CHAPTER 5

Model Simulation of Major Climate Features

Although a typical use of atmosphere-ocean general circulation model (AOGCM) output for climate impact assessment focuses on one particular region such as a river basin or one of the 50 United States, knowing model simulation overall accuracy on continental to global scales is important. Fidelity in simulating climate on the largest scales is a necessary condition for credible predictions of future climate on smaller scales. Model developers devote great effort to assessing the level of agreement between simulated and observed large-scale climate, both for the present day and for the two centuries since the Industrial Era began. Unlike physical theories of such fundamentally simple systems as the hydrogen atom, AOGCMs cannot promise precise accuracy for every simulated variable on all relevant space and time scales. Nevertheless, before applying a model to a practical question, users should demand reasonable overall agreement with observations, with the definition of "reasonable" in part subjective and dependent on the problem at hand. Here we provide an overview of how well modern AOGCMs satisfy this criterion.

5.1 MEAN SURFACE TEMPERATURE AND PRECIPITATION

Simulations of monthly near-surface air temperature and precipitation provide a standard starting point for model evaluation since these fields are central to many applications. The two fields also illustrate the difficulty in designing appropriate metrics for measuring model quality.

By most measures, modern AOGCMs simulate the basic structure of monthly mean near-surface temperatures quite well. The globally averaged annual mean value generally lies within the observed range (~286 to 287 K) of modern and preindustrial values; this agreement, however, is in part a consequence of the "final tuning" of the models' energy balance as described in Chapter 2 and by itself is not a stringent test of model quality. More relevant is consideration of space and time variations about the global annual mean (including the seasonal cycle). The overall correlation pattern between simulations and observations typically is 95 to 98%, and variation magnitudes typically agree within ±25% (Covey et al. 2003). This level of success has been retained in the latest generation of models that allow ocean and atmosphere to exchange heat and water without artificial adjustments (Randall et al. 2007). Nevertheless, as shown below, local errors in surface temperature that are clearly outside the bounds of observational uncertainty persist in the latest generation of models.

AOGCM simulations are considerably less accurate for monthly mean precipitation than for temperature. The space-time correlation between models and observations typically is only about 50 to 60% (Covey et al. 2003). As we discuss below, these poor correlations originate mainly in the tropics, where precipitation varies greatly over relatively small ranges of latitude and longitude. Strong horizontal gradients in the field lead to a significant drop in correlations with observations, even with only slight shifts in the modeled precipitation distribution. These modest correlations are relevant for precipitation at a particular location, but AOGCMs generally reproduce the observed broad patterns of precipitation amount and year-to-year variability (see Fig. 5.1 and Dai 2006). One prominent error is that models without flux adjustment typically fail to simulate the observed northwest-to-southeast orientation of a large region of particularly heavy cloudiness and precipitation in the southwest Pacific Ocean. Instead, these models tend to rotate this convergence zone into an east-west orientation, producing an unrealistic pair of distinct, parallel convection bands straddling the equator instead of a continuous Inter-Tropical Convergence Zone (ITCZ). The double-ITCZ error has been frustratingly persistent in climate models despite much effort to correct it.

Another discrepancy between models and observations appears in the average day-night cycle of precipitation. While the model's diurnal temperature cycle exhibits general agreement with observations, simulated cloud formation and precipitation tend to start too early in the day. Also, when precipitation is sorted into light, moderate, and heavy categories, models reproduce the observed extent of moderate precipitation (10 to 20 mm/day) but underestimate that 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 United States (Ruiz-Barradas and Nigam 2006)].

For illustration, we show examples from two of the U.S. models discussed in Chapter 4. In Fig. 5.1 (Delworth et al. 2006) and Fig. 5.2 (Collins et al. 2006a), simulated and observed maps of surface temperature and even precipitation appear rather similar at first glance. 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. CCSM3 temperature-difference maps exhibit the largest errors in the Arctic (note scale change in Fig. 5.2d), 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 above the surface (warm air overlying cold air). For precipitation, GFDL 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 CCSM3 results (e.g., a 28% underestimate of Amazon annual mean). AOGCM precipitation errors have serious implications for Earth system models with interactive vegetation, because such models use simulated precipitation to calculate plant growth (see Chapter 6). Errors of this magnitude would produce an unrealistic distribution of vegetation in an Earth system model, for example, by spuriously deforesting the Amazon basin.

In summary, modern AOGCMs generally simulate continental and larger-scale mean surface temperature and precipitation with considerable accuracy, but the models often are not reliable for smaller regions, particularly for precipitation.

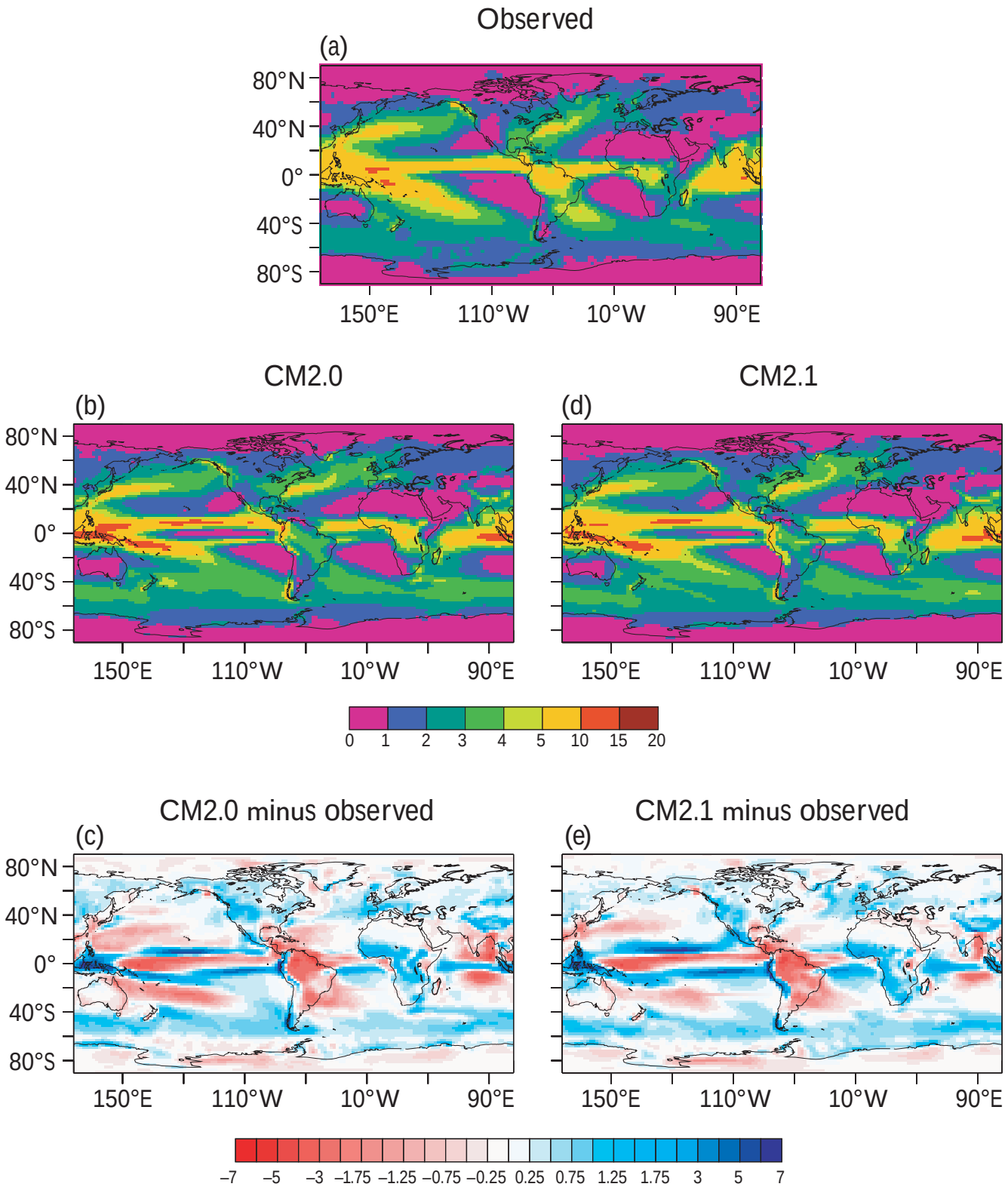


Figure 5.1a–e. Observed and GFDL Model-Simulated Precipitation (mm/day).

Observed image from P. Xie and P.A. Arkin 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin American Meteorological Society*, **78**, 2539–2558. [Other images from Fig. 17 in T.L. Delworth et al. 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674. Images reproduced with permission of the American Meteorological Society.]

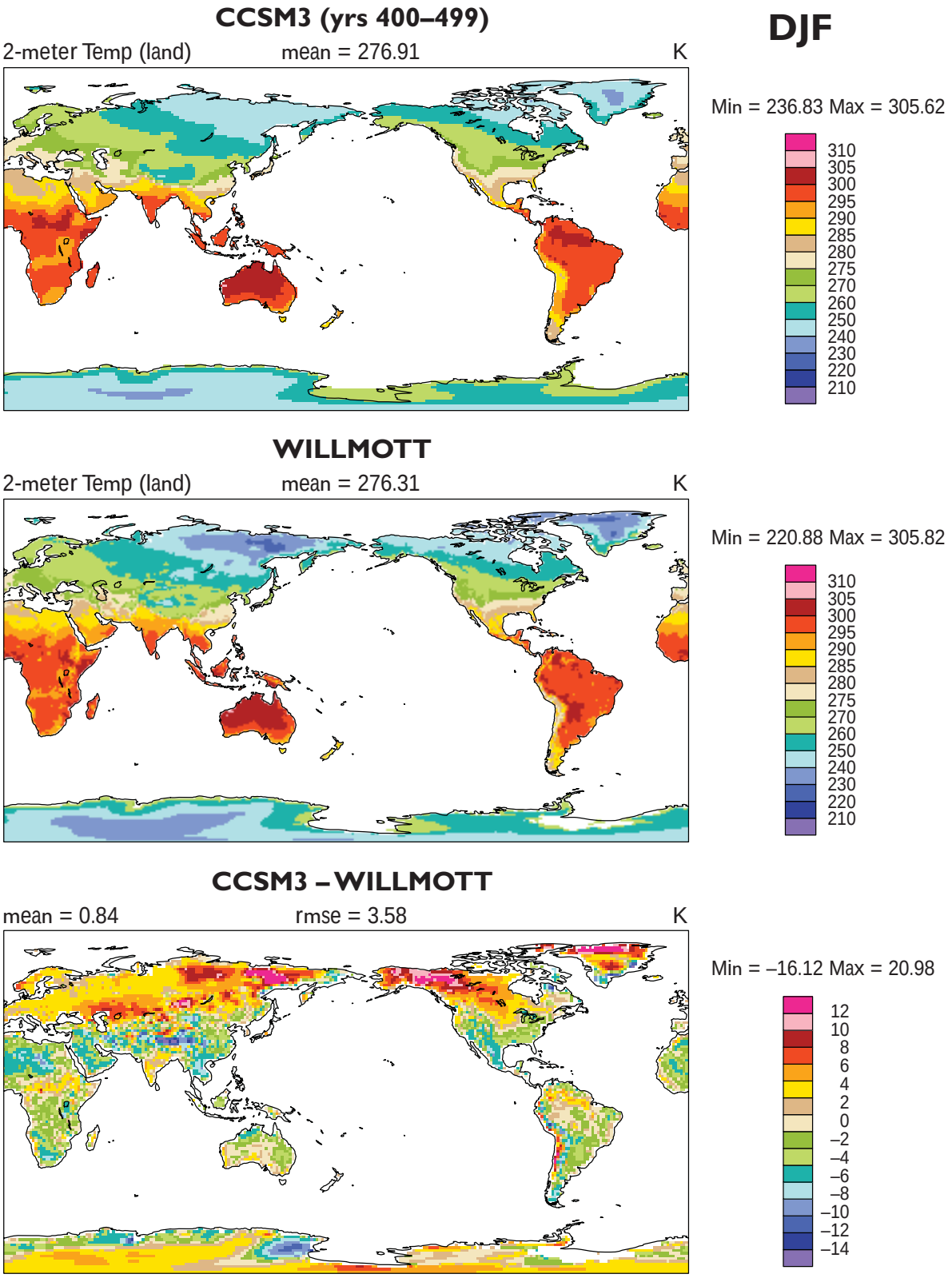


Figure 5.2a–c. CCSM3 December-January-February Simulated (top panel), Observed (middle panel), and Simulated-Minus-Observed (bottom panel) Near-Surface Air Temperature for Land Areas (°C).

Note change in scale from 5.2a to 5.2c. [Figures from W. Collins et al. 2006: The Community Climate System Model Version 3 (CCSM3). *J. Climate*, **19**(11), 2122–2143. Reproduced with permission of the American Meteorological Society.]

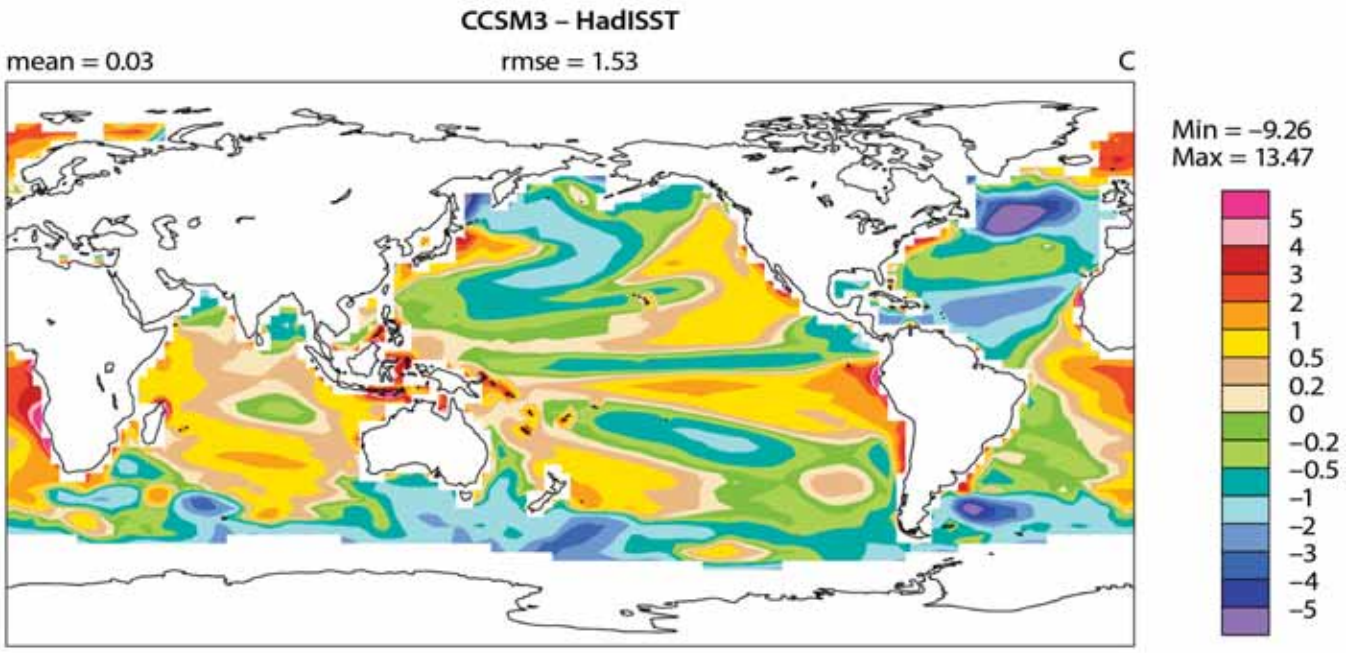


Figure 5.2d. CCSM3 Annual Mean Simulated-Minus-Observed Sea Surface Temperature (°C).

[Figure from W. Collins et al. 2006: The Community Climate System Model Version 3 (CCSM3). *J. Climate*, **19**(11), 2122–2143. Reproduced with permission of the American Meteorological Society.]

5.2 TWENTIETH CENTURY TRENDS

Modern AOGCMs are able to simulate not only the time-average climate but also changes (trends) in climate over the past 140 years. For example, Fig. 5.3 shows results from the three U.S. models and the "average" CMIP3 model. Plotted in the figure are curves of globally averaged annual mean near-surface temperature from model simulations and the observational value as determined from the U.K. Climatic Research Unit (CRU) gridded observational database. Two curves are plotted for the CMIP3 models. The first shows the average over all CMIP3 models, and the second, the average over only CMIP3 models that included the effects of volcanic eruptions. Results from individual U.S. models are shown for separate ensemble members (dotted lines) and for the average over all ensemble members (continuous lines). Individual members of a particular model ensemble differ from each other because they were run from different initial conditions. Precise initial conditions, especially deep-ocean temperature and salinity, are not known for 1860. The spread among individual simulations from the same model (the dotted-line curves) thus indicates uncertainty in model-simulated temperature arising from lack of knowledge about initial conditions.

These results demonstrate that modern climate models exhibit agreement with observed global mean near-surface temperature trends to within observational uncertainty, despite imprecise initial conditions and uncertain climate forcing and heat uptake by the deep ocean (Min and Hense 2006). Models achieve this agreement only if they include anthropogenic emissions of greenhouse gases and aerosols. No plausible combination of natural climate-forcing factors allows models to explain the global warming observed over the last several decades. Indirect solar effects [e.g., involving cosmic rays and clouds (Svensmark 2007)] are not generally included in AOGCM simulations. These effects have been proposed occasionally as causes of global warming, although over the past 20 years their trends would, if anything, lead to cooling

(Lockwood and Fröhlich 2007). Unless the models grossly underestimate the climate system's natural internally generated variability or are all missing a large unknown forcing agent, the conclusion is that most recent warming is anthropogenic (IPCC 2007b).

Nevertheless, total climate forcing during the 20th Century is not accurately known, especially the aerosol component (see Chapter 2). Aerosol forcing used in these simulations, however, is derived from aerosol parameterizations constrained by satellite and ground-based measurements of the aerosols themselves and was not designed to obtain a fit to observed global mean temperature trends. The observed trend in surface temperature can result from models with different aerosol forcing (Schwartz 2007). Thus, 20th Century temperature records cannot distinguish models that would warm by differing amounts for the same total forcing.

Note that climate sensitivity is not prescribed in AOGCMs. Instead, this sensitivity emerges as a result of a variety of lower-level modeling choices. In contrast to simple energy-balance models that predict only the global mean temperature using a limited representation of climate physics, an AOGCM's climate sensitivity is difficult to specify a priori. More fundamentally, AOGCMs, unlike simpler climate models, have far fewer adjustable parameters than the number of observations available for model evaluation (Randall et al. 2007). Thus, an AOGCM's multidimensional output can be compared to observations independent of this adjustment (e.g., using observed trends in regional temperature). Agreement between modeled and observed trends has been described for temperature trends on each inhabited continent (Min and Hense 2007); for trends in climate extremes, such as heat-wave frequency and frost-day occurrence (Tebaldi et al. 2006); and for trends in surface pressure and Arctic sea ice (see Chapter 9 in IPCC 2007), all of which complement comparisons between modeled and observed time-averaged climate discussed in the following sections.

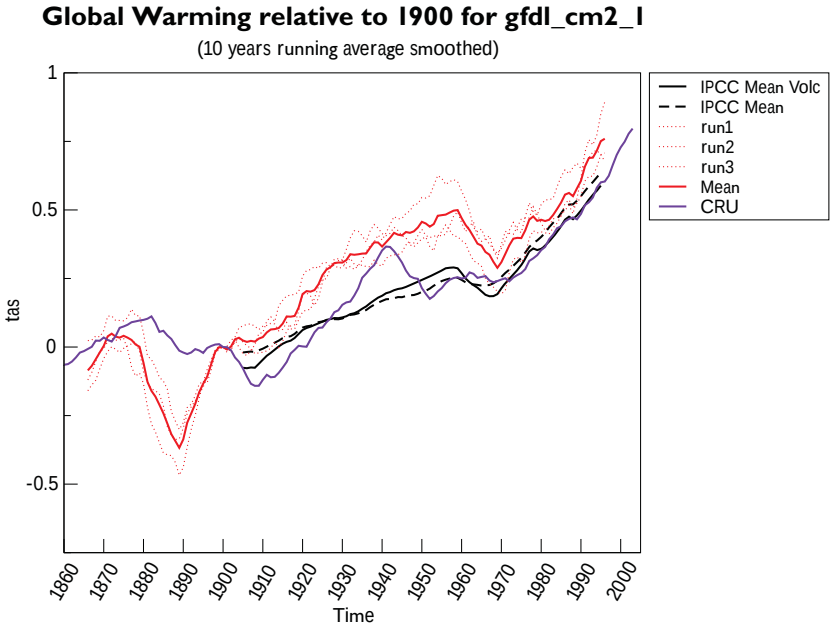


Figure 5.3a. Simulation of 20th Century Globally Averaged Surface Temperature from GFDL CM2.1.

"CRU" is the value based on the Climate Research Unit gridded observational dataset, "IPCC Mean" is the average value of all CMIP3 models, and "IPCC Mean Volc" is the average of all CMIP3 models that included volcanic forcing. Individual realizations of the CMIP3 20th Century experiment are denoted by the dotted curves labeled "run(1–3)," and the ensemble mean is marked "Mean."

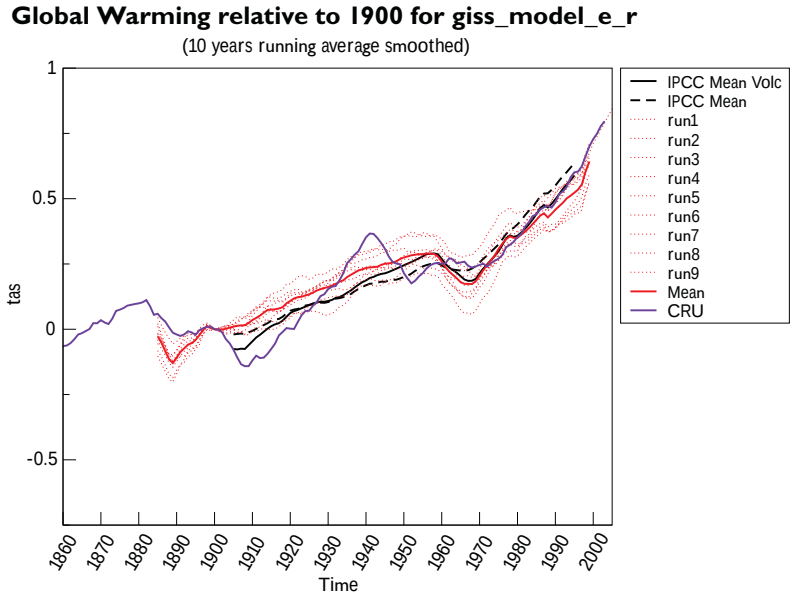


Figure 5.3b. Simulation of 20th Century Globally Averaged Surface Temperature from GISS Model E-R.

Curve labels are the same as in Fig. 5.3a.

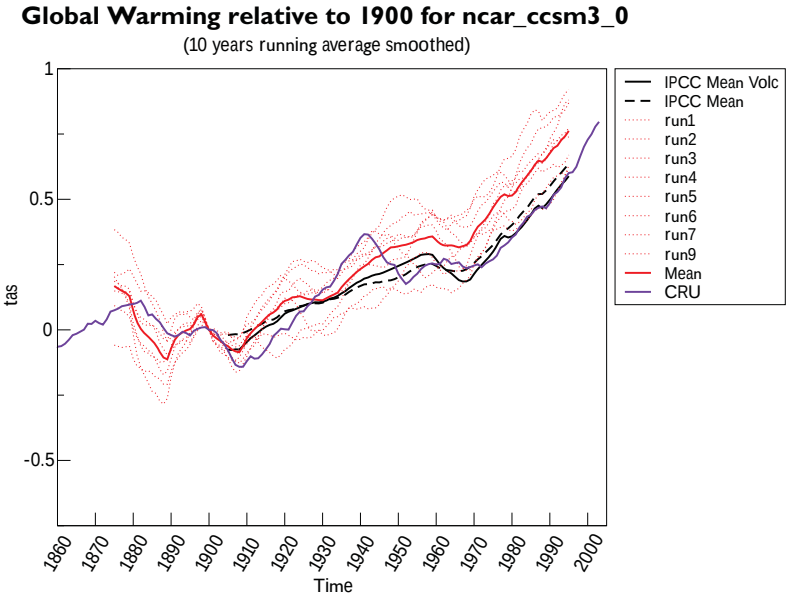


Figure 5.3c. Simulation of 20th Century Globally Averaged Surface Temperature from CCSM3.

Curve labels are the same as in Fig. 5.3a.

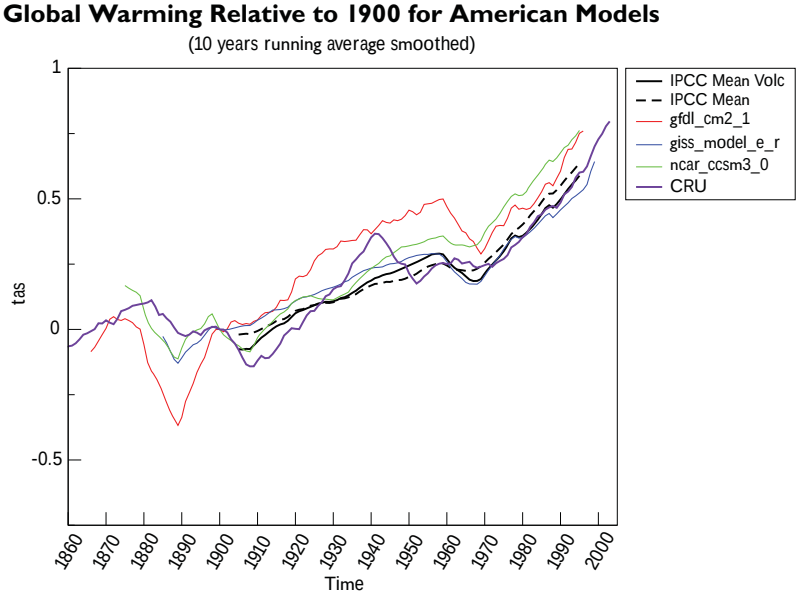


Figure 5.3d. Comparison of Simulations of 20th Century Globally Averaged Surface Temperature from the Three U.S. CMIP3 Models.

Model curves represent ensemble means for CCSM3 (ncar_ccsm3_0), GISS Model E-R (giss_e_r), and GFDL CM2.1 (gfdl_cm2_1). "CRU," "IPCC Mean," and "IPCC Mean Volc" labels are the same as in Fig. 5.3a.

As an example of 20th Century temperature trends on continental-to-global spatial scales and multidecadal time scales, Fig. 5.4 shows global maps for different time periods between 1880 and 2003 as observed and simulated by GISS ModelE (Hansen et al. 2006; also see Knutson et al. 2006). The figure shows general agreement between model and observations not only for the overall period but also for segments 1880 to 1940 and 1979 to 2003, which encompass periods of early and late 20th Century warming. For 1940 to 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) is slightly positive rather than slightly negative as observed. Part of this discrepancy may result from chaotic fluctuations within observed climate that the model cannot synchronize correctly due to inprecise knowledge of the initial conditions in the 19th Century period. These chaotic fluctuations generally are more important in regional trends than in the global average, where uncorrelated fluctuations in different regions tend to cancel. For both 20th Century warming periods, the model simulates, but underestimates, the high-latitude amplification of global warming. Additional discrepancies between AOGCMs and observations appear at smaller scales. For example, model-simulated trends do not consistently match the observed lack of 20th Century warming in the central United States (Kunkel et al. 2006).

5.2.1 Trends in Vertical Temperature

While models simulate the 20th Century warming observed at the surface, agreement is less obvious with tropospheric observations from satellites and weather balloons. This issue was the focus of CCSP SAP 1.1 (CCSP 2006). Since 1979 (beginning of the satellite record), globally averaged warming in the troposphere according to climate models is within the range of available observations. Within the tropics, the model-simulated troposphere warms more rapidly than observed (see CCSP 2006, Fig. 5.4 F–

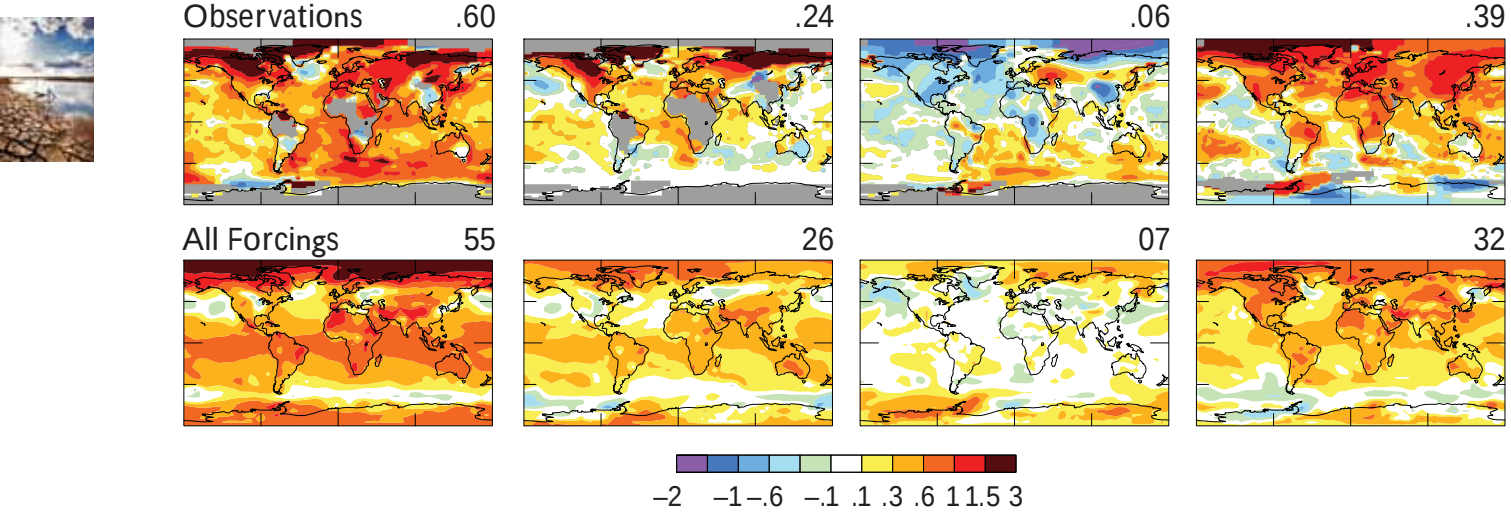


Figure 5.4. Near-Surface Temperature Changes as Observed (top panels) and as Simulated by GISS ModelE (bottom panels) for Selected Time Periods Between 1880 and 2003.

Numbers above upper right panel corners are global means. [Images from Fig. 9 in J. Hansen et al. 2007: Climate simulations for 1880–2003 with GISS ModelE. *Climate Dynamics*, **29**(7–8), 661–696. Reproduced with kind permission of Springer Science and Business Media.]

G). SAP 1.1 noted, however, that "Large structural uncertainties in the observations . . . make it difficult to reach more definitive conclusions regarding the significance and importance of model-data discrepancies" (CCSP 2006, p. 112 and Section 5.4).

Research since publication of SAP1.1 has continued to highlight uncertainties implicit in measuring the difference between surface and lower-atmospheric warming. For example, Thorne et al. (2007) found that the tropical atmosphere-to-surface warming ratio in both observations and model simulations is sensitive to the time period analyzed. Meanwhile, debate continues over the best way to process data from satellites (Christy et al. 2007) and weather balloons (Christy and Spencer 2005). AOGCMs continue to differ from most published observations on the ratio of atmosphere-to-surface warming in the tropics since the beginning of satellite observations (e.g., as shown by Thorne et al. 2007, Fig. 3), with the ratio being larger in the models than is seen in decadal observational trends.

Paradoxically, trends are more consistent between models and observations on interannual time scales. AOGCM simulation of tropical atmospheric warming involves mainly subgrid-scale parameterizations. As discussed in Chapter 2, these are not as trustworthy as explicitly computed processes, but internal variability [primarily due to El Niño–Southern Oscillation (ENSO)] provides a useful test of the models' ability to redistribute heat realistically. AOGCMs simulate very well the portion of tropical temperature trends due to interannual variability (Santer et al. 2005). In addition, explaining how atmospheric water vapor increases coincidentally with surface temperature is difficult (Trenberth, Fasullo, and Smith 2005; Santer et al. 2007; Wentz et al. 2007) unless lower tropospheric temperature also increases coincidentally with surface temperature. While deficiencies in model subgrid-scale parameterizations are certainly possible, trends in poorly documented forcing agents (see Chapter 4) may prove important in explaining the discrepancy over the longer time scales. Future research is required to resolve the issue because tropospheric observations at face value suggest a trend toward greater tropical instability, which has implications for many aspects of model projections in the tropics.

5.2.2 Model Simulation of Observed Climate Variability

The following sections discuss a number of specific climate phenomena directly or indirectly related to near-surface temperature, precipitation, and sea level. Numerous studies of climate change have focused on one or two of these phenomena, so a great deal of information (and occasional debate) has accumulated for each of them. Here we attempt to summarize the points that would best give users of AOGCM model output a general sense of model reliability or unreliability. Although the following sections individually note different types of climate variation, the reader should recognize that the total amount of natural climate variability forms background "noise" that must be correctly assessed to identify the "signal" of anthropogenic climate change. Natural variability in turn separates into an externally forced part (e.g., from solar energy output and volcanic eruptions) and internally generated variability just as weather varies on shorter time scales because of the system's intrinsic chaotic character. As noted above, long-term trends in both solar and volcanic forcing during the past few decades have had a cooling rather than warming effect. It follows that if global warming during this period is not anthropogenic, then the climate system's internal variation is the most likely alternative explanation.

Control runs of AOGCMs (in which no changes in external climate forcing are included) provide estimates of the level of internally generated climate variability. Control runs generally obtain realistic near-surface temperature variability on annual-to-decadal time scales, although they typically underestimate variability in areas of the Pacific and Indian Ocean where ENSO and the Pacific Decadal Oscillation (PDO) (see below) predominate (Stouffer, Hegerl, and Tett 2000). Unfortunately, the longest time periods that are directly relevant to separating natural from anthropogenic climate change are the least observed. Assessing variations of surface temperature for time periods longer than 50 to 100 years depends on paleodata such as ice-core composition and tree-ring thickness. Interpreta-

tion of these data is made difficult by sparse geographical coverage and also is complicated by natural variations in external climate forcing.

5.2.2.1 Extra-Tropical Storms

Climate models have developed from numerical weather-prediction models whose performance has been judged primarily on their ability to forecast midlatitude weather. The success of forecast models in their simulation of midlatitude cyclones and anticyclones has resulted in continuous growth in the value of numerical weather prediction. The ability of GCMs to generate realistic statistics of midlatitude weather also has been central in climate model development. This is true 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 simulation of global climate. Indeed, a defining feature of atmospheric general circulation models (AGCMs) is that they compute midlatitude eddy statistics and associated eddy fluxes through explicit computation of the life cycles of individual weather systems and not through some turbulence or parameterization theory. Computing the evolution of individual eddies may seem very inefficient when primary interest is in long-term eddy statistics, but the community clearly has judged for decades that explicit eddy simulation in climate models is far superior to attempts to develop closure theories for eddy statistics. The latter theories typically form the basis for Earth system models of intermediate complexity (EMICs), which are far more efficient computationally than GCMs but provide less convincing simulations.

Two figures illustrate the quality of simulated midlatitude eddy statistics from coupled AOGCMs used in IPCC AR4. Shown for the GFDL CM2.1 in Fig. 5.5a is wintertime variance of the north-south velocity component at 300 hPa (in the upper troposphere). This quantity represents the magnitude of variability in the upper troposphere associated with day-to-day weather. In Fig. 5.5b, the wintertime poleward eddy heat flux or covariance between temperature and north-south velocity is shown at 850 mb (in the lower troposphere). For these calculations, the monthly means were subtracted before computing variances. In each case, eddy statistics are compared to estimates of observed statistics obtained from NCEP/NCAR Reanalysis (B.Wyman, personal communication). When analyzing eddy statistics, the data are typically filtered to retain only those time scales, roughly 2 to 10 days, associated with midlatitude weather systems. The two quantities chosen here, however, are sufficiently dominated by these time scales that they are relatively insensitive to the monthly filtering used here. In winter, Northern Hemisphere storms are organized into two major oceanic storm tracks over the Pacific and Atlantic oceans. Historically, atmospheric models of horizontal resolutions of 200 to 300 km typically are capable of simulating midlatitude storm tracks with realism comparable to that shown in the figure. Eddy amplitudes often are a bit weak and often displaced slightly equatorward. In spectral models with resolution coarser than 200 to 300 km, simulation of midlatitude storm tracks typically deteriorates significantly (see, e.g., Boyle 1993). General improvements in most models in the CMIP3 database over previous generations of models, as described in Chapter 1, are thought to be partly related to the fact that most of these models now have grid sizes of 100 to 300 km or smaller. Although even-finer resolution results in better simulations of midlatitude-storm structure, including that of warm and cold fronts and interactions among these storms and coastlines and mountain ranges, improvements in midlatitude climate on large scales tend to be less dramatic and systematic. Other factors besides horizontal resolution are considered important for details of storm track structure. Such factors include distribution of tropical rainfall, which is sensitive to parameterization schemes used for moist convection, and interactions between stratosphere and troposphere, which are sensitive to vertical resolution. Roeckner et al. (2006), for example, illustrate the importance of vertical resolution for midlatitude circulation and storm track simulation.

Lucarini et al. (2006) provide a more detailed look at the ability of CMIP3 models to simulate the space-time spectra of observed eddy statistics. These authors view the deficiencies noted, which vary in detail from model to model, as serious limitations to model credibility. As indicated in Chapter 1, however, our ability is lim-

ited in translating measures of model biases into useful measures of model credibility for 21st Century projections, and the implications of these biases in eddy space-time spectra are not self-evident. Indeed, in the context of simulating eddy characteristics generated in complex turbulent flows in the laboratory (e.g., Pitsch 2006), the quality of atmospheric simulations, based closely on fluid dynamical first principles, probably should be thought of as one of the most impressive characteristics of current models. As an example of a significant model deficiency that plausibly can be linked to limitations in climate projection credibility, note that the Atlantic storm track, as indicated by the maximum velocity variance in Fig. 5.5a, follows a latitude circle too closely and the observed storm track has more of a southwest-northeast tilt. This particular deficiency is common in CMIP3 models (van Ulden and van Oldenborgh 2006) and is related to difficulty in simulating the blocking phenomenon in the North Atlantic with 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.

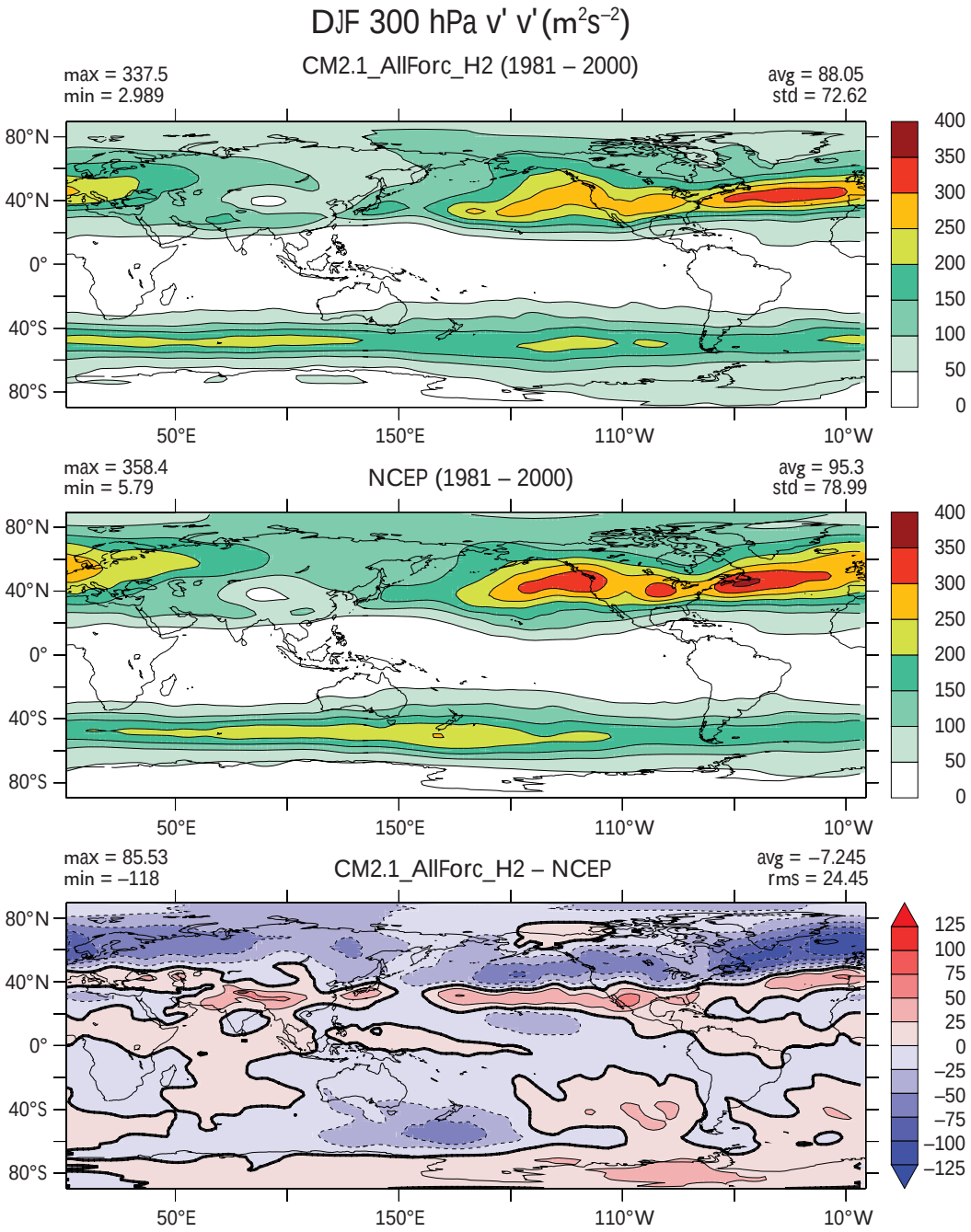


Figure 5.5a. Top: Variance of North-South Velocity at 300 hPa as Simulated by GFDL CM2.1 Model in Years 1981 to 2000 of One Realization of 20C3M Simulation, as Contributed to the CMIP3 Database. Units are m^2/s^2. Middle: Same quantity as obtained from NCEP/NCAR Reanalysis (Kalnay et al. 1996). Bottom: Model minus observations.

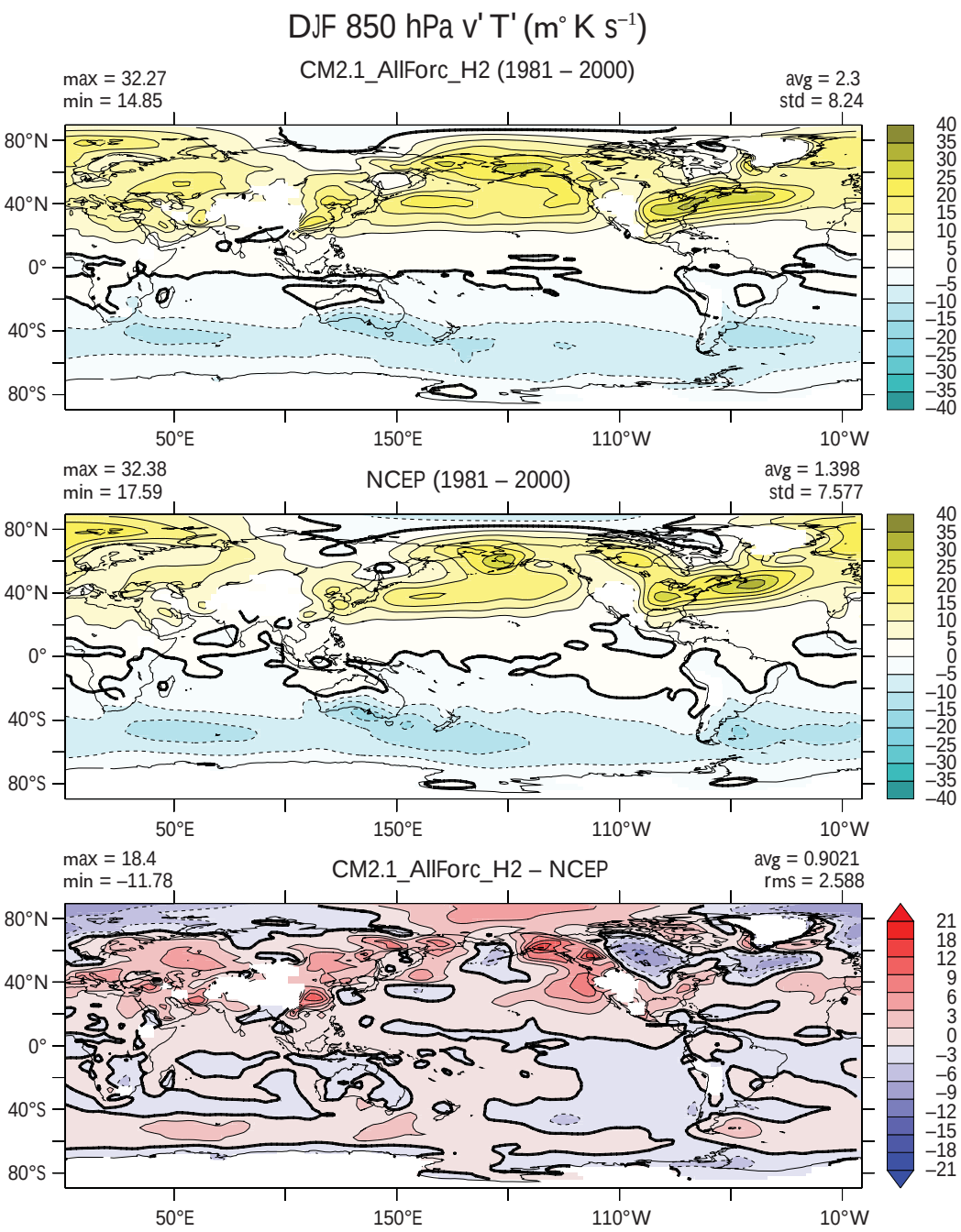


Figure 5.5b. Top: Covariance of North-South Velocity and Temperature at 850 hPa as Simulated by GFDL CM2.1 Model in Years 1981 to 2000 of One Realization of 20C3M Simulation, as Contributed to the CMIP3 Database.

Units are K-m/s. Middle: Same quantity as obtained from NCEP/NCAR Reanalysis (Kalnay et al. 1996). Bottom: Model minus observations.

5.2.2.2 Tropical Storms

Tropical storms (hurricanes in the Atlantic and typhoons in the Pacific and Indian oceans) are too small to be simulated reliably in the class of global-climate models currently used for climate projections. There is hope for simulating regional climate aspects that control the genesis of tropical depressions, however. Vitart and Anderson (2001), for example, identified tropical storm-like vortices in simulations with models of this type, demonstrating some skill in simulating the effects of El Niño on Atlantic storm frequency.

Simulations with atmospheric models are steadily moving to higher resolutions (e.g., Bengtsson, Hodges, and Esch 2007). The recent 20-km–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. Experi-

ence with tropical storm forecasting suggests that this resolution should be adequate for describing many aspects of the evolution of mature tropical storms and possibly the generation of storms from incipient disturbances, but probably not tropical storm intensity. A promising alternative approach is described by Knutson et al. (2007), in which a regional model of comparable resolution (18 km) is used in a downscaling framework (see Chapter 3) to simulate the Atlantic hurricane season. Given observed year-to-year variations in the large-scale atmosphere structure over the Atlantic Ocean, the model is capable of simulating year-to-year variations in hurricane frequency over a 30-year period with a correlation of 0.7 to 0.8. It also captures the observed trend toward greater hurricane frequency in the Atlantic during this period. These results suggest that downscaling using models of this resolution may be able to provide a convincing capability for tropical storm frequency projections into the future, although these projections still will rely on the quality of global model projections for changes in sea-surface temperature, atmospheric stability, and vertical shear.

5.2.2.3 Monsoons

A monsoonal circulation is distinguished by its seasonal reversal after the sun crosses the equator into the new summer hemisphere. Rain is most plentiful in, if not entirely restricted to, 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 (Privé and Plumb 2007). Variations in the monsoon's spatial extent from year to year determine which inland regions experience drought.

Over a billion people are dependent on the arrival of monsoon rains for water and irrigation for agriculture. The Asian monsoon during boreal summer is the most prominent example of a monsoon circulation dominating global rainfall during this season. However, the summer rainfall maximum and seasonal reversal of winds also indicate monsoon circulations in West Africa and the Amazon basin. In addition, during boreal summer, air flows off the eastern Pacific Ocean toward Mexico and the American Southwest while, over the Great Plains, moisture from the Gulf of Mexico brings an annual peak in rainfall. Thus, the climate in these regions also is described as monsoonal.

Because of the Asian monsoon's geographical extent, measures of the fidelity of Asian monsoonal simulations can differ depending on specific regional focus and the metrics being used. Kripalani et al. (2007) judged that 3/4 of the 18 analyzed coupled models match the timing and magnitude of the summertime peak in precipitation over East Asia between 100 and 145°E and 20 to 40°N evident in the NOAA-NCEP Climate Prediction Center's Merged Analysis of Precipitation (CMAP, Xie, and Arkin 1997). However, only half of these models were able to reproduce the gross observed spatial distribution of monsoon rainfall and its migration along the coast of China toward the Korean peninsula and Japan. Considering a broader range of longitude (40 to 180°E) that includes the Indian subcontinent, Annamalai, Hamilton, and Sperber (2007) found that 6 of 18 AOGCMs significantly correlated with the observed spatial pattern of CMAP precipitation from June through September. (These six models also produced relatively realistic simulation of ENSO variability, which is known to influence interannual variations in the Asian summer monsoon.) Kitoh and Uchiyama (2006) computed the spatial correlation and root-mean-square error of simulated precipitation over a similar region and found, for example, the GFDL models in the top tercile with a spatial correlation exceeding 0.8.

During boreal winter, Asian surface winds are directed offshore: from the northeast over India and the northwest over East Asia. Hori and Ueda (2006) provide correlations between observed spatial distributions of surface pressures and 850-mb zonal winds during the East Asian winter monsoon with winds and pressures simulated by nine CMIP3 models. Correlations for zonal winds, for example, vary from 0.96 to 0.75. Monsoonal simulations in these models clearly vary considerably in quality, more so perhaps than other circulation features. Observed year-to-year variability of the West African monsoon is related to remote ocean temperatures in the North and South Atlantic and Indian oceans (Rowell et al. 1992; Zhang and Delworth 2006) as well as to temperatures

in the nearby Gulf of Guinea. Cook and Vizy (2006) found that slightly more than half of 18 analyzed coupled models reproduced the observed precipitation maximum over land from June through August. Of these models, only six (including GISS ModelE-H and both GFDL models) reproduced the observed anticorrelation between Gulf of Guinea ocean temperature and Sahel rainfall.

The late 20th Century Sahel drought was a dramatic change in the Earth's hydrological cycle that plausibly must be simulated by climate models if we are to have any confidence in their ability to project future climate in this region. Atmospheric models, when run over observed oceanic temperatures, simulate this drought reasonably well (Hoerling et al. 2006). In these models, the drought is at least partly forced by warming of the Northern Hemisphere oceans, particularly the North Atlantic, with respect to Southern Hemisphere oceans, especially the Indian Ocean and Gulf of Guinea. Although the consensus is that these variations in ocean temperature gradients are at least partly due to natural variability, they may have been partly anthropogenically forced. Analysis of CMIP3 simulations of the 20th Century by Biasutti and Giannini (2006), supporting the earlier modeling study of Rotstayn and Lohmann (2002), suggests that aerosol forcing in these models played a part in generating this drought by cooling the North Atlantic with respect to other ocean basins. A small number of coupled models simulate droughts of the observed magnitude, including GFDL models (Held and Soden 2006), but why some models are more realistic in this regard than others is not understood.

Rainfall over the Sahel and Amazon are anticorrelated: when the Gulf of Guinea warms, rainfall generally is reduced over the Sahel but increases over South America. Amazon rainfall also depends on the eastern equatorial Pacific, and, during an El Niño, rainfall is reduced in the Nordeste region of the Amazon. Li et al. (2006) compare the hydrological cycle of 11 CGCMs over the Amazon during the late 20th and 21st centuries. Based on a comparison to CMAP rainfall, the GISS ModelE-R is among the best.

The ability of climate models to simulate Northern Hemisphere summer rainfall over the U.S. Great Plains and Mexico was summarized by Ruiz-Barradas and Nigam (2006). Models generally have more difficulty in simulating summer rainfall in the Great Plains than winter rainfall, and this disparity probably should be thought of as reflecting the quality of future rainfall projections as well. Strengths and weaknesses vary considerably across the models. As an example, GISS ModelE-H closely matches the annual precipitation cycle over the Great Plains and Mexico and is one of two models to simulate interannual precipitation variations significantly correlated with observed variability during the second half of the 20th Century.

Initial monsoon evaluations 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 interrupted temporarily, are crucial to forecasting the monsoon's impact on water supply. Simulating the diurnal cycle and the local hour of rainfall also is important to partitioning rainfall between runoff and transpiration, and these are important topics for future model evaluation. Transports of moisture by regional circulations beneath model resolution (such as low-level jets along the Rockies and Andes and tropical cyclones) contribute to the onshore transport of moisture. In general, models show some success at simulating gross seasonal features of various monsoon circulations, but studies are limited on variations of the smaller spatial and time scales important to specific watersheds and hydrological projections.

5.2.2.4 Madden-Julian Oscillations

The Madden-Julian Oscillation (MJO) consists of large-scale eastward-propagating patterns in humidity, temperature, and atmospheric circulation that strengthen and weaken tropical rainfall as they propagate around the Earth in roughly 30 to 60 days. This pattern often dominates tropical precipitation variability on time scales longer than a few days and less than a season, creating such phenomena as 1- to 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 a main impediment to more useful extended-range

weather forecasts in the tropics, and improved simulation of this phenomenon is considered an important metric for the credibility of climate models in the tropics.

Nearly all models capture the pattern's essential feature, with large-scale eastward propagation and with roughly the correct vertical structure. But propagation often is too rapid and amplitudes too weak. Recent surveys of model performance indicate that simulations of MJO remain inadequate. For example, Lin et al. (2006), in a study of many CMIP3 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 multimodel comparison study, state that "… commendable progress has been made in MJO simulations in the past decade, but the models still suffer from severe deficiencies …." As an example of recent work, Boyle et al. (2008) attempted, with limited success, to determine whether two U.S. CMIP3 models could maintain a preexisting strong MJO pattern when initialized with observations [from the Tropical Ocean Global Atmosphere–Coupled Ocean Atmosphere Response Experiment (called TOGA-COARE) field experiment].

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

A picture seems to be emerging that simulation difficulty may not be due to a single model deficiency but is a result of the phenomenon's complexity, given the long list of factors thought to be significant. In several multimodel studies such as Lin et al. (2006), a few models do perform well. However, without a clearer understanding of how these factors combine to generate the observed characteristics of MJO, maintaining a good simulation when the model is modified for other reasons is difficult, as is applying the understanding gained from one model's successful simulation to other models. Whether models with superior MJO simulations should be given extra weight in multimodel studies of tropical climate change is unclear.

5.2.2.5 El Niño–Southern Oscillation

By the mid-20th Century, scientists recognized that a local anomaly. in rainfall and oceanic upwelling near the coast of Peru was in fact part of a disruption to atmospheric and ocean circulations 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, abbreviated in combination as ENSO, are the largest source of tropical variability observed during recent decades. Because of the Earth's rotation, easterly winds along the equator cool the surface by raising cold water from below, which offsets heating by sunlight absorption (e.g., Clement et al. 1996). Cold water is especially close to the surface in the east Pacific, while warm water extends deeper in the west Pacific so upwelling has little effect on surface temperature there. The westward increase in temperature along the equator is associated with a decrease in atmospheric pressure, reinforcing the easterly trade winds. El Niño occurs when easterly trade winds slacken, reducing upwelling and warming the ocean surface in the central and east Pacific.

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 typically is 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 eastern equatorial Pacific, accurate projections of climate change in the 21st 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, long-term warming of the eastern equatorial Pacific relative to the surrounding ocean will favor a weaker Asian monsoon year after year, even in the absence of changes to the size and frequency of El Niño events.

In general, coupled models developed for 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 models assessed more recently by the El Niño Simulation Intercomparison Project (called ENSIP) and CMIP2 (Latif et al. 2001; AchutaRao and Sperber 2002), ENSO variability of ocean surface temperature is more realistic in CMIP3 simulations, 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 equatorial ocean circulation that has accompanied increases in computing speed. Table 5.1 shows horizontal and vertical resolution near the equator in oceanic components of the seven American coupled models whose output was submitted to CMIP3.

Table 5.1. Spacing of Grid Points at the Equator in the American Coupled Models Developed for AR4*

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

**Except for GISS models, spacing of grid points generally increases away from the equator outside the ENSO domain, so resolution is highest at the equator.*

Along the equator, oceanic waves that adjust the equatorial temperature and currents to changes in the wind are confined tightly 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 a half-degree resolution, while both GFDL models have equatorial resolution of a 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 cooling effect of upwelling is spread over a larger area, so the amplitude of the resulting surface temperature fluctuation is weakened. In fact, both the GISS AOM models and ModelE-R have unrealistic ENSO variations that are much smaller than observed (Hansen et al. 2007). This minimizes the influence of their simulated El Niño and La Niña 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 (Neelin et al. 1992), the AR4 coupled models generally simulate El Niño near the observed amplitude or even above (AchutaRao and Sperber 2006). The latter study compared sea-surface temperature (SST) variability within the tropical Pacific, calculated under preindustrial conditions. Despite its comparatively low two-degree latitudinal grid spacing, the GISS ModelE-H (among American models) most closely matches observed SST variability since the mid-19th Century, according to the HadISST v1.1 dataset (Rayner et al. 2003). The NCAR PCM also exhibits El Niño warming close to the observed magnitude. This comparison is based on spatial averages within three longitudinal bands, and GISS ModelE-H, along with NCAR models, exhibits its largest variability in the eastern band as observed. However, GISS ModelE-H underestimates variability since 1950, when the NCAR CCSM3 is closest to observa-

tions (Joseph and Nigam 2006). Although the fidelity of each model's ENSO variability depends on the specific dataset and period of comparison (c.f. Capotondi, Wittenberg, and Masina 2006; Merryfield 2006; van Oldenborgh, Philip, and Collins 2005), the general consensus is that 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 GFDL CM2.1. Most models (including GISS ModelE-H and both NCAR models but excluding GFDL models) exhibit the largest variability in the eastern band of longitude, but none of the CMIP3 models matches the observed variability at the South American coast where El Niño was identified originally (AchutaRao and Sperber 2006; Capotondi, Wittenberg, and Masina 2006). This possibly is because the longitudinal spacing of model grids is too large to resolve coastal upwelling and its interruption during El Niño (Philander and Pacanowski 1981). Biases in atmospheric models (e.g., underestimating persistent stratus cloud decks along the coast) also may contribute (Mechoso et al. 1995).

El Niño occurs every few years, albeit irregularly. The spectrum of anomalous ocean temperature shows a broad peak between 2 and 7 years, and multidecadal variations occur in event frequency and amplitude. Almost all AR4 models have spectral peaks within this range of time scales. Interannual power is distributed broadly within the American models, as observed, with the exception of NCAR CCSM3, which exhibits strong biennial oscillations (Guilyardi 2006).

Although models generally simulate the observed magnitude and frequency of events, reproducing 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 NCAR CCSM3 (Joseph and Nigam 2006). Warming in GFDL CM2.1 and GISS ModelE-H is nearly uniform throughout the year, while warming in 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 boreal spring and cool again late in the calendar year. GFDL CM2.1 and NCAR PCM1 have the weakest seasonal cycle among 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 CMIP3 models, amplitude of the seasonal cycle of equatorial ocean temperature generally varies inversely with the ENSO cycle's strength.

Several studies have compared mechanisms generating ENSO variability in CMIP3 models to those inferred from observations (e.g., van Oldenborgh, Philip, and Collins 2005; Guilyardi 2006; Merryfield 2006; Capotondi, Wittenberg, and Masina 2006). Models must simulate the change in ocean upwelling driven by changes in surface winds, which in turn are driven by regional contrasts in ocean temperature. In general, GFDL2.1 is ranked consistently among American models as providing the most realistic simulation of El Niño. This is not based primarily on its surface-temperature variability (which is slightly too large) but on its faithful simulation of the observed relationship between ocean temperature and surface wind, along with wind-driven ocean response. While SST variability in CMIP3 models is controlled by anomalies of either upwelling rate or temperature, these processes alternate in importance over several decades within GFDL CM2.1 as observed (Guilyardi 2006). Since the 1970s the upwelling temperature, rather than the rate, has been the predominant driver of SST variability (Wang 1995). A confident prediction of future El Niño amplitude requires both the upwelling rate and temperature, along with their relative amplitude, to be simulated correctly. This remains a challenge.

El Niño events are related to climate anomalies throughout the globe. Models with more realistic ENSO variability generally exhibit an anticorrelation with the strength of the Asian summer monsoon (e.g., Annamalai, Hamilton, and Spencer 2007), while 21st Century changes to Amazon rainfall have been shown to depend on 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 CMIP3 models by

Joseph and Nigam (2006). This relation is strongest during boreal winter, when tropical anomalies are largest. Anomalous circulations driven by rainfall over the warming equatorial Central Pacific radiate atmospheric disturbances into midlatitudes amplified within the north Pacific storm track (Sardeshmukh and Hoskins 1988; Held, Lyons, and Nigam 1989; Trenberth et al. 1998). To simulate ENSO's influence on North America, models must represent realistic rainfall anomalies in the correct season so the connection is amplified by 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 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 Niño over North America is nearly negligible in this model. The weak rainfall anomaly presumably is a result of unrealistic coupling between atmospheric and ocean physics. When SST instead is 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 myriad contributing processes and their changing importance in the observational record. Key aspects of coupling between ocean and atmosphere—the relation between SST and wind stress anomalies, for example—are the result of complicated interactions among resolved model circulations, along with parameterizations of 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 GCMs. 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, improved simulations of the ENSO cycle compared to previous generations (AchutaRao and Sperber 2006) suggest that additional realism can be expected in the future.

5.2.2.6 Annular Modes

The primary mode of Arctic interannual variability is the Arctic Oscillation (Thompson and Wallace 1998), which also is referred to as the northern annular mode (NAM) and 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. The variability modes are particularly important for attributing and projecting climate change; observed circulation changes in the past few decades (especially in the Southern Hemisphere) and model-projected changes in future circulation strongly resemble these structures.

Coupled climate models have shown skill in simulating NAM (Fyfe, Boer, and Flato 1999; Shindell et al. 1999; Miller, Schmidt, and Shindell 2006). In some cases, too much variability in the simulation of sea-level pressure is associated with NAM (Miller, Schmidt, and Shindell 2006). Global climate models also realistically simulate SAM (Fyfe, Boer, and Flato 1999; Cai, Whetton, and Karoly 2003; Miller, Schmidt, and Shindell 2006), although some details of SAM (e.g., amplitude and zonal structure) show disagreement among global climate model simulations and reanalysis data (Raphael and Holland 2006; Miller, Schmidt, and Shindell 2006).

In response to increasing concentrations of greenhouse gases and tropospheric sulfate aerosols in the 20th Century, the multimodel average exhibits a positive trend in the annular mode index in both hemispheres, with decreasing sea-level pressure over the poles and a compensating increase in midlatitudes most apparent in the Southern Hemisphere (Miller, Schmidt, and Shindell 2006). A variety of modeling studies also have shown that trends in stratospheric climate can affect the troposphere's annular modes (Shindell et al. 1999). Indeed, an important result from atmospheric modeling in recent years is the realization that

the stratospheric ozone hole has contributed significantly to observed trends in surface winds and sea-level pressure distribution in the Southern Hemisphere (Thompson and Solomon 2002; Gillett and Thompson 2003). The models, however, may not be trustworthy in their simulation of the relative magnitude of greenhouse gas and stratospheric ozone effects on the annular mode. They also may underestimate the coupling of stratospheric changes due to volcanic aerosols with annular surface variations (Miller, Schmidt, and Shindell 2006; Arblaster and Meehl 2006).

5.2.2.7 Other Modes of Multidecadal Variability

In the Arctic during the last century, two long-period warm events occurred, one between 1920 and 1950 and another beginning in the late 1970s. Wang et al. (2007) evaluated a set of CMIP3 models for their ability to reproduce the amplitudes of air temperature variability of this character. As examples, CCSM3 and GFDL-CM2 models contain variance similar to that observed in the Arctic region.

Multidecadal variability in the North Atlantic is characterized by the Atlantic Multidecadal Oscillation (AMO) index, which represents a spatial average of SST (Enfield, Mestas-Nuñez, and Trimble 2001). Kravtsov and Spannagle (2007) analyzed SST from a set of current generation climate models. Their analysis attempts to separate variability associated with internal ocean fluctuations from that associated with changes by anthropogenic contributions. By isolating the multidecadal period of several regions in the ensemble SST series through statistical methods, they found that models obtain the observed magnitude of the AMO (Kravtsov and Spannagle 2007).

In the midlatitude Pacific region, decadal variability generally is underrepresented in the ocean (e.g., volume transports as described by Zhang and McPhaden 2006), with some models approaching amplitudes seen in observations. Examination of complicated feedbacks between atmosphere and ocean at decadal and longer scales shows that, while climate models generally reproduce the SST pattern related to the Pacific Decadal Oscillation (PDO), observed correlations between 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 the competing effects of warm water inflow through the Indonesian archipelago, ENSO, and monsoons. The processes interact to varying degrees, challenging a model's ability to simulate all system aspects with observed relative emphasis. An index used to understand variability is the Indian Ocean Dipole pattern that combines information about SST and wind stress fields (Saji et al. 1999). While most models evaluated by Saji, Xie, and Yamagata (2005) were able to simulate the Indian Ocean's response to local atmospheric forcing in short time periods (semiannual), longer-period events such as the ocean's response to ENSO changes in the Pacific were not simulated well.

5.2.3 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). Melting glaciers and ice sheets in Greenland and western Antarctica 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. Key processes include sea ice, seasonally frozen ground, and permafrost (Lawrence and Slater 2005; Yamaguchi, Noda, and Kitoh 2005). Processes also include seasonal snow cover (Slater et al. 2001), which can have significant subgrid heterogeneity (Liston 2004), and clear-sky precipitation, especially in the Antarctic (King and Turner 1997; Guo, Bromwich, and Cassano 2003). Polar regions test the ability of models to handle extreme geophysical behavior such as longwave radiation in clear, cold environments (Hines et al. 1999; Chiacchio, Francis, and Stackhouse 2002; Pavolonis, Key, and Cassano 2004) and cloud microphysics in the relatively clean polar atmosphere (Curry et al. 1996; Pinto, Curry, and Intrieri 2001; Morrison and Pinto 2005). In addition, polar atmospheric

boundary layers can be very stable (Duynkerke and de Roode 2001; Tjernström, Zagar, and Svensson 2004; Mirocha, Kosovic, and Curry 2005), and their simulation remains an important area for model improvement.

For polar regions, much of simulated-variability assessment has focused on primary modes of polar interannual variability, along with the northern and southern annular modes. 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 daily variability in sea-level pressure in the Antarctic and Arctic. In both polar regions, they found that the ensemble was not able to reproduce many features of daily synoptic climatology, with only a small subset of models accurately simulating the frequency of primary synoptic weather patterns identified in global reanalysis datasets. 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 CCSM models part of a small, accurate subset. More encouraging results were obtained by Vavrus et al. (2006), who assessed the ability of seven global climate models to simulate extreme cold-air outbreaks in the Northern Hemisphere.

Attention also has been given to the ability of regional climate models to simulate polar climate. In particular, the Arctic Regional Climate Model Intercomparison Project (ARCMIP) engaged a suite of Arctic regional atmospheric models to simulate a common domain and period over the western Arctic (Curry and Lynch 2002). Rinke et al. (2006) evaluated spatial and temporal patterns simulated by eight 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 six ARCMP models. In general, surface pressure, air temperature, humidity, and wind speed all were well simulated, as were radiative fluxes and turbulent momentum flux. The research group also found that turbulent heat flux was poorly simulated and that, over an entire annual cycle, the accumulated turbulent heat flux simulated by models was many times larger than the observed turbulent heat flux (Fig. 5.6).

In global models, polar climate may be affected by errors in simulating other planetary regions, but much of the difference from observations and the uncertainty about projected climate change stem 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 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 soil freezing and thawing are being implemented in land surface models (see Chapter 2). Modeling soil freeze and thaw continues to be a challenging problem as characteristics of energy and water flowing through soil affect temperature changes. Such fluxes are poorly understood (Yamaguchi, Noda, and Kitoh 2005).

Frozen soil affects surface and subsurface hydrology, which influences the surface water's spatial distribution with attendant effects on other parts of the polar climate system such as carbon cycling (e.g., Gorham 1991; Aurela, Laurila, Tuovinen 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, at best, limited representation of subsurface water reservoirs (aquifers) and horizontal flow of water both at and below the 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 land ice sheets, whose melting could raise sea levels substantially. At present, climate models do not include ice-sheet dynamics (see Chapter 2), and thus cannot account directly for ways in which ice sheets might change, possibly changing heat absorption from the sun and atmospheric circulation in the vicinity of ice sheets.

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. This need for finer resolution may be satisfied by regional models simulating just a polar region. Because both northern and southern polar regions are within circumpolar atmospheric circulations (cf. Giorgi and Bi 2000 and Gutowski et al. 2007b), their coupling with other regions is more limited than in the case of midlatitude regions, which could allow polar-specific models that focus on Antarctic and Arctic processes, in part, to improve modeling of surface-atmosphere exchange processes (Fig. 5.6). Although each process above has been simulated in finer-scale, stand-alone models, their interactions as part of a climate system also need to be simulated and understood.

5.2.3.1 Sea Ice

Sea ice plays a critical role in the exchange of heat, mass, and momentum between ocean and atmosphere, and any errors in the sea-ice system will contribute to errors in other components. Two recent papers (Holland and Raphael 2006; Parkinson, Vinnikov, and Cavalieri 2006a, b) quantify how current models simulate the climate system's sea-ice process. Very limited observations make any evaluation of sea ice difficult. The primary observation available is sea-ice areal concentration. In some comparisons, sea-ice extent (the area where local ice concentration is greater than 15%) is used. For the past few decades, satellites have made it possible to produce a more complete dataset of observations. Observations of ice extent were fewer before that. Other quantities that might be evaluated include ice thickness, but, due to limited observations, comparisons with models are difficult and will not be discussed further here.

The seasonal pattern in ice growth and decay in polar regions for all the models is reasonable (Holland and Raphael 2006; see Fig. 5.7). However, a large amount of variability between models occurs in their representation of sea-ice extent in both Northern and Southern hemispheres. Generally, models do better in simulating the Arctic than the Antarctic region, as shown with Fig. 5.8. An example of the complex nature of reproducing the ice field is given in Parkinson, Vinnikov, and Cavalieri (2006a,b), which found that all models showed an ice-free region in winter to the west of Norway, as seen in observational data, but all also produced too much ice north of Norway. The authors 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 models correctly reproducing the observations. The set of models having the most fidelity in the Arctic is not the same as the set having the most fidelity in the Antarctic. This difference may be due to distinctive ice regimes in the north and south or to differences in simulations of oceanic or meteorological circulations in those regions.

Holland and Raphael (2006) examined carefully the variability in Southern Ocean sea-ice extent. As an indicator of ice response to large-scale atmospheric events, they compared data from a set of IPCC AR4 climate models to the atmospheric index SAM for the April–June (AMJ) period (see Table 5.2). The models show that ice variability does respond modestly to large-scale atmosphere forcing but less than the limited observations show. Table 5.2 uses the U.S. models to examine whether models exhibit the observed out-of-phase buildup of ice between the Atlantic and Pacific sectors (referred to as the Antarctic Dipole).

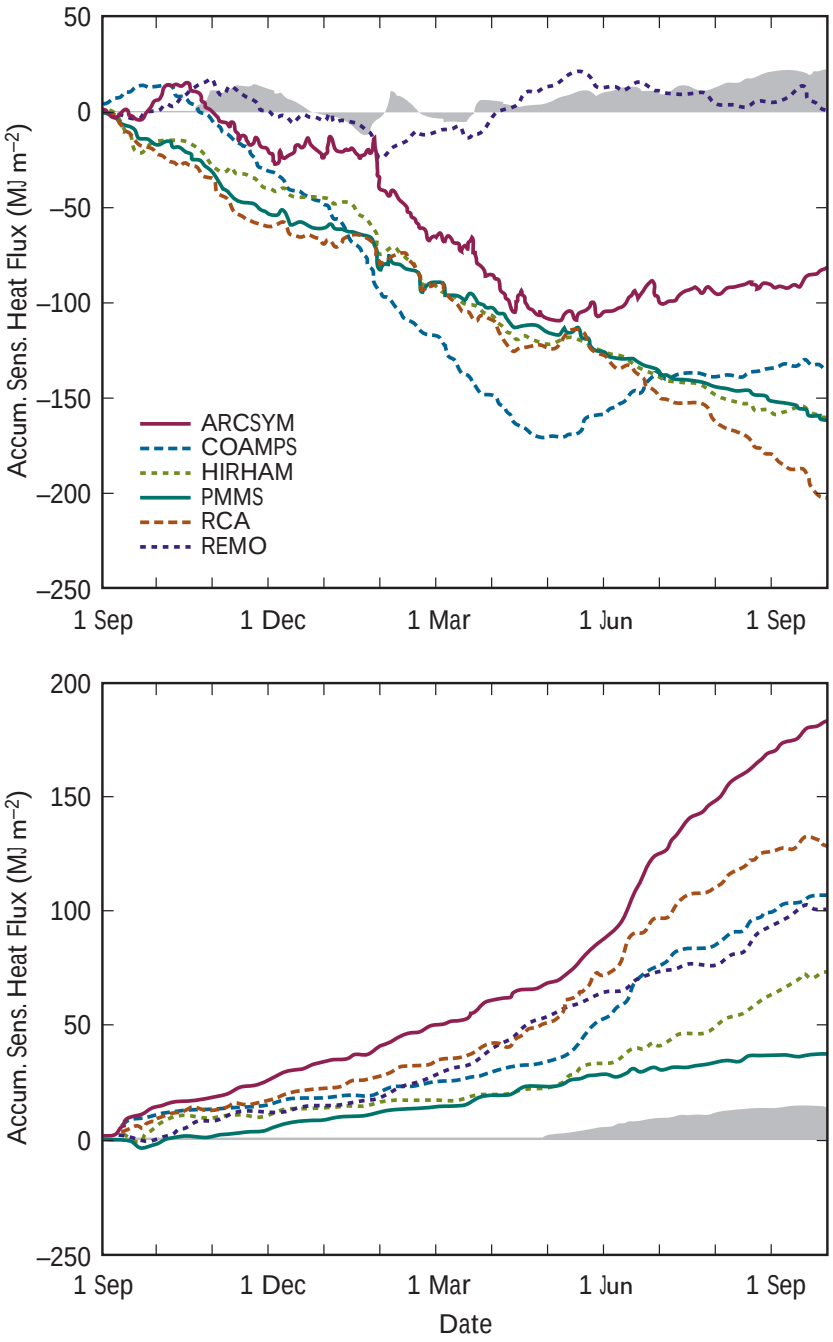


Figure 5.6. Cumulative Fluxes of Surface Sensible Heat (top panel) and Latent Heat (bottom) at the SHEBA Site.

Data are from six models simulating a western Arctic domain for Sept. 1997 through Sept. 1998 for ARCMIP. SHEBA observations are gray shaded regions; model results are shown by the individual curves identified in the key at the lower left of the upper panel. [Figure adapted from Fig. 10(c and d) in M. Tjernstrom et al. 2005: Modelling the Arctic boundary layer: An evaluation of six ARCMIP regional-scale models with data from the SHEBA project. *Boundary-Layer Meteorology*, **117**, 337–381. Reproduced with kind permission of Springer Science and Business Media.]

Table 5.2. Correlations of the Leading Mode of Sea-Ice Variability and Southern Annular Mode (SAM) for 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

The leading mode of sea-ice variability represents a shift of ice from the Atlantic to the Pacific sector. Bold values are significant at the 95% level, accounting for autocorrelation of the time series.

[Table modified from Table 1, p. 19, in M.M. Holland and M.N. Raphael 2006: Twentieth Century simulations of the Southern Hemisphere climate in coupled models. Part II: Sea ice conditions and variability. *Climate Dynamics*, **26**, 229–245. Reproduced with kind permission of Springer Science and Business Media.]

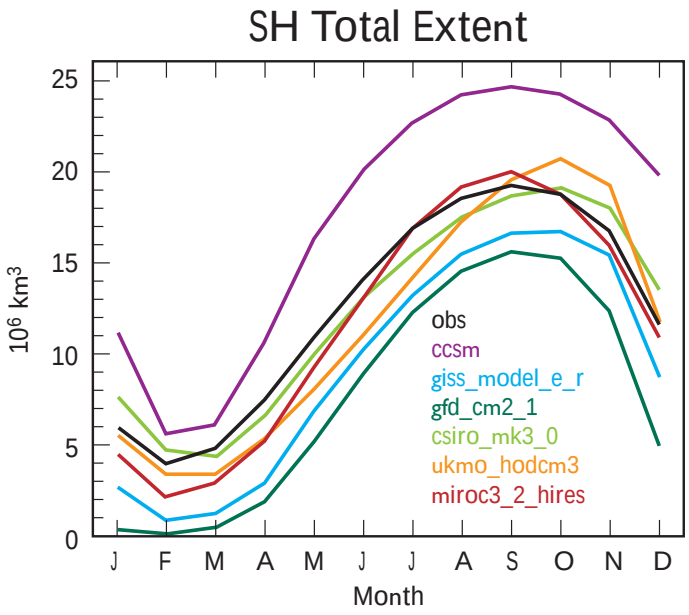


Figure 5.7. Annual Cycle of Southern Hemisphere Ice Extent.

It is defined as the area of ice with concentrations greater than 15%. Observations are identified by the black curve labeled "Obs," while the results from individual models are identified by the six colored curves. [From Fig. 1 in M.M. Holland and M.N. Raphael 2006: Twentieth Century simulations of the Southern Hemisphere climate in coupled models. Part II: Sea ice conditions and variability. *Climate Dynamics*, **26**, 229–245. Reproduced with kind permission of Springer Science and Business Media.]

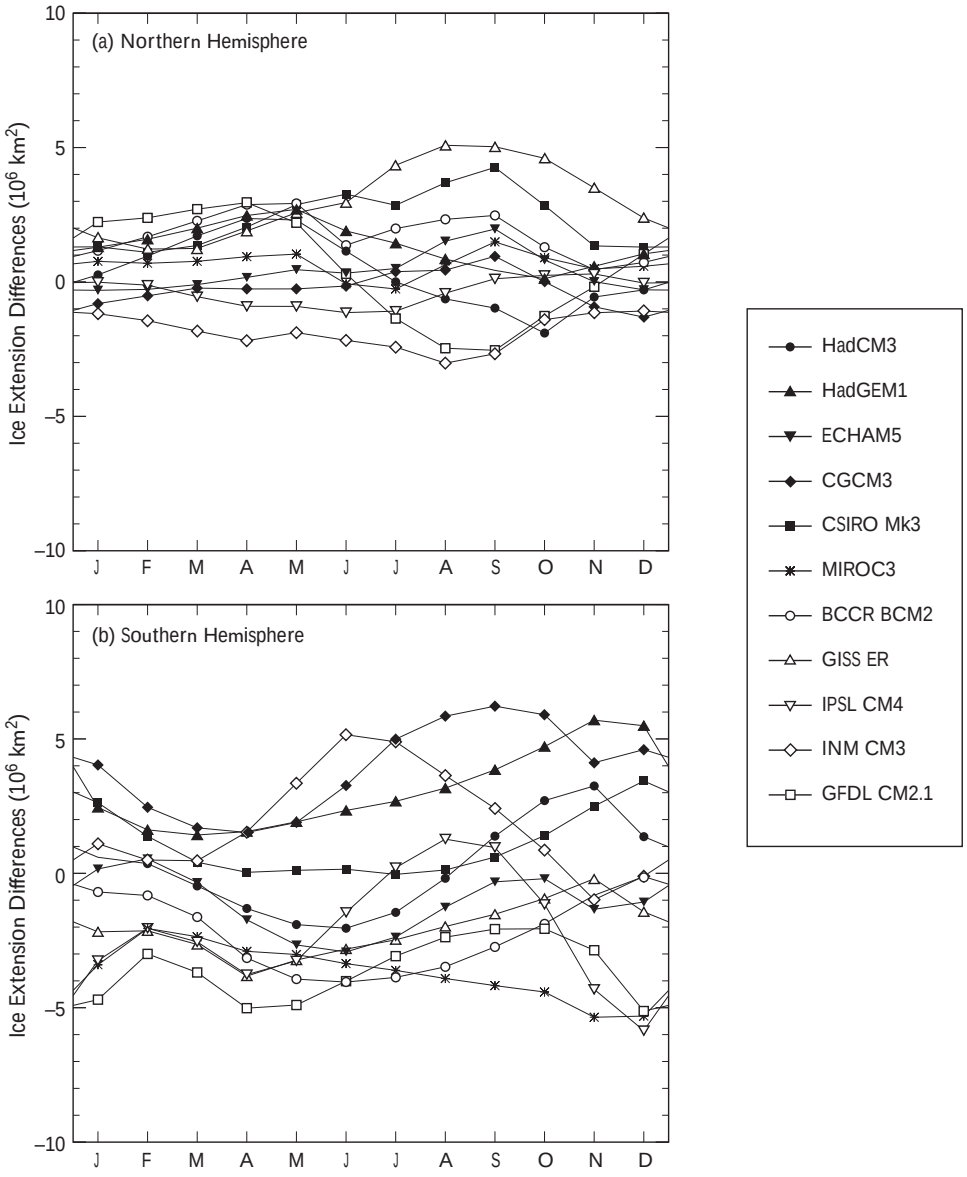


Figure 5.8. Difference Between Modeled 1979 to 2004 Monthly Average Sea-Ice Extents and Satellite-Based Observations (modeled minus observed).

Data are shown for each of 11 major GCMs for both (a) Northern Hemisphere and (b) Southern Hemisphere. [From Fig. 4 in C.L. Parkinson, K.Y. Vinnnikov, and D.J. Cavalieri 2006: Correction to evaluation of the simulation of the annual cycle of Arctic and Antarctic. *J. Geophysical Research*, **111**, C07012. Reproduced by permission of the American Geophysical Union (AGU).]

5.2.4 Ocean Structure and Circulation

Unlike the atmosphere, the amount of observational data available to evaluate ocean simulations is very limited for long time periods. Nevertheless, sufficient data exist to identify a set of ocean characteristics or metrics to evaluate ocean models for their climate simulation properties. The most important is sea-surface temperature, but other quantities that serve as good indicators of ocean realism in climate models are ocean heat uptake, meridional overturning and ventilation, sea-level variability, and global sea-level rise.

5.2.4.1 Sea-Surface Temperature

Sea-surface temperature (SST) plays a critical role in determining climate and the predictability of climate changes. Because of interactions in atmospheric and ocean circulations at the surface, errors in SSTs typically originate with deficiencies in both atmospheric and ocean model components. In general, more recent model versions show improvement over previous models when simulated SST fields are compared to observations. Figure 5.9 (Delworth et al. 2006) shows comparisons of simulated and observed mean SST fields of both the older GFDL CM2.0 and newer CM2.1 averaged over a 100-year period. The new model reduced a cold bias in the Northern Hemisphere from earlier simulations, resulting in both a more-realistic representation of atmospheric wind stress at the ocean surface and a modified treatment of subgrid-scale oceanic mixing. The CCSM3.0 model's improved SST simulation over CCSM2.0 results mainly from changes in representing processes associated with the mixed layer of upper ocean waters (Danabasoglu et al. 2006).

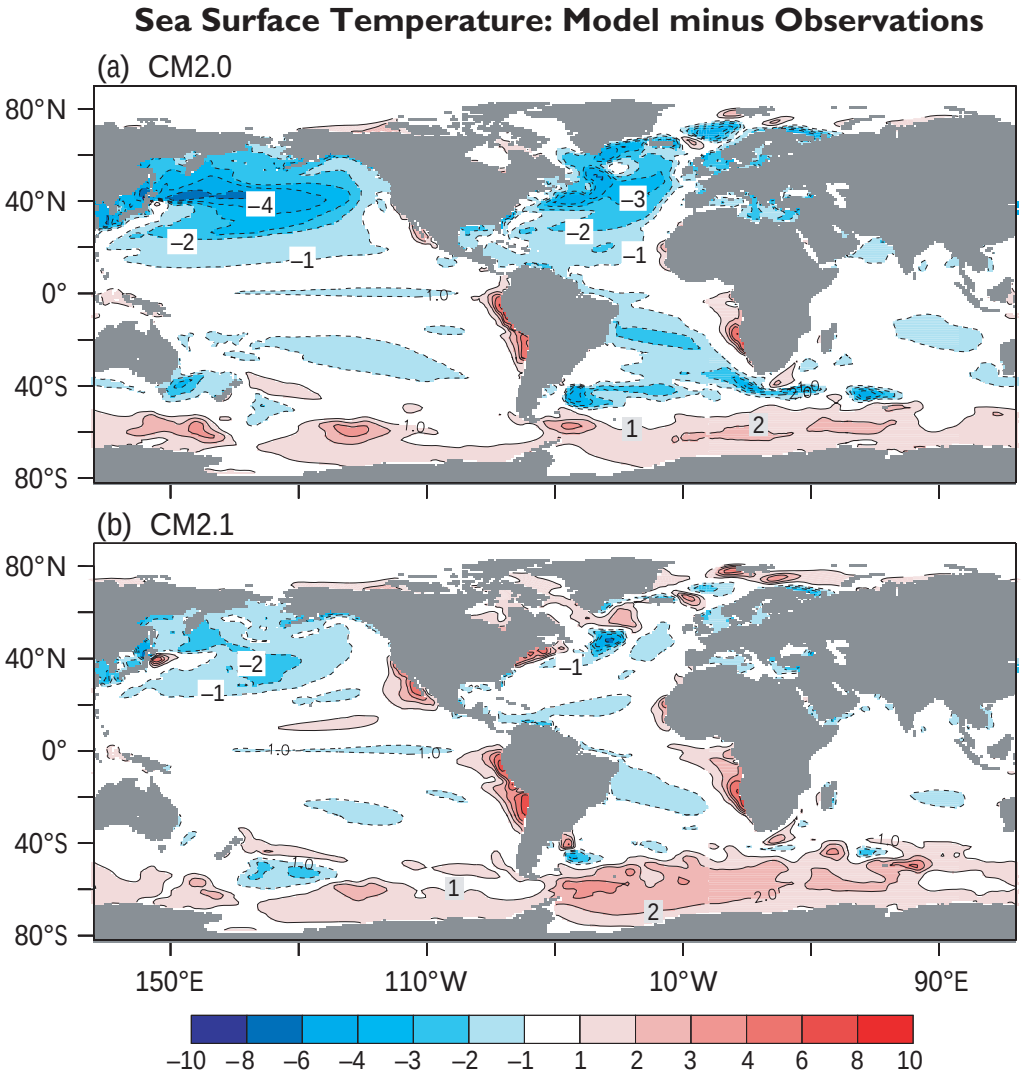


Figure 5.9. Maps of Simulation Errors in Annual Mean SST.

Units are Kelvin (K). Errors are computed as model minus observations from Reynolds SST data (provided by NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, from their Web site, www.cdc.noaa.gov). (a) CM2.0 (using model years 101 to 200). (b) CM2.1 (using model years 101 to 200). Contour interval is 1 K, except for no shading of values between 1 K and +1 K. [Images from T.L. Delworth et al. 2006: GFDL's CM2 global coupled climate models. Part 1: Formulation and simulation characteristics. *J. Climate*, **19**, 643–684. Reproduced by permission of the American Meteorological Society.]

In addition to SST mean values, 20th Century trends of SST changes also are significant for model evaluation, since ocean SST contributes the dominant signal to the observed global surface temperature trend. An intermodel comparison of 50-year tropical SST trends is shown in Fig. 5.10. Trends range from a low of 0.1°C/50 yrs to a high of about 0.6°C/50 yrs, with the observational trend estimate given as about 0.43°C/50 yrs. The figure also shows some randomness within a group of simulations run by the same model. For example, the two different GFDL model versions discussed above were each run for multiple realizations of the 20th Century. CM 2.0 simulations are noted by GFDL201, GFDL202, and GFDL203, and CM 2.1 simulations are noted by GFDL211, GFDL212, and GFDL213.

5.2.4.2 Meridional Overturning Circulation and Ventilation

The planetary-scale circulation transporting heat and freshwater throughout global oceans is referred to as global thermohaline circulation. The Atlantic portion is called the Atlantic meridional overturning circulation (AMOC). Tropical and warm waters flow northward via the Gulf Stream and North Atlantic Current. Southward flow occurs when water is subducted in regions around Labrador and Greenland; surface waters freshen, become denser, and flow down the slope to deeper depths. Similar processes occur at locations in the Southern Ocean. "Ventilation" is the name given to the process by which these dense surface waters are carried into the ocean interior. An important climate parameter is the rate at which this process occurs. The pattern of circulation may weaken,

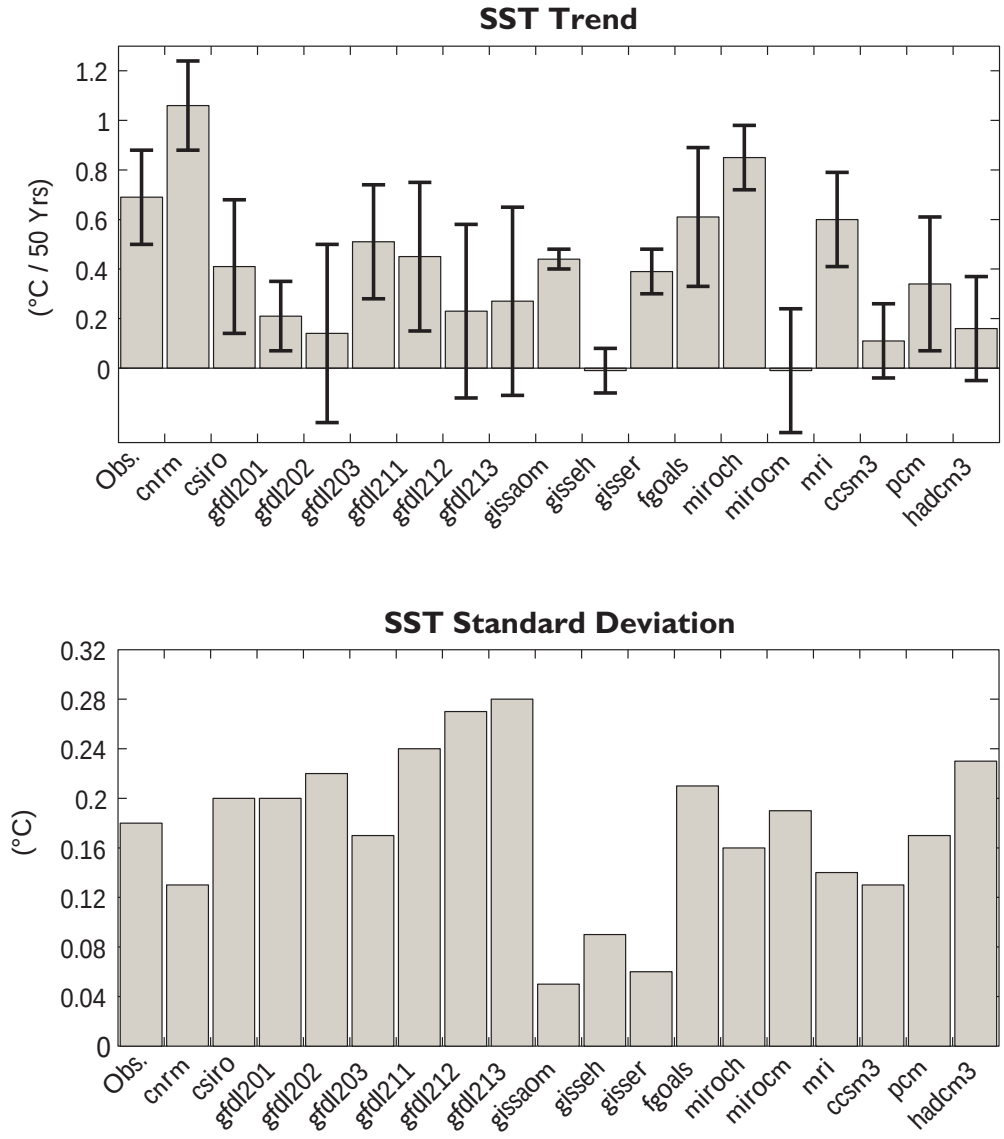


Figure 5.10. Trends and Standard Deviations of Tropical SST Between 1950 and 1999.

Observations are shown by the leftmost bar in each figure. All others are model results. Error bars show 95% significance levels for trends. [Images from Fig. 9 in D. Zhang and M.J. McPhaden 2006: Decadal variability of the shallow Pacific meridional overturning circulation: Relation to tropical sea-surface temperatures in observations and climate change models. *Ocean Modelling*, **15**, 250–273. Used with permission from Elsevier.]

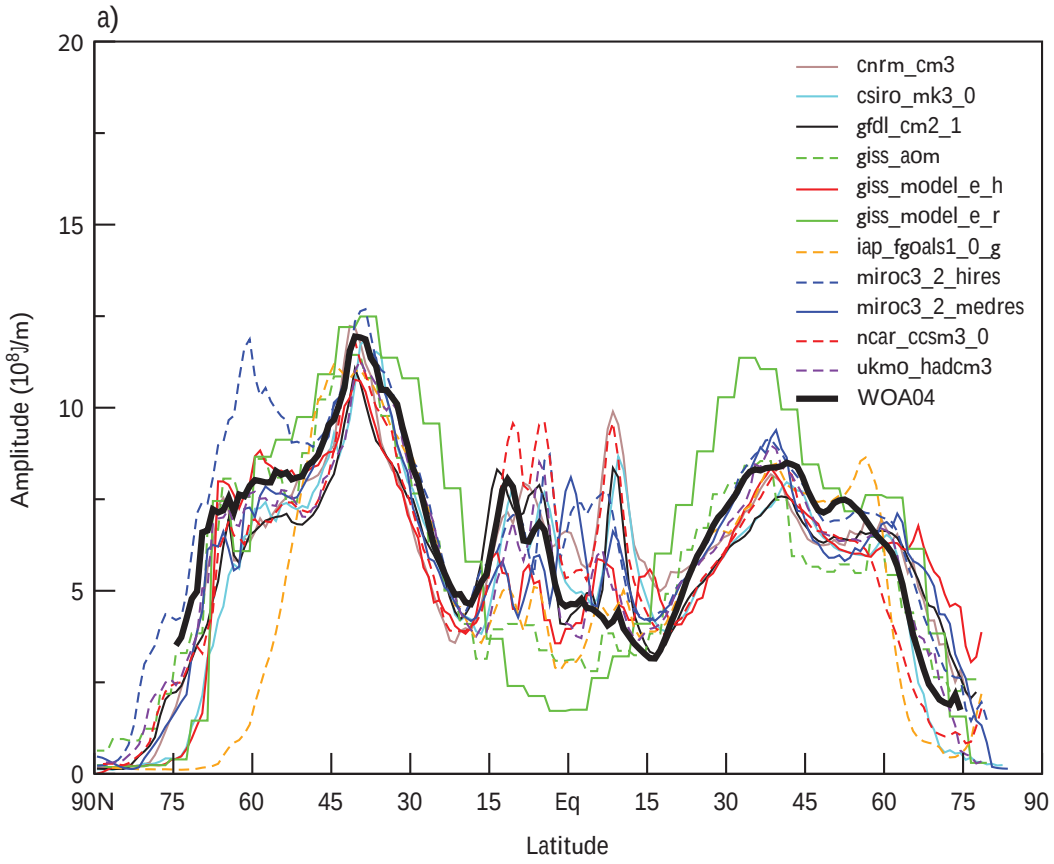
affecting the climate in the region surrounding the North Atlantic. Schmittner, Latif, and Schneider (2005) examined a small ensemble set of simulations to quantify uncertainty in model representation of 20th Century AMOC transports. To make their estimate, they evaluated global temperature, global salinity, pycnocline depth, surface temperature, surface salinity in the Atlantic (SST, SSS), and the overturning calculations at three Atlantic locations. Their results suggest that temperature is simulated most successfully on a large scale and that the overturning transports at 24°N are close (~18 Sv) to observed measurements (~15.8 Sv). However, the maximum mean overturning transports in these models are too high, between 21.2 and 31.7 Sv, when compared to the observed value (17.7 Sv). Several other CMIP3 models underestimated maximum transport. The authors do not attempt to explain why models are different from each other and from observations.

Another aspect of planetary-scale ocean circulation of interest is transport of mass by the Antarctic Circumpolar Current 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. Observed mean transport is around 135 Sv (Cunningham et al. 2003). Russell, Stouffer, and Dixon (2006, 2007) estimate passage flow for a subset of climate models. Simulated mean values show a wide range. For example, GFDL and GISS-EH models do fairly well in reproducing the observed average transport with values between 113 and 175 Sv. Once again, 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 provided by the atmosphere correlate with how well the transport reflects observed values.

5.2.4.3 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 exchange between the tropics and the extratropical regions 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. 2006a; Schmidt et al. 2006). The current models have significantly improved over the last generation in the Northern Hemisphere. Comparisons of simulated values to 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. In coupled ocean-atmosphere simulations, erroneous ocean heat transport is compensated by changes in atmospheric heat transport that give a more realistic total heat transport (Covey and Thompson 1989).

Heat Content. The global mean mass-weighted ocean temperature is called the ocean's heat content. Its time evolution is centrally important in determining how realistically the models reproduce heat uptake. The seasonal cycle and longer-term trends of heat content provide useful model metrics, although the seasonal cycle does not affect the deep ocean. An evaluation of temporally evolving ocean-heat content in the CMIP3 suite of climate models shows the models' abilities to simulate the zonally integrated annual and semiannual cycle in heat content. In the middle latitudes (Gleckler, Sperber, and AchutaRao 2006), the models do a reasonable job, although a broad spread of values is apparent for tropical and polar regions. This analysis showed that the models replicate the annual cycle's dominant amplitude along with its phasing in the midlatitudes (Figs. 5.11 a–b and 5.12 a–f). At high latitudes, comparisons with observations are not as consistent. Although the annual cycle and global trend are reproduced, model analyses (e.g., Hansen et al. 2005a, b) show they do not simulate decadal changes in estimates made from observations (Levitus et al. 2001). Part of the difficulty of comparisons at high latitudes and long periods is the paucity of observational data (Gregory et al. 2004).



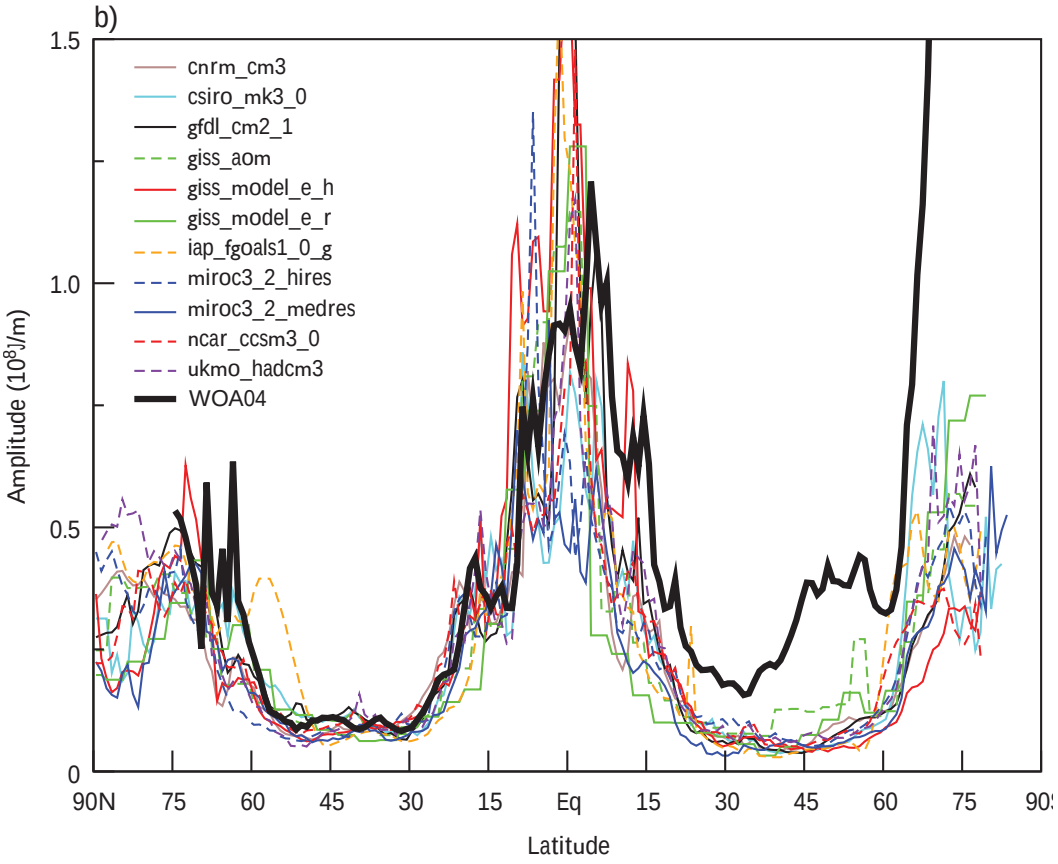


Figure 5.11a–b. Observed and Simulated Zonally Integrated Ocean Heat Content (0–250 m).

Observations are represented by the curve labeled "WOA04." All other curves are model results. (a) annual cycle amplitude (10^8 J/m^2) and (b) semiannual/annual (A2/A1). [From Fig. 1 in P.J. Gleckler, K.R. Sperber, and K. AchutaRao 2006: Annual cycle of global ocean heat content: Observed and simulated. J. *Geophysical Research*, **111**, C06008. Reproduced by permission of the American Geophysical Union (AGU).]

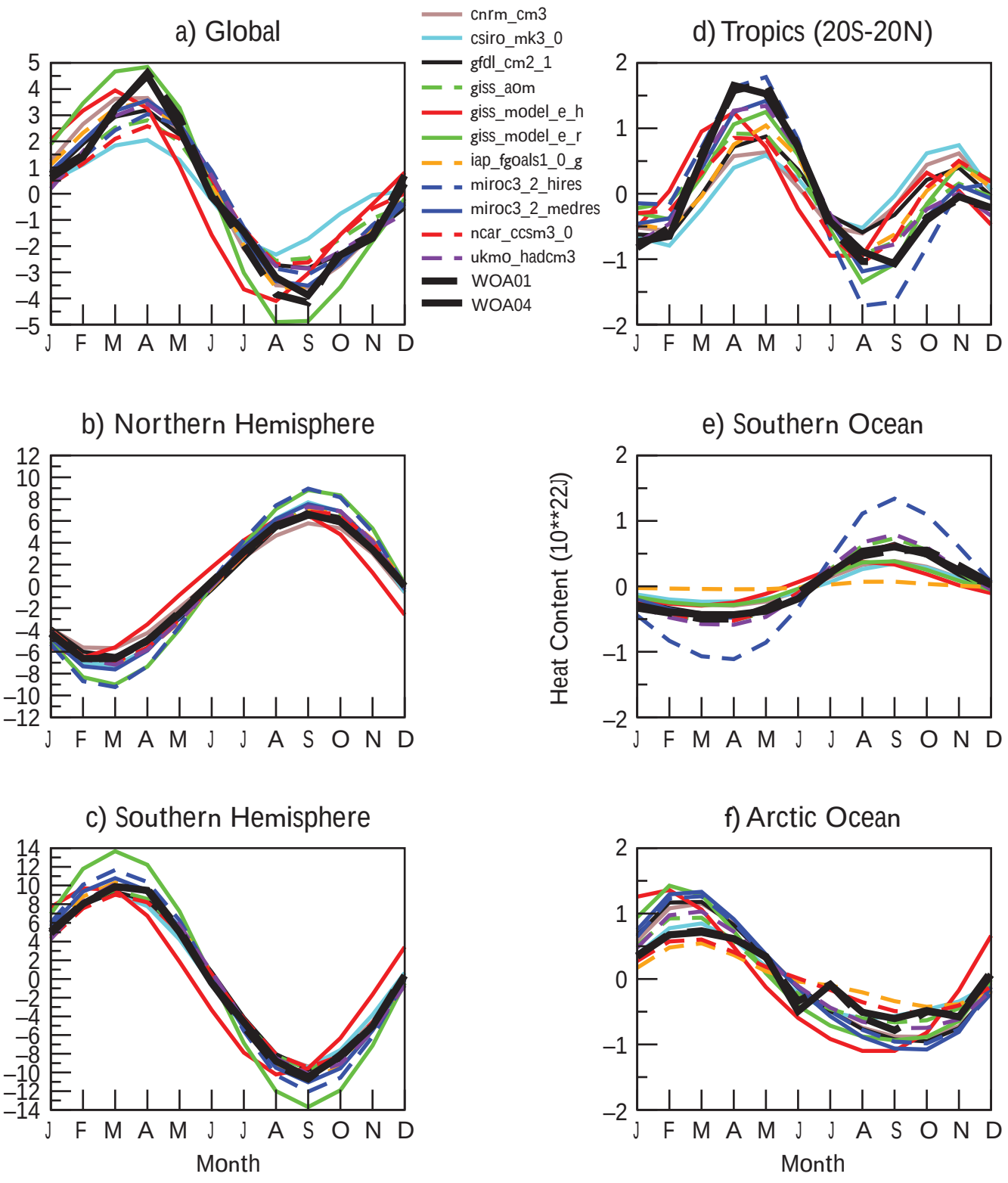


Figure 5.12a–f. Annual Cycle of Observed and Simulated Basin Average Global Ocean Heat Content (0–250 m).

Observations are represented by the curves labeled "WOA01" and "WOA04." Units are 10^{22} J. Arctic Ocean is defined as north of 60°N, and Southern Ocean is south of 60°S. [From Fig. 3 in P.J. Gleckler, K.R. Sperber, and K. AchutaRao 2006: Annual cycle of global ocean heat content: Observed and simulated. *J. Geophysical Research*, **111**, C06008. Reproduced by permission of the American Geophysical Union (AGU).]

5.2.5 Global Mean Sea-Level Rise

Two separate physical processes contribute to sea-level rising: (1) ocean thermal expansion from an increase in ocean heat uptake (steric component) and (2) addition of freshwater from precipitation, continental ice melt, and river runoff (eustatic component). Various ocean models handle freshwater fluxes in different ways. With the addition of a free surface in the current generation of ocean models, freshwater flux into oceans can be included directly (Griffies et al. 2001). The freshwater contribution is computed in quantities estimated by the climate model's atmosphere and ice-sheet components (e.g., Church, White, and Arblaster 2005; Gregory, Lowe, and Tett 2006). In general, 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 separate component is within the observed values' uncertainty. The reason for this is an open research question and may relate either to observational sampling or to incorrectly accounting for all eustatic contributions. The steric component to global mean sea-level rise is estimated at 0.40 ± 0.05 mm/yr from observations (Antonov, Levitus, and Boyer 2005). Models simulate a similar but somewhat smaller rise (Gregory, Lowe, and Tett 2006; Meehl et al. 2005). Significant differences also occur in the magnitudes of decadal variability between observed and simulated sea level. Progress is being made, however, over the previous generation of climate models. When atmospheric effects from volcanic eruptions are included, for example, current-generation ocean models capture the volcanoes' observed impact (a decrease in the global mean sea level). Figure 5.13 from Church, White, and Arblaster (2005) gives an example of a few models and their detrended estimate of the historic global mean sea level. It shows the influence of including additional atmospheric forcing agents in changing the ocean's steric height.

5.3 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 between 1983 and 2003 (www.flooddamagedata.org). Losses from the 1988 drought were estimated at $40 billion and the 2002 drought at $11 billion. Heat waves in 1995 resulted in 739 additional deaths in Chicago alone (Whitman et al. 1997). A large component of overall climate change impacts probably will arise from changes in the intensity and frequency of extreme events.

Modeling of extreme events poses special challenges since they are, by definition, rare. Although the intensity and frequency of extreme events are modulated by ocean and land surface state and by trends in the mean climate state, internal atmospheric variability plays a very large role, and the most extreme events arise from chance confluence of unlikely conditions. The very rarity of extreme events makes statistical evaluation of model performance less robust than for mean climate. For example, in evaluating a model's ability to simulate heat waves as intense as that in 1995, only a few episodes in the entire 20th Century approach or exceed that intensity (Kunkel et al. 1996). For such rare events, estimates of the real risk are highly uncertain, varying from once every 30 years to once every 100 years or more. Thus, a model that simulates these occurrences at a frequency of once every 30 years may be performing adequately, but its performance cannot be distinguished from that of the model that simulates a frequency of once every 100 years.

Although it might be expected that a change in mean climate conditions will apply equally to changes in 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 1930s, yet winters during that decade averaged third warmest since 1890; in fact, no significant correlation is shown between the number of records and U.S. wintertime temperature (Vavrus et al. 2006). Thus, the severest cold air outbreaks in the past do not necessarily coincide with cold winters. Another examination of model data showed that future changes in extreme temperatures differ from changes in 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.

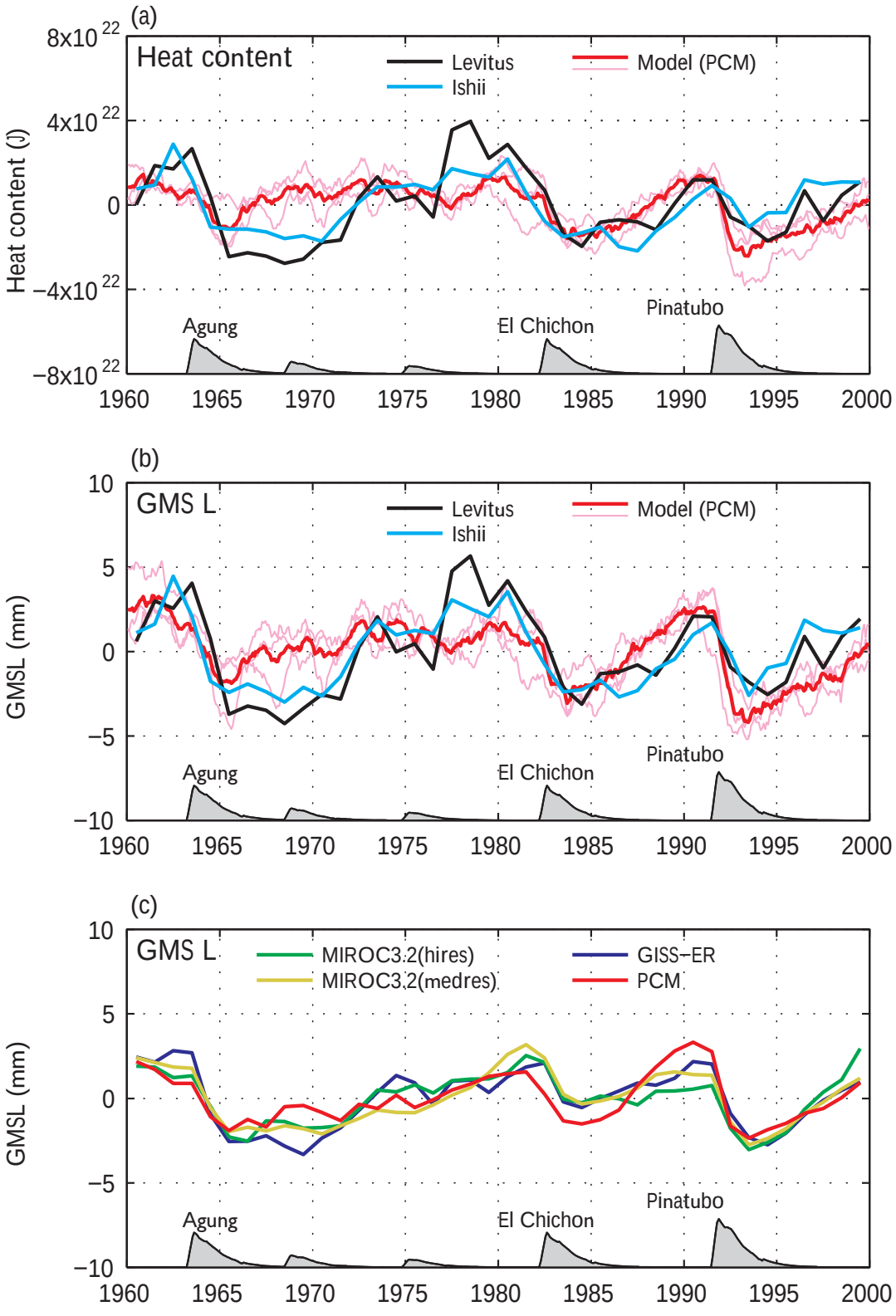


Figure 5.13. Observed and Modeled Global Ocean Heat Content (GOHC) and Global Mean Sea Level (GMSL) for 1960 to 2000.

The response to volcanic forcing, as indicated by differences between pairs of PCM simulations for GOHC (a) and GMSL (b) is shown for the ensemble mean (bold line) and the three ensemble members (light lines). Observational estimates 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 to 2000. For c, the ensemble mean (full-depth) GMSL for GISS-ER, MIROC3.2(hires), MIROC3.2(medres), and PCM models (after subtracting a quadratic) are shown. [From Fig. 2 in J.A. Church, N.J. White, and M. Arblaster 2005: Significant decadal-scale impact volcanic eruptions on sea level and ocean heat content. *Nature*, **438**(7064), 74–77. Used with permission from Nature Publishing Group.]

Evaluation of model performance with respect to extremes is hampered by incomplete data on historical frequency and severity of extremes. Frich et al. (2002) analyzed ten indicators of climate extremes and presented global results. However, many areas were missing due to lack of suitable station data, particularly in the tropics. Using some of these indices for comparisons between models and observations has become common. 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, averaging station data over areas representing a grid cell is necessary.

Several approaches are used to evaluate model performance for 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, every 6 years in New York City, every 13 years in Chicago, and every 200 years in Phoenix. A useful metric would be the extent to which a model is able to reproduce 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 global trend is an increase in the frequency of heavy precipitation, particularly during the last 20 to 30 years of the 20th Century. This trend is significant at the 95% confidence level for the period 1979 to 2003 and at the 99% confidence level for the period 1951 to 2003 (Trenberth et al. 2007). Another notable observed trend is an increase in the length of the frost-free season.

In some key respects, model simulation of temperature extremes probably is less challenging than simulating precipitation extremes, in large part due to the scales of these phenomena. The typical heat wave or cold wave covers a relatively large region, on 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 simulate directly the major processes causing temperature extremes while heavy precipitation is sensitive to parameterization of subgrid-scale processes, particularly convection (Chapter 2; Emori and Brown 2005; Iorio et al. 2004).

5.3.1 Droughts and Excessive Rainfall Leading to Floods

Recent analysis indicates 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, Brown, and Christidis 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 relationship between droughts and variability was covered above in Section 5.2.2.3 Monsoons.

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/day in Dai 2006) or percentile (e.g., 4th-largest precipitation event equivalent to 99th percentile of 365 daily values as in Emori et al. 2005). Recent studies of model simulations produced for CMIP3 provide information on the performance of the latest model generation.

Models generally tend to underestimate very heavy precipitation. This is shown in a comparison between satellite (TRMM) estimates of daily precipitation and model-simulated values within the 50°S–50°N latitude belt (Dai 2006). TRMM observations derive 7% of total precipitation from very heavy rainfall of 50 mm or more per day, in contrast to only 0 to 2% for the models. For the frequency of very heavy precipitation of 50 mm or more per day, TRMM data show a frequency of 0.35% (about once every 300 days), whereas it is 0.02 to 0.11% (once every 900 to 5000 days) for the models. A global analysis of model simulations showed that models produced too little precipitation in events exceeding 10 mm/day (Sun et al. 2006). Examining how many days it takes to accumu-

late two-thirds of annual precipitation, 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 indicates generally good agreement with observed annual maximum 5-day precipitation amounts 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 global models' low spatial resolution. Experiments with individual models show that increasing resolution improves the simulation of heavy events. For example, the fourth-largest precipitation event in a model simulation with a resolution of about 300 km averaged 40 mm over the conterminous United States, compared to an observed value of about 80 mm. When the resolution was increased to 75 km and 50 km, the fourth-largest event was still smaller than observed but by a much smaller amount (Iorio et al. 2004). A second important factor 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 must be indirectly represented in models (Chapter 2). One experiment showed that changes to this representation improve model performance and, when combined with high resolution of about 1.1° latitude, can produce quite-accurate simulations of the fourth-largest precipitation event on a globally averaged basis (Emori et al. 2005). Another experiment found that the use of a cloud-resolving model imbedded in a global model eliminated underestimation of heavy events (Iorio et al. 2004). A cloud-resolving model eliminates the need for convection parameterization but is very expensive to run. These sets of experiments indicate that the problem of heavy-event underestimation may be reduced significantly in future as increases in computer power allow simulations at higher spatial resolution and perhaps eventually the use of cloud-resolving models.

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

Most studies of observed precipitation extremes suggest that they 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 toward 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 CGCMs simulate an increasing trend in heavy rain over the tropical ocean.

5.3.2 Heat and Cold Waves

Analyses of simulations for IPCC AR4 by seven climate models indicate that they reproduce the primary features of cold air outbreaks (CAOs), with respect to location and magnitude (Vavrus et al. 2006). In the analyses, 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 two standard deviations below the gridpoint's winter mean temperature. Maximum frequencies of about four CAO days per winter are simulated over western North America and Europe, while minimal occurrences of less than one day per winter exist over the Arctic, northern Africa, and parts of the North Pacific. GCMs generally are accurate in their simulation of primary fea-

tures, with high pattern correlation to observations and maximum number of days meeting 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, models simulate a frequency of about 4 CAO days per year, in general agreement with the observed values of 3 to 4 days. Models underestimate frequency in the southeastern United States (mean simulated values range from 0.5 to 2 days vs 2 to 2.5 days in observations). This regional bias occurs in all models and reflects the inability of GCMs to penetrate Arctic air masses far enough southeastward over North America.

CMIP3 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). The simulations indicate that this is in general agreement with observations, except that there is no observed trend in heat waves. The modeled spatial patterns generally have larger positive trends in western North America than in eastern sections. For the United States, this is in qualitative agreement with observations showing that decreases in frost-free season and frost days are largest in the western United States (Kunkel et al. 2004; Easterling 2002).

Analysis of individual models provides a more detailed picture of model performance. In a simulation from PCM (Meehl, Tebaldi, and Nychka 2004), the largest trends for decreasing frost days occur in the western and southwestern United States (values greater than –2 days per decade). Trends near zero in the upper Midwest and northeastern United States show good agreement with observations. The biggest discrepancy between model and observations is over parts of the southeastern United States, where the model shows trends for decreasing frost days and observations show slight increases. This is thought to be a partial consequence of two large El Niño events in observations during this time period (1982–1983 and 1997–1998) when anomalously cool and wet conditions occurred over the southeastern United States and contributed to slight increases of frost days. The model's ensemble mean averages out effects from individual El Niño events, and thus frost-day trends reflect a more general response to forcings that occurred during the latter part of the 20th Century. An analysis of short-duration heat waves simulated by 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 southeastern United States for both model simulation and observations. The overall spatial pattern of heatwave intensity in the model matched closely with the observed pattern. In a four-member ensemble of simulations from HadCM3 (Christidis et al. 2005), the model showed a rather uniform pattern of increases in the warmest night for 1950 to 1999. Observations also show a global mean increase, but with considerable regional variations. In North America, 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.

Analysis of the magnitude of temperature extremes for California in a regional climate model simulation (Bell, Sloan, and Snyder 2004) shows mixed results. The hottest maximum in the model is 4°C less than observations, while the coldest minimum 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, RCM results are still somewhat deficient, perhaps reflecting the study region's very complex topography.

Models display some capability to simulate extreme temperature and precipitation events, but there are differences from observed characteristics. Models typically produce global increases in extreme precipitation and severe drought and decreases in extreme minimum temperatures and frost days, in general agreement with observations. Models have 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 the climate system's stochastic variability, assessing the significance of these discrepancies is difficult.