conditions are smoothly applied (Davies, 1976), or they are applied directly on the outer edge of the domain (Mesinger, 1977). Comparison between these schemes has shown no clear advantage of one over the other (Veljovic et al., 2010).

The effect of the boundary itself has been explored in so-called "Big-brother" experiments (Denis et al., 2002). This involves an RCM run at high resolution over a large domain: the big-brother run. To mimic coarse resolution driving data, a filter is applied to remove smaller scales from the big-brother run and the result is used as a boundary condition for the same RCM at the same high resolution but a smaller sub-domain: the little-brother run. Because the same model is used in the big-brother and the little-brother, the differences between the runs should be due to the initial and lateral boundary conditions. Comparisons show small differences, suggesting that RCMs can, in principle, simulate the small-scale climate features absent in the lateral boundary conditions.

The large-scale pattern of the atmosphere generated by the RCM should be consistent with that of the large-scale driving data. Some experiments have shown degradation by the RCM (Castro et al., 2005; Laprise et al., 2008) whereas (Veljovic et al., 2010) while others have shown that the RCM may actually improve the large-scale results, at least in terms of the upper level jet . (Castro et al., 2005) analyzed the spectra of RCM results at different resolution and domain size. They showed that as the grid size and domain size increased, the RCM retained less of the large scale data. In a forecasting context, Veljovic et al (2010) showed that RCM runs using ECMWF GCM ensemble forecasts as the lateral boundaries had better skill than the GCM forecast.

The location of the lateral boundaries can also have an important impact on RCM performance (Giorgi et al., 1996). For example, (Liang et al., 2001) and (Xue et al., 2007), using two different RCMs, found that domains which extended further south degraded the precipitation and low-level jet simulations possibly due to the uncertainties along the southern boundary.

The issue of the position of the lateral boundaries is also related to the issue of the domain size. When the lateral boundary is close to the area of interest, the downscaled climatology is strongly constrained by the driving boundary data, whereas further from the boundaries, small-scale features can be generated. This suggests that the size of the domain should be carefully chosen in order to achieve added value of the RCMs over the driver GCMs. In (Jones et al., 1995), four different 90-day RCM runs over Europe showed that the larger domain, which included the North Atlantic, diverged from the driving AOGCM data; the intermediate-size domain size produced more variability than the driving AOGCM, while the smallest domain produced less variability. Similar results were obtained by (Seth and Giorgi, 1998) and (Xue et al., 2007). By applying the Big-brother experiment, (Leduc et al., 2011) showed that the variance of the small-scale transient eddies of wind is underestimated in smaller domains and at upper levels, over North America summer, this underestimation was clearer in winter (Leduc and Laprise, 2009) when upper level winds are stronger. While the smaller scale features of precipitation improved in larger domains, the time correlation of precipitation decreased. Spectral nudging (von Storch et al., 2000) has been adopted by some RCMs to keep the large scales of the RCM solution close to that of the driving AOGCM .

9.6.3.4 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 that is 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., 2002b; 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., 2002b; Knutti et al., 2002; Sokolov et al., 2010; Stott and Forest, 2007; Tebaldi and Knutti, 2007b; Xiao et al., 1998). [Figure 9.32 to illustrate evaluation of climate model sensitivity based on 20th century observations; new results to be assessed as they become available.]

[INSERT FIGURE 9.32 HERE]

Figure 9.32: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure from (Stott and Forest, 2007) Probability distributions of Transient Climate Response estimated from climate change detection statistics and from calibrated EMICs. (Figure 8 and caption from (Stott and Forest, 2007)). Caption from: Probability distributions of TCR (expressed as warming rates over the century), as constrained by observed 20th century temperature change, and as calculated using HadCM3 (red), PCM (green), GFDL (blue) and from unweighted average of all three PDFs (turquoise). Coloured circles show each model's TCR. Also shown as diamonds are the TCRs of climate models forming part of the IPCC AR4 ensemble and the 5 to 95 percentile range derived by (Forest et al., 2006) using a large ensemble of EMICs is shown as the black bar with the star showing the median. [Data from (Knutti and Tomassini, 2008) to be incorporated.]

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, 2008a). 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.

9.6.4 Climate Sensitivity

[PLACEHOLER FOR FIRST ORDER DRAFT]

9.6.4.1 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 and the lapse-rate and in 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 × 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 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., 2010a).

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).

9.6.4.2 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).

9.6.4.3 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., 2010b). 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).

9.6.4.4 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.

9.6.4.5 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.

9.6.4.6 Role of Carbon-Cycle Feedbacks in Climate Sensitivity

[PLACEHOLDER FOR FIRST ORDER DRAFT: References include (Friedlingstein and Prentice, 2010), (Field et al., 2007),(Gregory et al., 2009),(Sitch et al., 2008),(Luo, 2007),(Thornton et al., 2009), possibly (Randerson et al., 2009), etc.]

9.6.4.7 Comparison Between Future Climate and Last Glacial Maximum

The AR4 reported on new attempts to relate the simulated LGM changes in tropical SST to global climate sensitivity, providing thereby a range of acceptable climate sensitivity valued from LGM proxy records (Hegerl et al., 2007). These studies tested either ensemble simulations with varying parameters or the PMIP2 multi-model ensemble, even though the latter was of small size (Crucifix, 2006). Edwards et al. (2007) proposed a synthesis of these studies (Figure 9.33) and discussed the value added and the limitation of combining constraint on past climate and of present day climate, such as the one developed in Annan and Hargreaves (Annan and Hargreaves, 2006). LGM Temperature changes in the tropics and in Antarctica have been shown to scale well with climate sensitivity, even though 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). These analyses have been extended to the LGM simulations realised as part of CMIP5 [reference needed]. It is thus possible to compare in a consistent way the major feedbacks operating in a warmer and a colder world for [zzz] models [Table xx]. [Feedbacks / future to be discussed]

[INSERT FIGURE 9.33 HERE]

Figure 9.33: [PLACEHOLDER FOR FIRST ORDER DRAFT] **[**Example of figure that could be drawn using CMIP5 results, compared to PMIP3 (the one included here) and to other ensemble simulations (new estimates should be available soon). It is only an example, but a completely new set of figures will be proposed when new results are available.] The legend for this figure is extracted from Edwards et al. (Edwards et al., 2007): Climate sensitivity as a function of tropical sea surface temperature change between the pre-industrial and the LGM for three model ensembles: the MIROC3.2 model with PMIP LGM boundary conditions (Annan, private communication); the CLIMBER-2 model with PMIP LGM boundary conditions; and CLIMBER-2 with additional dust and vegetation forcings (Schneider von Deimling, private communication). For comparison, five PMIP 2 coupled ocean-atmosphere GCMs are shown (Crucifix, 2006). The MIROC3.2 ensemble uses a simpler version of the model than PMIP 2. The vertical lines indicate the 1σ limits of reconstructed tropical SST change at the LGM from Ballantyne et al. (2005).

[The different constraints from data, either global or regional, suggest that model with climate sensitivity between [xxx] and [xxx] are in better agreement with paleo data; information about feedbacks to be included; information about ESM/OA to be included; and previous estimates (PMIP2) to be included.

Rapid synthesis of impact relative impact of different forcings and adjustment of tropical climate to be included if literature available. Conclusion on what it brings/future climate to be included.]

9.6.5 Discussion of Results in Context of Section 9.3 and Section 9.4

[PLACEHOLDER FOR FIRST ORDER DRAFT: this section will relate the mechanistic analysis of Section 9.6 back to the errors in the CMIP5 simulations assessed in Sections 9.3 and 9.4, i.e., can we explain some of the errors illustrated earlier? Where possible, examples of role of model complexity in model errors/biases to be included.]

1) Introductory examples of interactions of climate error and other types of simulation error
   a) Interaction of errors in surface stress (Ekman transport) and the oceanic carbon cycle
   b) Interactions of errors in rainfall and the terrestrial carbon cycle (e.g., Amazon sink)
2) Role of fast physics in setting biases, as established by Transpose AMIP
   a) Errors in mean climate, in particular the double ITCZ
   b) Errors in intraseasonal variability, i.e., MJO
   c) Errors in interannual to decadal variability, e.g., ENSO, and other major modes
   d) Errors in teleconnections, e.g., errors in standing waves due to errors in TWP
3) Insights from perturbed physics ensembles:
   a) Are models errors primarily parametric or structural in nature, or are other errors sources dominant?
   b) Information from new types of perturbed ensembles, e.g., coupled systems, on propagation of parametric uncertainty to other components of the climate system
   c) Information from perturbed ensembles in other components, e.g., land.
4) Insights from high-resolution modeling and RCMs
   a) To what extent are errors reduced with increasing resolution?
   b) Evidence for "irreducible" errors that resolution cannot mitigate due to uncertainties in parametric settings/structural formulation/initial conditions/boundary conditions.
   c) Evidence from eddy-permitting/resolving ocean models that identifies errors in thermohaline circulation, tracer mixing, and oceanic mixing timescales in conventional AOGCMs
   d) Is agreement and/or fidelity across a multi-model ensemble improved by increasing the resolution of the ensemble members?]

9.7 Relating Model Performance to Credibility of Projections

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 [for CMIP3 models; update], 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., 2008b; Reichler and Kim, 2008b), underlying key processes (Waugh and Eyring, 2008; Williams and Webb, 2009), and climate phenomena (Guilyardi et al., 2009b). 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 [tbc for CMIP5]. 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, agree better with observations in selected processes, phenomena or ECVs than others [tbc – the reverse might be true; to be specified].

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 range that has been observed. The ability of climate models to realistically reproduce observed climate processes, variability and interrelationships contributes to confidence in their ability to simulate future change in spite of the excursion into 'unknown territory'. Finally, 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, our

ability to directly evaluate long-term future climate projection is necessarily limited to inferences drawn from past performance.

What is clear is that 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 for many reasons. First, due to the longer time-scales, observations for verification are limited. Second, a conclusive link between model performance in the current climate and the response to climate change forcings has not been established. 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 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., 2005; 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 Section 12.4.1. There are some examples where the large inter-model variations in both mean climate and past trends are well correlated with comparably large inter-model variations in the model projections (Boe et al., 2009a; Boe et al., 2009b; Eyring et al., 2007; Hall and Qu, 2006; Mahlstein and Knutti, 2010). Hall and Qu (2006) showed that large inter-model variations in the seasonal cycle of the snow albedo feedback are strongly correlated with comparably large inter-model variations in snow albedo feedback strength on climate change timescales, see Figure 9.34 ((de Jong et al., 2009), (Hall and Qu, 2006)).

To date a robust approach to assigning weights to individual model projections of climate change has not been identified, and several studies point to the general difficulties in model weighting. Multiple studies have shown that different sets of performance metrics produce different rankings of models (Christensen et al., 2010; Gleckler et al., 2008; Schaller et al., 2011; Waugh and Eyring, 2008). For essential climate variables such as temperature, the correlations between the current mean climate state and the mean state in future projections are often weak (Knutti et al., 2010a; Raisanen et al., 2010; Whetton et al., 2007). Ultimately it is the realistic representation of processes on various spatial and temporal scales (e.g., monthly, annual, decadal, centennial), especially those related to important feedbacks in the climate system, that is expected to improve the credibility of model projections (Eyring et al., 2005; Knutti et al., 2010b). Hence a thorough evaluation of all aspects of climate models as carried out in this chapter appears prudent to guide the assessment of model quality.

In the end, weighting models based on performance metrics of present day climate must be considered carefully. It carries the risk of double counting observations that were already used in the model evaluation process, and the risk of obtaining biased or overconfident results if model weights are incorrectly specified or dependent on natural variability (Weigel et al., 2010), in particular in small ensembles. At present, statistical relationships among models that lead to refined distributions of climate sensitivity need to be supported by physical arguments in order to be credible (Klocke et al., 2011). But in cases where performance metrics are clearly related to inter-model spread in projections, and where the relationship between the observable and projection is robust and well understood, it might be feasible to weight models or to choose subsets of models (Allen et al., 2000; Forest et al., 2002a; Frame et al., 2006; Stott et al., 2006) [to be specified for CMIP5].

[INSERT FIGURE 9.34 HERE]

Figure 9.34: [PLACEHOLDER FOR FIRST ORDER DRAFT] Scatter plot of simulated springtime $\Delta\alpha_s/\Delta T_s$ values in climate change (ordinate) vs simulated springtime $\Delta\alpha_s/\Delta T_s$ values in the seasonal cycle (abscissa) in transient climate change experiments with 17 AOGCMs (α_s and T_s are surface albedo and surface air temperature, respectively). From Hall and Qu (2006). [to be replaced with promising new and not yet existing CMIP5 figure that relates model performance to projections.]

[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 since the first IPCC Assessment, and so in that sense current Earth System models are vastly 'better' than the models available when that assessment was written in the late 1980s. 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. So climate models of today are better 'in principle' than their predecessors; however, 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. So 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 compare different generations of models. This has been done by (Reichler and Kim, 2008a) who 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 results 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 and illustrates the ongoing improvement.

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 necessary that past climate, its variability and change, be well simulated. But this may not sufficient. Weather predictions, and seasonal climate predictions, can be verified on a regular basis, whereas climate projections spanning a century or more cannot, particularly as anthropogenic forcing is driving the climate system toward conditions not observed in the instrumental record. However, climate models are based on verifiable physical principles, and are able to reproduce many important aspects of past response to external forcing, and so they provide a scientifically sound preview of the climate to come.

[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: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure illustrating model improvement over time, based on some metric, or collection of metrics (as indicated by the coloured 'clouds'), computed for models participating in the various CMIP intercomparions. A figure along these lines was published by Reichler and Kim (2008), but something a bit different to be produced based on more recent results and better understanding of the properties of various model performance metrics.

[END FAQ 9.1 HERE]

Supplementary Material

[PLACEHOLDER FOR FIRST ORDER DRAFT] Material is largely unavailable at this point, drawing on results from CMIP5 for example.

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. *Proceedings of the National Academy of Sciences of the United States of America*, **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.

Alexander, M. A., D. J. Vimont, P. Chang, and J. D. Scott, 2010: The Impact of Extratropical Atmospheric Variability on ENSO: Testing the Seasonal Footprinting Mechanism Using Coupled Model Experiments. *Journal of Climate*, **23,** 2885-2901.

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***,** 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***,** 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., 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. D., 2011: **Climate model independence and agreement**. *Geophysical Research Letters*, **submitted**.

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.

Assmann, K. M., M. Bentsen, J. Segschneider, and C. Heinze, 2010: An isopycnic ocean carbon cycle model. *Geoscientific Model Development*, **3,** 143-167.

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.

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., and T. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, 581-584.
Baldwin, M. P., et al., 2001: The quasi-biennial oscillation. *Reviews of Geophysics*, **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., S. J., and W. T., 2011: A methodology for the comparison of blocking climatologies across indices, models and climate scenarios. *Climate*, **submitted**.
Barnier, B., et al., 2006: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dynamics*, **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.
Bartlein, P. J., et al., 2010: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics*.
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, C., L. Gray, A. Charlton-Perez, M. Joshi, and A. Scaife, 2009: Stratospheric Communication of El Nino Teleconnections to European Winter. *Journal of Climate*, 4083-4096.
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.
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.
Betts, R. A., et al., 2007: Projected increase in continental runoff due to plant responses to increasing carbon dioxide. *Nature*, **448,** 1037-U1035.
Biastoch, A., C. W. BÃ¶ning, J. Getzlaff, J.-M. Molines, and G. Madec, 2008: 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., and C. W. Hughes, 2009: Signature of the Atlantic meridional overturning circulation in sea level along the east coast of North America. *Geophys. Res. Lett.*, **36,** L02603-.
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.
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.
Blyth, E., J. Gash, A. Lloyd, M. Pryor, G. P. Weedon, and J. Shuttleworth, 2010: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data. *Journal of Hydrometeorology*, **11,** 509-519.
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*, -.

Bodas-Salcedo, A., et al., 2011: COSP: satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, **submitted**.

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*, -.

——, 2009b: September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience*, 341-343.

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.

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.

Braconnot, P., et al., 2007b: 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.

Branstator, G., and H. Y. Teng, 2010: Two Limits of Initial-Value Decadal Predictability in a CGCM. *Journal of Climate*, **23,** 6292-6311.

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.

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*, 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.

Bronnimann, S., J. Luterbacher, J. Staehelin, T. M. Svendby, G. Hansen, and T. Svenoe, 2004: Extreme climate of the global troposphere and stratosphere in 1940-42 related to El Nino. *Nature*, **431,** 971-974.

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.

Bryan, K., and L. J. Lewis, 1979: WATER MASS MODEL OF THE WORLD OCEAN. *Journal of Geophysical Research-Oceans and Atmospheres*, **84,** 2503-2517.

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.
Bulic, H., and C. Brankovic, 2007: ENSO forcing of the Northern Hemisphere climate in a large ensemble of model simulations based on a very long SST record. 231-254.
Buser, C., H. Kunsch, and C. Schar, 2010: Bayesian multi-model projections of climate: generalization and application to ENSEMBLES results. *Climate Research***,** 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. *Journal of Geophysical Research*, **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**.
Cagnazzo, C., and E. Manzini, 2009: Impact of the Stratosphere on the Winter Tropospheric Teleconnections between ENSO and the North Atlantic and European Region. *Journal of Climate*, **22,** 1223-1238.
Cao, C. S., and E. S. Titi, 2007: Global well-posedness of the three-dimensional viscous primitive equations of large scale ocean and atmosphere dynamics. *Annals of Mathematics*, **166,** 245-267.
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.
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.
Charlton, A., A. O'Neill, W. Lahoz, and A. Massacand, 2004: Sensitivity of tropospheric forecasts to stratospheric initial conditions. *Quarterly Journal of the Royal Meteorological Society***,** 1771-1792.
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.
Chepfer, H., M. Chiriaco, R. Vautard, and J. Spinhirne, 2007: Evaluation of MM5 optically thin clouds over Europe in fall using ICESat lidar spaceborne observations. *Monthly Weather Review*, **135,** 2737-2753.
Chepfer, H., S. Bony, D. Winker, M. Chiriaco, J. Dufresne, and G. Seze, 2008: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model. *Geophysical Research Letters***,** -.
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***,** 179-194.
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, S. Brown, D. J. Karoly, and J. Caesar, 2007: Human contribution to the lengthening of the growing season during 1950-99. *Journal of Climate*, **20,** 5441-5454.
Chung, E. S., B. J. Soden, and B. J. Sohn, 2010a: Revisiting the determination of climate sensitivity from relationships between surface temperature and radiative fluxes. *Geophysical Research Letters*, **37**.
Chung, E. S., D. Yeomans, and B. J. Soden, 2010b: An assessment of climate feedback processes using satellite observations of clear-sky OLR. *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*, **submitted**.
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.
CLIMAP, 1981: Seasonal reconstructions of the Earth's surface at the last glacial maximum. 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*, 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*, 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 formulation and atmospheric simulation of the Community Atmosphere Model version 3 (CAM3). *Journal of Climate*, **19,** 2144-2161.
——, 2006c: The Community Climate System Model version 3 (CCSM3). *J. Climate*, **19,** 2122-2143.
Connolley, W., and T. Bracegirdle, 2007: An antarctic assessment of IPCC AR4 coupled models. *Geophysical Research Letters*, -.
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*, 121-134.
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. *Clim. Dyn.*, **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,** -.
Cruz, F. T., A. J. Pitman, and Y. P. Wang, 2010: Can the stomatal response to higher atmospheric carbon dioxide explain the unusual temperatures during the 2002 Murray-Darling Basin drought? *Journal of Geophysical Research-Atmospheres*, **115**.
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.
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 J. Marshall, 2007: Effects of vertical variations of thickness diffusivity in an ocean general circulation model. *Ocean Modelling*, **18,** 122-141.
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.
Danilov, S., 2011: FESOM under Coordinated Ocean-ice Reference Experiment forcing. *Ocean Dynamics*, **accepted**.
Davies, H. C., 1976: LATERAL BOUNDARY FORMULATION FOR MULTILEVEL PREDICTION MODELS. *Quarterly Journal of the Royal Meteorological Society*, **102,** 405-418.
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.
De Szoeke, S. P., and S. P. Xie, 2008: The tropical eastern Pacific seasonal cycle: Assessment of errors and mechanisms in IPCC AR4 coupled ocean - Atmosphere general circulation models. *Journal of Climate*, **21,** 2573-2590.
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.
DeMott, C. A., D. A. Randall, and M. Khairoutdinov, 2010: Implied Ocean Heat Transports in the Standard and Superparameterized Community Atmospheric Models. *Journal of Climate*, **23,** 1908-1928.
Denis, B., R. Laprise, D. Caya, and J. Cote, 2002: Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment. *Climate Dynamics*, **18,** 627-646.
Deque, M., 2010: Regional climate simulation with a mosaic of RCMs. *Meteorologische Zeitschrift*, **19,** 259-266.
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.

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*, 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*, 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.
Donner, L. J., et al., 2011(in press): The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL Global Coupled Model CM3. *Journal of Climate*.
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*, 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.
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.
——, 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.
Eden, C., and R. J. Greatbatch, 2008: Towards a mesoscale eddy closure. *Ocean Modelling*, **20,** 223-239.
Eden, C., M. Jochum, and G. Danabasoglu, 2009: Effects of different closures for thickness diffusivity. *Ocean Modelling*, **26,** 47-59.
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.
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.
Eyring, V., et al., 2010a: Transport impacts on atmosphere and climate: Shipping. *Atmospheric Environment*, **44,** 4735-4771.
——, 2005: A strategy for process-oriented validation of coupled chemistry-climate models. *Bulletin of the American Meteorological Society*, 1117-+.
——, 2010b: Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models. *Atmospheric Chemistry and Physics*, 9451-9472.
——, 2007: Multimodel projections of stratospheric ozone in the 21st century. *Journal of Geophysical Research-Atmospheres*, -.
Faloona, I., 2009: Sulfur processing in the marine atmospheric boundary layer: A review and critical assessment of modeling uncertainties. *Atmospheric Environment*, **43,** 2841-2854.
Farda, A., M. Deque, S. Somot, A. Horanyi, V. Spiridonov, and H. Toth, 2010: Model ALADIN as regional climate model for Central and Eastern Europe. *Studia Geophysica Et Geodaetica*, 313-332.
Farneti, R., and P. R. Gent, 2011: The effects of the eddy-induced advection coefficient in a coarse-resolution coupled climate model. *Ocean Modelling*, **accepted**.
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.
Fedorov, A. V., and S. G. Philander, 2001: A Stability Analysis of Tropical Ocean-Atmosphere Interactions: Bridging Measurements and Theory for El Niño. *J. Climate*, **14,** 3086-3101.
Feldstein, S. B., and C. Franzke, 2006: Are the North Atlantic Oscillation and the Northern Annular Mode Distinguishable? *Journal of the Atmospheric Sciences*, **63,** 2915-2930.
Feng, J., and C. Fu, 2006: Inter-comparison of 10-year precipitation simulated by several RCMs for Asia. *Advances in Atmospheric Sciences*, 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.
Field, C. B., D. B. Lobell, H. A. Peters, and N. R. Chiariello, 2007: Feedbacks of terrestrial ecosystems to climate change. *Annual Review of Environment and Resources*, **32,** 1-29.
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**.
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.
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., P. Stone, A. Sokolov, M. Allen, and M. Webster, 2002a: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science***,** 113-117.
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, 2002b: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295,** 113-117.
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.: Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate simulations. *Ocean Modelling*, **In Press, Corrected Proof**.
——, 2011: Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate simulations. *Ocean Modelling*, **In Press, Corrected Proof**.
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***,** 159-177.
Fraedrich, K., and K. Müller, 1992: Climate anomalies in Europe associated with ENSO extremes. *Int. J. Climatol.*, **12,** 25-31.
Frame, D., D. Stone, P. Stott, and M. Allen, 2006: Alternatives to stabilization scenarios. *Geophysical Research Letters***,** -.
Friedlingstein, P., and I. C. Prentice, 2010: Carbon-climate feedbacks: a review of model and observation based estimates. *Current Opinion in Environmental Sustainability*, **2,** 251-257.
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.
Friend, A. D., et al., 2007: FLUXNET and modelling the global carbon cycle. *Global Change Biology*, **13,** 610-633.
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***,** -.
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**.
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.
Gangsto, R., M. Gehlen, B. Schneider, L. Bopp, O. Aumont, and F. Joos, 2008: Modeling the marine aragonite cycle: changes under rising carbon dioxide and its role in shallow water CaCO3 dissolution. *Biogeosciences*, **5,** 1057-1072.
Gates, W. L., et al., 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Am. Met. Soc.*, **80,** 29-55.
Gbobaniyi, E. O., 2010: Transferability of Regional Climate Models over different climatic domains, <st1:place w:st="on"><st1:city w:st="on">University of Cape.
Gbobaniyi, E. O., B. J. Abiodun, M. A. Tadross, B. C. Hewitson, and W. J. Gutowski, 2011 (in press): The coupling of cloud base height and surface fluxes: a transferability intercomparison. *Theoretical and Applied Climatology*.

——, 2011 (submitted): The coupling of cloud base height and surface fluxes: a transferability study over Cabauw. *International Journal of Climatology*.
Gebbie, G., and E. Tziperman, 2009: Predictability of SST-Modulated Westerly Wind Bursts. *Journal of Climate*, **22,** 3894-3909.
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.
Gedney, N., P. M. Cox, R. A. Betts, O. Boucher, C. Huntingford, and P. A. Stott, 2006: Detection of a direct carbon dioxide effect in continental river runoff records. *Nature*, **439,** 835-838.
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. *J. Phys. Oceanogr.*, **20,** 150-155.
Gent, P. R., and G. Danabasoglu, 2011: Climate Model Response to Increasing Southern Hemisphere Winds. *Journal of Climate*, **submitted**.
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.
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*, -.
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. *Climatic Change*, **100,** 787-795.
Giorgi, F., L. O. Mearns, C. Shields, and L. Mayer, 1996: A regional model study of the importance of local versus remote controls of the 1988 drought and the 1993 flood over the Central United States. *J. Climate*, **9,** 1150-1162.
Gleckler, P., K. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *Journal of Geophysical Research-Atmospheres*, -.
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*, -.
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.
Goosse, H., W. Lefebvre, A. de Montety, E. Crespin, and A. H. Orsi, 2009: Consistent past half-century trends in the atmosphere, the sea ice and the ocean at high southern latitudes. *Climate Dynamics*, **33,** 999-1016.
Greatbatch, R. J., 1994: A NOTE ON THE REPRESENTATION OF STERIC SEA-LEVEL IN MODELS THAT CONSERVE VOLUME RATHER THAN MASS. *Journal of Geophysical Research-Oceans*, **99,** 12767-12771.
Greatbatch, R. J., J. Lu, and K. A. Peterson, 2004: Nonstationary impact of ENSO on Euro-Atlantic winter climate. *Geophysical Research Letters*, **31**.
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., 2010: Long-term effect of volcanic forcing on ocean heat content. *Geophysical Research Letters*, -.
Gregory, J., and M. Webb, 2008: Tropospheric adjustment induces a cloud component in CO2 forcing. *Journal of Climate*, **21,** 58-71.
Gregory, J. M., C. D. Jones, P. Cadule, and P. Friedlingstein, 2009: Quantifying Carbon Cycle Feedbacks. *Journal of Climate*, **22,** 5232-5250.
Griffies, S. M., 2011: Dynamic sea level and the non-Boussinesq steric effect. *Ocean Modelling*, **submitted**.
Griffies, S. M., R. C. Pacanowski, M. Schmidt, and V. Balaji, 2001: Tracer conservation with an explicit free surface method for z-coordinate ocean models. *Monthly Weather Review*, **129,** 1081-1098.
Griffies, S. M., et al., 2000: Developments in ocean climate modelling. *Ocean Modelling*, **2,** 123-192.
——, 2005: Formulation of an ocean model for global climate simulations. *Ocean Sci.*, **1,** 45-79.
——, 2009a: Coordinated Ocean-ice Reference Experiments (COREs). *Ocean Modelling*, **26,** 1-46.
——, 2009b: Problems and Prospects in Large-Scale Ocean Circulation Models. *OceanObs09: Sustained Ocean Observations and Information for Society*, Venice, ESA.
Guilyardi, E., 2006: El Niño - mean state - seasonal cycle interactions in a multi-model ensemble. *Clim. Dyn.*, **26,** 229-348.
Guilyardi, E., P. Delecluse, S. Gualdi, and A. Navarra, 2003: Mechanisms for ENSO phase change in a coupled GCM. *J. Climate*, **16,** 1141-1158.

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-+.

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***,** 3047-3078.

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

Gutzler, D. S., et al., 2005: The North American Monsoon Model Assessment Project - Integrating numerical modeling into a field-based process study. *Bulletin of the American Meteorological Society*, **86,** 1423-1429.

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

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, 2010a: Impact of diurnal atmospher-ocean coupling on tropical climate simulations using a coupled GCM. *Climate Dynamics*, **34,** 905-917.

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

Hamilton, K., 1993: A general circulation model simulation of El Nino effects in the extratropical northern hemisphere stratosphere. *Geophys. Res. Lett.*, **20,** 1803-1806.

——, 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.

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.

Harrison, D. E., and G. A. Vecchi, 1997: Westerly Wind Events in the Tropical Pacific, 1986â“95*. *Journal of Climate*, **10,** 3131-3156.

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. *Bull. Am. Met. Soc.*, **90,** 1095-1107.

Haynes, J. M., Z. Luo, G. L. Stephens, R. T. Marchand, and A. Bodas-Salcedo, 2007: A Multipurpose Radar Simulation Package: QuickBeam. *Bulletin of the American Meteorological Society*, **88,** 1723-1727.

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

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***,** -.

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-+.

Held, I. M., and V. D. Larichev, 1996: A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta plane. *Journal of the Atmospheric Sciences*, **53,** 946-952.

Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen, 1995: THE PROJECT FOR INTERCOMPARISON OF LAND-SURFACE PARAMETERIZATION SCHEMES (PILPS) - PHASE-2 AND PHASE-3. *Bulletin of the American Meteorological Society*, **76,** 489-503.

Hendon, H., K. Sperber, D. Waliser, and D. J. Wheeler, submitted: Modelling monsoon intraseasonal variability: form theory to operational forecasting. *bulletin of the American Meteorological Society*.

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.
Hibler, W. D., 1979: A dynamic thermodynamic sea ice model. *J. Phys. Oceanogr.*, **9,** 815-846.
Hirschi, M., et al., 2011a: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, **4,** 17-21.
——, 2011b: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience***,** 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.
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**.
Hourdin, F., et al., 2010: AMMA-MODEL INTERCOMPARISON PROJECT. *Bulletin of the American Meteorological Society*, **91,** 95-+.
Huang, R. X., 1993: REAL FRESH-WATER FLUX AS A NATURAL BOUNDARY-CONDITION FOR THE SALINITY BALANCE AND THERMOHALINE CIRCULATION FORCED BY EVAPORATION AND PRECIPITATION. *Journal of Physical Oceanography*, **23,** 2428-2446.
Hunke, E. C., and J. K. Dukowicz, 1997: An elastic-viscous-plastic model for sea ice dynamics. *J. Phys. Oceanogr.*, **27,** 1849-1867.
Huntingford, C., P. M. Cox, L. M. Mercado, S. Sitch, N. Bellouin, O. Boucher, and N. Gedney, 2011: Highly contrasting effects of different climate forcing agents on ecosystem services. *Phil. Trans. Roy. Soc. A.*
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.
Ihara, C., Y. Kushnir, M. A. Cane, and V. H. de la Pena, 2009: Climate Change over the Equatorial Indo-Pacific in Global Warming. *Journal of Climate*, **22,** 2678-2693.
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***,** 2851-2868.
Ineson, S., and A. A. Scaife, 2009: The role of the stratosphere in the European climate response to El Nino. *Nature Geosci*, **2,** 32-36.
IOC, SCOR, and IAPSO, 2010: The international thermodynamic equation of seawater-2010: calculation and use of thermodynamic properties, 196 pp.
Irannejad, P., and A. Henderson-Sellers, 2007: Evaluation of AMIP II global climate model simulations of the land surface water budget and its components over the GEWEX-CEOP regions. *Journal of Hydrometeorology*, **8,** 304-326.
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.
Jansen, E., et al., 2007: Paleoclimate. *Climate Chane 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessement Report of the Intergovernmental Panel on Climate Change*, S. Solomon, et al., Eds., Cambridge University Press.
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*, -.
Joly, M., A. Voldoire, H. Douville, P. Terray, and J. F. Royer, 2007: African monsoon teleconnections with tropical SSTs in a set of IPCC4 coupled models. *Clim. Dyn.*, 1-32.
Jones, R. G., J. M. Murphy, and M. Noguer, 1995: SIMULATION OF CLIMATE-CHANGE OVER EUROPE USING A NESTED REGIONAL-CLIMATE MODEL .1. ASSESSMENT OF CONTROL CLIMATE, INCLUDING SENSITIVITY TO LOCATION OF LATERAL BOUNDARIES. *Quarterly Journal of the Royal Meteorological Society*, **121,** 1413-1449.
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*, 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.
——, 2010: Climate and carbon-cycle variability over the last millennium. *Climate of the Past*, **6,** 723-737.
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.
Kanzow, T., et al., 2009: Basinwide Integrated Volume Transports in an Eddy-Filled Ocean. *J. Phys. Oceanogr.*, **39,** 3091-3110.
——, 2010: Seasonal variability of the Atlantic meridional overturning circulation at 26.5°N. *J. Climate*, **23,** 5678–5698.
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*, -.
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.
Kawatani, Y., K. Hamilton, and S. Watanabe, 2011: The quasi-biennial oscillation in a double CO2 climate. *Journal of the Atmospheric Sciences*.
Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, **453,** 84-88.
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, 2011 (submitted): A Systematic Relationship between Intraseasonal Variability and <st1:place w:st="on"><st1:placename w:st="on">Mean <st1:placetype w:st="on">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, and T. Yasunari, submitted: Global monsoon, El Nino, and their linkage simulated by MIROC5 and the CMIP3 CGCMs.
Kim, H. J., B. Wang, and Q. H. Ding, 2008: The Global Monsoon Variability Simulated by CMIP3 Coupled Climate Models. *Journal of Climate*, **21,** 5271-5294.
Kim, S., and F.-F. Jin, 2010a: An ENSO stability analysis. Part I: results from a hybrid coupled model. *Climate Dynamics*, 1-15.

——, 2010b: An ENSO stability analysis. Part II: results from the twentieth and twenty-first century simulations of the CMIP3 models. *Climate Dynamics*, 1-19.
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*, 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., 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.
Klocke, D., R. Pincus, and J. Quaas, 2011: On constraining estimates of climate sensitivity. *Journal of Climate*, **submitted**.
Klocker, A., and T. J. McDougall, 2010: Influence of the Nonlinear Equation of State on Global Estimates of Dianeutral Advection and Diffusion. *Journal of Physical Oceanography*, **40,** 1690-1709.
Kloster, S., et al., 2010: Fire dynamics during the 20th century simulated by the Community Land Model. *Biogeosciences*, **7,** 1877-1902.
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? *Climatic 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*, -.
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.
Koltzow, M., S. Eastwood, and J. E. Haugen, 2003: Parameterization of snow and sea ice albedo in climate models.
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*, 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.
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. *Climatic Change*, **in press**.
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.
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.
Le Quere, C., et al., 2007: Saturation of the Southern Ocean CO2 sink due to recent climate change. *Science*, 1735-1738.
——, 2005: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, **11,** 2016-2040.
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*.
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. *Proceedings of the National Academy of Sciences of the United States of America*, **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.
Lemieux, J.-F., B. Tremblay, J. Sedlacek, P. Tupper, S. Thomas, D. Huard, and J.-P. Auclair, 2010: Improving the numberical convergence of viscous-plastic sea ice models with the Jacobian-free Newton-Krylov method. *Journal of Computational Physics*, **229,** 2840-2852.
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***,** 151-166.
Lenderink, G., and E. Van Meijgaard, 2008: Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, **1,** 511-514.
Lengaigne, M., J. P. Boulanger, C. Menkes, and P. Delecluse, 2004a: Westerly Wind Events in the Tropical Pacific and Their Influence on the Coupled Ocean-Atmosphere System: A Review. ***,** 1-21.
Lengaigne, M., J. P. Boulanger, C. Menkes, S. Masson, G. Madec, and P. Delecluse, 2002: Ocean response to the March 1997 Westerly Wind Event. *J. Geophys. Res*, **107,** 16,11-16,17.
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., J.-P. Boulanger, C. Menkes, G. Madec, P. Delecluse, E. Guilyardi, and J. Slingo, 2003: The March 1997 westerly wind event and the onset of the 1997/98 El Niño: understanding the role of the atmospheric response. *J. Climate*, **16,** 3330-3343.
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.
Lengaigne, M., et al., 2004b: Triggering of El Nino by westerly wind events in a coupled general circulation model. *Clim. Dyn.*, **23,** 601-620.
Lenton, A., F. Codron, L. Bopp, N. Metzl, P. Cadule, A. Tagliabue, and J. Le Sommer, 2009: Stratospheric ozone depletion reduces ocean carbon uptake and enhances ocean acidification. *Geophysical Research Letters***,** -.
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.
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.
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, 2011: 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.
Liang, X. Z., K. E. Kunkel, and A. N. Samel, 2001: Development of a regional climate model for US midwest applications. Part I: Sensitivity to buffer zone treatment. *Journal of Climate*, **14,** 4363-4378.
Lin, J.-L., 2007a: The Double-ITCZ Problem in IPCC AR4 Coupled GCMs: Ocean-Atmosphere Feedback Analysis. *Journal of Climate*, **20,** 4497-4525.
Lin, J. L., 2007b: 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.
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, 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.
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. *Quart. J. Roy. Met. Soc.*, **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.
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.
Luo, Y. Q., 2007: Terrestrial carbon-cycle feedback to climate warming. *Annual Review of Ecology Evolution and Systematics*, **38,** 683-712.
Lynch, A., P. Uotila, and J. Cassano, 2006: Changes in synoptic weather patterns in the polar regions in the twentieth and twenty-first centuries, part 2: Antarctic. *International Journal of Climatology***,** 1181-1199.
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***,** -.
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.
Manabe, S., 1969: CLIMATE AND THE OCEAN CIRCULATION1. *Monthly Weather Review*, **97,** 739-774.
Manzini, E., M. A. Giorgetta, M. Esch, L. Kornblueh, and E. Roeckner, 2006: The influence of sea surface temperatures on the northern winter stratosphere: Ensemble simulations with the MAECHAM5 model. *J. Climate*, **19,** 3863-3881.
Marchand, R., and T. Ackerman, 2010: An analysis of cloud cover in multiscale modeling framework global climate model simulations using 4 and 1 km horizontal grids. *Journal of Geophysical Research-Atmospheres*, **115**.
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. *Journal of Geophysical Research*, **114,** D18110.
Marshall, A. G., O. Alves, and H. H. Hendon, 2009: A Coupled GCM Analysis of MJO Activity at the Onset of El Niño. *Journal of the Atmospheric Sciences*, **66,** 966-983.
Marshall, D. P., and A. J. Adcroft, 2010: Parameterization of ocean eddies: Potential vorticity mixing, energetics and Arnold's first stability theorem. *Ocean Modelling*, **32,** 188-204.
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, D., and R. Knutti, 2011: Climate model genealogy. *Geophysical Research Letters*, **submitted**.
Masunaga, H., M. Satoh, and H. Miura, 2008: A joint satellite and global cloud-resolving model analysis of a Madden-Julian Oscillation event: Model diagnosis. *J. Geophys. Res.*, **113,** D17210.
Masunaga, H., et al., 2010: Satellite Data Simulator Unit: A Multisensor, Multispectral Satellite Simulator Package. *Bulletin of the American Meteorological Society*, **91,** 1625-1632.
Mathieu, P.-P., R. T. Sutton, B. Dong, and M. Collins, 2004: Predictability of Winter Climate over the North Atlantic European Region during ENSO Events. *Journal of Climate*, **17,** 1953-1974.
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-.
Mauritsen, T., 2011 (in preparation): The tuning of ECHAM6. *Journal of Advances in Modeling Earth Systems*.
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.
Maykut, G. A., and N. Untersteiner, 1971: Some results from a time dependent thermodynamic model of sea ice. *Journal of Geophysical Research*, **76,** 1550-1575.

McAvaney, B. J., et al., 2001: Model evaluation. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, et al., Eds., Cambridge University Press, 471-523.
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 J. A. Church, 1986: PITFALLS WITH THE NUMERICAL REPRESENTATION OF ISOPYCNIC AND DIAPYCNAL MIXING. *Journal of Physical Oceanography*, **16,** 196-199.
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.
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., 2007: Irreducible imprecision in atmospheric and oceanic simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **104,** 8709-8713.
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., 2007: The WCRP CMIP3 multimodel dataset - A new era in climate change research. *Bulletin of the American Meteorological Society*, **88,** 1383-+.
——, 2009b: DECADAL PREDICTION Can It Be Skillful? *Bull. Am. Met. Soc.*, **90,** 1467-1485.
Megann, A. P., A. L. New, A. T. Blaker, and B. Sinha, 2010: The Sensitivity of a Coupled Climate Model to Its Ocean Component. *Journal of Climate*, **23,** 5126-5150.
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.
Menendez, C., M. de Castro, A. Sorensson, J. Boulanger, and C. M. Grp, 2010: CLARIS Project: towards climate downscaling in South America. *Meteorologische Zeitschrift***,** 357-362.
Mercado, L. M., N. Bellouin, S. Sitch, O. Boucher, C. Huntingford, M. Wild, and P. M. Cox, 2009a: Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, **458,** 1014-1017.
——, 2009b: Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, **458,** 1014-U1087.
Mesinger, F., 1977: Forward-backward scheme, and its use in a limited area model. *Contr Atmos Phys*, **50,** 200-210.
Mesinger, F., et al., 2006: North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87,** 343-+.
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.
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.
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.
Mo, K. C., and G. H. White, 1985: TELECONNECTIONS IN THE SOUTHERN-HEMISPHERE. *Monthly Weather Review*, **113,** 22-37.
Mochizuki, T., et al., 2010: Pacific decadal oscillation hindcasts relevant to near-term climate prediction. *Proc. Natl. Acad. Sci. U. S. A.*, **107,** 1833-1837.
Moncrieff, M. W., D. E. Waliser, M. A. Shapiro, J. Caughey, G. R. Asrar, and L. A. Barrie, 2011: Year of Tropical Convection (YOTC): Concept and Research Strategy. *Bulletin of the American Meteorological Society*, **submitted**.
Morgenstern, O., et al., 2010: Anthropogenic forcing of the Northern Annular Mode in CCMVal-2 models. *J. Geophys. Res.*, **115,** D00M03-.
Moron, V., and I. Gouirand, 2003: Seasonal modulation of the El Niño-southern oscillation relationship with sea level pressure anomalies over the North Atlantic in October-March 1873-1996. *Int. J. Climatol.*, **23,** 143-155.
Morrison, H., and A. Gettelman, 2008: A new two-moment bulk stratiform cloud microphysics scheme in the community atmosphere model, version 3 (CAM3). Part I: Description and numerical tests. *Journal of Climate*, **21,** 3642-3659.
Moss, R. H., et al., 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463,** 747-756.

Mueller, P., 2010: Constructing climate knowledge with computer models. *Wiley Interdisciplinary Reviews: Climate Change*, **1,** 565-580.

Müller, W. A., and E. Roeckner, 2006: ENSO impact on midlatitude circulation patterns in future climate change projections. *Geophys. Res. Lett.*, **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***,** 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.

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.

Nicholls, N., 1989: Sea surface temperatures and Australian winter rainfall. *Journal of Climate|Journal of Climate*, **2,** 965-973.

Nikiema, O., and R. Laprise, 2010: Diagnostic budget study of the internal variability in ensemble simulations of the Canadian RCM. *Climate Dynamics***,** 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***,** 41-55.

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.

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.

Oden, J. T., and S. Prudhomme, 2002: Estimation of modeling error in computational mechanics. *J. Comput. Phys.*, **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.

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**.

Oreskes, N., K. Shraderfrechette, and K. Belitz, 1994: VERIFICATION, VALIDATION, AND CONFIRMATION OF NUMERICAL-MODELS IN THE EARTH-SCIENCES. *Science*, **263,** 641-646.

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, J., T. Raddatz, M. Claussen, V. Brovkin, and V. Gayler, 2009: Separation of atmosphere-ocean-vegetation feedbacks and synergies for mid-Holocene climate. *Global Biogeochemical Cycles*, **23,** -.

Paeth, H., et al., 2011: Progress in regional downscaling of west African precipitation. *Atmospheric Science Letters***,** 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. *Bull. Am. Met. Soc.*, **89,** 459-470.

Pechony, O., and D. T. Shindell, 2009: Fire parameterization on a global scale. *Journal of Geophysical Research-Atmospheres*, **114**.
Pennell, C., and *T. Reichler*, 2011: **On the Effective Number of Climate Models**. *Journal of Climate*.
Perlwitz, J., S. Pawson, R. Fogt, J. Nielsen, and W. Neff, 2008: Impact of stratospheric ozone hole recovery on Antarctic climate. *Geophysical Research Letters***,** -.
Petersen, A. C., 2000: Philosophy of climate science. *Bulletin of the American Meteorological Society*, **81,** 265-271.
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.
Piao, S. L., P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudre, D. Labat, and S. Zaehle, 2007: Changes in climate and land use have a larger direct impact than rising CO2 on global river runoff trends. *Proceedings of the National Academy of Sciences of the United States of America*, **104,** 15242-15247.
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., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences of the United States of America*, **106,** 8441-8446.
Pincus, R., C. Batstone, R. Hofmann, K. Taylor, and P. Glecker, 2008a: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *Journal of Geophysical Research-Atmospheres***,** -.
Pincus, R., C. P. Batstone, R. J. P. Hofmann, K. E. Taylor, and P. J. Glecker, 2008b: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *Journal of Geophysical Research-Atmospheres*, **113**.
Pincus, R., S. Platnick, S. A. Ackerman, R. S. Hemler, and R. J. P. Hofmann, 2011: Reconciling GCM-simulated and satellite-observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *Journal of Climate*, **submitted**.
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**.
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.
Pohlmann, H., J. H. Jungclaus, A. Koehl, D. Stammer, and J. Marotzke, 2009: Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *J. Climate*, **22,** 3926–3938.
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.
Raisanen, J., L. Ruokolainen, and J. Ylhaisi, 2010: Weighting of model results for improving best estimates of climate change. *Climate Dynamics*, **35,** 407-422.
Rammig, A., et al., 2010: Estimating the risk of Amazonian forest dieback. *New Phytologist*, **187,** 694-706.
Randall, D., et al., 2003a: Confronting models with data - The GEWEX cloud systems study. *Bulletin of the American Meteorological Society*, **84,** 455-469.
Randall, D. A., and B. A. Wielicki, 1997: Measurements, models, and hypotheses in the atmospheric sciences. *Bulletin of the American Meteorological Society*, **78,** 399-406.
Randall, D. A., M. F. Khairoutdinov, A. Arakawa, and W. W. Grabowski, 2003b: 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***,** 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 deterministic learning algorithm. *J. Comput. Phys.*, submitted.
Rayner, N. A., et al., 2003: Global analysis of sea surface temperature, sea ice, and night marine air temperature since the late ninteeth century. *Journal of Geophysical Research*, **108,** doi:10,1029/2002JD002670.
Reason, C. J. C., 1999: Interannual warm and cool events in the subtropical/mid-latitude South Indian Ocean region. *Geophysical Research Letters*, **26,** 215-218.
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. *J. Phys. Oceanogr.*, **12,** 1154-1158.
Reichler, T., and J. Kim, 2008a: How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, 303-311.
——, 2008b: How well do coupled models simulate today's climate? *Bull. Am. Met. Soc.*, **89,** 303-311.
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.
Rinke, A., et al., 2006: Evaluation of an ensemble of Arctic regional climate models: spatiotemporal fields during the SHEBA year. *Climate Dynamics*, **26,** 459-472.
Roads, J., et al., 2003: International Research Institute/Applied Research Centers (IRI/ARCs) regional model intercomparison over South America. *Journal of Geophysical Research-Atmospheres*, **108**.
Roberts, M., and D. Marshall, 1998: Do we require adiabatic dissipation schemes in eddy-resolving ocean models? *J. Phys. Oceanogr.*, **28,** 2050-2063.
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.
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.
Rojas, M., 2006: Multiply Nested Regional Climate Simulation for Southern South America: Sensitivity to Model Resolution. *Monthly Weather Review*, **134,** 2208-2223.
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*, in press.
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*, 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.
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*, 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. *Proceedings of the National Academy of Sciences of the United States of America*, 14778-14783.
——, 2007: Identification of human-induced changes in atmospheric moisture content. *Proceedings of the National Academy of Sciences of the United States of America*, 15248-15253.
——, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *International Journal of Climatology*, 1703-1722.
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.
Saunders, R., M. Matricardi, and P. Brunel, 1999: An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quarterly Journal of the Royal Meteorological Society*, **125,** 1407-1425.
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., 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.
Schaller, N., I. Mahlstein, J. Cermak, and R. Knutti, 2011: Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. *Journal of Geophysical Research*, **submitted**.
Scherrer, S. C., 2010: Present-day interannual variability of surface climate in CMIP3 models and its relation to future warming. *International Journal of Climatology*.
Schmittner, A., M. Latif, and B. Schneider, 2005: Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations. *Geophys. Res. Lett.*, **32,** L23710-.
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. *Journal of Geophysical Research*, **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.
Schwalm, C. R., et al., 2010: A model-data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research-Biogeosciences*, **115**.
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.
Scinocca, J. F., N. A. McFarlane, M. Lazare, Li, J., and D. Plummer, 2008: Technical Note: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, **8,** 7055–7074.
Sellers, P. J., et al., 1996: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. *J. 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. 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*, 4942-4961.
Seth, A., and F. Giorgi, 1998: The effects of domain choice on summer precipitation simulation and sensitivity in a regional climate model. *Journal of Climate*, **11,** 2698-2712.
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.
Shaw, T. A., and J. Perlwitz, 2010: The Impact of Stratospheric Model Configuration on Planetary-Scale Waves in Northern Hemisphere Winter. *Journal of Climate*, **23,** 3369-3389.

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**.
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.
Sigmond, M., and J. Fyfe, 2010: Has the ozone hole contributed to increased Antarctic sea ice extent? *Geophysical Research Letters***,** -.
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., P. M. Cox, W. J. Collins, and C. Huntingford, 2007: Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature*, **448,** 791-794.
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.
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.
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***,** 87-106.
Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317,** 796-799.
Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010a: Skilful multi-year predictions of Atlantic hurricane frequency. *Nature Geoscience*, **3,** 846-849.
Smith, P. C., N. De Noblet-Ducoudre, P. Ciais, P. Peylin, N. Viovy, Y. Meurdesoif, and A. Bondeau, 2010b: 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, 2010c: 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., and P. R. Gent, 2004: Anisotropic Gent-McWilliams parameterization for ocean models. *Journal of Physical Oceanography*, **34,** 2541-2564.
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.
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, H., 1971: On the representation of isentropic mixing in ocean models. *Journal of Physical Oceanography*, **1,** 233-234.
Solomon, S., R. Portmann, T. Sasaki, D. Hofmann, and D. Thompson, 2005: Four decades of ozonesonde measurements over Antarctica. *Journal of Geophysical Research-Atmospheres***,** -.
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***,** 112-126.
Son, S., et al., 2008: The impact of stratospheric ozone recovery on the Southern Hemisphere westerly jet. *Science***,** 1486-1489.
——, 2010: Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *Journal of Geophysical Research-Atmospheres***,** -.
SPARC-CCMVal, 2010: SPARC Report on the Evaluation of Chemistry-Climate Models.
Sperber, K., J. M. Slingo, and P. Inness, in press: Modelling the Madden-Julian Oscillation. Chapter 11. *Intraseasonal Variability of the Atmosphere-Ocean Climate System, 2nd edition*, W.-K.-M. L. a. D. E. Waliser, Ed., Praxis Publishing, Chichester, UK.

Sperber, K. R., and T. N. Palmer, 1996: Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. *Journal of Climate*, **9,** 2727-2750.
Sperber, K. R., and P. A. M. Grp, 1999: Are revised models better models? A skill score assessment of regional interannual variability. *Geophysical Research Letters*, **26,** 1267-1270.
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., et al., 2010: Monsoon Fact Sheet: CLIVAR Asian-Australian Monsoon Panel.
Stainforth, D. A., et al., 2005a: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433,** 403-406.
——, 2005b: 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.
Stevens, B., and D. H. Lenschow, 2001: Observations, experiments, and large eddy simulation. *Bulletin of the American Meteorological Society*, **82,** 283-294.
Stockli, R., et al., 2008: Use of FLUXNET in the community land model development. *Journal of Geophysical Research-Biogeosciences*, **113**.
Stott, P., J. Kettleborough, and M. Allen, 2006: Uncertainty in continental-scale temperature predictions. *Geophysical Research Letters***,** -.
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.
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**.
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.
Suzuki, R., S. K. Behera, S. Iizuka, and T. Yamagata, 2004: Indian Ocean subtropical dipole simulated using a coupled general circulation model. *Journal of Geophysical Research-Oceans*, **109**.
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.
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.
Taguchi, M., and D. L. Hartmann, 2006: Increased Occurrence of Stratospheric Sudden Warmings during El NiÃ±o as Simulated by WACCM. *Journal of Climate*, **19,** 324-332.
Takahashi, M., 1999: Simulation of the stratospheric quasi-biennial oscillation in a general circulation model. *Geophysical Research Letters*, **26,** 1307-1310.
Takayabu, Y. N., T. Iguchi, M. Kachi, A. Shibata, and H. Kanzawa, 1999: Abrupt termination of the 1997-98 El Nino in response to a Madden-Julian oscillation. *Nature*, **402,** 279-282.
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. *Reviews of Geophysics*, **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., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research-Atmospheres***,** 7183-7192.
Taylor, K. E., 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***,** 2053-2075.
——, 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.
Teng, Q., J. Fyfe, and A. Monahan, 2007: Northern Hemisphere circulation regimes: observed, simulated and predicted. 867-879.
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., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *Journal of Climate*, **13,** 1000-1016.
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.
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**.
Thornton, P. E., et al., 2009: Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model. *Biogeosciences*, **6,** 2099-2120.
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.
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.
Toniazzo, T., and A. A. Scaife, 2006: The influence of ENSO on winter North Atlantic climate. *Geophysical Research Letters*, **33**.
Tozuka, T., J. J. Luo, S. Masson, and T. Yamagata, 2007: Decadal modulations of the Indian Ocean dipole in the SINTEX-F1 coupled GCM. *Journal of Climate*, **20,** 2881-2894.
Trenberth, K. E., and J. M. Caron, 2000: The Southern Oscillation Revisited: Sea Level Pressures, Surface Temperatures, and Precipitation. *Journal of Climate*, **13,** 4358-4365.
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.
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***,** 467-481.
van Loon, H., and K. Labitzke, 1987: The Southern Oscillation. Part V: The Anomalies in the Lower Stratosphere of the Northern Hemisphere in Winter and a Comparison with the Quasi-Biennial Oscillation. *Monthly Weather Review*, **115,** 357-369.
van Meijgaard, E., U. Andrae, and B. Rockel, 2001: Comparison of model predicted cloud parameters and surface radiative fluxes with observations on the 100 km scale. *Meteorology and Atmospheric Physics*, **77,** 109-130.
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 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.
Vecchi, G. A., B. J. Soden, A. T. Wittenberg, I. M. Held, A. Leetmaa, and M. J. Harrison, 2006: 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.
Visbeck, M., J. Marshall, T. Haine, and M. Spall, 1997: Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Oceanogr.*, **27,** 381-402.
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.
von Storch, H., H. Langenberg, and F. Feser, 2000: A spectral nudging technique for dynamical downscaling purposes. *Monthly Weather Review*, **128,** 3664-3673.
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. E., et al., 2003: AGCM simulations of intraseasonal variability associated with the Asian summer monsoon. *Climate Dynamics*, **21,** 423-446.
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, 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., 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**.
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., 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.
Waugh, D., L. Oman, P. Newman, R. Stolarski, S. Pawson, J. Nielsen, and J. Perlwitz, 2009: Effect of zonal asymmetries in stratospheric ozone on simulated Southern Hemisphere climate trends. *Geophysical Research Letters***,** -.
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***,** -.
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**.
Wilkinson, J. M., R. J. Hogan, A. J. Illingworth, and A. Benedetti, 2008: Use of a Lidar Forward Model for Global Comparisons of Cloud Fraction between the ICESat Lidar and the ECMWF Model. *Monthly Weather Review*, **136,** 3742-3759.
Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences.* Academic Press, 467 pp.
Williams, K., and M. Webb, 2009: A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics***,** 141-157.
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.
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***,** 2955-2965.

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, 2011: *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.

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-.

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. *Nat. Geosci.*, **1,** 165-168.

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. *Glob. Biogeochem. Cyc.*, **12,** 345-360.

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.

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.

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, 2010a: Spatial Variability of Sea Level Rise in Twenty-First Century Projections. *Journal of Climate***,** 4585-4607.

Yin, J. J., R. J. Stouffer, M. J. Spelman, and S. M. Griffies, 2010b: Evaluating the Uncertainty Induced by the Virtual Salt Flux Assumption in Climate Simulations and Future Projections. *Journal of Climate*, **23,** 80-96.

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, T., T. Tozuka, and T. Yamagata, 2009: 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. *Geophys. Res. Lett.*, **37,** L01401.

——, 2010b: 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, 2010c: 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.

Zelle, H., G. J. van Oldenborgh, and H. Dijkstra, 2005: El Niño and greenhouse warming: results from ensemble simulations with the NCAR CCSM. *J. Climate*, **18,** 4669-4683.

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, X., et al., 2007: Detection of human influence on twentieth-century precipitation trends. *Nature***,** 461-U464.

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, 2010: 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.

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.

Tables

Table 9.1a: Prototype Table 1 - List of coupled AOGCMs that participated at CMIP5 / MMD PCMDI.
Selected model features of participating Earth System Models and important references. The models are listed by IPCC identification (Model ID) along with the calendar year of the first publication of results from each model. Also listed are respective sponsoring institutions, resolution information for atmosphere and ocean, features of the Carbon Cycle in ocean and land, details of the sea ice implementation, details of the terrestrial ecosystem implementation, and details on the coupled system. [This table would be aligned to fill a page in landscape mode, similar to AR4]

Model ID	Sponsor(s)	Used for decadal Pred	Atmosphere Resolution DOF Reference	Ocean Resolution DOF Reference	Carbon Cycle Ocean Resolution Land Reference	Sea Ice Details Reference	Land Details Reference	Coupled System Flux Adjustments Reference
Online at PCMDI (but not available yet)								
1. bcc-csm1-1	Beijing Climate Center (China)		T42L26 H8192 V26 BCC-AGC (from CAM3)	1° × 1°–0.3° H83520 V40 MOM4-L4 (from MOM4)				
2. CanESM2	Canadian Centre for Climate Modelling and Analysis (Canada)		T63L35 H8192 V35 CanAM4 Scinocca et al., 2008	1.41° × 0.94° H49152 V40 CanOM4 (from NCOM)	CMOC Christian et al., 2010 CTEM Arora et al., 2009	Flato and Hibler, 1992	CLASS Verseghy, 1991	No flux adjustments CanESM2, 2011
3. GISS-E2-R	NASA / GISS (USA)		2° × 2.5° H12960 V40	1.41° × 0.94° H51840 V32				
4. hadgem2-es	Met Office Hadley Centre (UK)		T63L35 H27648 V38 HadGEM2-AO	0.83° × H77760 V40 HadGEM2-AO				No flux adjustments

[Alternative Table]

Table 9.1b: Prototype Table 2 - Significant properties of the AOGCMs used for CMIP5.
Selected model features of participating Earth System Models and important references. The models are listed by IPCC identification (Model ID) along with the calendar year of the first publication of results from each model and the respective sponsoring institutions. Also listed are references on the implementation of atmospheric dynamics, physics and chemistry, information on the implementation of oceanic dynamics, biogeochemistry and sea ice, information on the respective terrestrial ecosystem model, and details on the coupled system. [This table is a more detailed table that would go in portrait mode over two pages]

Model ID	Sponsor(s)	Atmosphere			Ocean			Land	Coupled System
		Dynamics	Physics	Chemistry/Aerosols	Dynamics	Biochemistry	Sea-ice	TES/CC	
Model A	Institute B	T63L30		interactive aerosols	1 × 0.4 deg	Carbon			
	(country C)	author a, year x	author b, year y	author a, year x	author a, year x	author a, year x	author a, year x	author a, year x	author a, year x

Table 9.2: [PLACEHOLDER FOR FIRST ORDER DRAFT] Observationally-based estimates of atmospheric Essential Climate Variables evaluated in Chapter 9; alternates are shown in parentheses. [to be updated]

ECV	DATASET	Period	Notes	Reference
Temperature	CRU		2 meter, land	
Temperature			SST	
Temperature	RSS		Lower-troposphere (TLT)	
Precipitation	GPCP			
Precipitation				
Radiation	CERES		Top-of-atmosphere longwave and shortwave	
Water vapour	RSS			
Total column ozone	TOMS	1980–1999		

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: Johanna Baehr, Alejandro Bodas-Salcedo, Bode Gbobaniyi, Stephen Griffies, Elizabeth Hunke, Tatiana Ilyina, Stephen A. Klein, Reto Knutti, Felix Landerer, Florian Rauser, Mark Rodwell, Adam A. Scaife, John Scinocca, Hideo Shiogama, Ken Sperber, Bjorn Stevens, Keith Williams, Tim Woollings

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

Date of Draft: 15 April 2011

Notes: TSU Compiled Version

Figures

[PLACEHOLDER FOR FIRST ORDER DRAFT: Figures provided here are intended to be representative of the *sort* of figures that will be included. Additional figures, with better balance between sections will be provided. Awaiting results to construct (and cite) a more comprehensive set of figures for this chapter.]

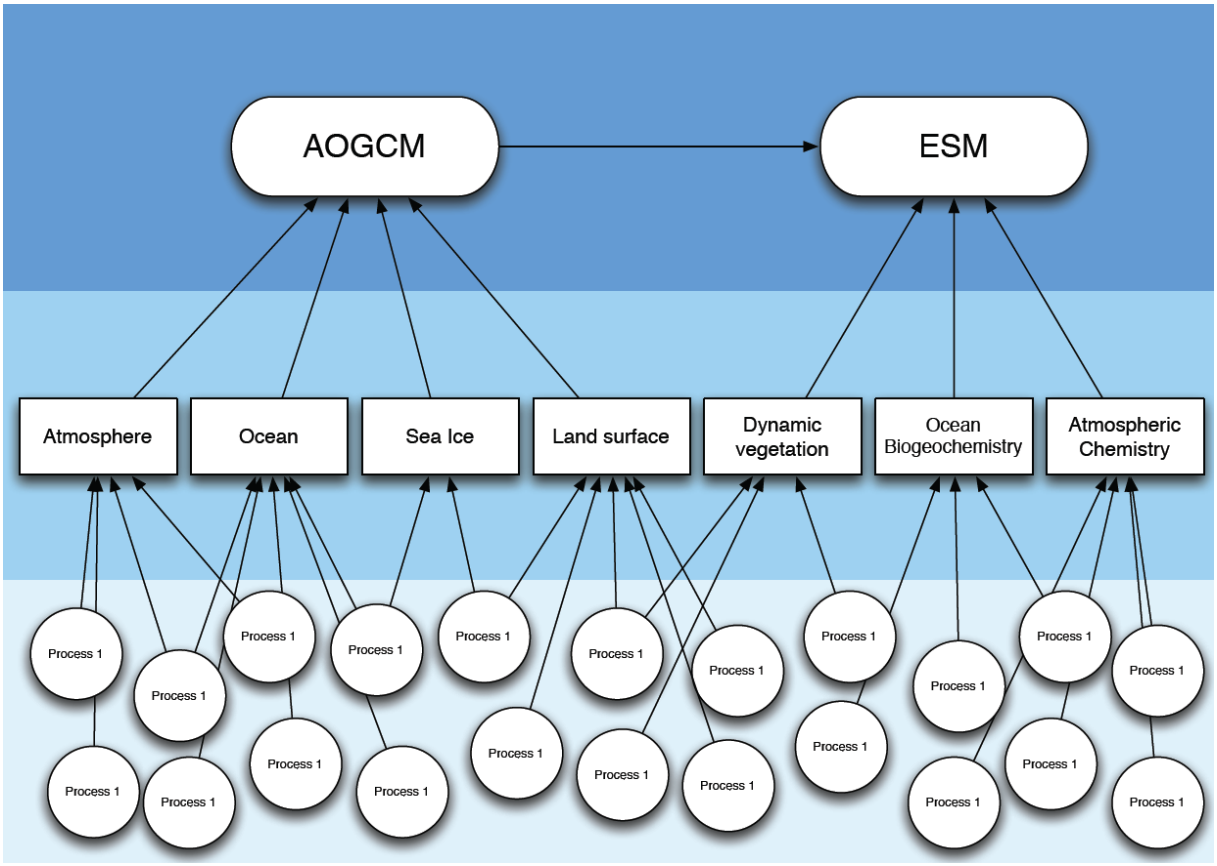


Figure 9.1: A simple figure that illustrates the three levels of model components and that highlights the different components of Atmosphere Ocean General Circulation Models (AOGCMs) and Earth System Models (ESMs).

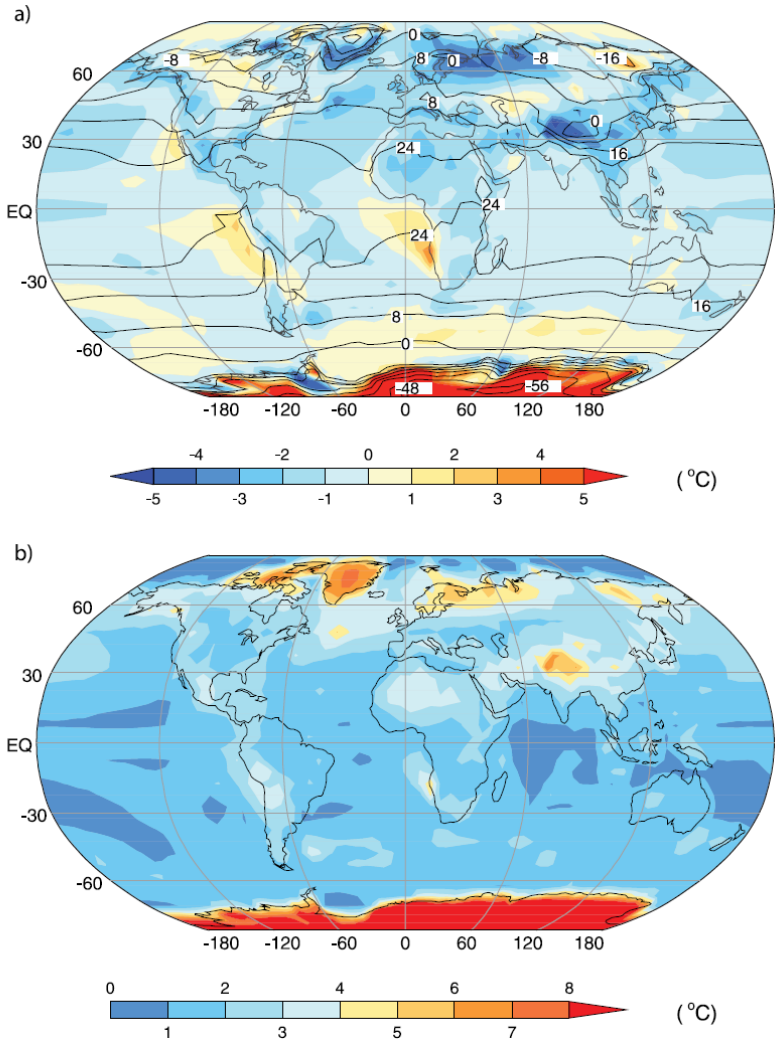


Figure 9.2: [PLACEHOLDER FOR FIRST ORDER DRAFT] [AR4 Figure 8.2.] (a) Observed climatological annual mean SST and, over land, surfaceair temperature (labelled contours) and the multi-model mean error in these temperatures, simulated minus observed (colour-shaded contours). (b) Size of the typical model error, as gauged by the root-mean-square error in this temperature, computed over all AOGCM simulations available in the MMD at PCMDI. The Hadley Centre Sea Ice and Sea Surface Temperature (HadISST; Rayner et al., 2003) climatology of SST for 1980 to 1999 and the Climatic Research Unit (CRU; Jones et al., 1999) climatology of surface air temperature over land for 1961 to 1990 are shown here. The model results are for the same period in the 20th-century simulations. In the presence of sea ice, the SST is assumed to be at the approximate freezing point of seawater (–1.8°C).

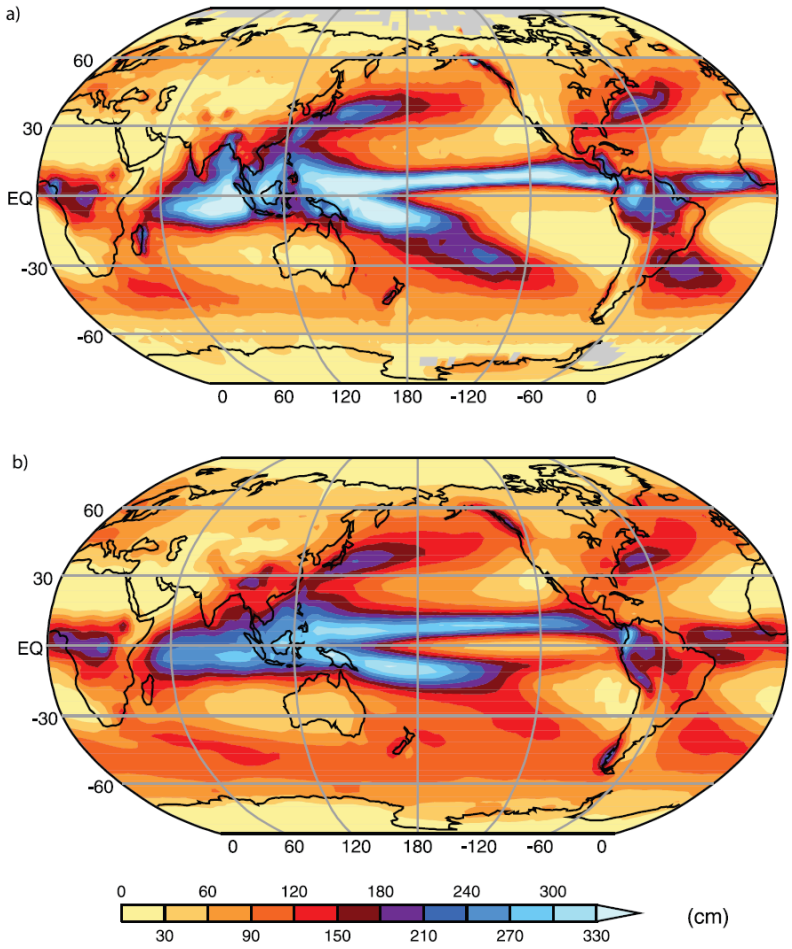


Figure 8.5. *Annual mean precipitation (cm), observed (a) and simulated (b), based on the multi-model mean. The Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997) observation-based climatology for 1980 to 1999 is shown, and the model results are for the same period in the 20th-century simulations in the MMD at PCMDI. In (a), observations were not available for the grey regions. Results for individual models can be seen in Supplementary Material, Figure S8.9.*

Figure 9.3: An extension of Figure 8.5 of the AR4 (above), showing 4 panels: (a) observations (with differences between 2 estimates shown in contours), (b) the multi-model means of the CMIP3 models, (c) the CMIP5 models, […] or presented as zonal averages.

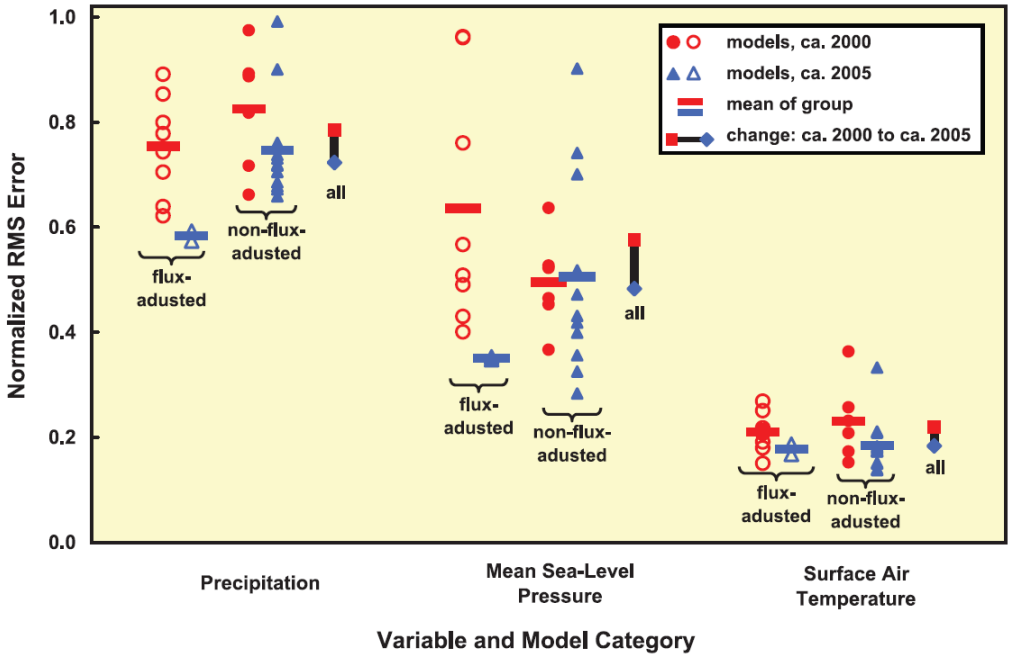


Figure 9.4: Normalised RMS error in simulation of climatological patterns of monthly precipitation, mean sea level pressure and surface air temperature. Recent AOGCMs (circa 2005) are compared to their predecessors (circa 2000 and earlier). Models are categorised based on whether or not any flux adjustments were applied. The models are gauged against the following observation-based datasets: Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997) for precipitation (1980–1999), European Centre for Medium Range Weather Forecasts 40-year reanalysis (ERA40; Uppala et al., 2005) for sea level pressure (1980–1999) and Climatic Research Unit (CRU; Jones et al., 1999) for surface temperature (1961–1990). Before computing the errors, both the observed and simulated fields were mapped to a uniform 4° x 5° latitude-longitude grid. For the earlier generation of models, results are based on the archived output from control runs (specifically, the first 30 years, in the case of temperature, and the first 20 years for the other fields), and for the recent generation models, results are based on the 20th-century simulations with climatological periods selected to correspond with observations. (In both groups of models, results are insensitive to the period selected.) [Taken from Figure 8.11 of Chapter 8 of the AR4, This figure would be substantially modified, comparing CMIP3 and CMIP5 (rather than CMIP2 and CMIP3). All individual models would be color coded and identified. Results would be shown for a number of ECVs (Ts, pr, clt, OLR, water vapour, etc.) Perhaps only models without flux correction would be shown (i.e., most). This is an alternate to a “portrait plot” of relative model performance.]

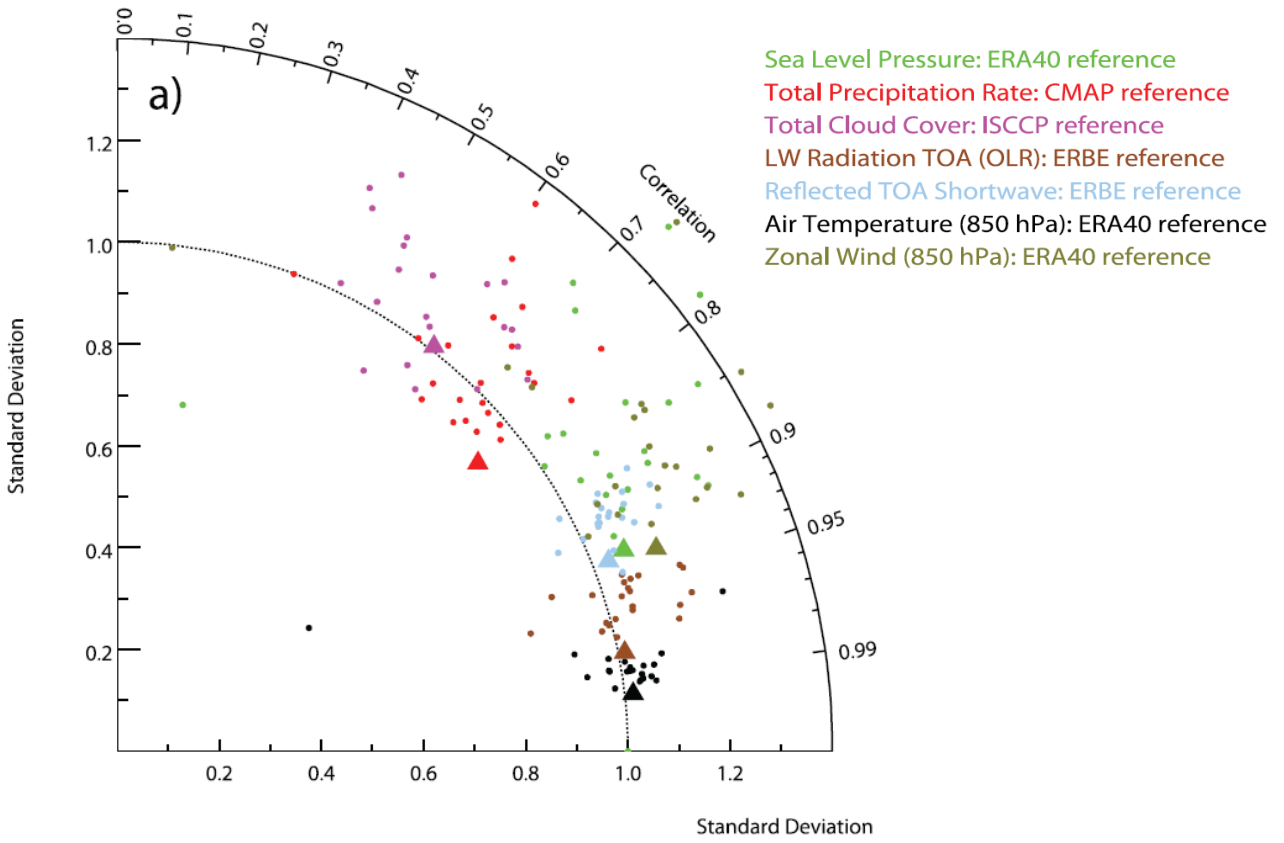


Figure 9.5: Multivariable Taylor diagrams of the 20th century CMIP3 annual cycle climatology (1980–1999) for NHEX (20°N–90°N). Color identified for each variable, with dots representing individual models and triangles denote multi-model mean. Standard deviation is normalized by the observed value for each variable. Taken from Gleckler et al. 2008 [Models to be updated with CMIP5 as well as all observations, and results to be shown for the global annual cycle. Individual models to be identified.]

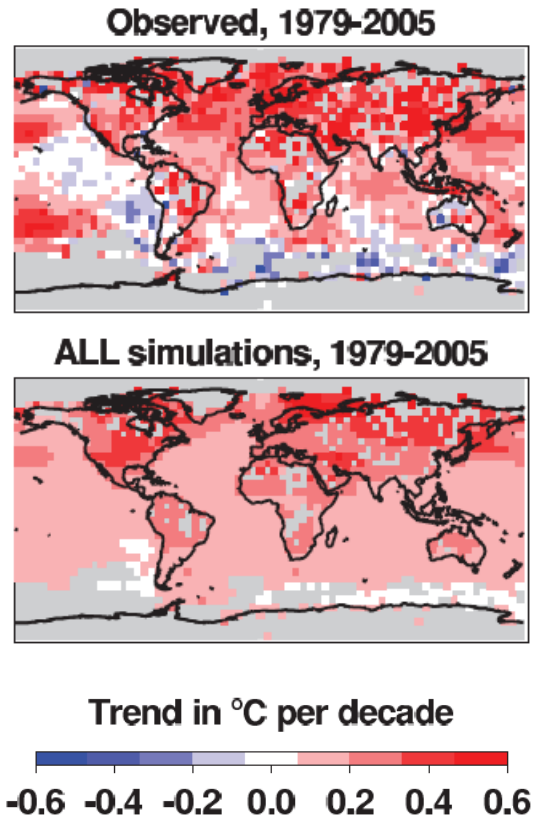


Figure 9.6: Taken from the AR4 Figure 9.6 (D&A Chapter), the above figure shows local linear trends of surface temperature from 1979–2005. The proposed figure would add similar panels for water vapour (1988–2005). Inter-model s.d. to be shown with contours.

Figure 9.7: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure to be constructed consists of two scatter plots, one for surface temperature, and one for water vapour. On each plot, the y-axis corresponds to the amplitude of the global average linear trend, and on the x-axis, the spatial pattern correlation of the simulated and observed linear trends.

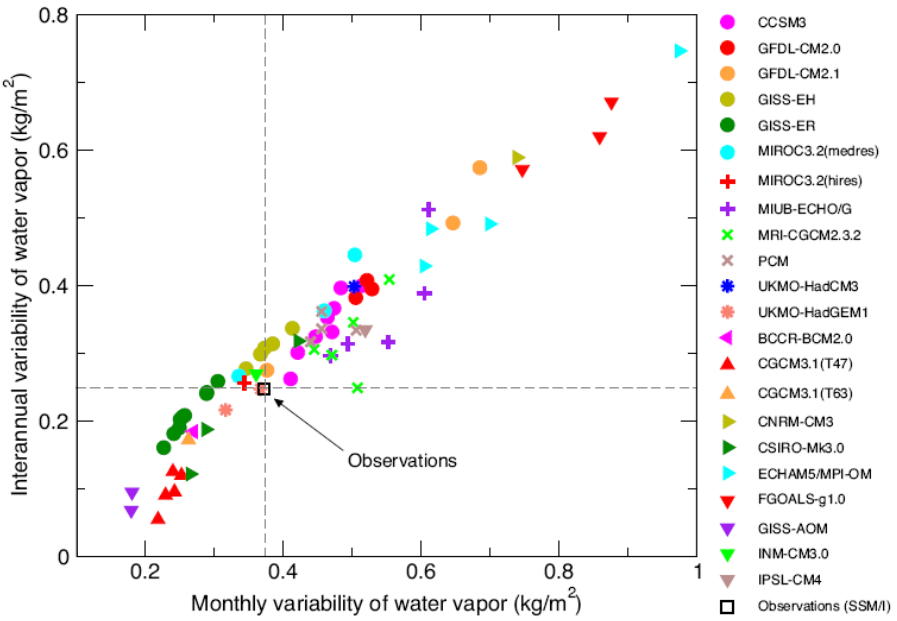


Fig. 1. Comparison of the simulated and observed temporal variability of atmospheric water vapor. Observations are from the SSM/I dataset (25, 26); model data are from 71 realizations of 20th century climate change performed with 22 different models (see *SI Appendix*). All variability calculations rely on monthly mean values of $\langle W \rangle$, the spatial average of total atmospheric moisture over near-global oceans. Model and observational $\langle W \rangle$ data were first expressed as anomalies relative to climatological monthly means over the period 1988–1999 and then linearly detrended. We computed temporal standard deviations from both the unfiltered and filtered anomaly data. The latter were smoothed by using a filter with a half-power point at ≈2 years. The raw and filtered standard deviations provide information on monthly and interannual-timescale variability, respectively. All calculations were over the 144-month period from January 1988 to December 1999 (the period of maximum overlap between the SSM/I data and most 20CEN simulations). The dashed gray lines are centered on the observations.

Figure 9.8: Taken from Santer et al. (2009). [PLACEHOLDER FOR FIRST ORDER DRAFT: Similar to the above except replacing the y-axis with the spatial pattern correlation of the observed and simulated monthly variability. One figure for SST, and one for water vapour.]

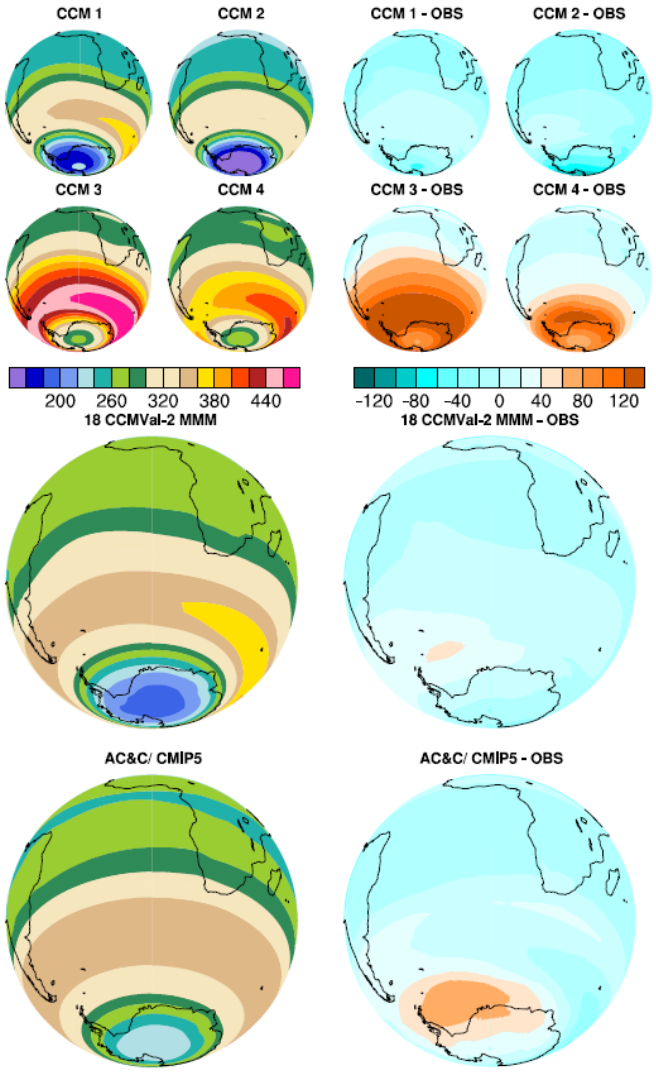


Figure 9.9: Simulated present-day (1980–1999 average) absolute September to November total column ozone (a) and the bias of it from TOMS data (b). The four models that deviate most from observations (two low-biased and two high-biased) are shown in (a, b) to illustrate the spread. The average of 18 CCMVal-2 models in the large panels is shown in (d,e), and the AC&C / SPARC ozone database in (f, g). Changes in total column ozone over Antarctica (averaged from 60°S–90°S) from 1960 to 2000 for the above datasets are shown in panel (h). Ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere. Modified from Eyring et al., J. Clim, (2011), in preparation. [Update with CMIP5 models that have interactive stratospheric chemistry.]

Figure 9.10: [PLACEHOLDER FOR FIRST ORDER DRAFT: Figure to be constructed summarizing results from paleoclimate comparisons.]

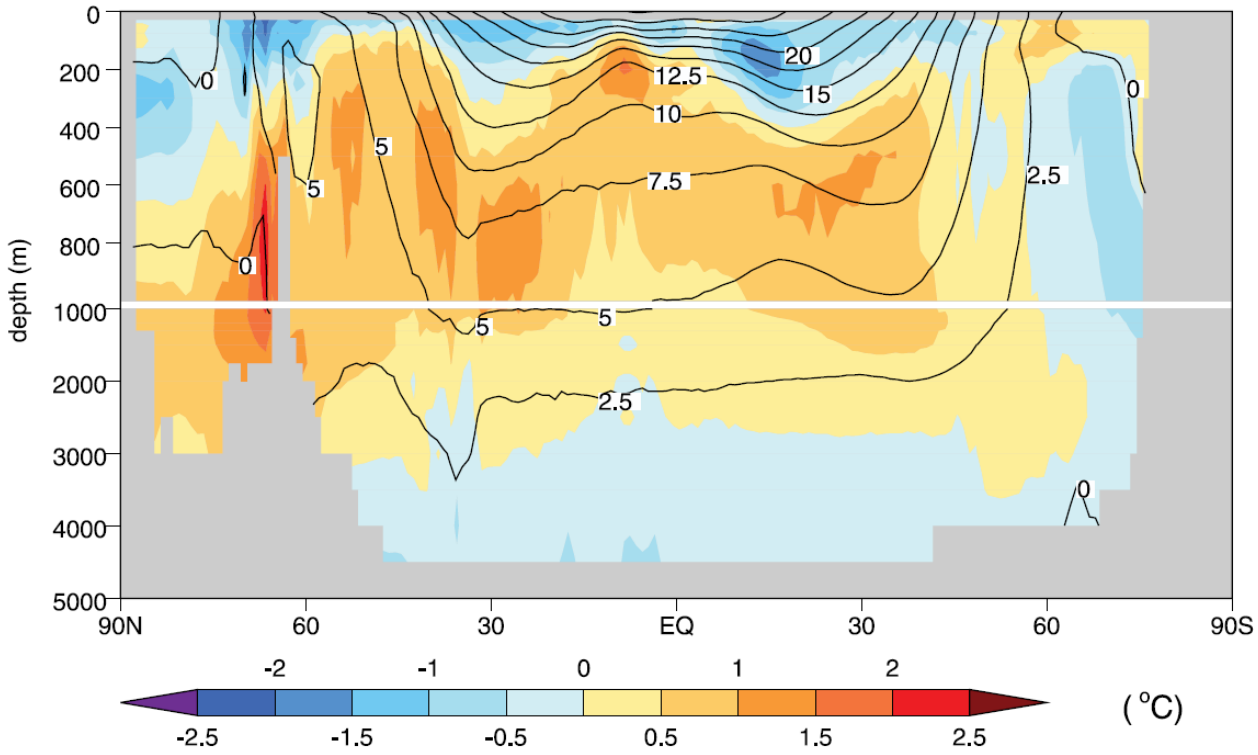


Figure 9.11: [PLACEHOLDER FOR FIRST ORDER DRAFT] From AR4, to be redone from CMIP5: Figure 8.9. Time-mean observed potential temperature (°C), zonally averaged over all ocean basins (labelled contours) and multi-model mean error in this field, simulated minus observed (colour-filled contours). The observations 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 MMD at PCMDI.

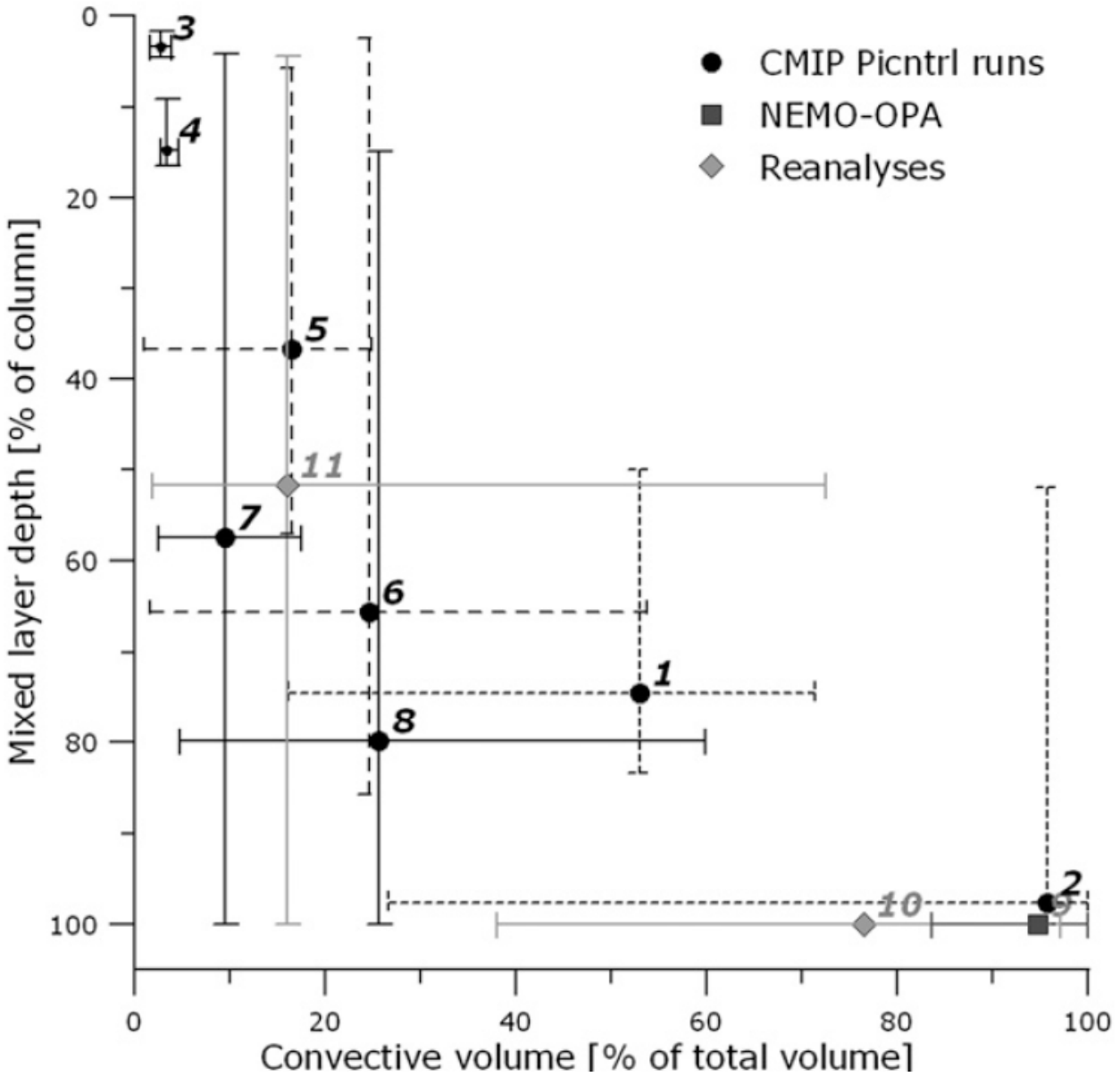


Figure 9.12: [PLACEHOLDER FOR FIRST ORDER DRAFT] [From (de Jong et al., 2009) to show ability of models to capture the mixed-layer depth in a crucial region for ocean uptake of heat or carbon. Another region could be shown.]

Figure 9.13: [PLACEHOLDER FOR FIRST ORDER DRAFT: For CMIP5 we will show results in the form of a scatter plot. One axis will show near-global ocean annual cycle RMSE errors of SSH and the other axis will be the same for OHC. Note these will be constructed from independent observations, namely altimetry and in-situ temperature measurements.]

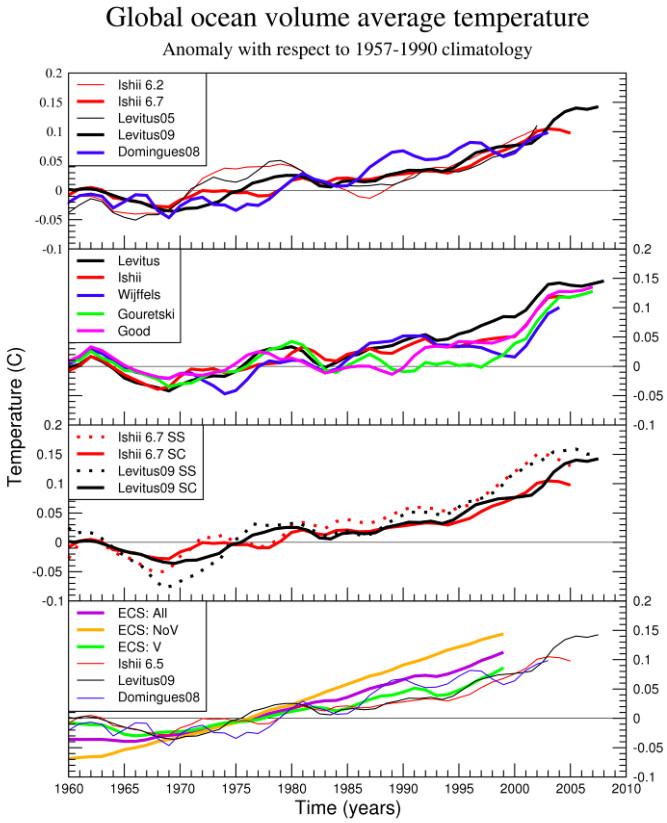


Figure 9.14: [Taken from Gleckler et al., D&A paper, in preparation] Global average evolution of upper ocean (0–700 m) volume average temperature. a) New and old estimates of Ishii et al. (2003, 2009) and Levitus et al. (2005, 2009), b) the impact of using different XBT bias correction methods demonstrated within a common analysis framework c) Ishii et al. (2009) and Levitus et al. (2009) spatially complete (SC) and subsampled (SS) as described in text, d) Recent observed estimates and the CMIP3 ensemble common signal (ECS) for the subset of models including volcanic eruptions, those without, and all models. All times series are departures from 1957–1990 climatologies (°C), and those in a) and c) are computed from spatially complete data. For visual clarity, all observational data are 5-year running averages. The 20CEN simulations in CMIP3 end in 1999 or 2000. [The emphasis of this figure is on observational uncertainty, with multi-model results only shown in panel (d). This information can be re-worked to focus more directly on model-obs comparison.]

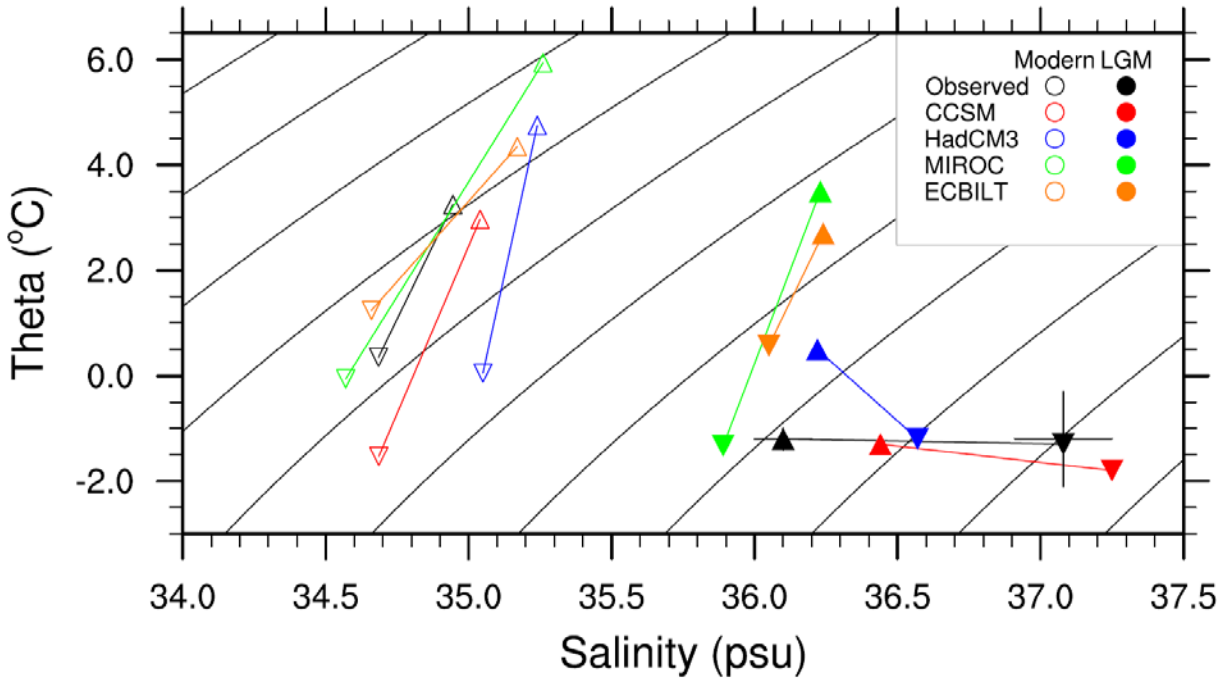


Figure 9.15: Temperature and salinity for modern (open symbols) and LGM (filled symbols) as estimated from data (with error bars) at ODP sites (Adkins et al., 2002) and predicted by the PMIP2 models. Site 981 triangles) is located in the North Atlantic (Feni Drift, 55_N, 15_W, 2184 m). Site 1093 (upside 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).

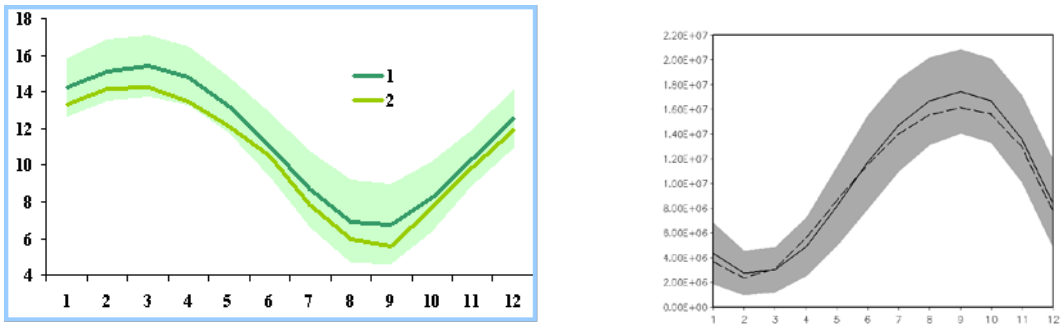


Figure 9.16: [PLACEHOLDER FOR FIRST ORDER DRAFT] Baseline climate (1980–1999) model mean sea ice extent seasonal cycle in the Northern (left) and Southern (right) hemispheres as simulated by CMIP3 models (1 and solid) and observed (2 and dashed). The shaded are shows the intermodel standard deviation. 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). [To be replaced by a similar figure for CMIP5 (for the AR5 baseline period 1986–2005).]

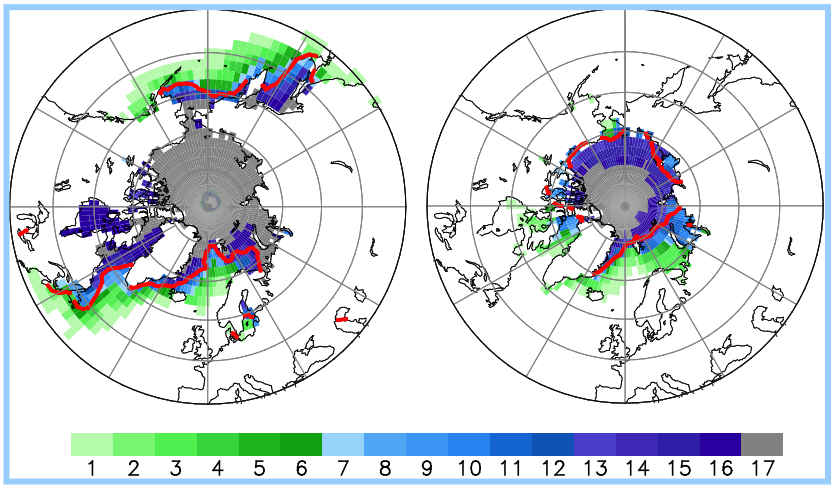


Figure 9.17: [PLACEHOLDER FOR FIRST ORDER DRAFT] Baseline climate (1980–1999) sea ice distribution in the Northern Hemisphere (upper panels) and the Southern Hemisphere (lower panels) simulated by 17 of CMIP3 AOGCMc for March (left) and September (right). For each 2.5° × 2.5° longitude-latitude grid cell, the figure indicates the number of 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). [To be replaced by a similar figure for CMIP5 (same period as above – thus comparable to that in AR4 Figure 8.10).]

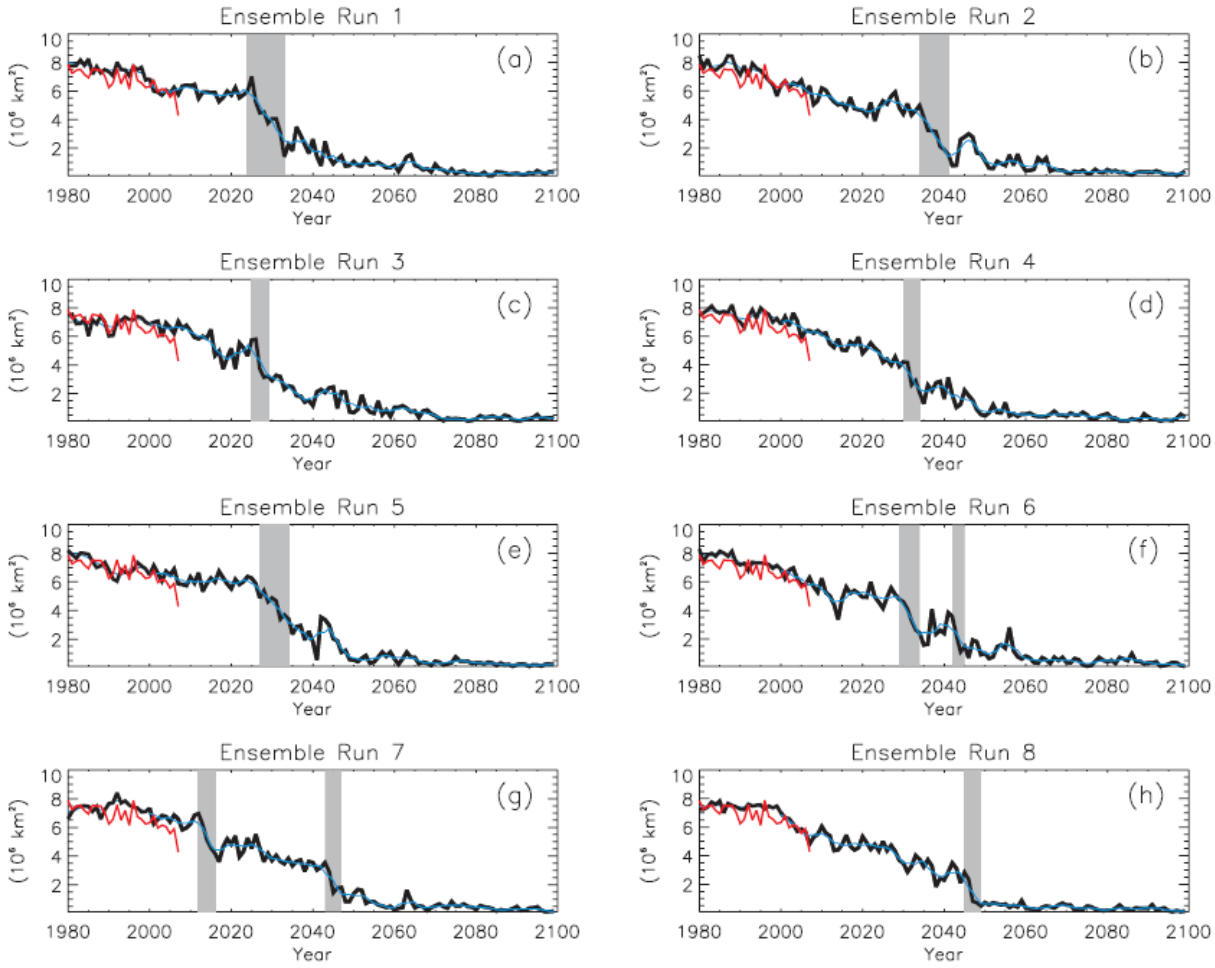


Plate 2. Time series of September ice extent from the (a–h) eight IPCC-AR4 CCSM3 ensemble members. The blue line shows the 5-year running mean time series. The observed time series from 1979 to 2007 is shown in red. Grey shading indicates abrupt events as defined in the text.

Figure 9.18: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure showing CMIP5 model ability to generate natural (unforced) variability of sea ice extent. [Placeholder figure from Holland et al., 2008.]

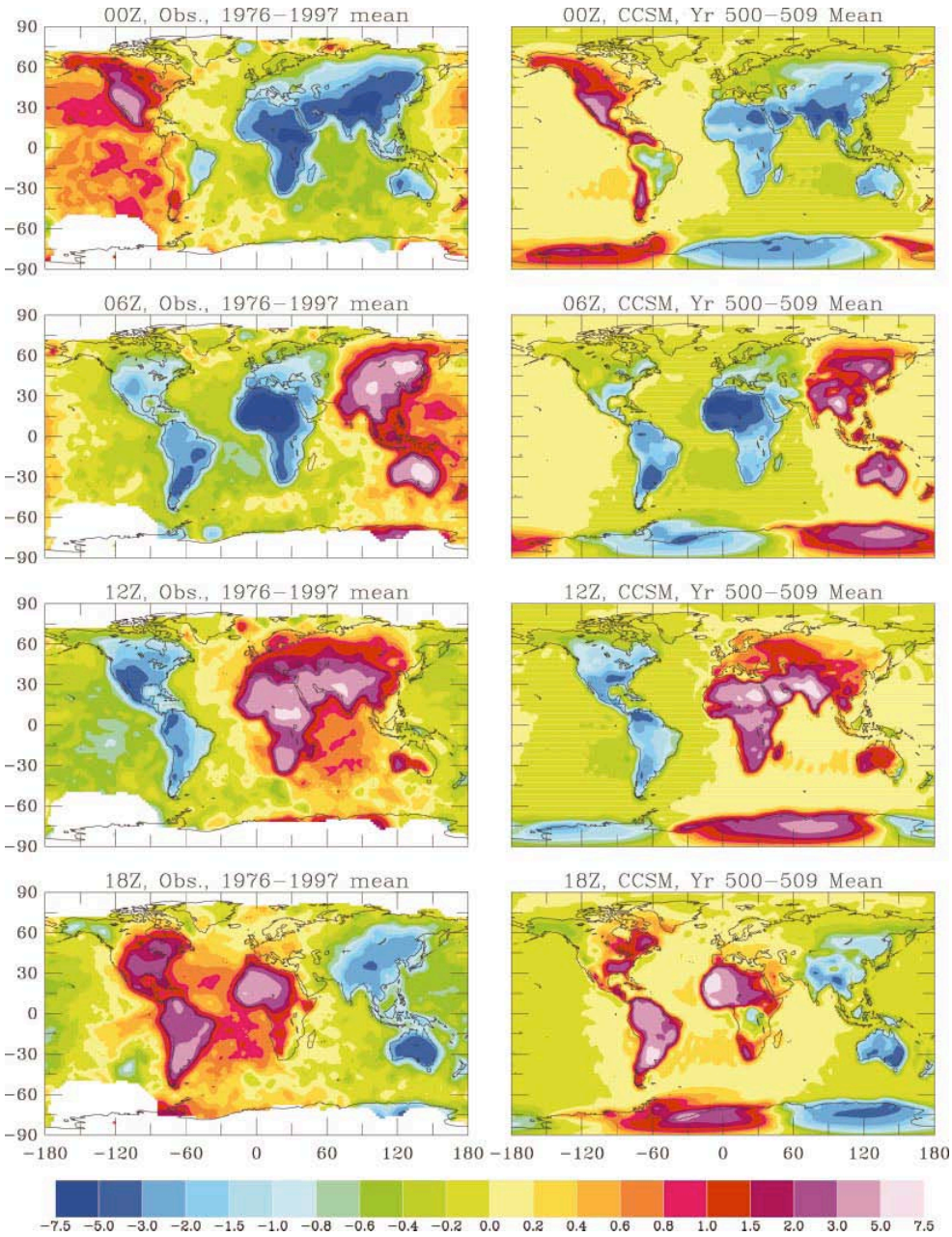


Figure 9.19: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure for diurnal cycle of temperature (from Dai and Trenberth, 2004). This could be replaced by a harmonic analysis of phase and amplitude, which would reduce it to two panels, or by selected line graphs, which would allow for all models to be shown.

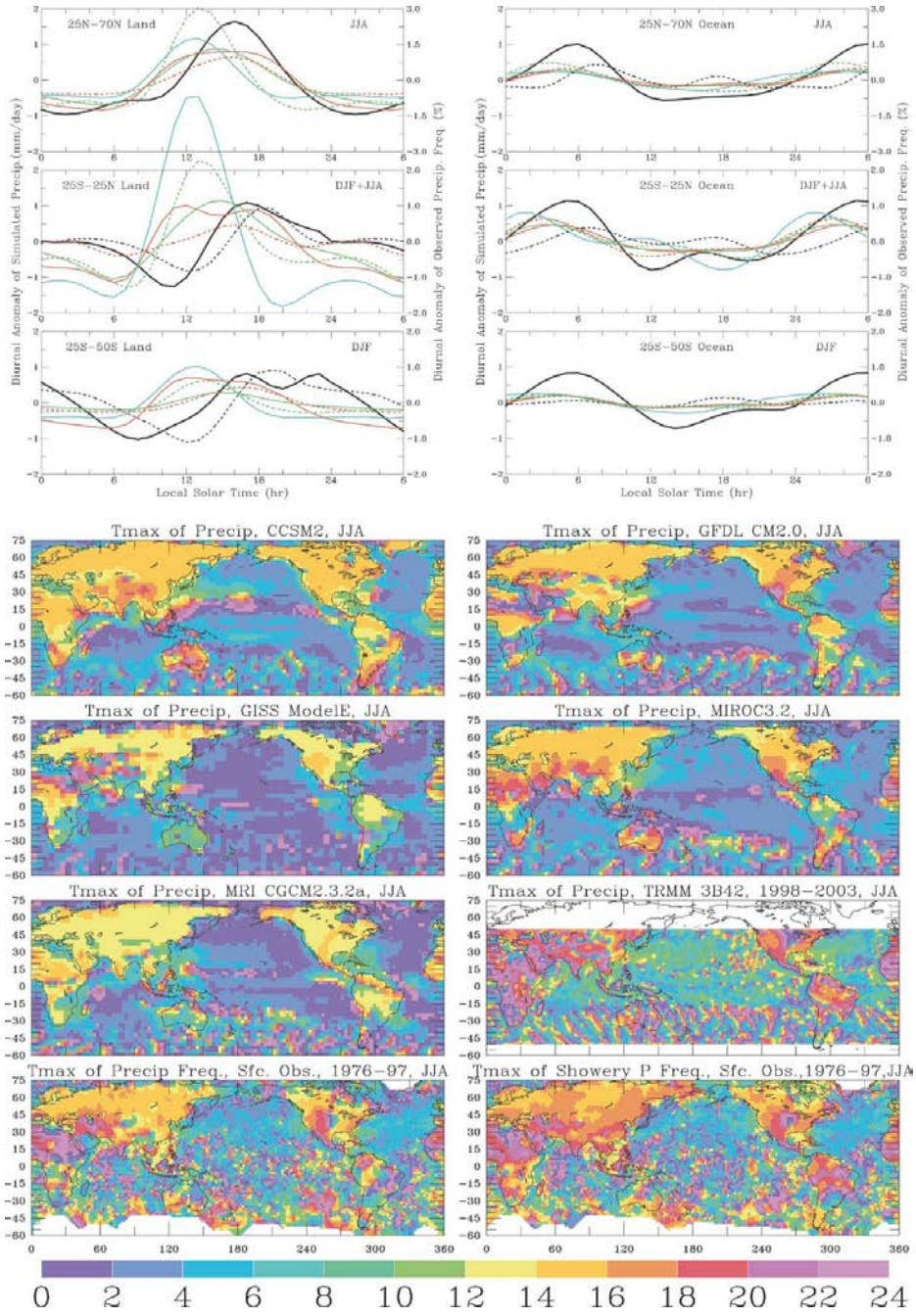


Figure 9.20: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure for the evaluation of diurnal cycle of precipitation (example figure from Dai, 2006). Line plots are proposed for main report, maps for supplementary material. Those figures could be replaced by a harmonic analysis of phase and amplitude.

1

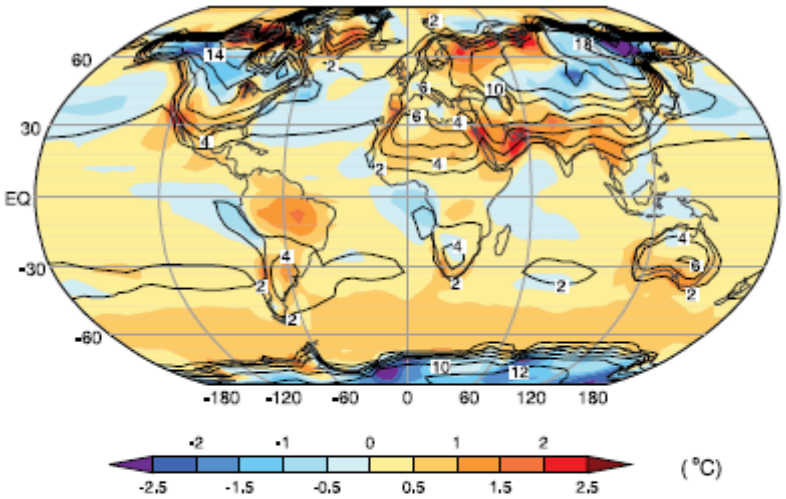


Figure 8.3. *Observed standard deviation (labelled contours) of SST and, over land, surface air temperature, computed over the climatological monthly mean annual cycle, and the multi-model mean error in the standard deviations, simulated minus observed (colour-shaded contours). In most regions, the standard deviation provides a measure of the amplitude of the seasonal range of temperature. The observational data sets, the model results and the climatological periods are as described in Figure 8.2. Results for individual models can be seen in the Supplementary Material, Figure S8.2.*

2

3 **Figure 9.21:** [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure illustrating CMIP5 models' ability to
4 simulate general features of the seasonal cycle. Placeholder figure is from the AR4.
5

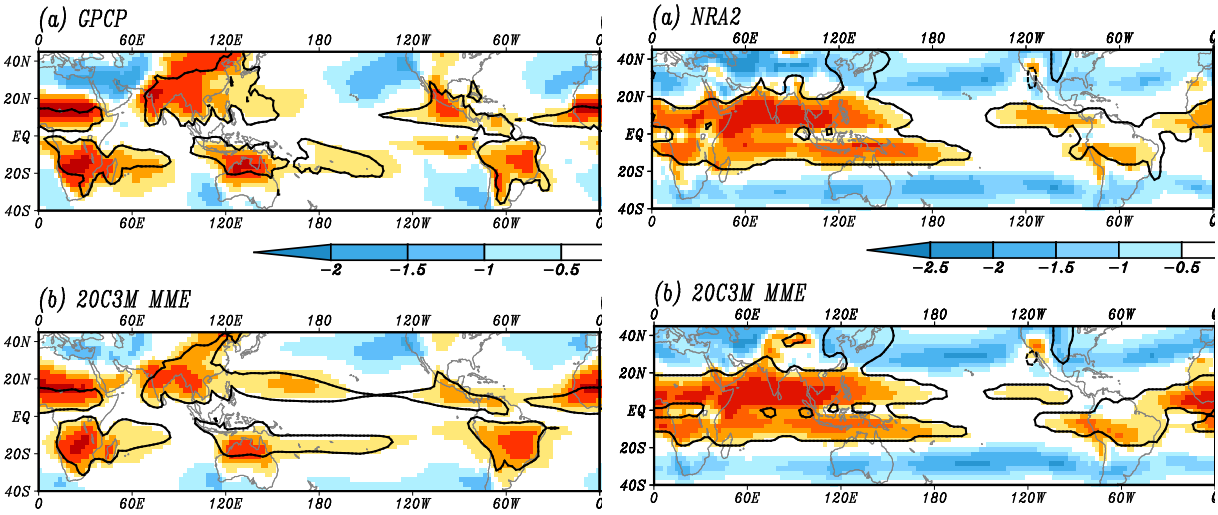


Figure 9.22: [PLACEHOLDER FOR FIRST ORDER DRAFT: Proposition with CMIP5 / CMIP3 results] The approximate extent of the global monsoon domain (solid line) and monsoon intensity (shading) are shown for precipitation (a and b) and 850hPa wind speed (c and d). The monsoon domain is defined where the local summer-minus-winter precipitation rate (850hPa windspeed) exceeds 2.5 mm/day (2.5 m/s). (a) and (c) are based on GPCP precipitation and NCEP/DOE Reanalysis-2, respectively. (c) and (d) show the multi-model mean from the CMIP3 20c3m simulations. After Kim et al. (2011)

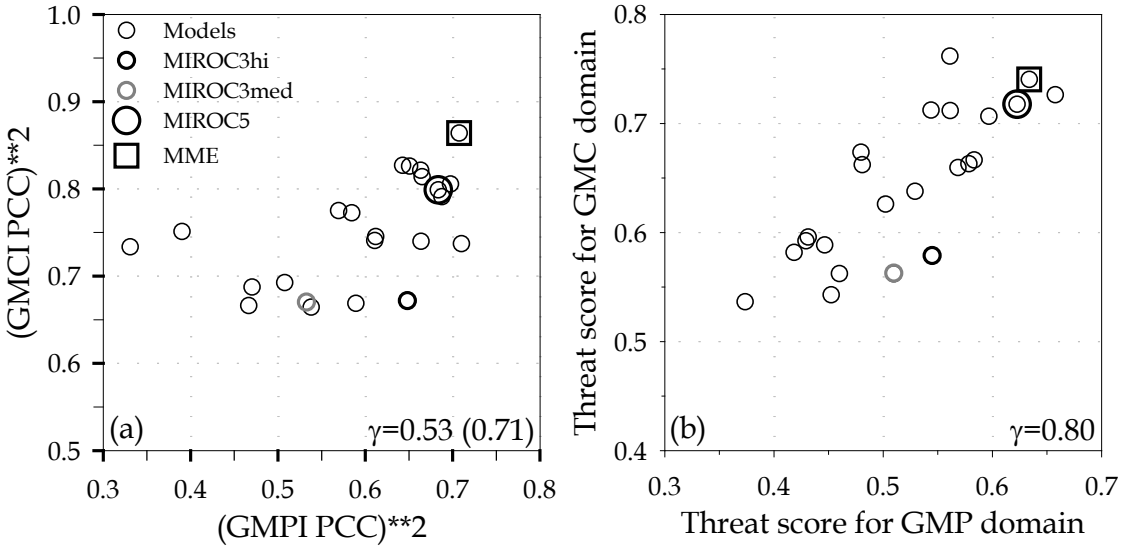


Figure 9.23: [PLACEHOLDER FOR FIRST ORDER DRAFT: Proposition this kind of figure with CMIP5 results] Evaluation of the CGCMs' performance on the climatological global monsoon intensity (left) and domain (right). The regression coefficient is shown in lower-right corner of each panel. The domain used is 0°–360°E, 40°S–45°N. The threat score has a range of 0–1, with 1 indicating perfect agreement with observations. After Kim et al. (2011)

1

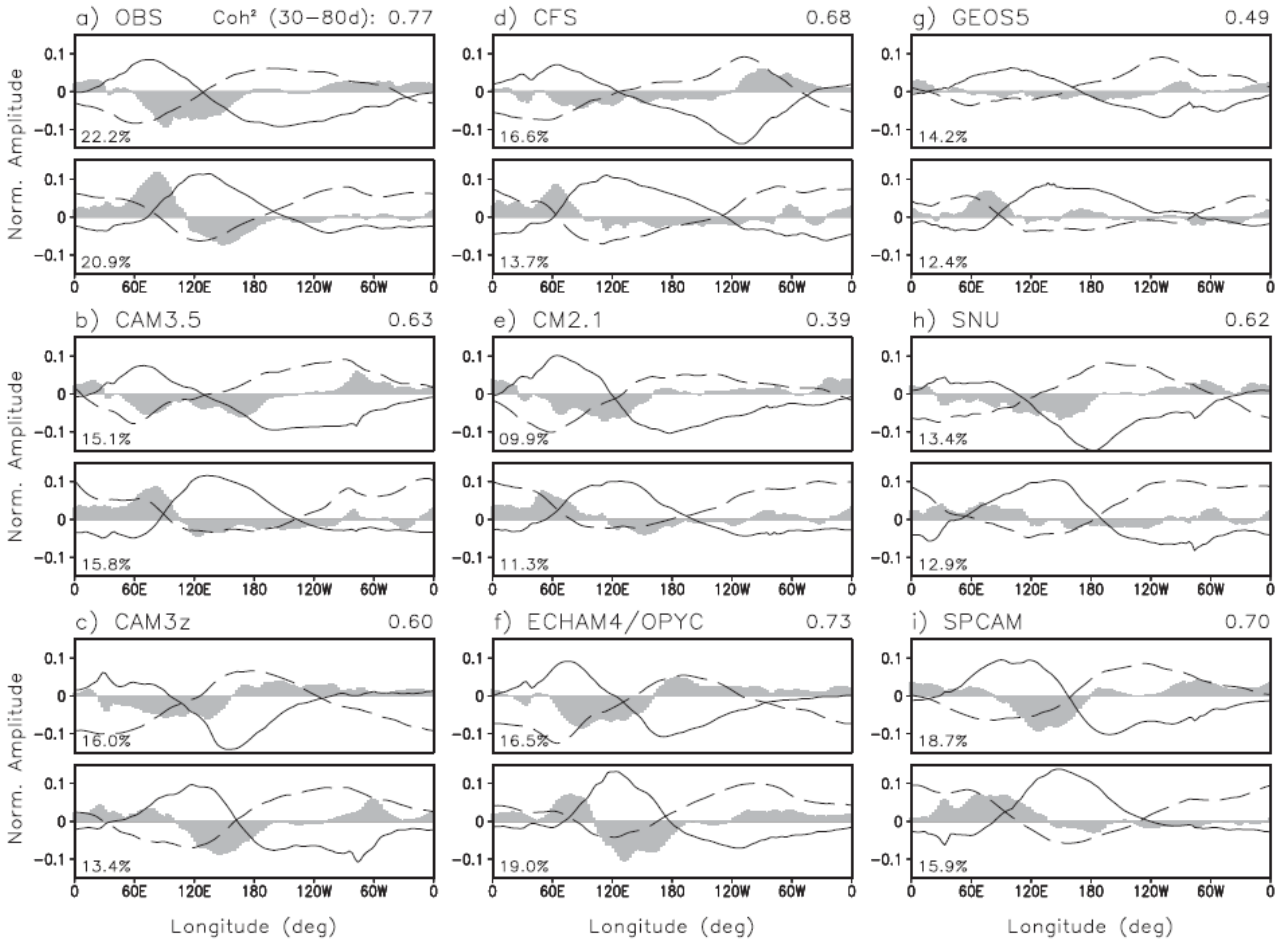


FIG. 6. First two CEOF modes of 20–100-day 15°S–15°N averaged 850-hPa and 200-hPa zonal wind and OLR for the (a) NCEP/NCAR and AVHRR, (b) CAM3.5, (c) CAM3z, (d) CFS, (e) CM2.1, (f) ECHAM4/OPYC, (g) GEOS5, (h) SNU, and (i) SPCAM models. The total variance explained by each mode is shown in the lower left of each panel. The mean coherence squared between principal components of two modes within a 30–80-day period is given above the upper panel. Sign and location (upper or lower) of each mode are arbitrarily adjusted to be similar to observation. The mode having the largest percentage variance explained is the first mode.

2

3 **Figure 9.24:** [PLACEHOLDER FOR FIRST ORDER DRAFT] (From Kim et al., 2009) illustrating model
4 performance in simulating MJO.

5

Figure 9.25: [PLACEHOLDER FOR FIRST ORDER DRAFT] [tbd]

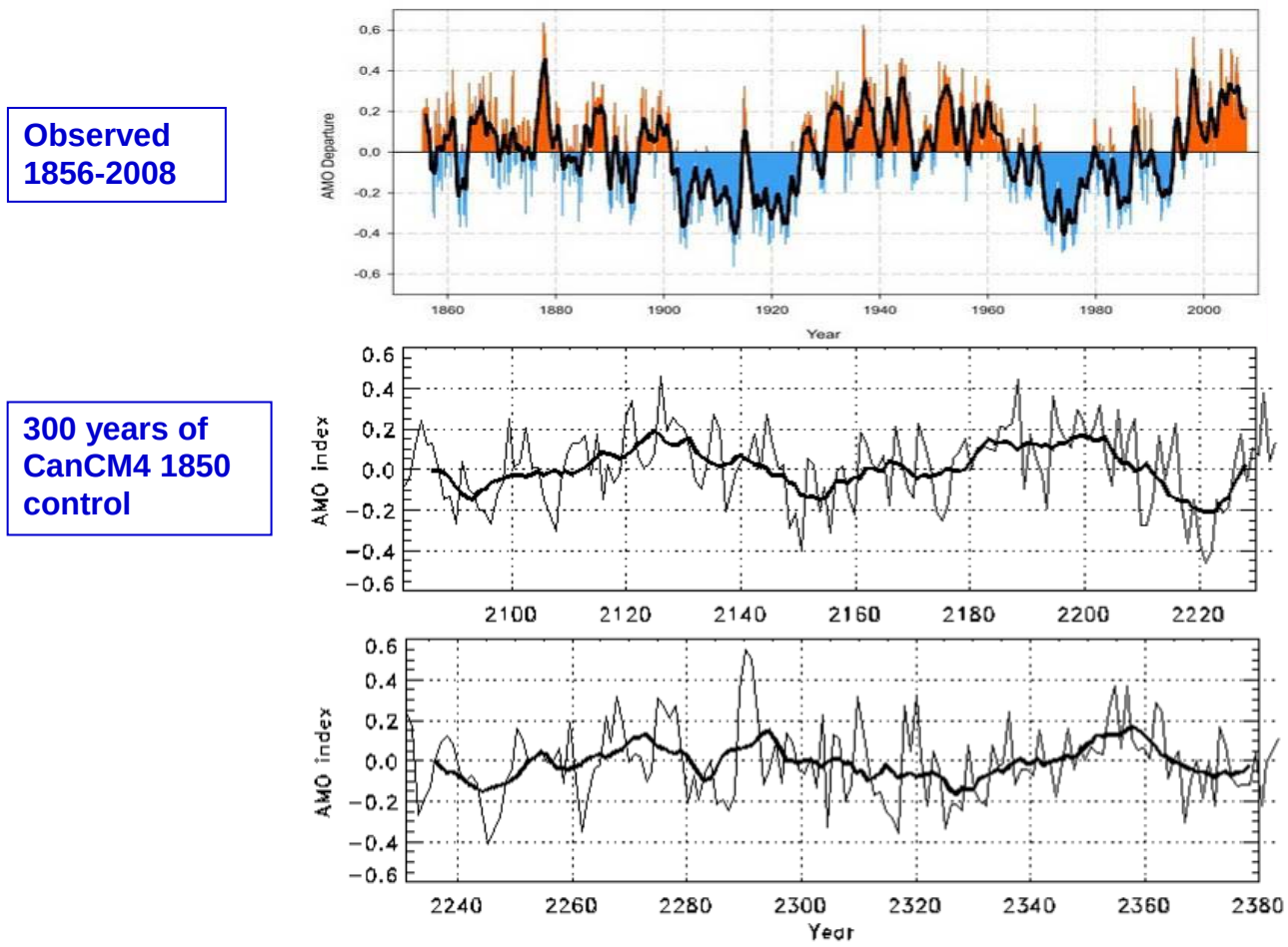


Figure 9.26: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure illustrating models' ability to simulate Atlantic multidecadal variability. The placeholder figure merely shows this for one particular model (CCCma).

1

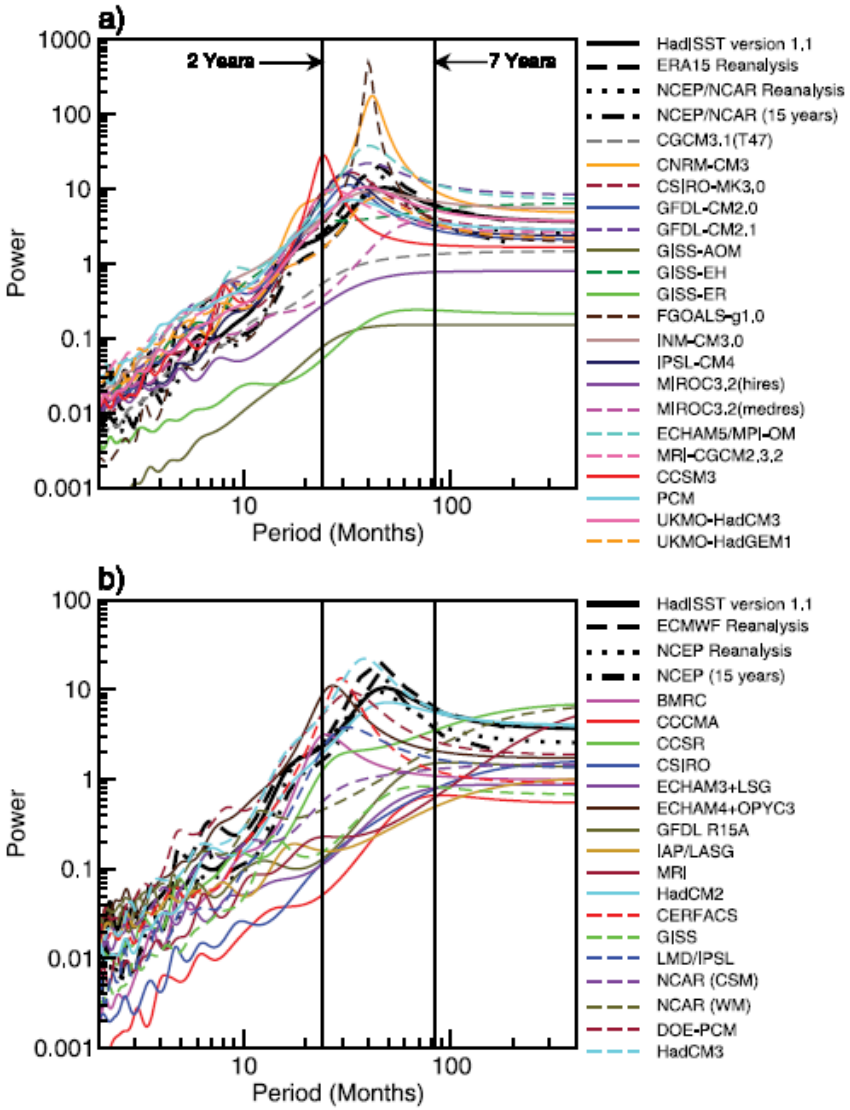
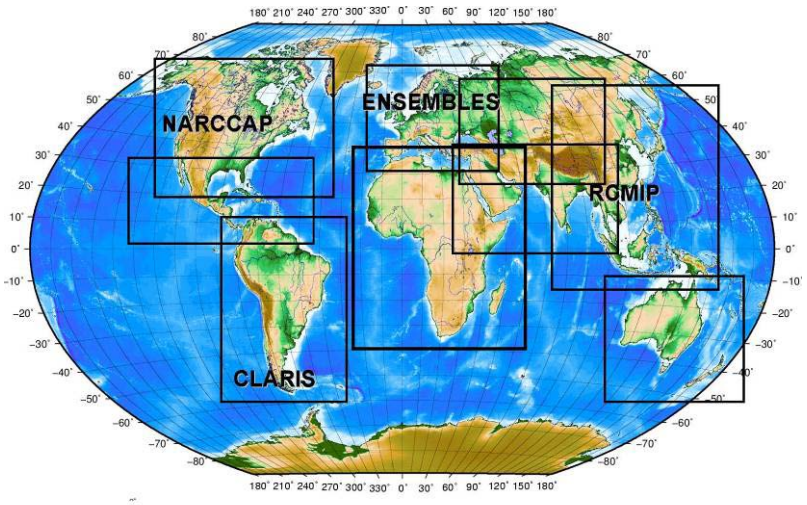


Figure 8.13. *Maximum entropy power spectra of surface air temperature averaged over the $NINO_3$ region (i.e., 5°N to 5°S, 150°W to 90° W) for (a) the MMD at the PCMDI and (b) the CMIP2 models. Note the differing scales on the vertical axes and that ECMWF reanalysis in (b) refers to the European Centre 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).*

2

3 **Figure 9.27:** [PLACEHOLDER FOR FIRST ORDER DRAFT] ENSO power spectra for different models.
4 [Placeholder figure taken from AR4 – to be updated]
5

1



2

3 **Figure 9.28:** [PLACEHOLDER FOR FIRST ORDER DRAFT] Map of recent coordinated downscaling
4 study regions. From Giorgi et al., WMO Bulletin 2009, 58:3, 175-183. [NB: Need copyright or redrawing.]
5

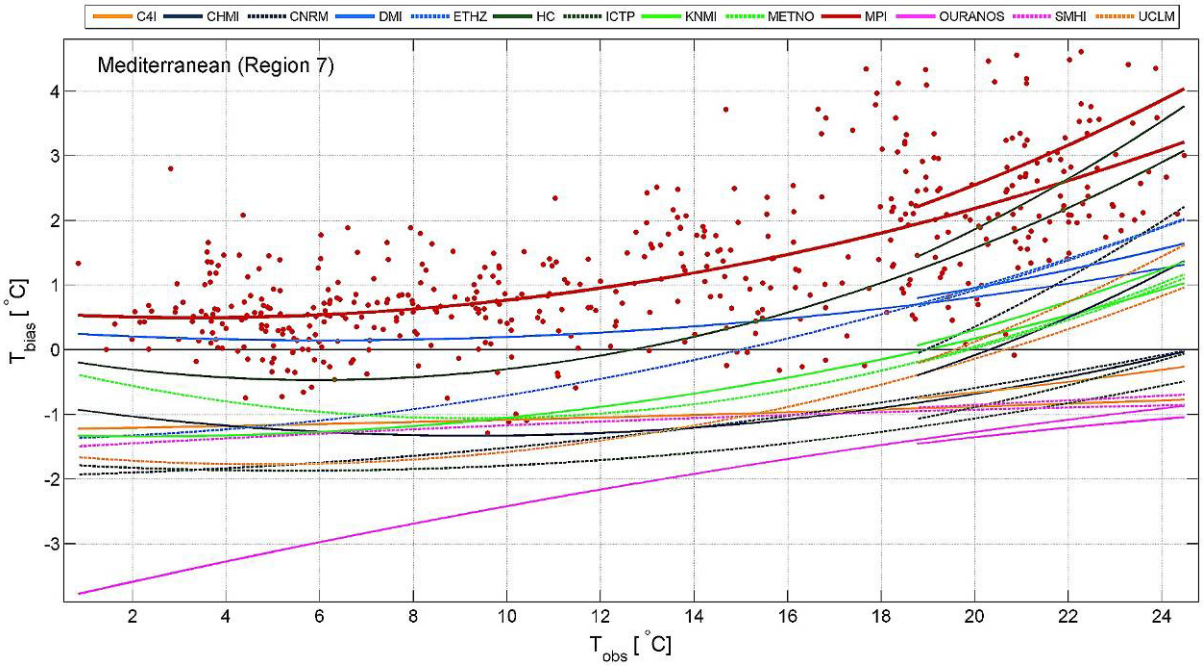


Figure 9.29: Monthly mean model temperature bias vs. observed monthly mean temperature for Region MD (Mediterranean) for the period 1961–2000. Raw data shown for MPI model, while full curves represents polynomial fit to underlying data. Dashed curve segments based on fit excluding the 25% warmest of all months are added to the right. Taken from (Christensen et al., 2008)

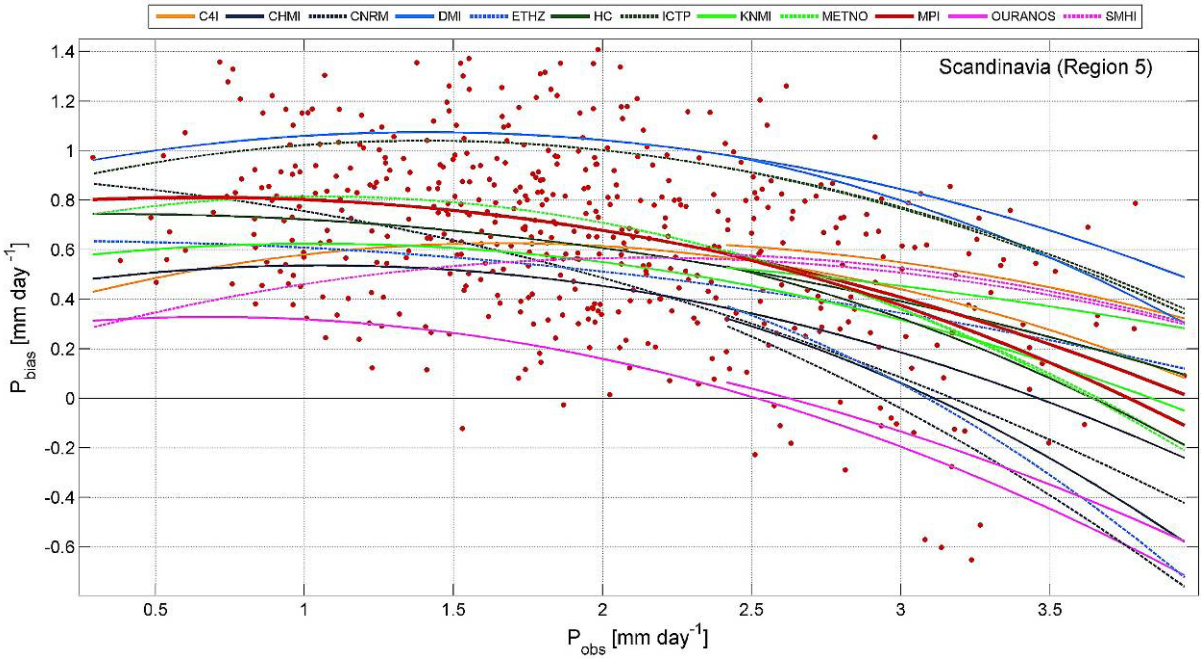


Figure 9.30: Monthly mean model precipitation bias vs. observed monthly mean precipitation for Region SC (Scandinavia) for the period 1961–2000. Raw data shown for MPI model, while full curves represent polynomial fit to underlying data. Dashed curve segments based on fit excluding the 25% wettest of all months are added to the right. No model precipitation data available for the UCLM PROMES experiment due to a temporary error in the data archive. Taken from (Christensen et al., 2008)

Figure 9.31: [PLACEHOLDER FOR FIRST ORDER DRAFT] Global display of regional model skill, as a synthesis figure. Preferably [tbc] for the same regions as those displayed in [Figure 9.28]. Generic results for various regions (sample, appropriately, results from ENSEMBLES (Europe), NARCCAP (North America), RMIP (Asia), CLARIS (South America), CORDEX (multiple regions)). [So far, ENSEMBLES accommodates the largest set of models for a specific region and features the longest hindcasts; Figure to be created, no suitable placeholder available.]

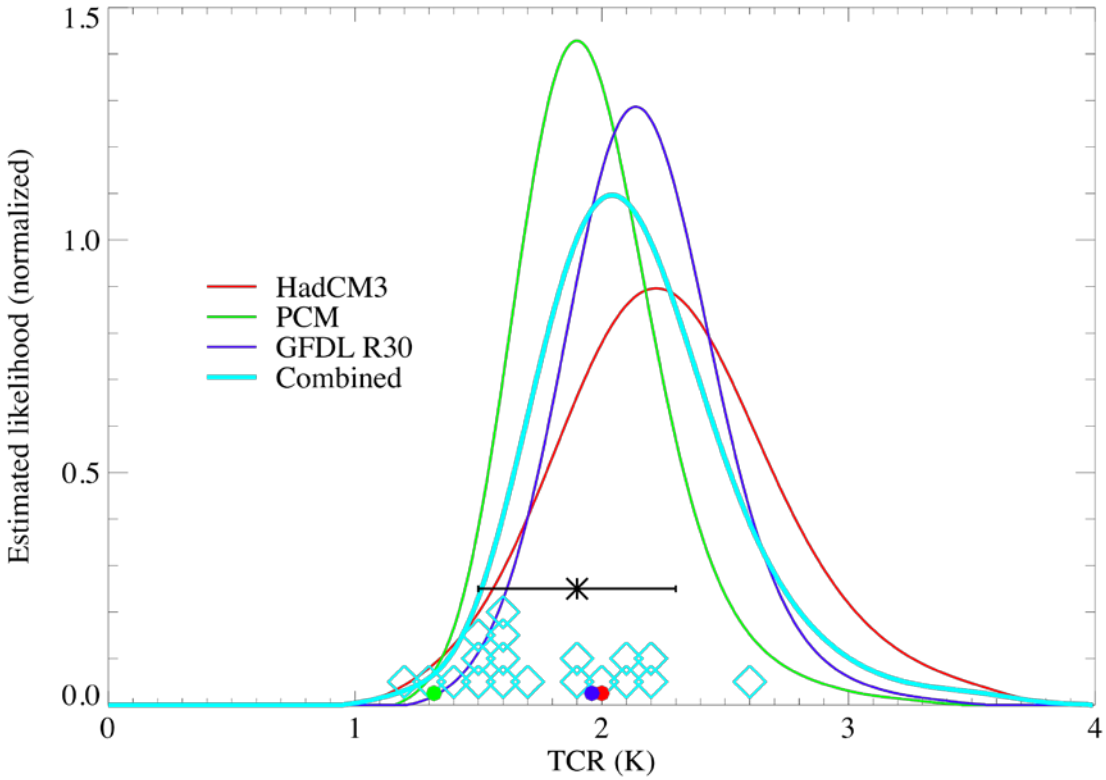


Figure 9.32: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure from (Stott and Forest, 2007) Probability distributions of Transient Climate Response estimated from climate change detection statistics and from calibrated EMICs. (Figure 8 and caption from (Stott and Forest, 2007)). Caption from: Probability distributions of TCR (expressed as warming rates over the century), as constrained by observed 20th century temperature change, and as calculated using HadCM3 (red), PCM (green), GFDL (blue) and from unweighted average of all three PDFs (turquoise). Coloured circles show each model's TCR. Also shown as diamonds are the TCRs of climate models forming part of the IPCC AR4 ensemble and the 5 to 95 percentile range derived by (Forest et al., 2006) using a large ensemble of EMICs is shown as the black bar with the star showing the median. [Data from (Knutti and Tomassini, 2008) to be incorporated.]

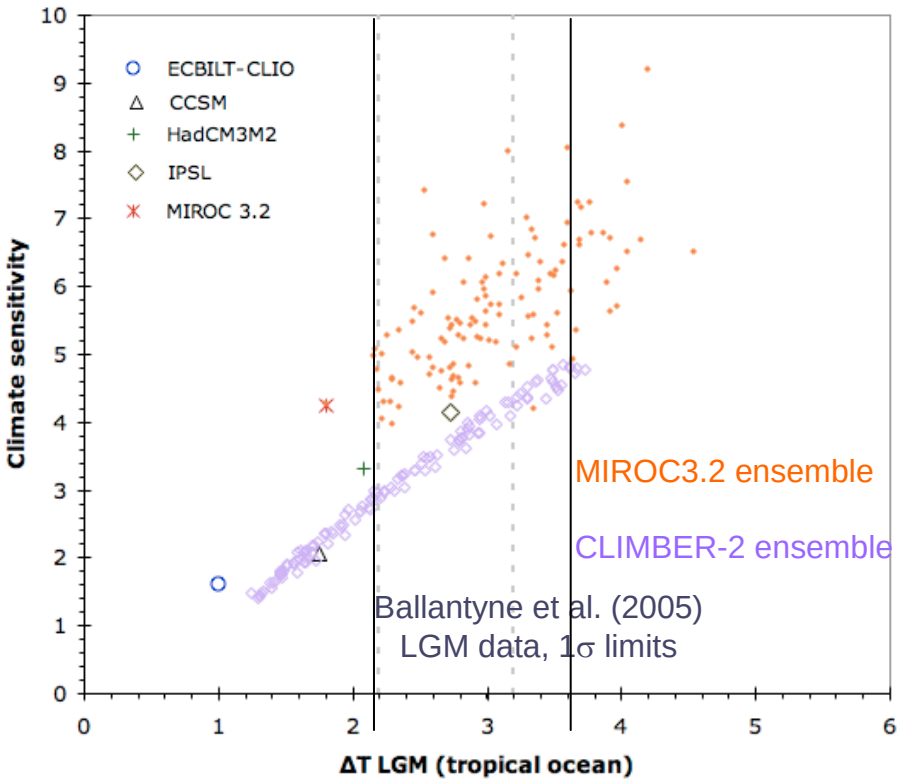


Figure 9.33: [PLACEHOLDER FOR FIRST ORDER DRAFT] **[**Example of figure that could be drawn using CMIP5 results, compared to PMIP3 (the one included here) and to other ensemble simulations (new estimates should be available soon). It is only an example, but a completely new set of figures will be proposed when new results are available.] The legend for this figure is extracted from Edwards et al. (Edwards et al., 2007): Climate sensitivity as a function of tropical sea surface temperature change between the pre-industrial and the LGM for three model ensembles: the MIROC3.2 model with PMIP LGM boundary conditions (Annan, private communication); the CLIMBER-2 model with PMIP LGM boundary conditions; and CLIMBER-2 with additional dust and vegetation forcings (Schneider von Deimling, private communication). For comparison, five PMIP 2 coupled ocean-atmosphere GCMs are shown (Crucifix, 2006). The MIROC3.2 ensemble uses a simpler version of the model than PMIP 2. The vertical lines indicate the 1σ limits of reconstructed tropical SST change at the LGM from Ballantyne et al. (2005).

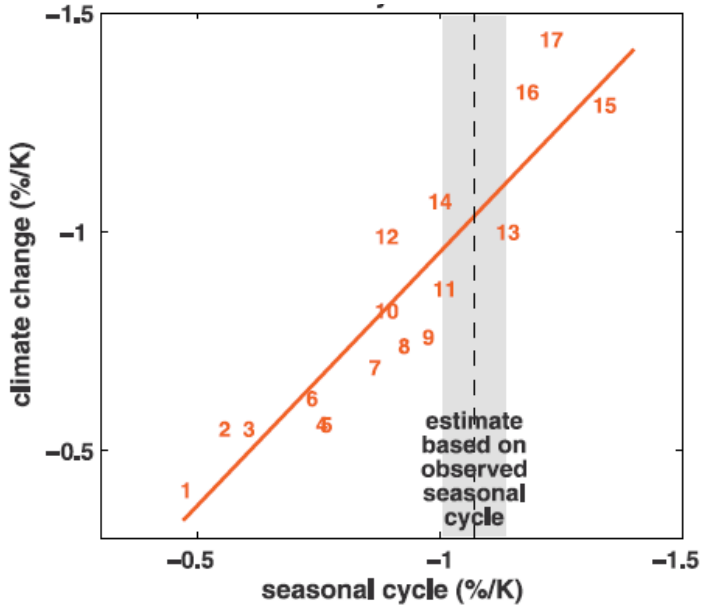
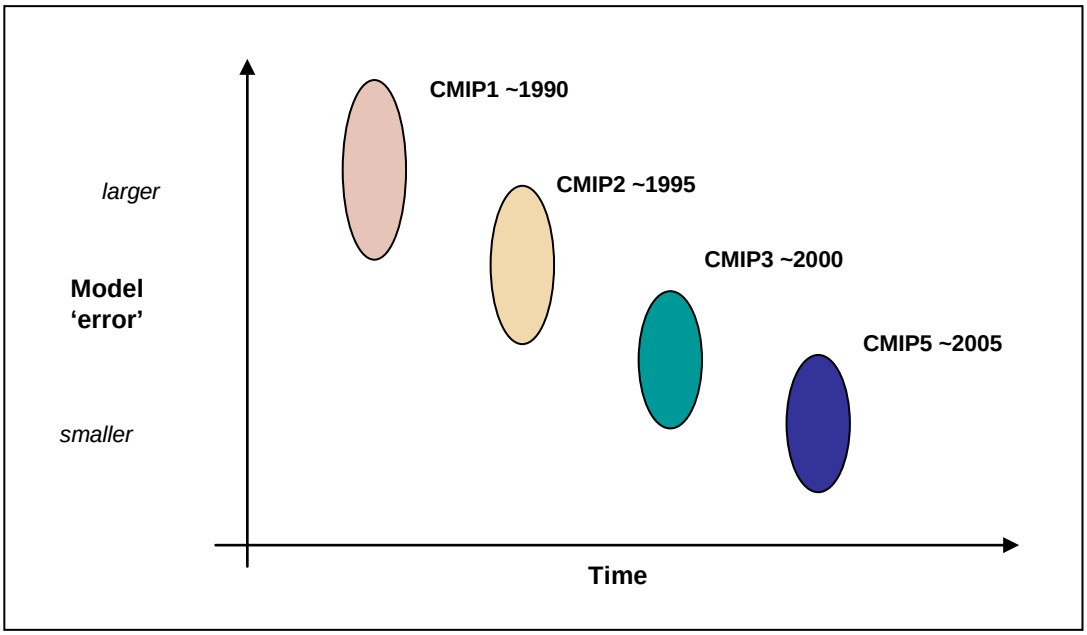


Figure 9.34: [PLACEHOLDER FOR FIRST ORDER DRAFT] Scatter plot of simulated springtime $\Delta\alpha_s/\Delta T_s$ values in climate change (ordinate) vs simulated springtime $\Delta\alpha_s/\Delta T_s$ values in the seasonal cycle (abscissa) in transient climate change experiments with 17 AOGCMs (α_s and T_s are surface albedo and surface air temperature, respectively). From Hall and Qu (2006). [to be replaced with promising new and not yet existing CMIP5 figure that relates model performance to projections.]



FAQ 9.1, Figure 1: [PLACEHOLDER FOR FIRST ORDER DRAFT] Figure illustrating model improvement over time, based on some metric, or collection of metrics (as indicated by the coloured 'clouds'), computed for models participating in the various CMIP intercomparions. A figure along these lines was published by Reichler and Kim (2008), but something a bit different to be produced based on more recent results and better understanding of the properties of various model performance metrics.