Chapter V: Model simulation of major climate features

Mean climate

Monthly mean near-surface temperature is well simulated by modern AOGCMs. This success occurs despite the fact that nearly all models now allow the ocean and atmosphere to exchange heat and water without explicitly forcing agreement with observation by artificial adjustment to air-sea fluxes. Figure V A quantifies the extent of agreement between simulations by several models and observations for both temperature and precipitation (the triangular points will be discussed in Chapter VI below). Each model's temperature or precipitation simulation produces a single point on the diagram, but in the figure, the ranges of results from all the models are shown as shaded areas.

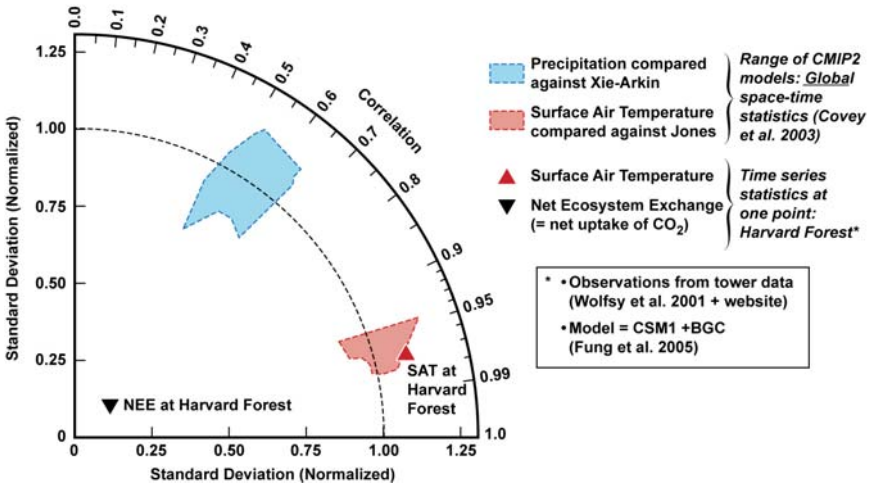


Figure V. A. Taylor Diagram of CMIP3 models

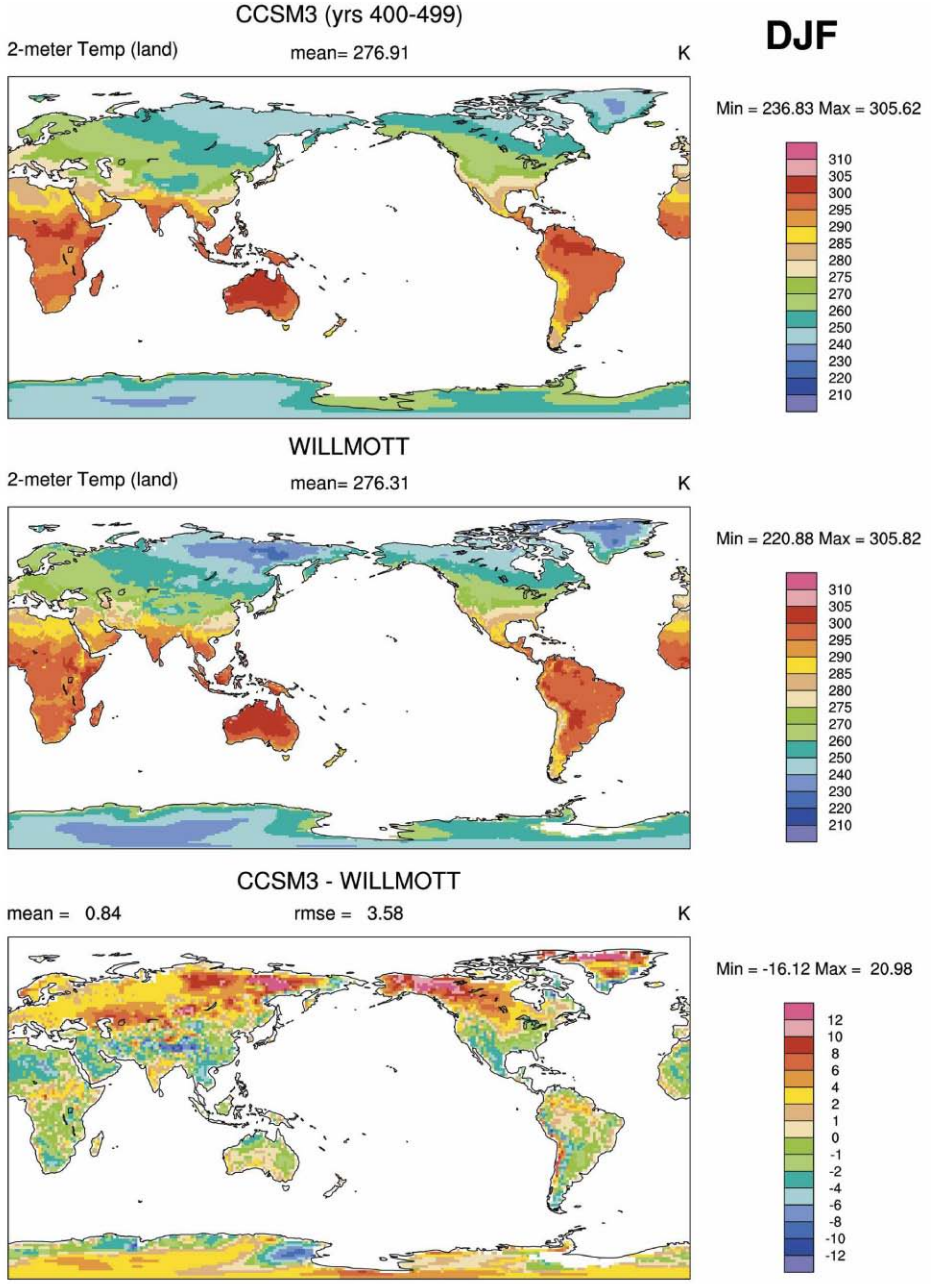
This type of diagram (Taylor 2001) displays the overall space-time correlation between simulated and observed variables as an angular coordinate. A 100% perfect correlation would place a point along the horizontal direction to the right, while zero correlation would place a point along the upward vertical direction. Looking at the red-shaded area that depicts the range of near-surface temperature simulations, one sees a remarkable 95–98% correlation with observations. The second independent (radial) coordinate in the diagram gives the ratio of simulated to observed amplitude for the variations that are being correlated. A value of 1.0 indicates perfect agreement of the amplitudes. In this coordinate system, complete agreement between simulation and observation in both dimensions would place a point where the dashed semicircle and the horizontal line intersect. The distance from this point to the actual point for any given model is proportional to the combined root-mean-square model error in both space and time dimensions. Temperature points for all of the models lie very close to complete agreement with observation—indeed nearly within the uncertainty range of the observations themselves (Covey *et al*., 2003).

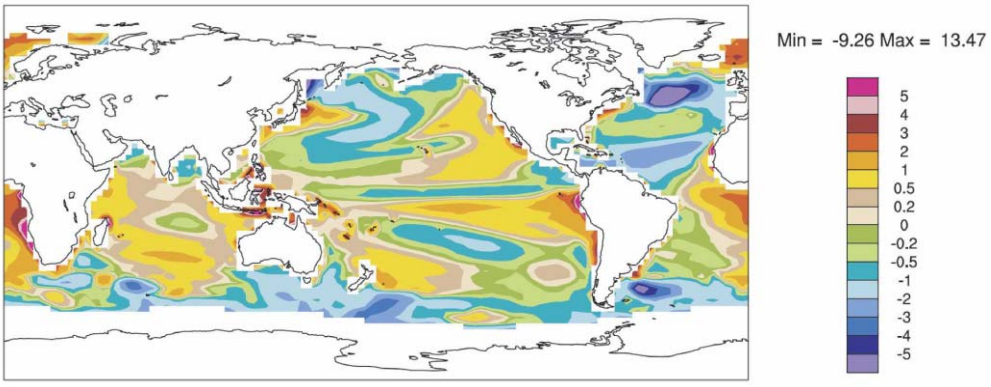
For monthly mean precipitation, AOGCM simulations are considerably less precise than for temperature. The figure shows that overall space-time correlation between models and observations is ~50–60%. Qualitative examination of latitude-longitude maps shows that AOGCMs generally reproduce the observed broad patterns of precipitation amount and year-to-year variability (A. Dai, 2006: Precipitation characteristics in eighteen coupled climate models, J. Climate, in press). The most prominent error is that models without flux adjustment fail to simulate the observed northwest-to-southeast orientation of a large region of particularly heavy cloudiness and precipitation in the Southwest Pacific Ocean (the Southwest Pacific Convergence Zone or SPCZ). Instead, these models produce an unrealistic set of Inter-Tropical Convergence Zones in two parallel lines straddling the Equator: a “double ITCZ” pattern. The double-ITCZ error has been frustratingly persistent in climate models despite much effort to correct it. The average day-night cycle of temperature and precipitation in AOGCMs exhibits general agreement with observations, although simulated cloud formation tends to start too early in the day. Another discrepancy between models and observations appears upon sorting precipitation into light, moderate and heavy categories. Models reproduce the observed extent of moderate precipitation (10-20 mm/day) but underestimate the extent of heavy precipitation and overestimate the extent of light precipitation

(Dai 2006). Additional model errors appear when precipitation is studied in detail for particular regions, e.g. within the US (Ruiz-Barradas, A., and S. Nigam, 2007).

Taking examples from two of the US model families discussed in Chapter IV, one finds that AOGCM-simulated and observed maps of surface temperature and even precipitation appear rather similar at first sight. Constructing simulated-minus-observed "difference maps," however, reveals monthly and seasonal mean temperature and precipitation errors up to ~10°C and 7 mm / day respectively at some points (Figs V B, W. Collins *et al*., 2006; and V C Dellworth *et al*., 2006).

Figure V.B 1–4. CCSM3 annual mean simulated-minus-observed sea surface temperature [°C]





1

The CCSM3 temperature difference maps exhibit the largest errors in the Arctic (note scale change in last frame), where continental wintertime near-surface temperature is overestimated. AOGCMs find this quantity particularly difficult to simulate because, for land areas near the poles in winter, models must resolve a strong temperature inversion (warm air overlying cold air).

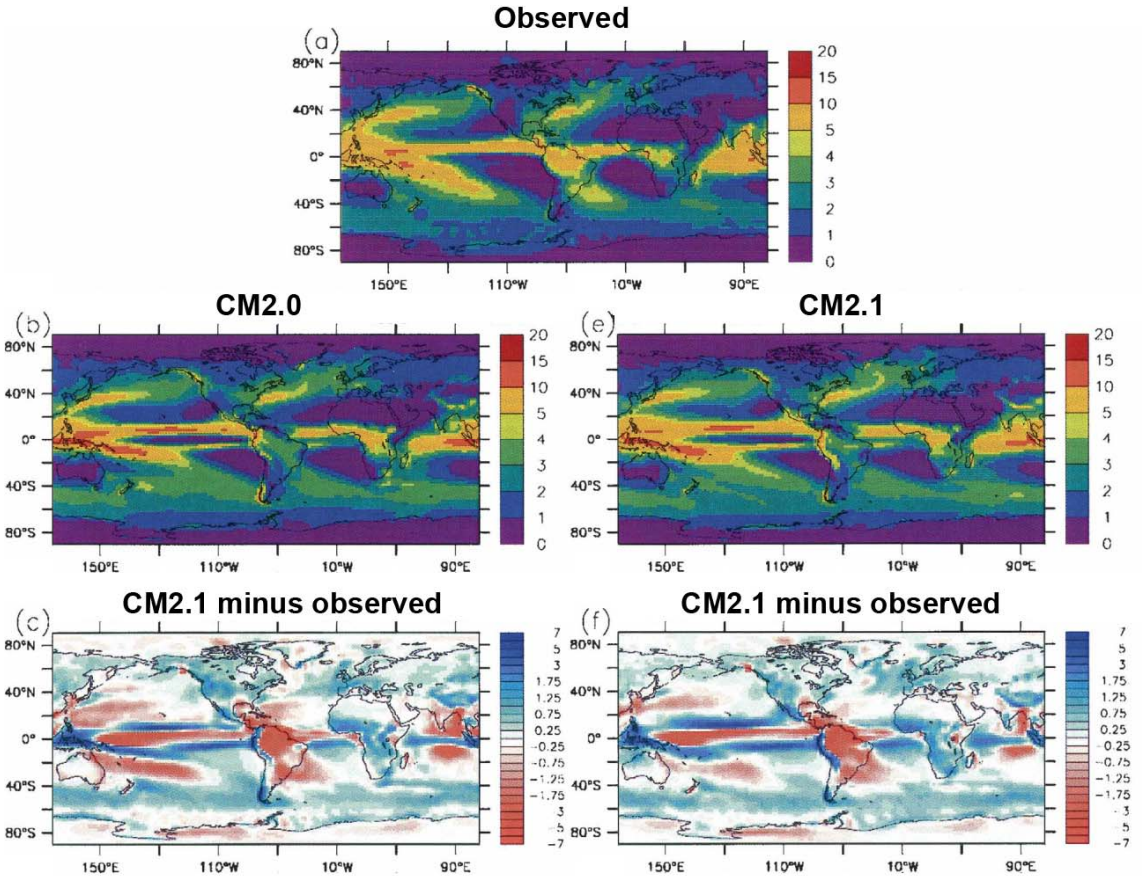


Figure V. C 1-5. **Observed and model-simulated precipitation [mm/day]**

The GFDL precipitation difference maps reveal significant widespread errors in the tropics, most notably in the ITCZ region discussed above and in the Amazon River basin, where precipitation is underestimated by several millimeters per day.

Similar precipitation errors appear in the following table of CCSM3 results Table V 1.

Table V 1 CCSM3 Precipitation by region(Collins, et al, 2006)

Region	CCSM3-simulated precip	Error
Southeast USA (30-40°N, 80-100°W)	2.4 mm/day	–24%
Amazon basin (10°S-10°N, 60-80°W)	4.5 mm/day	–28%
Southeast Asia (10-30°N, 80-110°E)	3.1 mm/day	–24%

AOGCM precipitation errors have serious implications for "Earth system" models with interactive vegetation, because such models use the simulated precipitation to calculate plant growth (see Chapter VI below). Errors of the magnitude shown above would produce an unrealistic distribution of vegetation in an Earth system model, e.g. by spuriously deforesting the Amazon basin.

In summary, modern AOGCMs generally simulate large-scale mean climate with considerable accuracy, but the models are not reliable for aspects of mean climate in some regions, especially precipitation.

20th century trends

Modern AOGCMs are able to simulate not only the time-average climate but also changes (trends) of climate during over the past century or more. For example, Figure V D shows results from the three US models and "average" CMIP3 models. All parts of the figure display the same curves of annual mean globally averaged near-surface temperature as observed by the UK Climatic Research Unit (CRU), as well as simulated by the average over all CMIP3 models and the average over only those CMIP3 models that included the effects of volcanic eruptions. Results from individual US models are shown both for separate ensemble members (dotted lines) and for the average over all ensemble members (continuous lines). Separate ensemble members were run under a variety of initial conditions. The precise initial conditions, especially deep ocean temperature and salinity, are not known for 1860; the spread among the dotted-line curves thus indicates uncertainty in model-simulated temperature arising from our lack of knowledge of initial conditions.

These results demonstrate that modern climate models typically exhibit good agreement with observed near-surface temperature trends for the global mean (Min and Hense, 2006). Global warming during the past few decades is successfully simulated by the models only if they include anthropogenic emissions of greenhouse gases and aerosols. Min and Hense, (in press) show the same is true for most individual continents. Observed trends in climate extremes such as heat-wave frequency and frost-day occurrence are also simulated with basic reliability by the latest generation of AOGCMs (Tebaldi et al., in press).

Figure V D 1. Twentieth century globally averaged surface temperature simulation from GFLD CM2.1

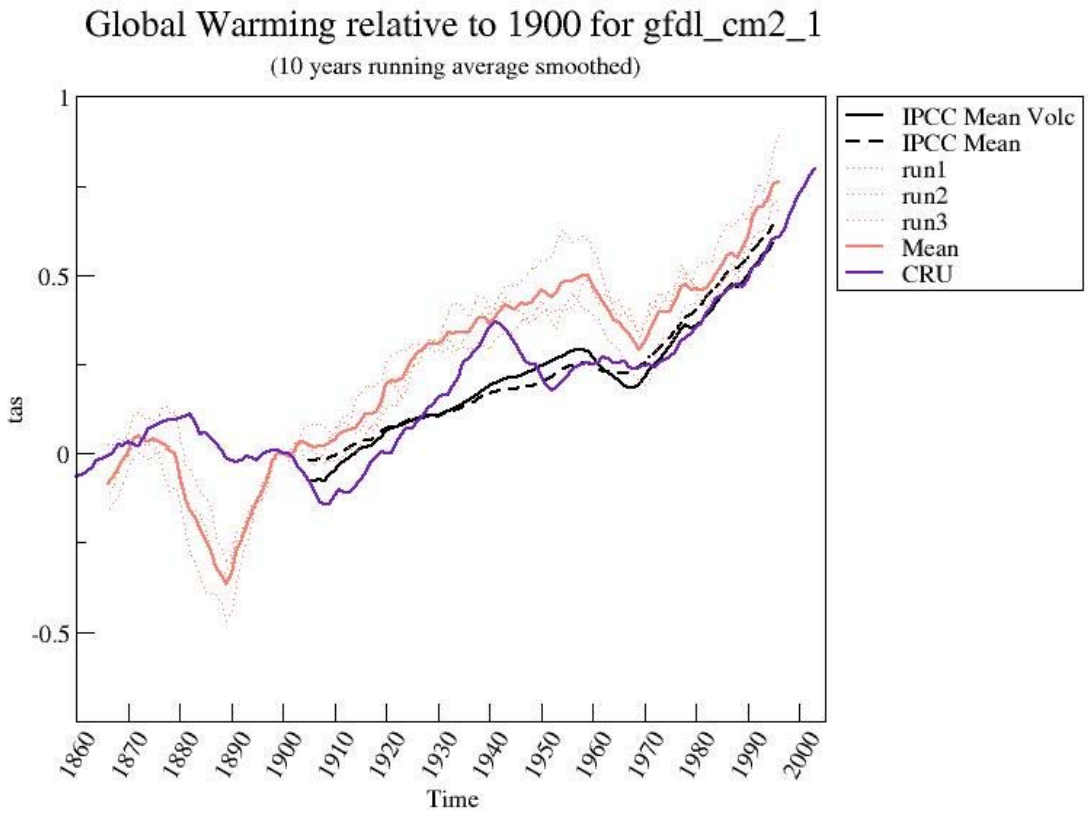


Figure V D 2 Twentieth century globally averaged surface temperature simulation from GISS Model E-r

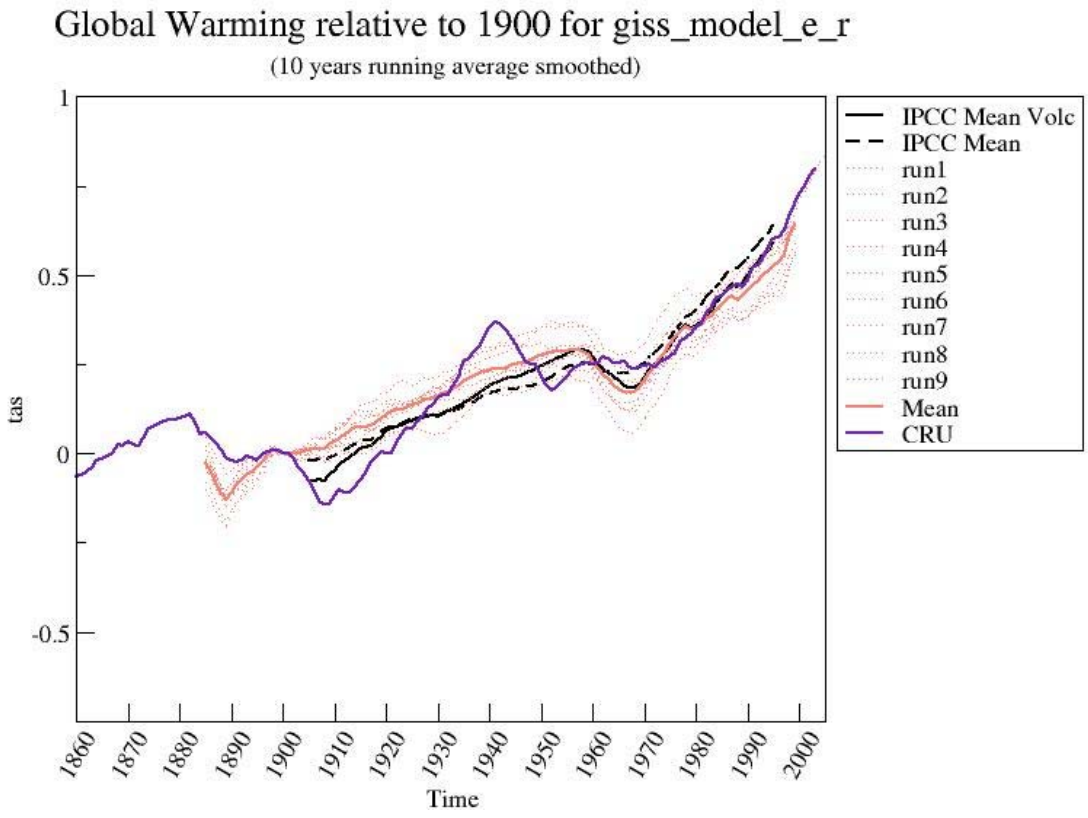


Figure V D 3 Twentieth century globally averaged surface temperature simulation from CCSM3

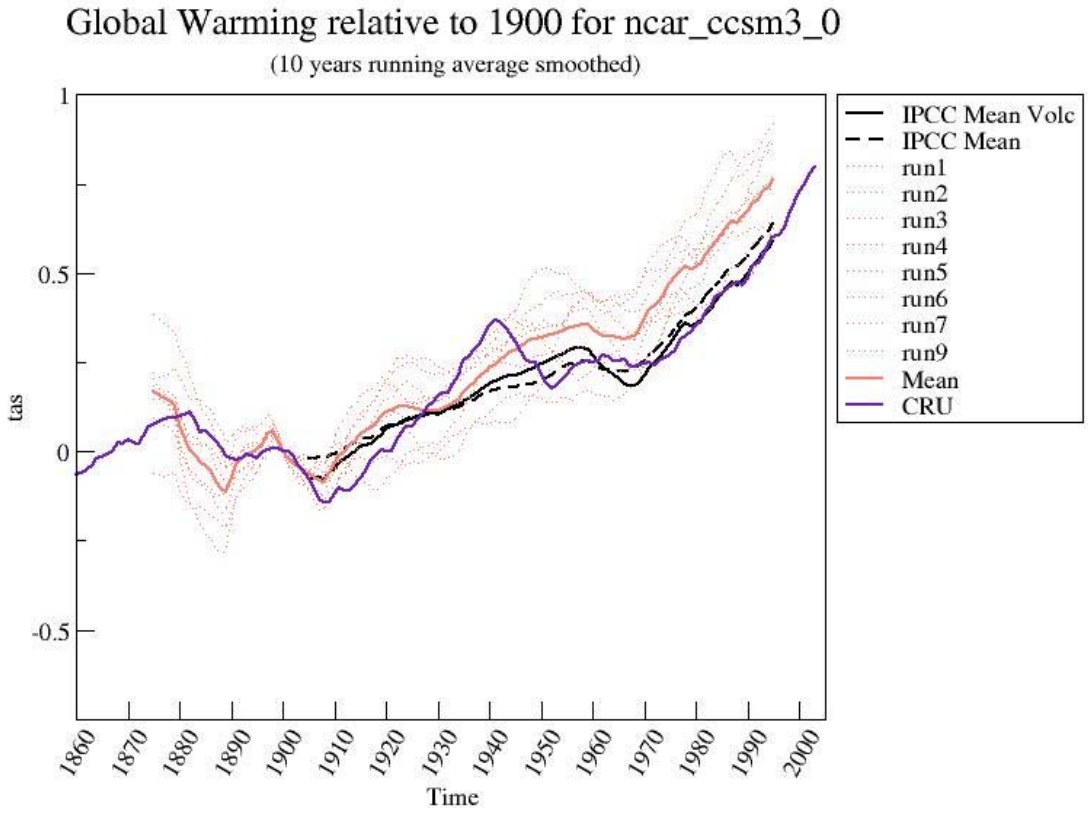
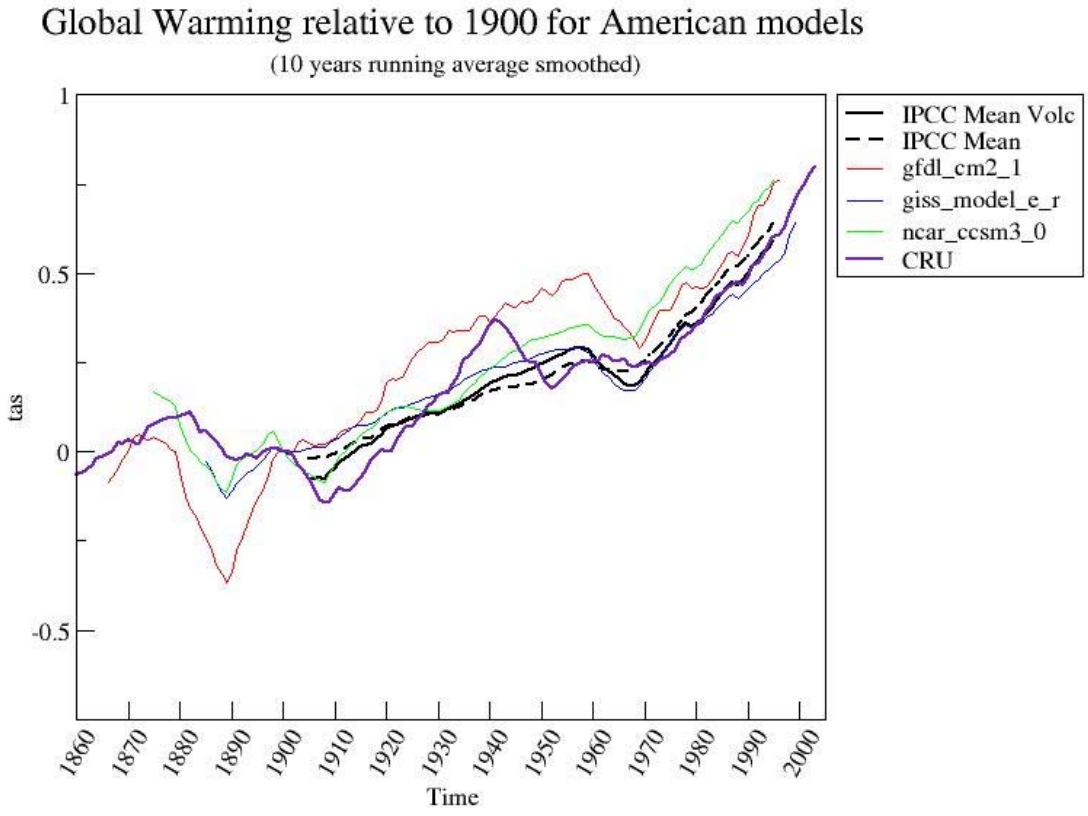


Figure V D 4 Twentieth century globally averaged surface temperature simulation from the three US CMIP3 models and the average of all CMIP3 models that include4d volcanic effects



At smaller scales the model simulation of trends can be less accurate. For example, model-simulated trends do not consistently match the observed lack of 20th century warming in the Central US **(**Kunkel et al., in press). The evolution of large-scale patterns, however, can be simulated with fair detail by modern climate models. For example, the longitude-latitude map of trends from GISS modelE agrees reasonably well with the observed spatial distribution Fig V E (Hansen *et al*., 2006) .

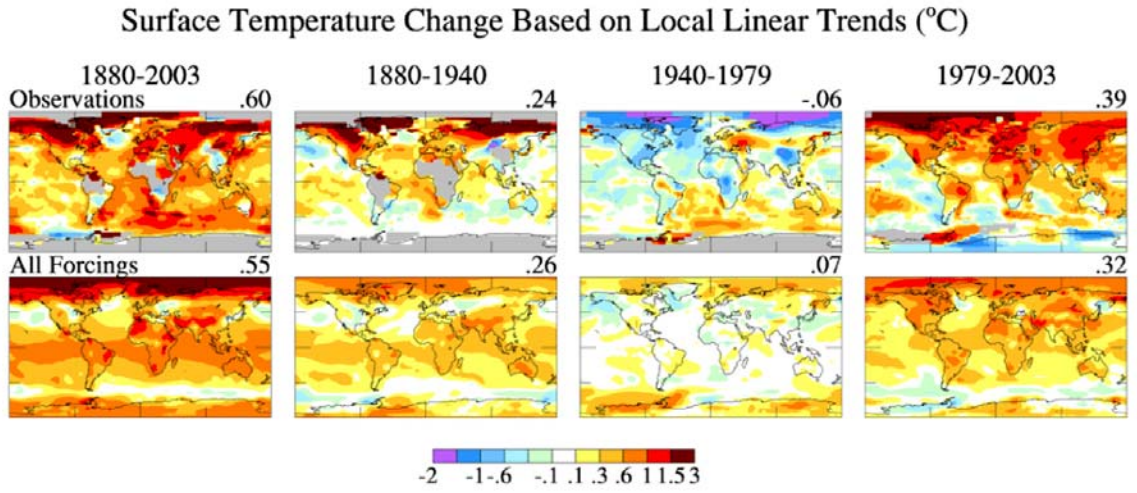


Figure V. E. The figure shows general agreement between the model and observations not only for the overall period 1880-2003, but also for the segments 1880–1940 and 1979–2003, which encompass periods of early and late 20th century warming.

Amplification of warming and cooling at high northern latitudes is the most obvious feature in the observations. For the period 1940–1979, the model simulates only a small change in global mean temperature in agreement with observations, but it fails to simulate the strong north polar cooling observed for this period. As a result, the model-simulated global mean temperature change (upper right corner of each frame) for 1940–1979 is slightly positive rather than slightly negative as observed. For both 20th century warming periods, the model simulates but underestimates the high-latitude amplification of global warming.

Finally, the CCSM3 simulates 4.7 cm of global mean sea level rise during the 20th century (Meehl *et al*. 2006). The actual value of sea level rise is 3–5 times as large, but the model does not include melting glaciers and ice sheets, and therefore it simulates only the part of sea level rise due to expansion of ocean water from heating.

A number of specific climate phenomena in addition to near-surface temperature, precipitation and sea level are discussed in the following sections. These are important for practical applications of climate models because they directly affect near-surface temperature and precipitation patterns (and thereby indirectly affect the evolution of sea level, together with many other features of climate).

Annular Modes

The primary mode of Arctic interannual variability is the Arctic Oscillation (Thompson and Wallace 1998), which is also referred to as the Northern Annual Mode (NAM) and which is related to the North Atlantic Oscillation (Hurrell 1995). The primary mode of Antarctic interannual variability is the Southern Annular mode (SAM) (Thompson and Wallace 2000), also known as Antarctic Oscillation. Coupled global climate models have shown skill in simulating the NAM (Fyfe et al. 1999, Shindell et al. 1999, Miller et al. 2006), although in some cases too much of the variability in sea level pressure is associated with the NAM in these models (Miller et al. 2006). Global climate models also realistically simulate the SAM (Fyfe et al. 1999, Cai et al. 2003, Miller et al. 2006), although some details of the SAM (e.g. amplitude and zonal structure) show

disagreement between global climate model simulations and reanalysis data (Raphael and Holland 2006; Miller et al. 2006).

In response to increasing concentrations of greenhouse gases and tropospheric sulfate aerosols in the 20th century, the multi-model average exhibits a positive annular trend in both hemispheres, with decreasing sea-level pressure (SLP) over the poles and a compensating increase in mid-latitudes (Miller et al. 2006). However, the models underestimate the coupling of stratospheric changes (from volcanic aerosols) to annular variations at the surface, and may not simulate the appropriate response to increasing GHGs (Miller et al. 2006)) and changes in stratospheric ozone (Arblaster and Meehl, 2006).

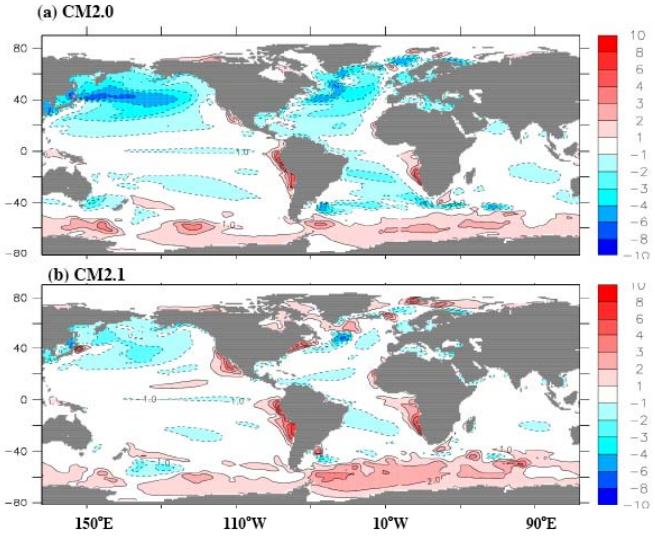
.

Ocean structure and circulation

A set of ocean characteristics or metrics (sea surface temperature, ocean heat uptake, meridional overturning and ventilation, sea level variability and global sea level rise) is used to describe the realism of the ocean in the climate models.

Sea surface temperature: The sea surface temperature (SST) plays a critical role in the determination of the climate and the predictability of the changes. In general, when the simulated fields of SST are compared to observational fields there is improvement in the models' representation of the mean SST Figure V F(Delworth *et al*., 2006) compares the CM2.0 and CM2.1 mean SST field averaged over a period of 100 years to the Reynolds SST observational climatology. With an improved atmospheric core and a different viscosity parameter value, the later version (CM2.1) of the GFDL climate model produces a reduced cold bias in the northern hemisphere.

Figure V F Maps of errors in simulation of annual mean sea-surface temperature (SST). Units are K. The errors are computed as model minus observations, where the observations are from the ReynoldsSST data (provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/). (a) CM2.0 (using model years 101-200). (b) CM2.1 (using model years 101-200). Contour interval is 1K, except that there is no shading for values between 1 K and +1 K.



The CCSM3.0 model also has improved its simulation of SST primarily in the handling of the processes associated with the mixed layer of the upper ocean waters (Danabasoglu *et al*., 2005). The improvement in the representation of the SST is apparent especially in the eastern tropical Pacific (see Figure V G). An inter-model comparison of the 50 year global SST trend for each model is shown in Figure V H. The SST trends range from a low of 0.1°C/50yrs to a high of about 0.6°C/50yrs, with the observational trend estimate given as about 0.43°C/50yrs. The figure also shows that within a group, the estimates significantly vary. This distribution of values in SST trends shows that improvements in any model's representation of SST are dependent on both advances in the ocean and atmospheric components.

Figure V G Differences in annual-mean surface temperature between CCSM2 and the HadISST data set (Rayner *et al*. 2003) (top); corresponding differences for CCSM3 (bottom) (Collins, et al, 2005).

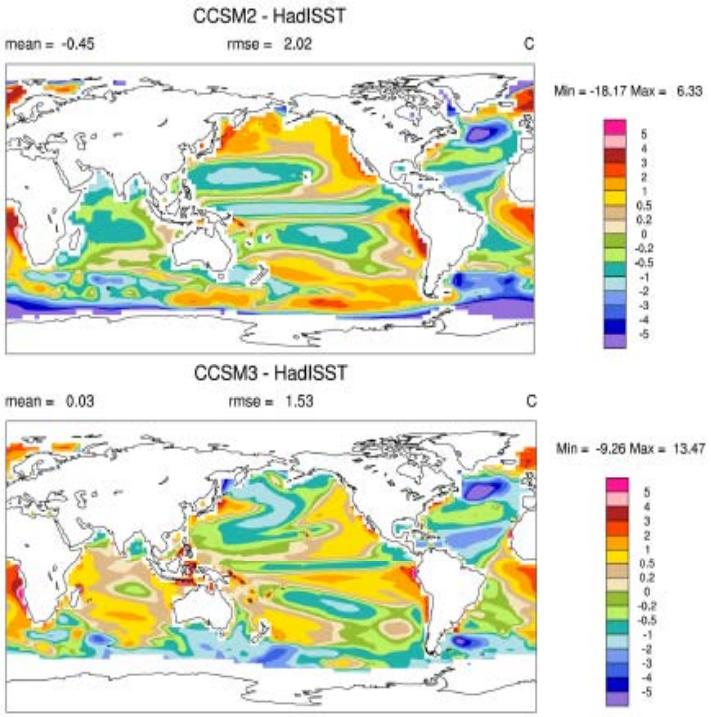
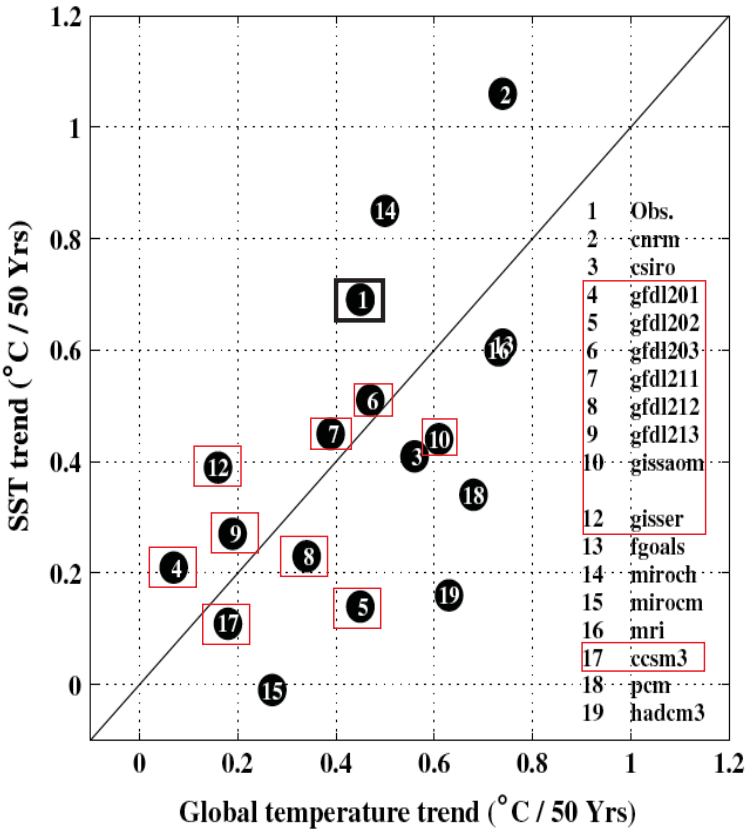


Figure V H Scatter plot of the SST trends averaged in the central and eastern tropical Pacific (9 S–9 N and 90–180 W), and global mean surface temperature trends. Correlation of the model results is 0.58, of higher magnitude than the 95% significance level of 0.46. The 1:1 line is drawn for clarity. The red boxes denote US Climate models and the black box is the relationship computed from observations. (**Zhang & McPhaden, 2006)**



Meridional overturning circulation and ventilation: The circulation process related to the transportation heat and freshwater throughout the global oceans is referred to as thermohaline circulation. The Atlantic portion of this process is called the Atlantic meridional overturning circulation (AMOC). Tropical and warm waters flow northward via the Gulf Stream and North Atlantic Current. The southward flow occurs when water is sub-ducted in the regions of the Labrador Sea and Greenland Seas and occurs when the freshening of the surface waters become denser and flow down the slope to deeper depths. Similar processes occur at locations in the Southern Ocean. Ventilation is the process by which these dense surface waters are carried into the interior of the ocean. The important climate parameter is the rate at which this process occurs, the so-called "ventilation rate". It has been suggested that this pattern of circulation if it becomes weaker (i.e. less warmer water flowing towards Europe) will impact the climate. It is thus important to understand how well the ocean component simulates the observed estimates of these overturning processes.

Schmittner et al. [2005] examined the performance of the models in reproducing the observed meridional overturning in 4 of the 5 US models. The authors examined a small ensemble set of simulations to quantify the uncertainty in the models' representation of 20^{th} century AMOC transports. To make their estimate, they evaluated the global temperature (T), the global salinity (S), the pycnocline depth (D), the surface temperature and surface salinity in the Atlantic (SST, SSS), and calculations of the overturning at 3 locations ~in the Atlantic. Their results suggest that temperature is simulated the most successfully on the large scale and that the overturning transports at 24°N are close (~18Sv) to the observed measurements (~15.8Sv). However, the maximum mean overturning transports in these models are too high (23.2, 31.7, 27.7, and 30.9 Sv: *Schmittner et al*. [2005] and 21.2 Sv from *Bryan et al.* [2006]) than the observed value (17.7 Sv). Table V 2 shows a reduced version of Table 1 from *Schmittner et al*. [2005] that shows the root mean errors (RMS) for the various quantities as compared to observations. The authors do not attempt to explain why the models are different from each other and from observations, rather, that there is a broad range in the value of these metrics for a set of climate models.

Table V 2 Model Errors

Model	T_{global}	S_{global}	D_{global}	SST_{NAtl}	SSS_{NAtl}	D_{NAtl}	$AMOC_{24N}$ (15.8) (SV)
GFDL-2.0	0.20	0.43	0.57	0.34	0.53	0.75	0.16 (18.3)
GISS-AOM	0.66	0.75	2.29	0.43	0.79	3.48	0.22 (19.2)
GISS-EH	0.31	0.76	1.57	0.61	1.12	1.85	0.34 (21.1)
GISS-ER	0.69	0.82	2.06	0.65	1.11	2.40	0.13 (17.9)

From Schmittner et al. [2005] Table 1. RMS Errors for the Individual Models; RMS errors are normalized by the standard deviation of the observations unless otherwise stated. *Schmittner et al.* 2005; "Observation-based estimates of the AMOC at 24 N from *Ganachaud and Wunsch* [2000] *and Lumpkin and Speer* [2003], at 48 N from *Ganachaud* [2003], and its maximum value in the North Atlantic from *Smethie and Fine* [2001] *and Talley et al.* [2003], as well as temperature, salinity, and pycnocline depth observations from the World Ocean Atlas 2001 [*Conkright et al.*, 2002] are used to evaluate the climate models."

The global overturning circulation can also be quantified by also examining the realism of the transports through the Drake Passage. The passage, between the tip of South America and the Antarctic Peninsula provides a constrained passage to measure the flow between two large ocean basins. The observed mean transport is around 135 Sv. *Russell et al.* [2006] estimate the flow in the passage for a subset of the climate models (Table V 3). There is a wide range in the simulated mean values. The interaction between the atmospheric and ocean component models appears to be important in reproducing the observed transport. The strength and location of the zonal wind stress correlates with how well the transport reflects observed values.

TABLE V 3

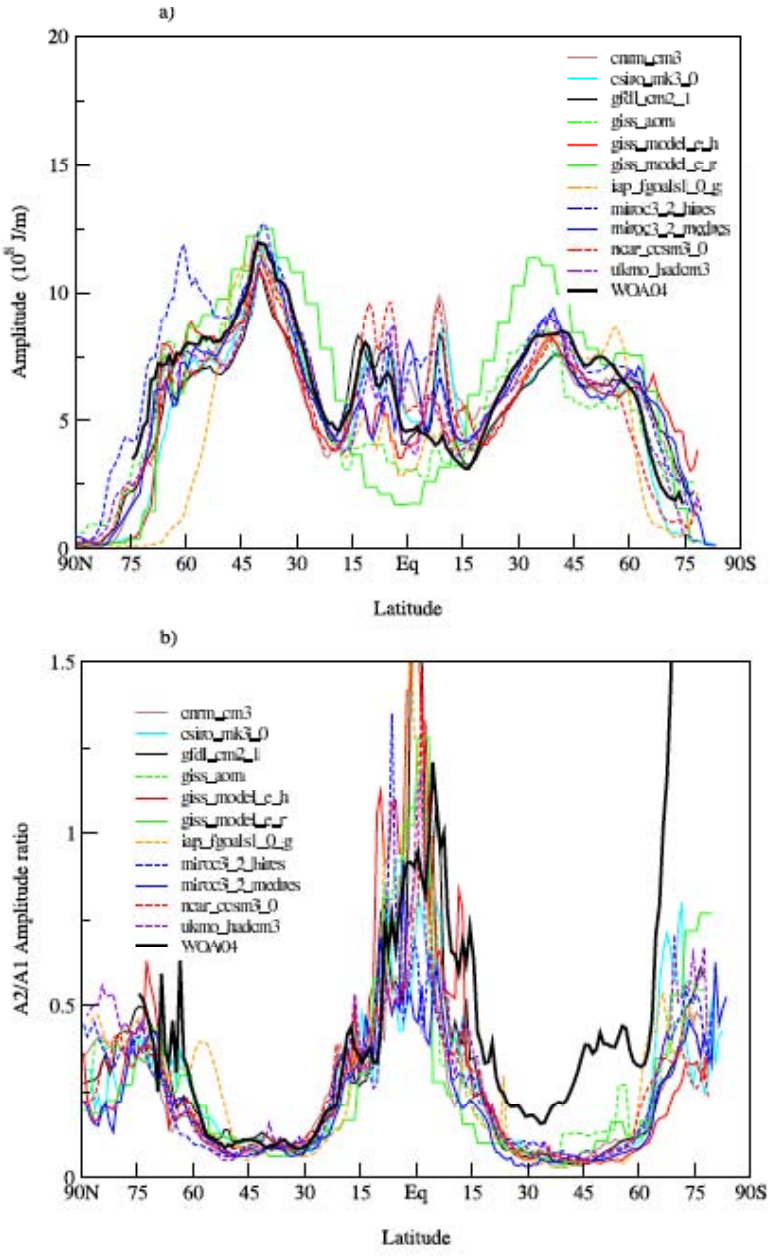
Model	ACC (Sv)	$d\rho/dy$ (kg m^{-3})	Total τ_x 10^{12} N	Max τ_x (N m^{-2})	Lat of max τ_x
Observational estimate	135	0.58	6.5	0.161	52.4
GISS-ER	266	0.62	4.3	0.107	46.0
GISS-AOM	202	0.38	2.9	0.166	43.5
GFDL-CM2.1	135	0.58	6.1	0.162	51.0
GFDL-CM2.0	113	0.56	4.5	0.149	46.0
GISS-EH	-6	0.43	3.6	0.096	46.0

Reduced From Table 1 *Russell et al.* [2006] Various parameters related to the strength of the ACC. The ACC transport is the integral of the zonal velocity across the Drake Passage at 69°W. The density gradient (dρ/dy) is the zonally averaged density difference between 65° and 45°S. The total ACC-related wind stress (τ_x total) is the integral of the zonal wind stress over the Drake Passage channel (54°–64°S). The maximum westerly wind stress (τ max) is the maximum of the zonally averaged wind stress that is located at the latitude given by Lat τ max. The observed ACC strength is from *Cunningham et al.* (2003). The observed density gradient is calculated from the World Ocean Atlas 2001 (*Conkright et al.* 2002). The observed wind data are from the NCEP long-term mean *(Kistler et al.* 2001). NA indicates data not archived at PCMDI

Northward Heat Transport: A common metric used to quantify the realism in ocean models is the northward transport of heat. This integrated quantity (from top to bottom and across latitude bands) gives an estimate of how heat moves within the ocean and is important in balancing the overall heat budget of the Earth. The calculations for the ocean's northward heat transport in the current generation of climate models show that the models reasonably represent the observations (*Delworth et al*. 2006, *Collins et al*. 2006, and *Schmidt et al*. 2006). The current models have significantly improved over the last generation in the Northern Hemisphere. Comparisons of the simulated values to the observed values for the North Atlantic are within the uncertainty of the observations. In the Southern Hemisphere, the comparisons in all the models are not as good, with the Indian Ocean transport estimates contributing to a significant part of the mismatch.

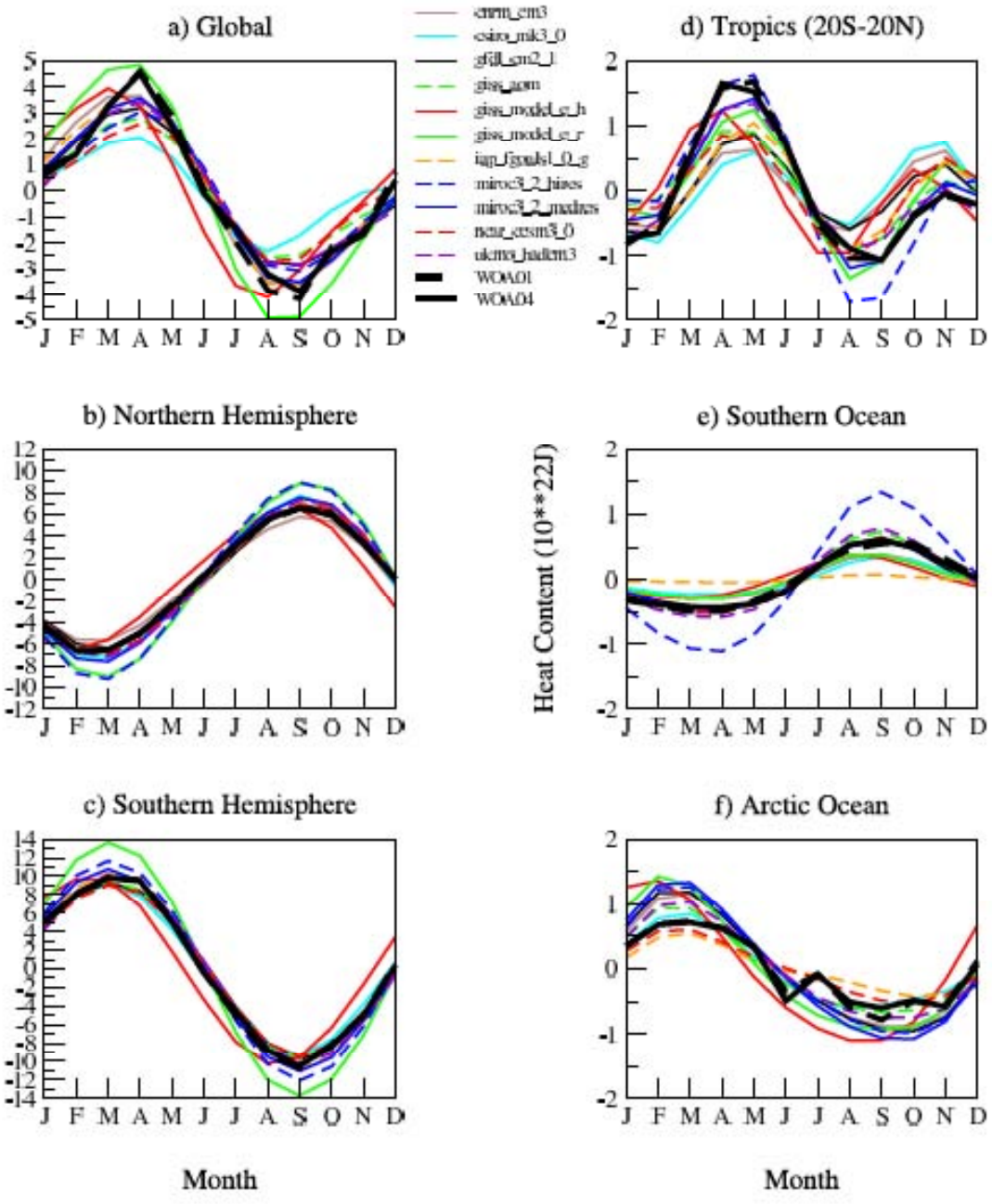
Heat Content: Related to the heat transport is the ocean's heat content itself. This can be thought of how realistically the models reproduce the uptake of heat. An evaluation of the temporally evolving ocean heat content in the suite of climate models for the AR4 shows the models abilities to simulate the zonally integrated annual and semi-annual cycle in heat content. In the middle latitudes [*Gleckler et al.* 2006], the models do a reasonable job while there is a broad spread of values for the tropical and polar regions. This analysis showed that the models replicate the dominant amplitude of the annual cycle along with its phasing in the mid-latitudes **[**Fig V I]. At high latitudes, the comparisons with observations are not as consistent. While the annual cycle and global trend are reproduced, analyses of the models [e.g. *Hansen et al.* 2005] show that they do not simulate the decadal changes in estimates made from observations [*Levitus et al.* 2001]. Part of the difficulty of comparisons at high latitudes and at long periods is the paucity of observational data [*Gregory et al.* 2004].

Figure V I 1



From Glecker *et al.* 2006 .**Figure 1.** Observed (WOA04) and simulated zonally integrated ocean heat content (0–250 m): (a) annual cycle amplitude (10^8J/ m^2) and (b) semiannual/annual (A2/A1).

Figure V I 2



From Glecker *et al.*, 2006, **Figure 3**. Annual cycle of observed (WOA04) and simulated basin average global ocean heat content (0–250 m). Units are 10^{22}J. Arctic Ocean is defined as north of 60 N, and Southern Ocean is south of 60 S.

Global mean sea level rise: Two separate physical processes contribute to the sea level rising: 1) the thermal expansion of the ocean from an increase in the heat uptake by the ocean (steric component) and 2) the addition of freshwater from precipitation, continental ice melt, and/or river runoff (the eustatic component). The current ocean component of all the models except the GISS models, conserve volume. In practice, the first component can be easily computed from a model's primary variables. The second contribution maybe considered as a freshwater flux into the ocean. The various ocean models handle the process in different ways. With the addition of a free surface in the current generation of ocean models, the freshwater flux into the oceans can be included directly [*Griffies et al*. 2001]. In other cases, the mass or freshwater contribution is computed by quantities estimated by land/ice sheet components of the climate model [e.g. *Church et al.*, 2005, *Gregory et al.*, 2006]. In general, the state-of-the-art climate models underestimate the combined global mean sea level rise as compared to tide gauge and satellite altimeter estimates while the rise for each of the separate components is within the uncertainty of the observed values. The reason for this is an open research question and may relate to either observational sampling or not correctly accounting for the all the eustatic contributions. The steric component to the global mean sea level rise is estimated to be 0.40+/-0.05mm/yr from observations [*Antonov et al.* 2005]. The models simulate a similar, but somewhat smaller rise [*Gregory et al.*, 2006, *Meehl et al.* 2005]. There are also significant differences in the magnitudes of the decadal variability between the observed and the simulated sea level or SSH. It most be noted, however, that progress is been made over the previous generation of climate models. When atmospheric volcanic contributions are included, for example, ocean models of the current generation capture the observed impact on the ocean (a decrease in the global mean sea level). Figure V J from *Church et al*. 2006 gives an example of a few models and their de-trended estimate of the historic global mean sea level that shows the influence of including the additional atmospheric constituents in changing the steric height of the ocean.

Figure V J

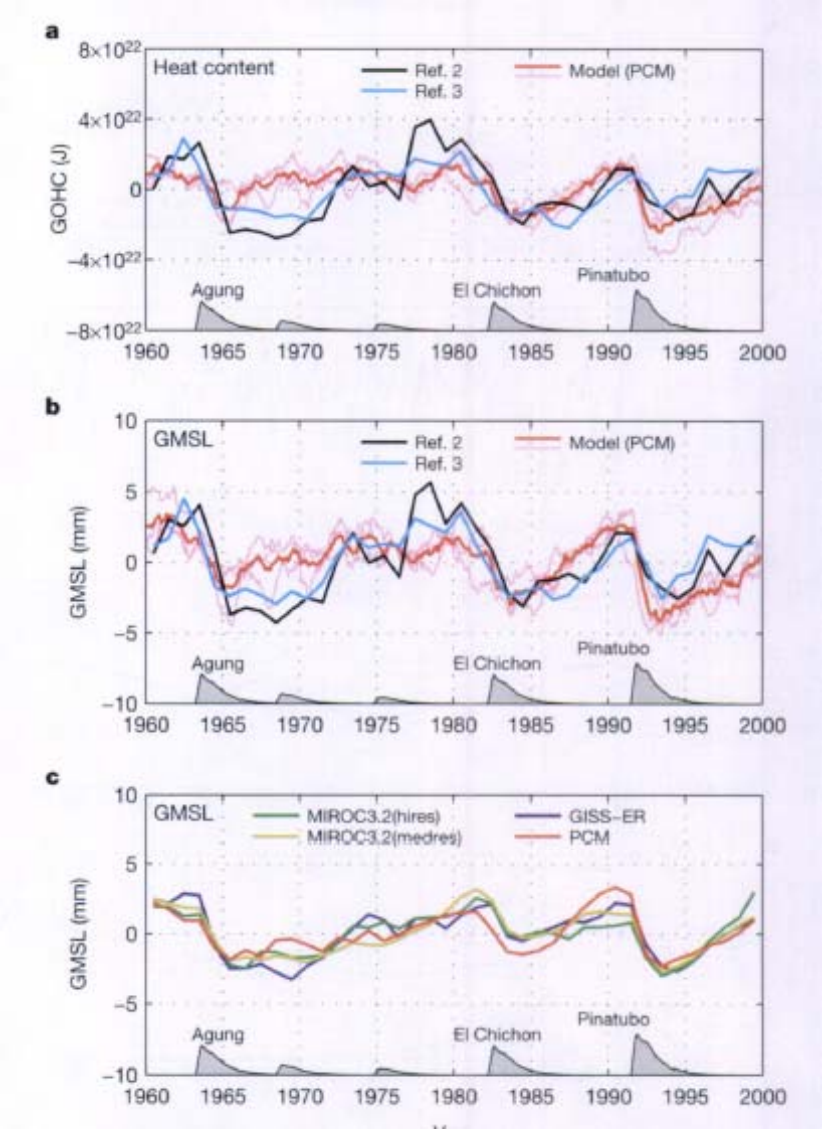


Figure 2 | Observed and modelled GOHC and GMSL for the period 1960–2000. The response to volcanic forcing, as indicated by the differences between the pairs of PCM simulations for GOHC (**a**) and the GMSL (**b**) is shown for the ensemble mean (bold line) and the three ensemble members (light lines). The observational estimates[2,3] of GOHC and GMSL are shown by the black and blue bold lines. For **a** and **b**, all results are for the upper 300 m only and have been detrended over the period 1960–2000. **c**, The ensemble mean (full depth) GMSL for the GISS-ER, MIROC3.2(hires), MIROC3.2(medres) and the PCM models (after subtracting a quadratic) are shown.

C. Simulation of specific climate dynamical features

Extratropical storms

Climate models have developed from numerical weather prediction models whose performance has been primarily judged on their ability to forecast mid-latitude weather. The success of these models in their simulation of midlatitude cyclones and anticyclones has resulted in the continuous growth in the value of numerical weather prediction. The ability of general circulation models to generate realistic statistics of midlatitude weather has also been central in the development of climate modeling. This is not only because midlatitude weather is important in its own right, but also because these storms are the primary mechanism by which heat, momentum, and water vapor are transported by the atmosphere, making their simulation crucial for the simulation of the atmospheric climate.

Indeed, it can be thought of as the defining feature of Atmospheric General Circulation Models (AGCMs) that they compute midlatitude eddy statistics and the associated eddy fluxes through explicit computation of the life cycles of individual weather systems and not through some turbulence or closure theory. It may seem very inefficient to compute the evolution of individual eddies when primarily interested in the long term statistics of the eddies, but it is has been the clear judgment of the community for decades that the explicit simulation of these eddies in climate models is far superior to the attempts that have been made to date in developing closure theories for the eddy statistics. The latter theories typically form the basis for EMICs (Earth System Models of Intermediate Complexity), which are far more efficient computationally than GCMs, but provide less convincing simulations.

Two figures illustrate the quality of the simulations of midlatitude eddy statistics that coupled AOGCMs of the horizontal resolution used in AR4 are capable of generating. Shown for the GFDL CM2.1 in Fig. V K 1 is the wintertime variance of the north-south component of velocity at 300 hPa (in the upper troposphere) and in Fig. V K 2 the wintertime poleward eddy heat flux, or the covariance between temperature and north-south velocity, at 850mb (in the lower troposphere). When analyzing eddy statistics it is often useful to filter the flow fields to retain only those time scales, roughly 2-10 days, associated with midlatitude weather systems, but the two quantities

chosen here are dominated by these time scales to a sufficient degree that they are relatively insensitive to filtering. Here we have simply removed the monthly means before computing variances. In each case, the eddy statistics are compared to the estimates of the observed statistics obtained from the NCEP-NCAR reanalysis (B.Wyman, personal communication).

In winter, Northern Hemisphere storms are organized into two major oceanic storm tracks over the Pacific and Atlantic oceans. Historically, it has been found that atmospheric models of resolutions of about 200-300 kms are typically capable of simulating the midlatitude storm tracks with comparable realism to that shown in the figure. The eddy amplitudes are often a bit weak and often displaced slightly equatorward, especially in Southern hemisphere summer (although the model shown here has a weaker Southern hemisphere bias than most models). In models with resolution coarser than 200-300kms, the simulation of the midlatitude storm tracks typically deteriorates significantly (see for example, Boyle 1993). It is thought to be important for the general improvement in model simulations described in Chapter 1 that most of the models in the CMIP3 database are now utilizing this 200-300km resolution. While finer resolution results in better simulations of the structure of midlatitude storms, including the structure of warm and cold fronts as well as the interaction between these storms and coastlines and mountain ranges, the improvements in the midlatitude climate on large scales tend to be less dramatic and systematic. Other factors besides horizontal resolution are considered to be important for the details of storm track structure, including the distribution of tropical rainfall, which is sensitive to the closure schemes utilized for moist convection, interactions between the stratosphere and the troposphere, which are sensitive to vertical resolution. Roeckner *et al* (2006) illustrate the importance of vertical resolution for the midlatitude circulation and storm track simulation.

A more detailed look at the ability of the AR4 models to simulate the space-time spectra of the observed eddy statistics is provided by Lucarini, *et* al,(2006). These authors view the deficiencies noted, which vary in detail from model to model, as serious limitations to the credibility of the models. But, as indicated in Chapter 1, our ability to translate measures of model biases into useful measures of model credibility is limited, and the implications of these biases in the space-time spectra of the eddies is not self-evident. Indeed, in the context of the simulation of the eddy characteristics generated in complex turbulent flows in the laboratory (e.g., Dimotakis , 2005) the

quality of these atmospheric simulations, closely based on fluid dynamical first principles, should probably be thought of as one of the most impressive characteristics of current models.

As an example of a significant model deficiency that can plausibly be linked to limitations in the credibility of the climate projections, note that the Atlantic storm track, as indicated by the maximum in velocity variance in Fig 5.1, is too zonally oriented, the observed stormtrack having more of an southwest-northeast tilt. This particular deficiency is common in the CMIP-3 models (van Ulden and van Oldenborgh, 2006) and is related to the difficulty in simulating the phenomenon of blocking in the North Atlantic with the correct frequency and amplitude. Van Ulden and van Oldenborgh make the case that this bias is significant for the quality of regional climate projections over Europe.

.

Figure V K 1: Top: variance of north-south velocity at 300hPa as simulated by the GFDL CM2.1 model in years 1981-2000 of one realization of the 20C3M simulation, as contributed to the CMIP3 database. Units are m^2/s^2. Middle: The same quantity as obtained from the NCEP-NCAR reanalysis (ref). Bottom: model minus observations.

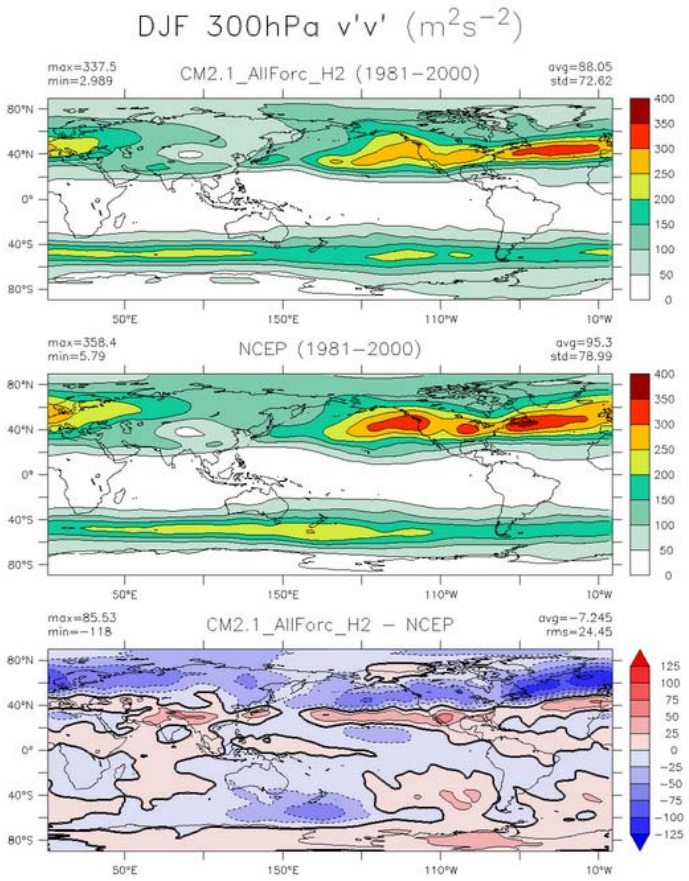
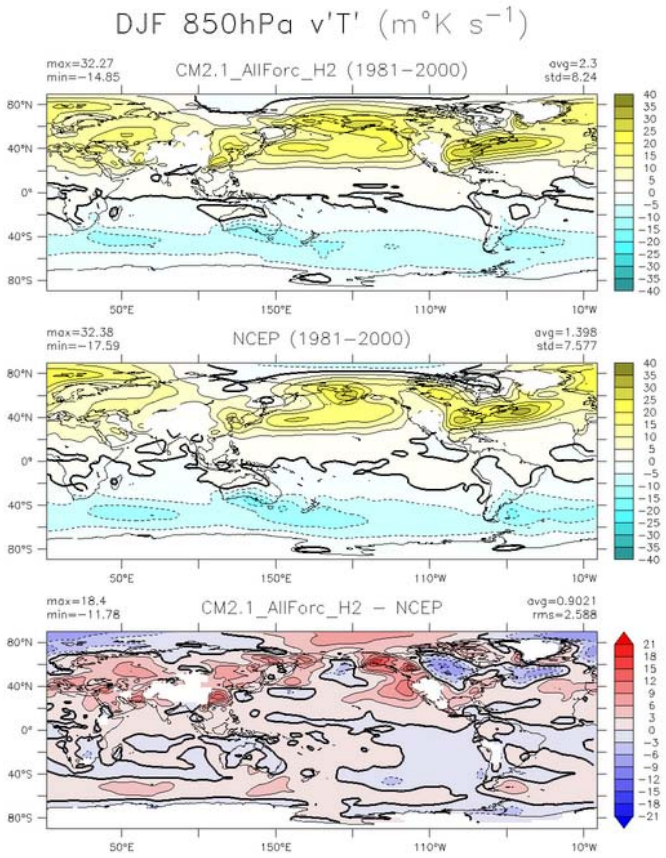


Figure V K 2: Top: covariance of north-south velocity and temperature at 850hPa as simulated by the GFDL CM2.1 model in years 1981-2000 of one realization of the 20C3M simulation, as contributed to the CMIP3 database. Units are Km/s. Middle: The same quantity as obtained from the NCEP-NCAR reanalysis (ref). Bottom: model minus observations.



Monsoons

The word `monsoon' derives from the Arabic word for season, and a monsoonal circulation is distinguished by its seasonal reversal after the sun crosses the equator into the new summer hemisphere. Rain is largest, if not entirely restricted, to the summer within monsoonal climates, when continental rainfall is supplied mainly by evaporation from the nearby ocean. This limits the reach of monsoon rains to the distance over which moisture can be transported onshore (Prive and Plumb 2007). Variations in the spatial extent of the monsoon from year to year determine which inland regions experience a drought.

Historical theories for the monsoon emphasize the influence of the contrast between land and ocean (Webster et al. 1998). Land responds more quickly to solar heating than the ocean, where heating is mixed over a deeper layer. Air is driven by this temperature contrast toward the warm land, where it ascends and precipitates moisture before returning offshore. Conversely, land cools more rapidly during winter when the sun is in the opposite hemisphere, and this drives air offshore toward the warmer ocean where it rises. While a coastal sea breeze is also driven by the temperature contrast between land and ocean, the monsoon is distinguished by its continental scale. The onshore flow is so extensive that it is deflected by the earth's rotation. Over the Arabian Sea, for example, surface air flows toward the east and northeast during the Northern Hemisphere summer, rather than traveling directly north toward the Asian continent.

While the monsoon takes its name from a language spoken by traders

around the Arabian Sea, this circulation reaches far beyond the periphery of the Indian Ocean, and local cultures have their own words for the monsoon: for example, Mei-yu in China, Chang-ma in Korea, and Bai-yu in Japan. Over a billion people are dependent upon the arrival of the monsoon rains for water and irrigation for agriculture. The Asian monsoon during NH summer is the most prominent example of the monsoon circulation, dominating global rainfall during this season. However, the seasonal reversal of winds and summer rainfall maximum also indicate monsoon circulations in West Africa and the Amazon basin. In addition, during NH summer, air flows off the eastern Pacific Ocean toward Mexico and the American southwest, while over the Great Plains of the United States, moisture from the Gulf of Mexico brings an annual peak in rainfall. Thus, the climate in these regions is also described as monsoonal.

Because of the geographic extent of the Asian monsoon, the fidelity of climate model simulations is weighed according to metrics from a variety of regions. Kripalani et al. (2007) judged that three-quarters of the eighteen analyzed coupled models (including the GFDL CM2.0 and 2.1 models, along with the NCAR PCM and GISS modelE-R) match the timing and magnitude of the summertime peak in precipitation over East Asia between 100 and 145E and 20 to 40N that is evident in the NOAA NCEP Climate Prediction Center Merged Analysis of Precipitation (CMAP, Xie and Arkin 1997). However, only half of these models (including both GFDL CGCMs) were able to reproduce the observed spatial distribution of monsoon rainfall, and its extension along the coast of China toward the Korean peninsula and Japan. Considering a broader range of longitude (40-180E) that includes the Indian subcontinent, Annamalai et al. (2007) found that only six of eighteen CGCMs (including both GFDL models) were significantly correlated with the observed spatial pattern of CMAP precipitation during June through

September. These six models also included a realistic simulation of ENSO variability, which is known to influence interannual variations in the Asian summer monsoon. Kitoh and Uchiyama (2007) computed the spatial correlation and root-mean-square error of simulated precipitation over a similar region and found the GFDL models in the top tercile with a spatial correlation exceeding 0.8, while the GISS modelE-R correlation was just under 0.5.

During NH winter, the Asian surface winds are directed offshore: from the northeast over India, and the northwest over East Asia. The two American models included in the comparison of the simulated East Asian winter monsoon by Hori and Ueda (2006), GFDL CM2.0 and GISS modelE-R, generally reproduce the observed spatial distribution of sea level pressure and 850 mb zonal wind.

In response to increasing greenhouse gases, models project increasing summer precipitation during the 21st century (Kripalani et al. 2007 Kimoto 2005). However, the circulation strength in both winter and summer is expected to weaken (Kimoto 2005, Ueda et al. 2006), consistent with simple physical arguments by Held and Soden (2006). The latter is also consistent with a study of previous generation models where interannual fluctuations in low-latitude rainfall increased, indicating increasingly severe seasonal departures from the mean (R\"ais\"anen 2002).

Observed variability of the West African monsoon is related to variations of ocean temperature in the Gulf of Guinea. The drying of the Sahel during the late 20th century, and the attendant societal impacts, is related to the inland extent of the monsoonal circulation. Cook and Vizy (2006) found that slightly over half of the 18 analyzed coupled models reproduced the observed maximum in

precipitation over land during June through August. Of these models, only six reproduced the anti-correlation between Gulf of Guinea ocean temperature and Sahel rainfall. The GISS modelE-H and both GFDL models were among the most realistic.

It is unresolved whether the late-20th century Sahel drought is due to natural or human influences. Hoerling et al. (2006) surveyed the average response of eighteen coupled model to conclude that anthropogenic forcings during this period account for only a small fraction of rainfall variations observed in the Sahel. In contrast, Biasutti and Giannini (2006), contrast Sahel rainfall between simulations with observed 20th century forcings (such as greenhouse gas and aerosol concentrations), nineteenth century (pre-industrial) conditions, and increasing greenhouse gases. They suggest that the observed late 20th century trend was externally forced, predominately by anthropogenic aerosols. This conclusion is based upon the average behavior of the models considered. It is supported in particular by the GFDL and GISS models. It is currently unclear how to resolve these contrasting conclusions, because they are based upon different methods and comparisons of models. Both studies agree that the Sahel drought is the result of ocean warming in the Gulf of Guinea, compared to the NH subtropical Atlantic. What remains unresolved is whether forcing by greenhouse gases and aerosols has changed the contrast in ocean temperature between these two regions.

Rainfall over the Sahel and Amazon are anti-correlated: when the Gulf of Guinea warms, rainfall is generally reduced over the Sahel but increases over South America. Amazon rainfall also depends upon the eastern equatorial Pacific, and during an El Nino, rainfall is reduced in the Nordeste region of the Amazon. Li et al. (2006) compare the hydrological cycle of eleven CGCMs over the Amazon during the

late 20th and twenty-first centuries. Based upon a comparison to CMAP rainfall, the GISS modelE-R is among the best, with the GFDL CM2.1 and NCAR CCSM3 models similarly ranked. Despite this fidelity, the models make disparate predictions for the 21st century. In the GISS modelE-R, the equatorial Pacific warms more in the west, resembling a La Nina event. This, together with warming in the Gulf of Guinea, is associated with an increase in Amazon rainfall. While the NCAR CCSM3 predicts a comparable increase, the GFDL CM2.1 exhibits a decrease and lengthening of the Amazon dry season.

The studies of Li et al. (2006) along with Ammamalai et al. (2007) note that future changes in the South American and Asian monsoons are intimately tied to the response of El Nino in the 21st century. Expected temperature changes in the eastern equatorial Pacific are discussed in ENSO section. Here, we note that a consensus is yet to emerge, adding to uncertainty in monsoon projections.

The ability of climate models to simulate NH summer rainfall over the US Great Plains and Mexico was summarized by Ruiz-Barradas and Nigam (2006). Among the American models, the GISS modelE-H matches the annual cycle of precipitation over the Great Plains and Mexico most closely. It is also one of two models to simulate interannual variations in precipitation that are significantly correlated with observed variability during the second half of the 20th century. The observed predominance of moisture import from the Gulf of Mexico compared to local evaporation is most closely reproduced by the NCAR PCM. Moisture import is excessive in the GISS modelE-H, whereas as evaporation contributes too large a fraction in the GFDL CM2.1.

Initial evaluations of the monsoon simulated by the most recent generation of climate models have emphasized the seasonal time scale.

However, subseasonal variations, such as break periods when the monsoon rains are temporarily interrupted, are crucial to forecasts and the impact of the monsoon upon water supply. Simulation of the diurnal cycle, and the local hour of rainfall, is also important to the partitioning of rainfall between runoff and transpiration, and these are important topics for future model evaluation. Transports of moisture by regional circulations beneath the resolution of the model (such as low-level jets along the Rockies and Andes and tropical cyclones) contribute to the onshore transport of moisture. In general, the models show success at simulating the gross seasonal features of the various monsoon circulations, but variations on smaller spatial and time scales that are important to specific watersheds and hydrological projections need to be evaluated.

Tropical storms

Tropical storms (hurricanes in the Atlantic and typhoons in the Pacific and Indian Oceans) are of too small a scale to be reliably simulated in the class of global climate models currently used for climate projections. There is hope for qualitatively useful simulations of the climatology of incipient tropical depressions, however. The work of Vitart and Anderson (2001) is an example of evidence for signficant information content concerning tropical storm-like vortices in simulations with models of this type, using the model's ability to simulate the effects of El Nino on Atlantic storm frequency as a guide.

The recent 20km resolution simulation with an atmospheric model over prescribed ocean temperatures by Oouchi et al (2006) is indicative of the kinds of modeling that will be brought to bear on this problem in the next few years. Experience with tropical storm forecasting suggests that this resolution should be adequate for describing many aspects of the evolution of nature tropical storms, and possibly the generation of storms from incipient disturbances, but probably not tropical storm intensity.

An alternative very promising approach is described by Knutson et al (2007), in which a regional model of comparable resolution (18 km) is used in a downscaling framework to simulate the Atlantic hurricane season. Given the observed year-to year variations in the large-scale structure of the atmosphere over the Atlantic ocean, the model is capable of simulating the year-to-year variations in hurricane frequency over a 30-year period with a correlation of 0.7-0.8 and also captures the observed trend towards greater hurricane frequency over this period in the Atlantic. These results suggest that models of this resolution may be able to provide a convincing downscaling capability for tropical storm frequency projections into the future, although these projections will still rely on the quality of the global model projections for changes in sea surface temperature, atmospheric stability, and vertical shear. The behavior of the El Nino Southern Oscillation into the future will be a key element affecting changes in those aspects of the large-scale structure of the atmosphere over the Atlantic that control tropical storm formation and tracks.

Polar climates

Changes in polar snow and ice cover affect the Earth's albedo and thus the amount of insolation heating the planet (e.g., Holland and Bitz 2003, Hall 2004, Dethloff *et al*. 2006). Concern has also emerged about potential melting of glaciers and ice sheets in Greenland and Antarctica that could produce substantial sea-level rise (Arendt *et al*. 2002, Braithwaite and Raper 2002, Alley *et al*. 2005). Polar regions thus require accurate simulation for projecting future climate change and its impacts.

Polar regions present unique environments and, consequently, challenges for climate modeling. The obvious are processes involving frozen water. While not unique to polar regions, they are more pervasive there. These processes include seasonally frozen ground and permafrost (Lawrence and Slater 2005, Yamaguchi *et* al. 2005) and seasonal snow cover (Slater *et al*. 2001), which can have significant sub-grid heterogeneity (Liston 2004), and clear-sky precipitation, especially in the Antarctic (King and Turner 1997, Guo *et al*. 2003). Polar radiation also has important characteristics that test the ability of models to handle extreme geophysical behavior,

such as longwave radiation in clear, cold environments (Hines *et al*. 1999, Chiacchio *et al*. 2002, Pavolonis *et al*. 2004) and cloud microphysics in the relatively clean polar atmosphere (Curry *et al*. 1996, Pinto *et al*. 2001, Morrison and Pinto 2005). In addition, polar atmospheric boundary layers can be very stable (Duynkerke and de Roode 2001, Tjernström *et al*. 2004, Mirocha *et al*. 2005), and stable boundary layers remain an important area for model improvement.

Confidence in climate model projections of future climate is greatly increased if it can be shown that climate models can accurately simulate the current climate state, and much effort has gone into this type of analysis (e.g. Collins et al. 2006, Delworth et al. 2006). In particular climate models should be able to reproduce both long-term and short-term variations in climate including daily, seasonal, interannual, and decadal variability. For polar regions, much of the assessment of simulated interannual variability has focused on the primary modes of polar interannual variability, the Northern and Southern Annular Modes. Assessment of simulated annular modes appears in Section B. of this chapter.

Less attention has been given to the ability of global climate system models to simulate shorter-duration climate and weather variability in polar regions. Uotila *et al*. (2007) and Cassano *et al*. (2007) evaluated the ability of an ensemble of 15 global climate-system models to simulate the daily variability in sea level pressure in the Antarctic and Arctic. In both polar regions, it was found that the 15-model ensemble was not able to reproduce the daily synoptic climatology, with only a small subset of the models accurately simulating the frequency of the primary synoptic weather patterns identified in global reanalysis data sets. The U.S. models discussed in detail in Chapter 2 of this report spanned the same range of accuracy as non-U.S. models, with GFDL and NCAR GCM versions part of the small, accurate subset. Vavrus *et al*. (2006) assessed the ability of seven global climate models to simulate extreme cold-air outbreaks in the Northern Hemisphere, and found that the spatial pattern of the outbreaks was accurately reproduced in the models, although some details differed.

Attention has also been given to the ability of regional climate models to simulate polar climate. In particular, the Arctic Regional Climate Model Intercomparison Project (ARCMIP) (Curry and Lynch 2002) engaged a suite of Arctic regional atmospheric models to simulate a common domain and period over the western Arctic. Rinke *et al*. (2006) evaluated the spatial and temporal patterns simulated by 8 ARCMIP models, and found that the model ensemble agreed well with global reanalyses, despite some large errors for individual models. Tjernstrom *et al*. (2005)

evaluated near-surface properties simulated by 6 ARCMP models. In general surface pressure, air temperature, humidity, and wind speed were all well simulated, as were radiative fluxes and turbulent momentum flux. Tjernstrom *et al*. (2005) found that turbulent heat flux was poorly simulated, and that over an entire annual cycle the accumulated turbulent heat flux simulated by the models was an order of magnitude larger than the observed turbulent heat flux (Fig. PA-1). In both Tjernstrom *et al.* (2005) and Rinke *et al.* (2006), the U.S. models performed about the same as their European counterparts.

Although simulations of polar climate display agreement with observed behavior, as indicated above, there remains room for improvement. In global models, polar simulation may be affected by errors in simulating other regions of the planet, but much of the difference from observations and uncertainty about projected climate change stems from current limitations in polar simulation. These limitations include missing or incompletely represented processes and poor resolution of spatial distributions.

As with other regions, model resolution affects simulation of important processes. In the polar regions, surface distributions of snow depth vary markedly, especially when snow drifting occurs. Improved snow models are needed to represent such spatial heterogeneity (e.g., Liston 2004), which will continue to involve scales smaller than resolved for the foreseeable future. Frozen ground, whether seasonally frozen or occurring as permafrost, presents additional challenges. Models for permafrost and seasonal freezing and thawing of soil are being implemented in land surface models (see Chapter 2, Land Surface Models). Modeling soil freeze and thaw continues to be a challenging problem as characteristics of energy and water flow through the soil affect temperature changes, and such fluxes are poorly understood (Yamaguchi *et al*. 2005**)**.

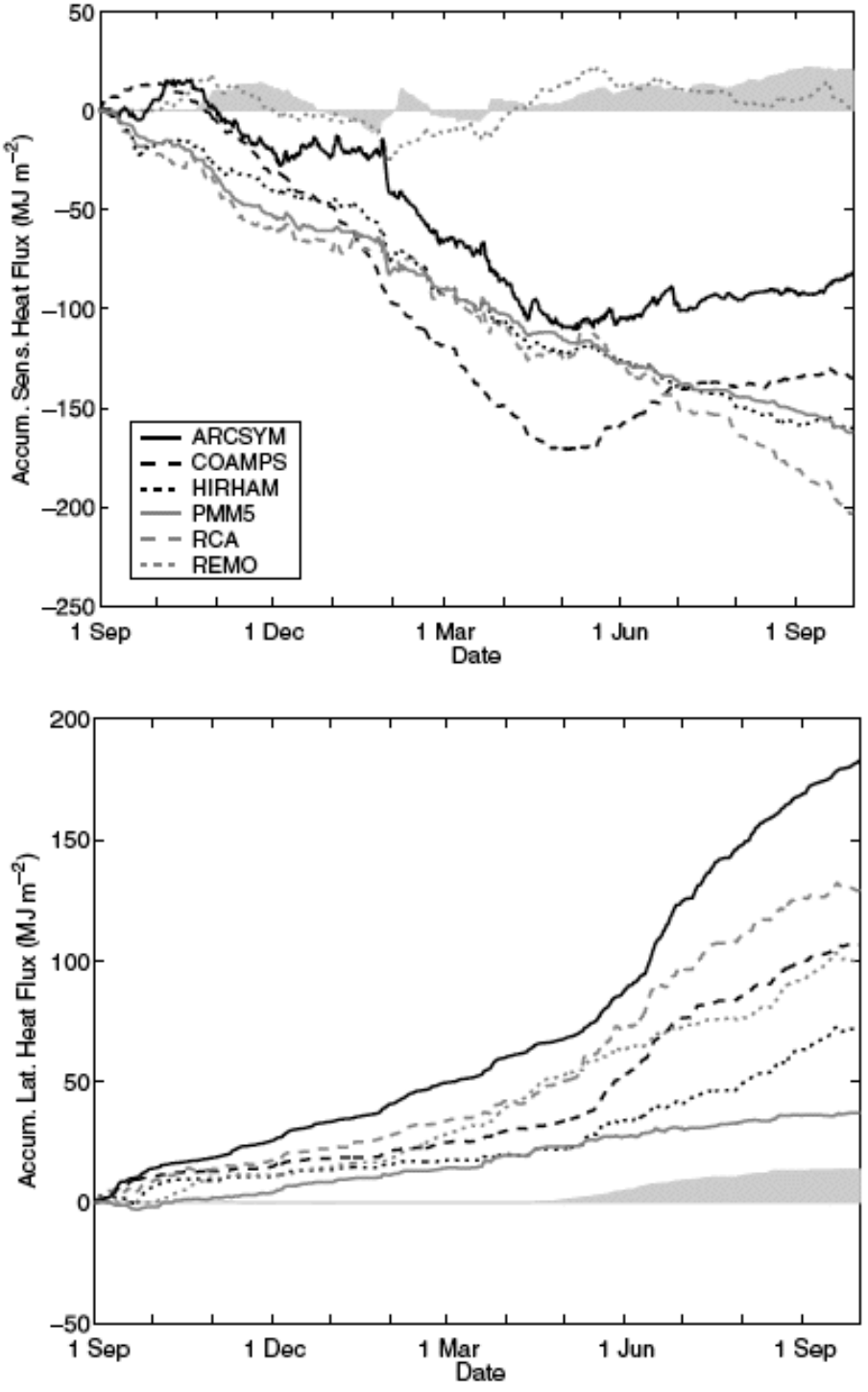
Frozen soil affects surface and subsurface hydrology, which influences the spatial distribution of surface water with attendant effects on other parts of the polar climate system such as carbon cycling (e.g., Gorham 1991, Aurela *et al*. 2004), surface temperature (Krinner 2003), and atmospheric circulation (Gutowski *et al*. 2007). The flow of fresh water into polar oceans potentially alters their circulation, too. Surface hydrology modeling typically includes limited, at best, representation of subsurface water reservoirs (aquifers) and horizontal flow of water at both the surface and below surface. These features limit the ability of climate models to represent changes in polar hydrology, especially in the Arctic.

Vegetation has been changing in the Arctic (Callaghan *et al*. 2004) and projected warming, which may be largest in regions where snow and ice cover retreat, may produce further changes in vegetation (e.g., Lawrence and Slater 2005). Current models use static distributions of vegetation, but dynamic vegetation models will be needed to account for changes in land-atmosphere interactions influenced by vegetation.

A key concern in climate simulations is how projected anthropogenic warming may alter ice sheets on land, whose melting could raise sea levels substantially. At present, climate models do not include ice-sheet dynamics and thus cannot account directly for how ice sheets might change, possibly changing heat absorption from the sun and atmospheric circulation in the vicinity of the ice sheets.

How well each of the processes above is represented in climate simulation depends in part on model resolution. Distributions of snow, ice sheets, surface water, frozen ground and vegetation have important spatial variation on scales smaller than the resolutions of typical contemporary climate models. Finer resolution is thus needed. Part of this need may be satisfied by regional models simulating just a polar region. Because both the northern and southern polar regions are within circumpolar atmospheric circulations, their synoptic coupling with other regions is more limited than is the case with midlatitude regions, where the westerlies rapidly move synoptic systems in and out of a region (e.g., Wei *et al*. 2002), which could allow polar-specific models that focus on ant/arctic processes, in part to improve modeling of surface-atmosphere exchange processes (Fig. V L). While each of the above processes have been simulated in finer scale, stand-alone models, their interactions as part of a climate system also need to be simulated and understood.

Fig. V L. Cumulative fluxes of surface sensible heat (top panel) and latent heat (bottom) at the SHEBA site from six models simulating a western Arctic domain for Sept. 1997 – Sept. 1998 for ARCMIP. SHEBA observations are the gray vertical bars; model identifications are given by the key in the upper panel. Adapted from Tjernstrom *et al*. (2005).



Sea ice plays a critical role in the exchange of heat, mass, and momentum between the ocean and atmosphere and any errors in the sea-ice system will contribute to errors in the other components. Two recent papers [*Holland and Raphael* 2006 and *Parkinson et al.* 2006] quantify how the current models simulate the sea-ice process of the climate system. Very limited observations make any evaluation of sea ice difficult. The primary observation available is sea ice concentration. In some comparisons, sea ice extent (ice concentration greater than 15%), is used. Satellites have made it possible for a more complete data set of observations for the past few decades. Prior to satellite measurements becoming available, observations of ice extent were fewer. Other quantities that might be evaluated include ice thickness. Such comparisons are difficult because of the limited number of observations and will not be discussed.

Ice Concentration and extent: Both of these studies indicate that the seasonal pattern in ice growth and decay in the polar regions for all the models is reasonable [*Holland and Raphael* 2006] (**Figure V M**). However, there is a large amount of variability between the models in their representation of the sea ice extent in both the northern and southern hemispheres. Generally, the models do better in simulating the Arctic region than in their simulation of the Antarctic region as shown with **Figure V N].** An example of the complex nature of reproducing the ice field is given in *Parkinson et al.* [2006]. They found that all the models showed an ice-free region in winter to the west of Norway, as seen in observational data, but all the models also produced too much ice north of Norway. They suggest that this is because the North Atlantic Current is not being simulated correctly. In a qualitative comparison, Hudson Bay is ice covered in winter in all the models correctly reproducing the observations. The set of models are not consistent in their "fidelity" between the Northern and Southern regions and maybe due, partly, to how the parameters are defined in the sea ice models.

Holland and Raphael [2006] examined the variability in the Southern Ocean sea ice extent extensively. As an indicator of the ice response to large scale atmospheric events, they compared a set of IPCC AR4 climate models sea ice response to the atmospheric index, the Southern Annular Mode (SAM) for the April-June (AMJ) period (Table V 4). The models show that the ice variability does respond modestly to the large scale atmosphere forcing but less than limited observations show. Two of the models also exhibit the out-of-phase buildup of ice between the Atlantic and Pacific sectors (the Antarctic Dipole) to some degree.

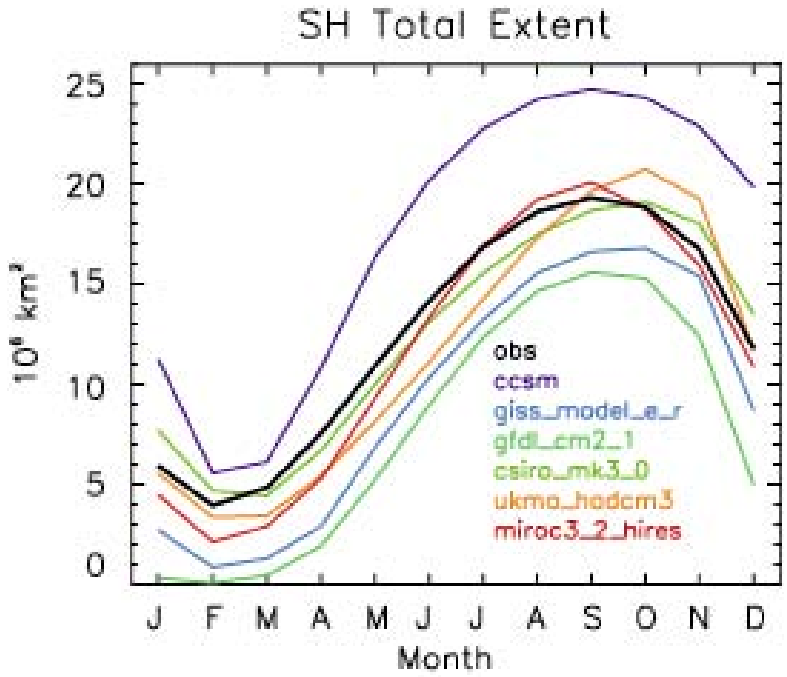
Table V 4

MODIFIED FROM Table 1 *Holland and Raphael* [2006] Correlations of the leading mode of sea ice variability and the southern annular mode (SAM) for the observations and model simulations

	AMJ SAM and high-pass filtered fields	AMJ SAM and detrended fields
Observations	**0.47**	**0.47**
CCSM3	**0.40**	**0.44**
GFDL-CM2.1	**0.39**	0.19
GISS-ER	0.30	0.20

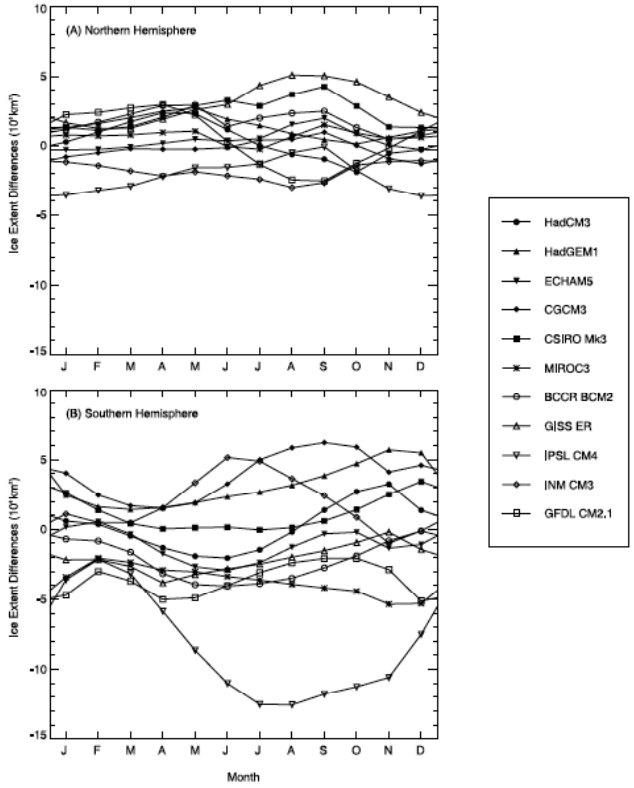
Bold values are significant at the 95% level accounting for the autocorrelation of the timeseries

Figure V M



From *Holland and Raphael* 2006. Fig. 1 The annual cycle of southern hemisphere ice extent defined to be the area of ice with concentrations greater than 15%

Figure V N



From Parkinson *et al*. 2006: Figure 4. Difference between the modeled 1979–2004 monthly average sea ice extents and the satellite-based observations (modeled minus observed), for each of 11 major GCMs, for both the (a) Northern Hemisphere and (b) Southern Hemisphere.

Modes of variability

The Madden-Julian Oscillation: (MJO) is a characteristic pattern in the tropical atmosphere. It has taken on special prominence in research on simulating the tropical atmosphere. This phenomenon consists of large-scale eastward propagating patterns in humidity, temperature, and atmospheric circulation which strengthen and weaken tropical rainfall as they propagate around the Earth in roughly 30-60 days. This pattern often dominates intraseasonal (within season) variability of tropical precipitation on time scales longer than a few days, creating such phenomena as 1-2 week breaks in Asian monsoonal rainfall and weeks with enhanced hurricane activity in the Eastern North Pacific and the Gulf of Mexico. Inadequate prediction of the evolution of these propagating structures is considered one the main impediments to more useful extended-range weather forecasts in the tropics, and improved simulation of this phenomenon is considered by some a litmus test for the credibility of climate models in the tropics

Recent surveys of model performance indicate that simulations of the MJO remain inadequate. For example, Lin *et al* (2006), in a study of many of the models in the CMIP-3 models, conclude that "… current GCMs still have significant problems and display a wide range of skill in simulating the tropical intraseasonal variability", while Zhang *et al*. (2005) in another multi-model comparison study, state that "… commendable progress has been made in MJO simulations in the past decade, but the models still suffer from severe deficiencies …" Nearly all models do capture the essential feature of the pattern, with large-scale eastward propagation and with roughly the correct vertical structure. But the propagation is often too rapid and the amplitudes too weak. As an example of recent work, Klein (2007?) studies whether two of the US IPCC models can maintain a pre-existing strong MJO pattern when initialized with observations (from the TOGA-COARE field experiment), with limited success. Controlled experiments have suggested that for models to simulate MJO, the instability of the atmospehre must be allowed to accumulate to a certain amount before convective storms are triggered, and sufficient mesoscale statiform heating from convective systems should exist in the upper troposphere (Wang and Schlesinger 1999). These processes are however poorly understood in current climate models.

The difficulty in simulation of the MJO is related to the multi-scale nature of the phenomenon: the propagating pattern is itself of large enough scale that it should be resolvable by climate models, but the convection and rainfall modulated by this pattern, and feeding back and energizing it, occur on much smaller, unresolved, scales. In addition to this dependence on the parameterization of tropical convection, a long list of other effects has been shown by models and/or observational studies to be important for the MJO. These include the pattern of evaporation generated as the MJO propagates through convecting regions, feedback from cloud-radiative interactions, intraseasonal ocean temperature changes, the diurnal cycle of convection over the ocean, as well as the vertical structure of the latent heating, including especially the proportion of shallow cumulus congestus clouds and deep convective cores in the different phases of the oscillation (Lin *et al*. 2004).

A picture seems to be emerging that the difficulty in simulation may not be due to a single model deficiency but a result of the complexity of the phenomenon, given this long list of factors thought to be significant. In several of the multi-model studies, such as Lin *et al* (2006) a few of the models do perform well, but without a clearer understanding of how these factors combine to generate the observed characteristics of the MJO, it is difficult to maintain a good simulation as the model is modified for other reasons, and it is difficult to transfer one model's successful simulation to other models. It also remains unclear whether the models with superior MJO simulations should be given extra weight in multi-model studies of climate change in the tropics.

The El Nino – Southern Oscillation (ENSO) El Nino was named originally in the 19th century by Peruvian sailors to note the early arrival of a warm current from equatorial latitudes (Philander 1990). Every few years, a springtime northerly current arrives prematurely around Christmas (Yu and McPhaden 1999), bringing heavy rains to coastal Peru and a temporary decline in the anchovy harvest. By the mid 20th century, scientists recognized that this local anomaly was in fact part of a disruption to the atmospheric circulation across the entire Pacific basin. During El Niño, atmospheric mass migrates west of the dateline as part of the Southern Oscillation, reducing surface pressure and drawing rainfall into the central and eastern Pacific (Rasmussen and Wallace 1983).

Together, El Niño and the Southern Oscillation, often abbreviated in combination as ENSO, are the largest source of tropical variability observed during recent decades.

Changes along the equatorial Pacific have been linked to global disruptions of climate (Ropelewski and Halpert 1987). During an El Niño event, the Asian monsoon is typically weakened, along with rainfall over eastern Africa, while precipitation increases over the American southwest. El Niño raises the surface temperature as far poleward as Canada, while changes in the North Pacific Ocean are linked to decadal variations in ENSO (Trenberth and Hurrell 1994). In many regions far from the eastern equatorial Pacific, accurate projections of climate change in the twenty-first century depend upon the accurate projection of changes to El Niño. Moreover, the demonstration that ENSO alters climate across the globe indicates that even changes to the *time-averaged* equatorial Pacific during the 21st century will influence climate far beyond the tropical ocean. For example, a long-term warming of the eastern equatorial Pacific relative to the surrounding ocean will favor a weaker Asian monsoon, even in the absence of changes to the size and frequency of El Niño events.

Incident sunlight is largest on the equator, but in the eastern Pacific, the ocean is colder than at neighboring latitudes. Because of the Earth's rotation, easterly winds along the equator cool the surface by raising cold water from below, which offsets heating by the absorption of sunlight (e.g. Clement *et al* 1996). In contrast, warm water extends deeper to the west so upwelling has little effect upon the surface temperature of the West Pacific, where the warmer ocean is consistent with the strong, equatorial solar heating. The westward increase of temperature along the equator is associated with a decrease in atmospheric pressure, reinforcing the easterly Trade winds.

Theoretical arguments offer conflicting projections of tropical Pacific climate during the twenty-first century. One projection is for the equatorial temperature contrast to increase, so that the average state more closely resembles La Niña, marked by unusually cold ocean temperatures and enhanced upwelling in the East Pacific, the opposite to El Niño (Clement *et al* 1996; Cane *et al* 1997). According to this argument, an increase in net radiation into the ocean resulting from an increase in greenhouse gas concentration is partially offset by the upwelling of cold water. This compensation is stronger in the east than in the west, where the surface layer of warm water extends to greater depth. There is evidence for an observed trend toward a La Niña state (Cane *et al* 1997),

but the trend remains ambiguous because of the large decadal variations in the ENSO cycle. Another theory is based upon the origin of upwelling water along the equator within descending surface water in higher latitudes. Liu *et al* (1998) suggest that as these higher latitudes warm, the temperature of the upwelling water will increase, reducing its ability to offset radiative warming at the surface. A third theory suggests that as the tropical atmosphere becomes more stable in response to surface warming, the tropical circulation will weaken (Knutsen and Manabe 1998; see also Meehl and Washington 1996). This will draw less cold water to the surface, preferentially warming the East Pacific. Until recently, many coupled ocean-atmosphere models projected larger warming of the East Pacific and a drift of mean conditions toward an ENSO state.

Below, we summarize the most recent model comparisons, emphasizing those studies carried out as part of the IPCC AR4. Our conclusions are based upon model behavior from a worldwide collection of coupled ocean-atmosphere models, although we illustrate many of the scientific issues using models from American laboratories. The coupled models are designed for prediction of global climate over decades and centuries and are not tuned to optimize their simulation of ENSO *per se,* unlike many of the more simple dynamical and statistical models currently used for operational forecasts of ENSO over a period of several months. Nonetheless, we find that the global models as a group exhibit realistic simulations of present-day seasonal variations and ENSO variability, and represent a marked improvement compared to previous generations of coupled models. However, among the most realistic models, there is little consensus on the anticipated change to either the mean state of the tropical Pacific (particularly the east-west difference in ocean temperature along the equator) or the amplitude and frequency of ENSO variability. This introduces uncertainty in the projected climate response within regions throughout the globe influenced by El Niño.

In general, coupled models developed for the CMIP3 are far more realistic than those of a decade ago, when ENSO variability was comparatively weak, and some models lapsed into permanent El Niño states (Neelin *et al.,* 1992). Even compared to the models assessed more recently by ENSIP and CMIP2 (Latif *et al.,* 2001; AchutaRao and Sperber 2002), ENSO variability of ocean surface temperature is more realistic, although sea level pressure and precipitation anomalies show little recent improvement (AchutaRao and Sperber 2006). Part of this progress is the result of increased resolution of the equatorial ocean circulation that has accompanied inevitable increases in

computing speed. Table V 5 shows the horizontal and vertical resolution of the seven American coupled models whose output was submitted to AR4.

Table V 5

Spacing of grid points at the equator in the American coupled models developed for AR4. Except for the GISS models, spacing of grid points generally increases away from the equator outside of the domain of ENSO, so that resolution is highest on the equator.

Model:	Longitude	Latitude	Vertical Levels
GFDL CM2.0	1	1/3	50
GFDL CM2.1	1	1/3	50
GISS AOM	5	4	13
GISS modelE-H	2	2	16
GISS modelE-R	5	4	13
NCAR CCSM3	1.125	0.27	27
NCAR PCM	0.94	0.5	32

.....

Along the equator, oceanic waves that adjust the equatorial temperature and currents to changes in the wind are tightly confined to within a few degrees of latitude. To simulate this adjustment, the ocean state is calculated at points as closely spaced as 0.27 degrees of latitude in the NCAR CCSM3. NCAR PCM has half degree resolution, while both GFDL models have equatorial resolution of one-third of a degree. This degree of detail is a substantial improvement compared to previous generations of models. In contrast, the GISS AOM and modelE-R calculate equatorial temperatures at grid points separated by four degrees of latitude. This is broad compared to the latitudinal extent of cold temperatures observed within the eastern Pacific (the `cold tongue' region), which are the result of a narrow band of cold water rising to the surface along the equator. In the coarse resolution models, changes to the upward flow are spread over the dimensions of the grid

box, which is broader than the observed upwelling. The cooling effect of this rising water is spread over a larger area, so that the amplitude of the resulting temperature fluctuation at the surface is weakened. In fact, both the GISS AOM and modelE-R models have unrealistic ENSO variations that are much smaller than observed (Hansen *et al* 2007). This minimizes the influence of their simulated El Nino and La Nina events on climate outside the equatorial Pacific, and we will not discuss these two models further in this section.

In comparison to previous generations of global models, where ENSO variability was typically weak, the AR4 coupled models generally simulate El Nino near the observed amplitude, or even above (Neelin *et al* 1992; AchutaRao and Sperber 2006). The latter study compared sea surface temperature (SST) variability within the tropical Pacific calculated under pre-industrial conditions. Despite its comparatively low two-degree latitudinal grid spacing, the GISS modelE-H among the American models most closely matches observed SST variability since the mid-19th century, according to the HadISST v1.1 data set (Rayner *et al* 2003). The NCAR PCM also exhibits El Niño warming close to the observed magnitude. This comparison is based upon spatial averages within three longitudinal bands, and GISS modelE-H along with the NCAR models exhibit their largest variability in the eastern band as observed. However, GISS modelE-H underestimates variability since 1950, when the NCAR CCSM3 is closest to observations (Joseph and Nigam 2006). While the fidelity of each model's ENSO variability depends upon the specific data set and period of comparison (c.f. Capotondi *et al.*, 2006; Merryfield 2006, van Oldenborgh *et al.*, 2005), the general consensus is that the GISS modelE-H, both NCAR models, and GFDL CM2.0 have roughly the correct amplitude, while variability is too large by roughly one-third in the GFDL CM2.1. While most models (including GISS modelE-H and both NCAR models, but excluding the GFDL models) exhibit the largest variability in the eastern band of longitude, none of the AR4 models match the observed variability at the South American coast, where El Nino was originally identified (AchutaRao and Sperber 2006; Capotondi *et al.*, 2006). This is possibly because the longitudinal spacing of the model grids is too large to resolve coastal upwelling, and its interruption during El Niño (Philander and Pacanowski 1981). Biases in the atmospheric model, including underestimate of the persistent stratus cloud decks along the coast, may also contribute (Mechoso *et al.*, 1995).

El Niño occurs every few years, albeit irregularly. The spectrum of anomalous ocean temperature

shows a broad peak between two and seven years, and there are multi-decadal variations in event frequency and amplitude. Almost all of the AR4 models have spectral peaks within this range of time scales. Interannual power is broadly distributed within the American models, as observed, with the exception of the NCAR CCSM3 which exhibits strong biennial oscillations.

While the models generally simulate the observed magnitude and frequency of events, reproduction of their seasonality is more elusive. Anomalous warming typically peaks late in the calendar year, as originally noted by South American fisherman. Among American models, this seasonal dependence is simulated only by the NCAR CCSM3 (Joseph and Nigam 2006). Warming in the GFDL CM2.1 and GISS modelE-H are nearly uniform throughout the year, while warming in the NCAR PCM is largest in December but exhibits a secondary peak in early summer. The mean seasonal cycle along the equatorial Pacific also remains a challenge for the models. Each year, the east Pacific cold tongue is observed to warm during NH spring and cool again late in the calendar year. The GFDL CM2.1 and NCAR PCM1 have the weakest seasonal cycle among the American models, while GISS modelE-H, GFDL 2.0 and NCAR CCSM3 are closest to the observed amplitude (Guilyardi 2006). Among the worldwide suite of AR4 models, the amplitude of the seasonal cycle of equatorial ocean temperature generally varies inversely with the strength of the ENSO cycle.

Anticipation of twenty-first changes to El Nino remains uncertain, because of a lack of consensus among the models. Among fifteen models forcecd by increasing carbon dioxide, three exhibit statistically significant increases in amplitude, while five exhibit a decrease, compared to their variability under pre-industrial conditions (Merryfield, 2006). Even when only the most realistic models are surveyed (including the GFDL CM2.1), identified according to a detailed examination of their mechanisms of variability (described below), no consensus emerges. No significant change in event period is found either (Guilyardi 2006). These trends are inferred based upon the response to a doubling or quadrupling of carbon dioxide, compared to a pre-industrial climate. This forcing is strong compared to forcing over the 20th century in which one might hope to infer trends in El Nino from the observational record. The occurrence of the two largest El Nino events late in the 20th century has been attributed to increasing greenhouse gas concentrations (Trenberth and Hoar 1997; Knutsen and Manabe 1998), although this remains unsettled because of large variations in the

tropical Pacific within the multi-decadal instrument record (Rajagopalan *et al.,* 1997).

Changes in the climate of the tropical Pacific (as opposed to trends in El Nino variability) are also inconsistent (van Oldenborgh *et al* 2006). Of particular interest is the relative warming along the equator, because this is related to the strength of the tropical circulation, which creates regional changes throughout the globe. The ostensible consensus among the most recent generation of models (including both American and international modeling centers) is that the eastern Pacific will warm by about a half degree Celsius more compared to the west (see Figure 10.16 of Meehl et al. 2007).
This is small compared to the currently observed difference of a few degress. When only the most realistic models are surveyed (including the two GFDL models), the warming is nearly uniform across the Pacific (van Oldenborgh et al. 2005). This behavior is consistent with a previous generation of global models, surveyed as part of CMIP2 (Collins et al. 2005). When the model predictions were weighted by the realism of each model, the multi-model average warming was nearly uniform, with only a small probability of greater warming in the east. In summary, warming along the equatorial Pacific is expected to be uniform or slightly larger to the east, but this contrast is on the order of differences among the models. This translates into an uncertainty in the climate in regions outside the tropical Pacific affected by ENSO.

The lack of consensus among model projections for the 21st century may result from the combination of physical mechanisms contributing to observed variability, and the difficulty of simulating them individually along with their relative importance. There is evidence that the importance of certain mechanisms changed in the middle of the 1970's (Wang 1995), so it is unclear what the correct emphasis should be. In addition, positive feedbacks, inferred from the observations, may exacerbate unrealistic features in the models, contributing further to model error.

Several studies have assessed the mechanisms contributing to variability among the AR4 models. Confidence in the models' projection of climate within the tropical Pacific during the twenty-first century depends upon accurate simulation of mechanisms of variability observed at present. El Niño occurs when the upwelling of cold water to the surface is interrupted within the equatorial eastern Pacific and South American coast. This can occur because the rate of upwelling decreases, or

alternatively because the temperature of the upwelling water increases. This subsurface temperature is related to the depth of the thermocline, within which the water temperature falls off sharply with depth. During El Niño, the thermocline deepens, so that upwelling water originating in the cold water below now begins its rise within the relatively warm layer above (Wyrtki 1975). In addition, the slowing of the easterly Trade winds reduces the rate of upwelling (Bjerknes 1969), which at the surface reduces the export of water from the cold tongue toward the West Pacific. Within the weaker surface current, water has more time to be warmed by the sun and overlying atmosphere. El Niño is a coupled phenomenon because the winds that change the upwelling of cold water to the surface depend upon the ocean temperature itself. Because the easterlies are driven partly by the temperature contrast between the cold east Pacific and the warmer ocean west of the dateline, warming in the east reinforces the slackening of the easterly Trade winds.

The coupling between ocean temperature and equatorial winds is typically inferred by regressing wind stress upon temperature averaged within the ENSO domain. The observed wind anomaly is westerly and strongest slightly to the west of a warm ocean anomaly, as expected based upon simple theoretical models (Gill 1980; Lindzen and Nigam 1987, Yu and Neelin 1997). The model wind anomalies are typically displaced farther west than observed, and are excessively confined to the equator (Capotondi *et al.,* 1987). The NCAR PCM regression is roughly half the observed strength, while among the American models, the NCAR CCSM3 and GFDL CM2.1 come closest to observations (Van Oldenborgh *et al.,* 2005). The GISS modelE-H exhibits reasonable coupling in the Central Pacific, but almost no coupling toward South America.

The converse response of SST to wind anomalies is diagnosed by evaluating various terms in the equation for the evolution of ocean temperature (van Oldenborgh 2005; Capotondi *et al.,* 2006). Changes in the temperature of upwelling water are observed to be important in the eastern Pacific (Capotondi *et al.,* 2006). This feedback is reproduced by the GFDL CM2.0 and NCAR CCSM3 models, although with somewhat low amplitude, possibly because the climatological upwelling is weak. (The model output necessary for this diagnosis was not available for the GFDL CM2.1, NCAR PCM, and GISS modelE-H.) While a decrease in the rate of upwelling is crucial to observed warming in the Central Pacific, this feedback is weak in the GFDL CM2.0, and absent in the NCAR CCSM3. The ocean feedback to wind anomalies is also diagnosed by regressing the evolution of

ocean temperature upon various mechanisms represented within the ocean heat budget (van Oldenborgh *et al.*,2005). The NCAR PCM has very strong feedbacks of upwelling rate and temperature in response to wind anomalies, which compensate for its weak wind response to anomalous SST. The GFDL CM2.1 generally reproduces the observed regression relations. In contrast, van Oldenborgh *et al.* (2005) note that regression analysis of GISS modelE-H is noisy and difficult to interpret. It is not clear at this point how GISS modelE-H compensates for its weak wind response to ocean temperature anomalies in order to create ENSO temperature variability near the observed magnitude and location. This lack of transparency calls its projection of future changes into question.

In general, GFDL2.1 is consistently ranked among American models as the most realistic simulation of El Nino (van Oldenborgh *et al.,* 2005; Guilyardi 2006; Merryfield 2006). This is based not only on its surface temperature variability (which in fact is slightly too high), but on its faithful simulation of the observed relationship between ocean temperature and surface wind, along with the wind-driven ocean response. While SST in many models is consistently dominated either by anomalies of upwelling strength or else temperature, these processes alternate in importance over several decades within the GFDL CM2.1 as observed (Guilyardi 2006). Since the 1970's, the upwelling temperature has been the predominant feedback (Wang 1995).

While GFDL CM2.1 predicts a reduced ENSO amplitude in response to increased greenhouse forcing, there is no consensus even among the most highly regarded models. Philip and Van Oldenborgh (2006) find that while both upwelling feedbacks amplify as the greenhouse gas concentration increases, damping processes (due to cloud radiation, for example) also become more effective. A robust prediction of future El Niño amplitudes requires both the upwelling feedback and damping along with their relative amplitude to be simulated consistently, which remains a challenge.

El Niño events are related to climate anomalies throughout the globe. Models with more realistic ENSO variability generally exhibit an anti-correlation with the strength of the Asian summer monsoon (e.g. Annamalai *et al.,* 2006), while 21st century changes to Amazon rainfall have been shown to depend upon projected trends in the tropical Pacific (Li *et al.,* 2006). El Niño has a long-

established relation to North American climate (Horel and Wallace 1981), assessed in the AR4 models by Joseph and Nigam (2006). This relation is strongest during NH winter, when the tropical anomalies are largest. Anomalous circulations driven by rainfall over the warming equatorial Central Pacific radiate atmospheric disturbances into mid-latitudes that are amplified within the North Pacific storm track (Sardeshmukh and Hoskins 1988; Held *et al.*, 1989; Trenberth *et al.*, 1998). To simulate the influence of ENSO upon North America, the models must simulate realistic rainfall anomalies and in the correct season in order that the connection is amplified by the wintertime storm tracks. The connection between equatorial Pacific and North American climate is simulated most accurately by the NCAR PCM model (Joseph and Nigam 2006). In the GFDL CM2.1, North American anomalies are too large, consistent with the model's excessive El Niño variability within the equatorial Pacific. The connection between the two regions is realistic if the model's tropical amplitude is accounted for. In the GISS model, anomalous rainfall during ENSO is small, consistent with the weak tropical wind stress anomaly cited above. The influence of El Nino over North America is nearly negligible in this model. The weak rainfall anomaly is presumably a result of unrealistic coupling between the atmospheric and ocean physics. When SST is instead prescribed in this model, rainfall calculated by the GISS modelE AGCM over the American southwest is significantly correlated with El Niño as observed.

Realistic simulation of El Niño, and its global influence, remains a challenge for coupled models, because of the myriad processes contributing and their changing importance in the observational record. Key aspects of the coupling between the ocean and atmosphere, the relation between SST and wind stress anomalies, for example, are the result of complicated interactions between the resolved model circulations, along with parameterizations of the ocean and atmospheric boundary layers and moist convection. Simple models identify parameters controlling the magnitude and frequency of El Niño, such as the wind anomaly resulting from a change in SST (e.g., Zebiak and Cane 1987; Fedorov and Philander 2000), offering guidance to improve the realism of fully coupled GCM's. However, in a GCM, the coupling strength is emergent rather than prescribed, and it is often unclear *a priori* how to change the coupling. Nonetheless, the improved simulations of the ENSO cycle compared to previous generations (AchutaRao and Sperber 2006) suggest that additional realism can be expected in the future. This optimism arises in part from the extensive and

unprecedented model comparisons carried out as part of the AR4, where the flaws identified in current models may point toward future solutions.

Multi-decadal variability

The Earth's climate varies naturally on multi-decadal scales due to the internal dynamics of the system. These changes are apparent from accurate measurements taken over decades to centuries. From the 1950s onward, an unprecedented volume of observations has been collected that contributes to the understanding of the changes to our climate. The satellite era, beginning in the late 1960's has further expanded the available data and contributed greatly to the set of measurements that are used in this area of research. Further, retrospective research efforts are able to deduce earlier changes to the climate through the analyses of climate indicators such as tree rings and ice cores.

To understand the long period changes in the Earth's climate system, scientists primarily use a set of indices that reduce a large amount of data to a small set of time series. For example, in the tropical Pacific, an index referred to as "Nino 3" is the average sea surface temperature (SST) between 5°N-5°S and150°W-90°W, and indicates variations associated with El Nino and the climate of the tropical Pacific. Other indices, such as the North American Oscillation Index, use sea surface pressure differences at two locations, one in Iceland and one near the Azores (Jones et al. 1997, Hurrell 1995) to examine large-scale shifts in atmospheric pressure systems. Long period measurements of precipitation, such as over the Sahel (20°N-10°N, 20°W-10°E) (Janowiak 1988) also are used understand decadal variability. These analyses can be used to assess the realism of internal or natural variability of the climate models. In addition to whether actual events have been modeled correctly, the climate models are evaluated also in terms of whether the statistical properties of the observed variability are well represented. Previous sections have described some of the low frequency behavior of the climate models (e.g. ENSO, annular modes, polar climates, ice models).

All the models have their own unique intrinsic or natural variability due to the various model design decisions that have been made. The models also tend to differ regionally in their simulation skill.

For example, some are better at simulating the North Atlantic, while others have more skill in the tropics. Often, this regional skill is serendipitous and emerges unexpectedly from attempts to improve simulation of processes that operate globally. A set of examples are given to provide an overview of the general abilities of the current climate models to reproduce decadal and longer variability.

In the Arctic, during the last century, there have been two long period warm events, one between 1920 and 1950 and another beginning after 1979. Wang et al. (2007) evaluated a set of IPCC Fourth Assessment models as to the models' ability are to reproduce the amplitudes of air temperature variability of the mid-century. The CCSM3 and GFDL-CM2 models contain similar variance with the observational variance in the Arctic region. Other models under-represented the natural variability.

Multi-decadal variability in the North Atlantic is characterized by the Atlantic Multidecadal Oscillation (AMO) index which represents a spatial average of SST (Enfield et al. 2001). Kravtsov and Spannagle (2007) analyzed SST from a set of current generation climate models. Their analysis attempts to separate the variability that is associated with internal fluctuations of the ocean from that associated with changes in the atmospheric component due to anthropogenic contributions. By isolating the multi-decadal period of several regions in the ensemble SST series through statistical methods, they found that models, on average, correlate well with the AMO (Figure 7, 8 from Kravtsov and Spannagle, 2007).

In the mid-latitude Pacific region, the decadal variability is generally under-represented in the ocean (e.g. volume transports as described by Zhang and McPhaden, 2006, Figure 3), with some of the models approaching the amplitudes seen in the observations. Examination of complicated feedbacks between the atmosphere and ocean at the decadal and longer scales show that the while the climate models generally reproduce the pattern in SST related to the Pacific Decadal Oscillation (PDO), observed correlations between the PDO and tropical SST are not seen in the models (e.g. Alexander et al. 2006).

One of the most difficult areas to simulate is the Indian Ocean, because of competing effects of warm water inflow through the Indonesian archipelago, ENSO, monsoons, etc). The processes interact to varying degrees, challenging a model's ability to simulate all aspects of the system with the observed relative emphasis. An index used to understand the Indian Ocean's variability is the Indian Ocean Dipole pattern that combines information about the SST and wind stress fields of the Indian Ocean (Saji et al., 1999). While most of the models evaluated by Saji et al. (2006) were able to simulate the Indian's Ocean response to local atmospheric forcing on short time periods (semi-annual), the longer period events such as the ocean's response to ENSO changes in the Pacific, were not simulated well.

Extreme events

Flood-producing precipitation, drought, heat waves, and cold waves have severe impacts on North America. Flooding resulted in average annual losses of $3.7 billion during 1983-2003 (**http://www.flooddamagedata.org/**). Losses from the 1988 drought were estimated at $40 billion and the 2002 drought at $11 billion. The heat waves in 1995 resulted in 739 excess deaths in Chicago alone (Whitman *et al*., 1997). It is probable that a large component of the overall impacts of climate change will arise from changes in the intensity and frequency of extreme events.

The modeling of extreme events poses special challenges since they are, by definition, rare in nature. Although the intensity and frequency of occurrence of extreme events are modulated by the state of the ocean and land surface and by trends in the mean climate state, internal variability of the atmosphere plays a very large role and the most extreme events arise from the chance confluence of unlikely conditions. Their very rarity makes statistical evaluation of model performance less robust than for the mean climate. For example, if one wanted to evaluate the ability of a model to simulate heat waves as intense as the 1995 event in Chicago, there are only a few episodes in the entire 20th century that approach or exceed that intensity (Kunkel *et al*., 1996). For such rare events, there is substantial uncertainty in the real risk, varying from once every 30 years to once every 100 years or more. Thus, a model that simulates such events at a frequency of once every 30 years may be

performing adequately, but it cannot be distinguished in its performance from a model simulating such an event at a frequency of once every 100 years.

Although one might expect that a change in mean climate conditions will apply equally to changes in the extremes, this is not necessarily the case. Using as an example the 50 state record low temperatures, the decade with the largest number of records is the 1930's, yet winters during this decade averaged as the third warmest since 1890; in fact, there is no significant correlation between the number of records and U.S. wintertime temperature (Vavrus *et al*., 2006). Thus, the severest cold air outbreaks in the past have not necessarily been coincident with cold winters. Another examination of model data showed that the future changes in extreme temperatures differed from changes in the mean temperature in many regions (Hegerl *et al*., 2004). This means that climate model output must be analyzed explicitly for extremes by examining daily (or even finer) resolution data, a resource-intensive effort.

The evaluation of model performance with respect to extremes is hampered by incomplete data on the historical frequency and severity of extremes. A study by Frich *et al*., (2002) described a set of indices suitable for performing global analyses of extremes and presented global results. However, many areas were missing due to lack of suitable station data, particularly in the tropics. It has become common to use some of these indices for comparisons between models and observations. Another challenge for model evaluation is the spatially-averaged nature of model data, representing an entire grid cell, while station data represent point observations. For some comparisons, it is necessary to average the station data over areas representing a grid cell.

There are several approaches toward the evaluation of model performance of simulation of extremes. One approach examines whether a model reproduces the magnitude of extremes. For example, a daily rainfall amount of 100 mm or more is expected to occur about once every year in Miami, once every 6 years in New York City, once every 13 years in Chicago, and once every 200 years in Phoenix. To what extent is a model able to reproduce the absolute magnitudes and spatial variations of such extremes? A second approach examines whether a model reproduces observed trends in extremes. Perhaps the most prominent observed trend in the U.S. is an increase in the

frequency and intensity of heavy precipitation, particularly during the last 20–30 years of the 20th century. Another notable observed trend is an increase in the length of the frost-free season.

In some key respects, it is likely that the model simulation of temperature extremes is less challenging than of precipitation extremes, in large part due to the scales of these phenomena. The typical heat wave or cold wave covers a relatively large region, of the order of several hundred miles or more, or a number of grid cells in a modern climate model. By contrast, heavy precipitation can be much more localized, often extending over regions of much less than 150 km, or less than the size of a grid cell. Thus, the modern climate model can directly simulate the major processes causing temperature extremes while heavy precipitation is sensitive to the parameterization of subgrid scale processes, particularly convection (Chapter 2; Emori *et al*., 2005; Iorio *et al*., 2004).

Droughts, particularly over North America and Africa

Recent analysis indicates that there has been a globally-averaged trend toward greater areal coverage of drought since 1972 (Dai *et al*., 2004). A simulation by the HadCM3 model reproduces this dry trend (Burke *et al*., 2006) only if anthropogenic forcing is included. A control simulation indicates that the observed drying trend is outside the range of natural variability. The model, however, does not always correctly simulate the regional distributions of areas of increasing wetness and dryness.

The simulation of specific regional features remains a major challenge for models. Globally, one of the most significant observed changes is the shift to more frequent and more severe droughts in the Sahel region of Africa since about 1970. Lau *et al*., (2006) find that only eight CGCMs produce a reasonable Sahel drought signal, while seven CGCMs produce excessive rainfall over the Sahel during the observed drought period. Even the model with the highest prediction skill of the Sahel drought could only predict the increasing trend of severe drought events but not the beginning and duration of the events. Hoerling *et al*. (2006) also finds that the AR4 models fail to simulate the drying and furthermore uses the model results to suggest that the observed drying was not due to anthropogenic forcing. However, two GFDL models are successful in reproducing the drying and analysis of those models suggests that the drying is of anthropogenic origin (Held *et al*. 2005).

Biasutti and Giannini (2006) interpret these results as an indication that the drying was a combination of decadal-scale internal variability superimposed on longer timescale changes associated with anthropogenic forcing. The differences between modeled and observed regional patterns may then be due to the randomness of natural variability, but may also result from inadequate representation of regional processes and feedbacks.

Excessive rainfall leading to floods Several different measures of excessive rainfall have been used in analyses of model simulations. A common one is the annual maximum 5-day precipitation amount, one of the Frich *et al*. (2002) indices. This has been analyzed in several recent studies (Kiktev *et al*. 2003; Hegerl *et al*. 2004; Tebaldi *et al*. 2006). Other analyses have examined thresholds of daily precipitation, either absolute (e.g. 50 mm per day in Dai 2006) or percentile (e.g. 4th largest precipitation event equivalent to 99th percentile of the 365 daily values as in Emori *et al*. 2005). Recent studies of model simulations produced for the IPCC AR4 provide information on the performance of the latest generation of models.

There is a general tendency for models to underestimate very heavy precipitation. This is shown in a comparison between satellite (TRMM) estimates of daily precipitation and model-simulated values within the 50S-50N latitude belt (Dai 2006). The TRMM observations derive 7% of the total precipitation from very heavy rainfall of 50 mm or more per day, in contrast to only 0-2% for the models. For the frequency of very heavy precipitation of 50 or more mm per day, the TRMM data show a frequency of 0.35% (about once every 300 days), whereas it is 0.02-0.11% (once every 900 to 5000 days) for the models. A global analysis of model simulations showed that the models produced too little precipitation in events exceeding 10 mm per day (Sun *et al*. 2006). Examining how many days it takes to accumulate 2/3's of the annual precipitation, the models generally show too many days compared to observations over North America, although a few models are close to reality. In contrast to the general finding of a tendency toward underestimation, a study (Hegerl *et al*. 2004) of two models (HadCM3 and CGCM2) indicates generally good agreement with the observed annual maximum 5-day precipitation amount over North America for HadCM3 and even somewhat of an overestimation for CGCM2.

This model tendency to produce rainfall events less intense than observed appears to be due in part to the low spatial resolution of global models. Experiments with individual models show that increasing the resolution improves the simulation of heavy events. For example, the 4th largest precipitation event in a model simulation with a resolution of approximately 300 km averaged 40 mm over the conterminous U.S., compared to an observed value of about 80 mm. When the resolution was increased to 75 km and 50 km, the 4th largest event was still smaller than observed, but by a much smaller amount (Iorio *et al*. 2004). A second factor that is important is the parameterization of convection. Thunderstorms are responsible for many intense events, but their scale is smaller than the size of model grids and thus they must be indirectly represented in models (Chapter 2). One experiment showed that changes to this representation improves model performance and, when combined with high resolution of about 1.1 deg latitude, can produce quite accurate simulations of the 4th largest precipitation event on a globally-averaged basis (Emori 2005). Another experiment found that the use of a cloud-resolving model imbedded in a global model eliminated the underestimation of heavy events (Iorio *et al*. 2004). A cloud-resolving model eliminates the need for a parameterization of convection, but is very expensive to run. These sets of experiments indicate that the problem of heavy event underestimation may be significantly reduced in the future as increases in the computer power allows simulations at higher spatial resolution and perhaps eventually the use of cloud-resolving models.

The improved model performance at higher spatial resolutions provide motivation for use of regional climate models when only a limited area is of interest, such as North America. The spatial resolution of these models is sufficient to resolve the major mountain chains; some of these models thus display considerable skill in areas where topography plays a major role in the spatial patterns. For example, they are able to reproduce rather well the spatial distribution of the magnitude of the 95th percentile of precipitation (Leung *et al*. 2003), the frequency of days with more than 50 mm and 100 mm (Kim and Lee 2003), the frequency of days over 25 mm (Bell *et al*. 2004), and the annual maximum daily precipitation amount (Bell *et al*. 2004) over the western U.S. Kunkel *et al*. (2002) found that an RCM's simulation of the magnitude of extreme events over the U.S. varied spatially and depended on the duration of the event being examined; there was a tendency for overestimation in the western U.S. and good agreement or underestimation in the central and eastern U.S.

Most studies of observed precipitation extremes suggest that such extremes have increased in frequency and intensity during the latter half of the 20th century. A study by Tebaldi *et al*., (2006) indicates that models generally simulate a trend towards a world characterized by intensified precipitation, with a greater frequency of heavy-precipitation and high-quantile events, although with substantial geographical variability. This is in agreement with observations. Wang and Lau (2006) find that the CGCMs simulate an increasing trend in heavy rain over the tropical ocean.

Heat and cold waves

Analysis of simulations produced for the IPCC AR4 by seven climate models indicates that they reproduce the primary features of cold air outbreaks (CAOs), with respect to location and magnitude (Vavrus *et al*., 2006). In their analysis, a CAO is an episode of at least 2 days duration during which the daily mean winter (December-January-February) surface temperature at a gridpoint is 2 standard deviations below the gridpoint's winter mean temperature. Maximum frequencies of about four CAO days/winter are simulated over western North America and Europe, while minimal occurrences of less than one day/winter exist over the Arctic, northern Africa, and parts of the North Pacific. The GCMs are generally accurate in their simulation of primary features, with a high pattern correlation with observations and the maximum number of days meeting the CAO criteria around 4 per winter. One favored region for CAOs is in western North America, extending from southern Alaska into the upper Midwest. Here, the models simulate a frequency of about 4 CAO days per year, in general agreement with the observed values of 3-4 days. The models underestimate the frequency in the southeastern United States: mean simulated values range from 0.5 to 2 days *versus* 2 to 2.5 days in observations. This regional bias occurs in all the models and reflects the inability of GCMs to penetrate Arctic air masses far enough southeastward over North America.

The IPCC AR4 model simulations show a positive trend for growing season, heat waves and warm nights and a negative trend for frost days and daily temperature range (maximum minus minimum) (Tebaldi *et al*. 2006). They indicate that this is in general agreement with observations, except that there is no observed trend in heat waves. The modeled spatial patterns have generally larger positive trends in western North America than in eastern sections. For the U.S., this is in qualitative agreement with observations which show that the decreases in frost-free season and frost days are largest in the western U.S. (Kunkel *et al*. 2004; Easterling *et al*. 2002).

Analysis of individual models provides a more detailed picture of model performance. In a simulation from the PCM (Meehl *et al*. 2004), the largest trends for decreasing frost days occurs in the western and southwestern USA (values greater than –2 days per decade), and trends near zero in the upper Midwest and northeastern USA, in good agreement with observations. The biggest discrepancy between model and observations is over parts of the southeastern USA where the model shows trends for decreasing frost days and the observations show slight increases. This is thought to be a partial consequence of the two large El Nino events in the observations during this time period (1982–83 and 1997–98) where anomalously cool and wet conditions occurred over the southeastern USA and contributed to slight increases of frost days. The ensemble mean from the model averages out effects from individual El Nino events, and thus the frost day trends reflect a more general response to the forcings that occurred during the latter part of the 20^{th} century. An analysis of short-duration heat waves simulated by the PCM (Meehl and Tebaldi, 2004) indicates good agreement with observed heat waves for North America. In that study, heat waves were defined by daily minimum temperature. The most intense events occurred in the southeast U.S. for both the model simulation and observations. The overall spatial pattern of heat wave intensity in the model matched closely with the observed pattern. In a four-member ensemble of simulations from the HadCM3 (Christidis *et al*. 2005), the model shows a rather uniform pattern of increases in the warmest night for 1950-1999. The observations also show a global mean increase, but with considerable regional variations. In North America, the observed trends in the warmest night vary from negative in the south-central sections to strongly positive in Alaska and western Canada, compared to a rather uniform pattern in the model. However, this discrepancy might be expected, since the observations probably reflect a strong imprint of internal climate variability that is reduced by ensemble averaging of the model simulations.

An analysis of the magnitude of temperature extremes for California in a regional climate model simulation (Bell *et al*. 2004) show mixed results. The hottest maximum in model is 4°C less than observations, while coldest min is 2.3°C warmer. The number of days >32°C is 44 in the model compared to an observed value of 71. This could result from the lower diurnal temperature range in the model (15.4°C observed vs. 9.7°C simulated). While these results are better than the driving GCM, the RCM results are still somewhat deficient, perhaps reflecting the very complex topography of the region of study.

Models display some capability to simulate extreme temperature and precipitation events, but there are differences from observed characteristics. They typically produce global increases in extreme precipitation and severe drought, and decreases in extreme minimum temperatures and frost days, in general agreement with observations. There is a general, though not universal, tendency to underestimate the magnitude of heavy precipitation events. Regional trend features are not always captured. Since the causes of observed regional trend variations are not known in general and such trends could be due in part to stochastic variability of the climate system, it is difficult to assess the significance of these discrepancies.