



Assessment of surface air temperature over the Arctic Ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations

Jiping Liu,¹ Zhanhai Zhang,² Yongyun Hu,³ Liqi Chen,⁴ Yongjiu Dai,⁵ and Xiaobo Ren⁶

Received 12 September 2007; revised 3 December 2007; accepted 29 January 2008; published 21 May 2008.

[1] The surface air temperature (SAT) over the Arctic Ocean in reanalyses and global climate model simulations was assessed using the International Arctic Buoy Programme/Polar Exchange at the Sea Surface (IABP/POLES) observations for the period 1979–1999. The reanalyses, including the National Centers for Environmental Prediction Reanalysis II (NCEP2) and European Centre for Medium-Range Weather Forecast 40-year Reanalysis (ERA40), show encouraging agreement with the IABP/POLES observations, although some spatiotemporal discrepancies are noteworthy. The reanalyses have warm annual mean biases and underestimate the observed interannual SAT variability in summer. Additionally, NCEP2 shows an excessive warming trend. Most model simulations (coordinated by the International Panel on Climate Change for its Fourth Assessment Report) reproduce the annual mean, seasonal cycle, and trend of the observed SAT reasonably well, particularly the multi-model ensemble mean. However, large discrepancies are found. Some models have the annual mean SAT biases far exceeding the standard deviation of the observed interannual SAT variability and the across-model standard deviation. Spatially, the largest inter-model variance of the annual mean SAT is found over the North Pole, Greenland Sea, Barents Sea and Baffin Bay. Seasonally, a large spread of the simulated SAT among the models is found in winter. The models show interannual variability and decadal trend of various amplitudes, and can not capture the observed dominant SAT mode variability and cooling trend in winter. Further discussions of the possible attributions to the identified SAT errors for some models suggest that the model's performance in the sea ice simulation is an important factor.

Citation: Liu, J., Z. Zhang, Y. Hu, L. Chen, Y. Dai, and X. Ren (2008), Assessment of surface air temperature over the Arctic Ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations, *J. Geophys. Res.*, 113, D10105, doi:10.1029/2007JD009380.

1. Introduction

[2] Surface air temperature (SAT) change is a primary measure of global climate change, since it integrates changes in the surface energy budget and general circulation [e.g., Hansen *et al.*, 1999; Jones and Mann, 2004]. The SAT over the Arctic Ocean is often considered a harbinger of global climate change, since global climate models predict that the impacts of greenhouse warming will be amplified

there due to the decline of the Arctic sea ice, particularly during the cold season (the opening of sea ice promotes large heat fluxes from the ocean to the atmosphere) [e.g., Serreze *et al.*, 2000; Walsh *et al.*, 2002; Serreze and Francis, 2006]. The main energy balance over the Arctic Ocean is between radiative energy balance and advective transport of heat into the region, which depend strongly on the SAT. As the ocean surface cools, sea ice forms, which inhibits heat loss from the ocean, allowing the SAT to fall rapidly below the freezing point and reducing the radiative heat loss. As the SAT rises above the freezing point, sea ice melts, which enhances air-sea heat exchanges. Thus, sea ice variations and associated feedbacks are highly sensitive to the SAT [e.g., Barry *et al.*, 1993].

[3] Earth's climate is changing, with the global temperature having risen 0.6°C in the past three decades [e.g., Hansen *et al.*, 2006]. Recent warming is not in doubt and appears to extend into the Arctic Ocean [e.g., Serreze *et al.*, 2000]. However, because of the lack of long-term observations, it is difficult to assess quantitatively how the SAT has been changing in the Arctic Ocean. Recently, some efforts have tried to remedy this situation by making use of satellite

¹State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.

²Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, China.

³Department of Atmospheric Sciences, School of Physics, Peking University, Beijing, China.

⁴Key Laboratory of Global Change and Marine-Atmospheric Chemistry, Third Institute of Oceanography, SOA, Xiamen, China.

⁵School of Geography, Beijing Normal University, Beijing, China.

⁶Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.

thermal infrared measurements [e.g., *Comiso*, 2003], and buoys, drifting and meteorological stations [e.g., *Martin and Munoz*, 1997; *Rigor et al.*, 2000]. The resulting data sets from such efforts are big improvements over previous records.

[4] Using surface temperature derived from satellite thermal infrared measurements for cloud-free conditions during 1981–2001, *Comiso* [2003] showed a warming trend of 0.33°C/decade over the Arctic sea ice. Using the compiled buoys, drifting and meteorological stations, *Rigor et al.* [2000] found significant warming trends in the Arctic Ocean during winter and spring, with values as high as 2°C/decade in the eastern Arctic Ocean during spring. Associated with the warming, the Arctic sea ice has declined dramatically for the past three decades [e.g., *Parkinson and Cavalieri*, 2002; *Liu et al.*, 2004; *Stroeve et al.*, 2005; *Serreze et al.*, 2007], which directly and indirectly causes wide-ranging impacts [e.g., *Arctic Climate Impact Assessment (ACIA)*, 2004]. Due to the significant decline of the Arctic sea ice, rises of the SAT in response to the increase of greenhouse gases are expected to be pronounced in the Arctic Ocean [e.g., *Meehl and Washington*, 1990; *IPCC*, 2001; *ACIA*, 2004; *Stroeve et al.*, 2005].

[5] Nevertheless, the complicated atmosphere-sea ice-ocean interactions make the projection of future climate change particularly challenging in the Arctic Ocean. Because of the significant warming trend, and the potential role of the Arctic Ocean in rapid climate change, it is important to 1) gauge the accuracy of the SAT over the Arctic Ocean in the most widely used reanalyses so that we can be assured whether or not they are of sufficient quality to assess Arctic climate variability, and 2) evaluate the simulations of the SAT over the Arctic Ocean in coupled global climate models (CGCMs) so that the representation of physical processes in the Arctic Ocean can be improved and uncertainties in the projections of future climate change can be reduced.

[6] In this paper, we assessed how well the current day state-of-the-art reanalyses and CGCMs are reproducing the annual mean, seasonal cycle, variability and trend of the observed SAT over the Arctic Ocean for the late 20th century (where sea ice changes are largest), providing information to climate modelers that can help to further improve physical parameterizations and numerical methods related to modeling the climate of the Arctic Ocean.

2. Data

[7] To assess the reanalyses and CGCM simulations of SAT over the Arctic Ocean, we used the International Arctic Buoy Programme/Polar Exchange at the Sea Surface (IABP/POLES) SAT observations, which is derived from extensive buoys, manned drifting stations, and coastal stations deployed in the Arctic Ocean. Compared to observations from the Russian North Pole drift stations, the IABP/POLES SAT data set has higher temporal correlations and lower root mean square errors than previous SAT data sets, and provides better temperature estimates in the Arctic Ocean and marginal ice zones (see Table 3 in *Rigor et al.* [2000]). The IABP/POLES provides SAT on a 100 km equal area scalable earth [EASE, *Armstrong and Brodzik*, 1995] grid for the Arctic during 1979–2004.

[8] Using the IABP/POLES observations, we evaluated two numerical weather prediction reanalyses (National Centers for Environmental Prediction Reanalysis II (NCEP2), and European Centre for Medium-Range Weather Forecast 40-year Reanalysis (ERA40)), and the most comprehensive set of CGCM simulations coordinated by the International Panel on Climate Change (IPCC) for its Fourth Assessment Report (AR4). To make comparisons of the reanalyses and IPCC AR4 model simulations consistent with IABP/POLES, a Cressman interpolation [*Cressman*, 1959] is used to translate the reanalyses and model simulations onto the IABP/POLES EASE 100km grid.

[9] NCEP2 is on a T62 gaussian grid with 192 (longitude) by 94 (latitude) points, from 1979 to present. NCEP2 fixed some known processing errors in the original NCEP/NCAR (National Center for Atmospheric Research) reanalysis and improved the parameterizations of some physical processes. Relevant to the northern high latitudes, changes include: updated sea ice boundary conditions, new snow cover analysis scheme, fixed problems in humidity diffusion and ocean albedo, and better parameterizations for the planetary boundary layer, shortwave radiation, convection, cloud-top cooling and cloud-tuning coefficients [see *Kanamitsu et al.* [2002], for details]. The spatial resolution of ERA40 is 2.5 by 2.5 degree, covering the period 1957–2002. ERA40 also corrected some problems in the previous ERA15 such as the severe cold bias in the surface and near-surface temperatures during winter and spring in the northern high latitude through improved parameterizations (i.e., soil freezing, surface snow cover and sea ice, see <http://www.ecmwf.int/research/era/> for more on the merits of the ERA40).

[10] More than a dozen climate modeling groups worldwide develop CGCMs, whose ability to simulate the current climate has improved measurably over the past two decades [e.g., *Meehl et al.*, 2005]. In support of the IPCC AR4, these modeling groups performed the most comprehensive suite of coordinated experiments during 2004–2005, including the climate of the 20th century simulations with observed anthropogenic and/or natural forcing and the 21st century simulations with the prescribed IPCC SRES scenarios [*IPCC*, 2001]. The resulting multi-model data set is a unique and valuable resource that will enable international scientists to assess models' performance in response to a variety of forcings for 20th and 21st century climate and climate change. Our assessment is based on the monthly mean outputs of twenty-three CGCMs achieved by the Program for Climate Model Diagnosis and Intercomparison. Only one simulation of the climate of the 20th century (20C3M) for each model is included in this assessment, because some modeling groups do not provide multi-member ensembles. As shown in Table 1, the IPCC AR4 models have different resolutions and coupling strategies. Since the various models differ in their parameterizations of physical processes in the atmosphere, ocean, and sea ice components, the simulations are obviously model-dependent (more detailed information can be found at http://www.pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php). Because the IABP/POLES observations start from the year 1979, and some model simulations

Table 1. List of IPCC AR4 Models Used in This Study

Models	Country	Resolution	Flux Adjustment
BCC-CM1	China	Atm: T63; Ocn: T63	Heat/momentum
BCCR-BCM2.0	Norway	Atm: T63; Ocn: 1.5 lat × 1.5 lon	None
CCCMA-CGCM3.1m	Canada	Atm: T47; Ocn: 1.85 lat × 1.85 lon	Heat/water
CCCMA-CGCM3.1h	Canada	Atm: T63; Ocn: 1.4 lat × 1.4 lon	Heat/water
CNRM-CM3	France	Atm: T42; Ocn: 2 lat × 2 lon	None
CSIRO-MK3.0	Australia	Atm: T63; Ocn: 0.84 lat × 1.875 lon	None
GFDL-CM2.0	United States	Atm: 2 lat × 2.5 lon; Ocn: 1 lat × 1 lon	None
GFDL-CM2.1	United States	Atm: 2 lat × 2.5 lon; Ocn: 1 lat × 1 lon	None
GISS-AOM	United States	Atm: 3 lat × 4 lon; Ocn: 3 lat × 4 lon	None
GISS-EH	United States	Atm: 4 lat × 5 lon; Ocn: 1 lat × 1 lon	None
GISS-ER	United States	Atm: 4 lat × 5 lon; Ocn: 4 lat × 5 lon	None
IAP-FGOALS1.0g	China	Atm: 2.8 lat × 2.8 lon; Ocn: 1 lat × 1 lon	None
INM-CM3.0	Russia	Atm: 4 lat × 5 lon; Ocn: 2 lat × 2.5 lon	Water
IPSL-CM4	France	Atm: 2.5 lat × 3.75 lon; Ocn: 2 lat × 2 lon	None
MIROC3.2h	Japan	Atm: T106; Ocn: 0.1875 lat × 0.28125 lon	None
MIROC3.2m	Japan	Atm: T42; Ocn: 1.4 lat × 1.4 lon	None
MIUB-ECHOg	Germany	Atm: T30; Ocn: T42	Heat and water
MPI-ECHAM5	Germany	Atm: T63; Ocn: 1.5 lat × 1.5 lon	None
MRI-CGCM2.3.2a	Japan	Atm: T42; Ocn: 2 lat × 2.5 lon	Heat/water/momentum
NCAR-CCSM3	United States	Atm: T85; Ocn: 1 lat × 1 lon	None
NCAR-PCM1	United States	Atm: T42; Ocn: 1 lat × 1 lon	None
UKMO-HadCM3	United Kingdom	Atm: 2.5 lat × 3.75 lon; Ocn: 1.25 lat × 1.25 lon	None
UKMO-HadGEM1	United Kingdom	Atm: 1.25 lat × 1.875 lon; Ocn: 1 lat × 1 lon	None

finish in the year 1999, here our assessment is confined to the period 1979–1999.

3. Results

3.1. Annual Mean

[11] The domain selected for this assessment encompasses the Arctic Ocean and peripheral seas, including Sea of Okhotsk, the Bering Sea, Hudson Bay, Baffin Bay and Davis Strait (see the shaded regions in Figure 2). Figure 1a shows the annual mean surface air temperature over the Arctic Ocean from the IABP/POLES, reanalyses, and IPCC AR4 models averaged for the period 1979–1999. IABP/POLES has annual mean SAT of -9.36°C . Compared to IABP/POLES, the reanalyses have warm biases. The warm bias of ERA40 (1.48°C) is two times larger than that of NCEP2 (0.7°C). The annual mean SAT varies greatly from -20.67°C to -0.65°C across the IPCC AR4 models, with an average of -10.16°C , which is fairly close to the IABP/POLES value. Most IPCC AR4 models reasonably reproduce the IABP/POLES observations, except BCC-CM1 and IAP-FGOALS1.0g. Specifically, BCC-CM1 shows extremely large warm biases, whereas IAP-FGOALS1.0g exhibits extremely large cold bias, far exceeding the standard deviation of the interannual IABP/POLES SAT variability (0.47°C). Encouragingly, CNRM-CM3, GFDL-CM2.1, GISS-EH, GISS-ER, IPSL-CM4, MIROC3.2m, MPI-ECHAM5, NCAR-CCSM3 and UKMO-HadCM3 have biases comparable to or less than the standard deviation of the observed interannual SAT variability. The standard deviation across the twenty-three models is 3.83°C , which is reduced to 2.47°C after excluding BCC-CM1 and IAP-FGOALS1.0g. On the annual basis, almost half of the IPCC AR4 models have biases that are comparable to or greater than the across-model standard deviations.

[12] Figure 2a shows the spatial distribution of the annual mean IABP/POLES surface air temperature. The observed SAT decreases poleward from the northern Atlantic and Pacific, with the lowest SAT found in the Canadian Archipelago. The intense temperature gradient in the northern Atlantic and southern Northwest Passage represents the marginal ice zones. Compared to IABP/POLES, NCEP2 has warm biases in the central Arctic Ocean and marginal ice zones, and cold biases in the northern Greenland Sea, Kara Sea and Beaufort Sea (Figure 2b). By contrast, ERA40 shows warm biases almost everywhere (Figure 2c). For both reanalyses, the largest SAT bias in the Arctic Ocean north of 70°N is found in the Laptev Sea. The SAT biases in the marginal ice zones are about three times larger than that in the central Arctic Ocean, since the marginal ice zones are characterized by more active atmosphere-sea ice-ocean interactions. The SAT differences between the average of the IPCC AR4 models and IABP/POLES show that the multi-model ensemble mean tends to be colder by $\sim 1^{\circ}\text{C}$ over much of the Arctic Ocean and by $\sim 2\text{--}4^{\circ}\text{C}$ over the Barents Sea and western Greenland Sea (Figure 2d). Exceptions are Hudson Bay, Davis Strait and the northern Bering Sea, where the multi-model ensemble mean is warmer. As shown in Figure 3, twelve of the twenty-three models (BCCR-BCM2.0, CCCMA-CGCM3.1m, CCCMA-CGCM3.1h, CSIRO-MK3.0, GFDL-CM2.0, GISS-EH, GISS-ER, IAP-FGOALS1.0g, MIUB-ECHOg, MRI-CGCM2.3.2a, NCAR-PCM1, and UKMO-HadGEM1) simulate SAT that are colder than the IABP/POLES observations over much of the Arctic Ocean, although some models show warm biases in the marginal ice zones. By contrast, four of the twenty-three models (BCC-CM1, GISS-AOM, INM-CM3.0, and MIROC3.2h) simulate SAT that are warmer than the observations almost everywhere. The rest of the models (CNRM-CM3, GFDL-CM2.1, IPSL-CM4, MIROC3.2m, MPI-ECHAM5, NCAR-CCSM3, and UKMO-HadCM3) show mixed features. As reflected by

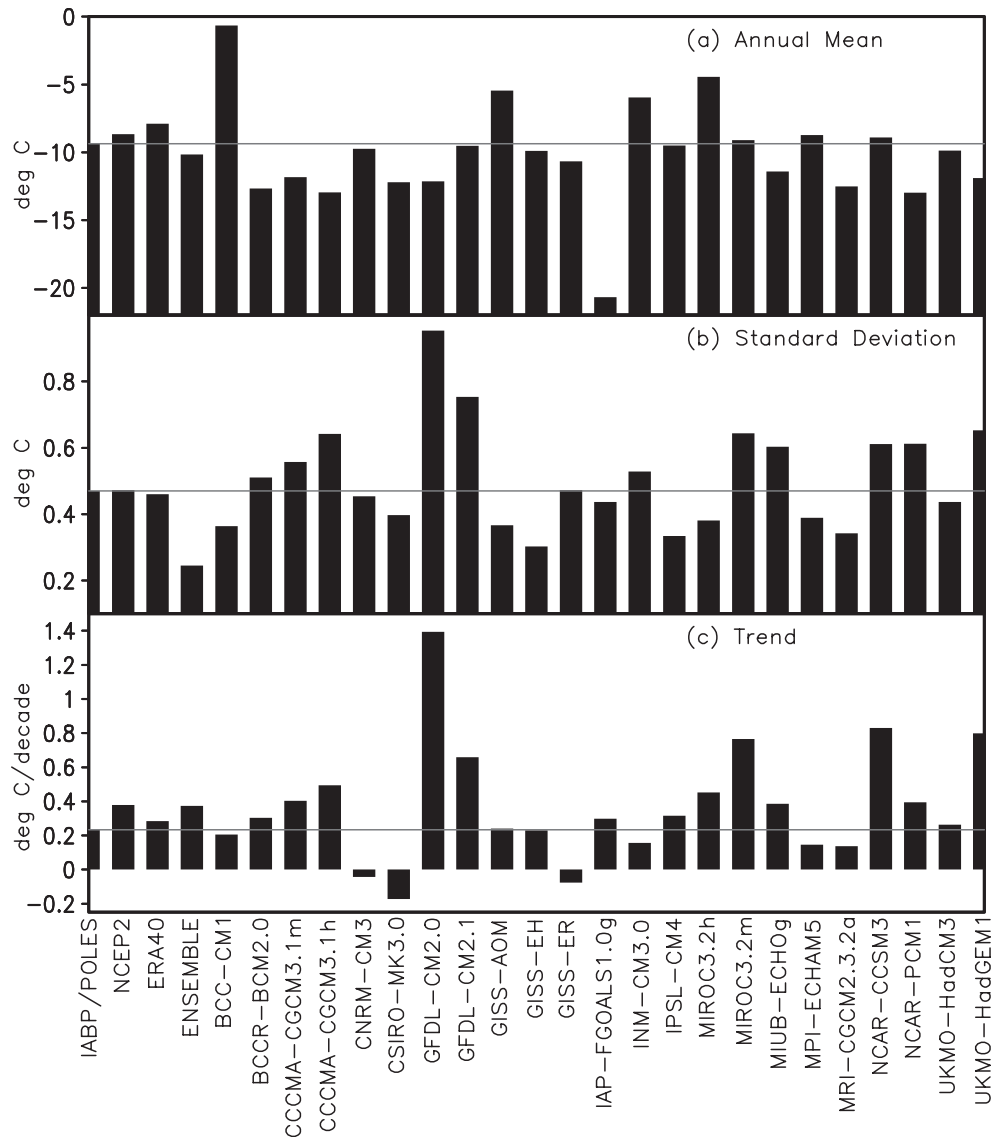


Figure 1. (a) Annual mean ($^{\circ}\text{C}$), (b) standard deviation of interannual variability ($^{\circ}\text{C}$), and (c) trend ($^{\circ}\text{C}/\text{decade}$) of surface air temperature over the Arctic Ocean during 1979–1999 for the IABP/POLES, reanalyses, and IPCC AR4 models.

the spatial distribution of the across-model standard deviations (Figure 3), the IPCC AR4 models have largest SAT discrepancies over the North Pole, Greenland Sea, Barents Sea, and Baffin Bay, where the values exceed 5°C .

[13] CCCMA and MIROC provide two versions of their models with different resolution. The SAT bias pattern of the two CCCMA simulations is similar, although the high-resolution simulation has colder SAT in the Barents Sea and Greenland Sea than the low-resolution counterpart. By contrast, for the two MIROC simulations, unlike the low-resolution simulation, the high-resolution simulation shows warm biases everywhere. Moreover, the magnitude of the SAT biases in the high-resolution simulation is much larger than the low-resolution counterpart. Thus, increasing resolution does not guarantee improved SAT simulations over the Arctic Ocean, although parameters in the high-resolution models might not be tuned properly.

[14] GFDL provides two versions of its models with different dynamical core in the atmospheric component. Compared to CM2.0, CM2.1 reduces the magnitude of the SAT biases extending from the Barents Sea, through the central Arctic Ocean, to the Sea of Okhotsk, although the sign of the SAT biases is reversed over much of the central Arctic Ocean.

[15] GISS also provides three versions of its model. In general, ER and EH, show similar SAT bias pattern as well as comparable SAT bias magnitude even though they differ in their ocean models and resolution. By contrast, AOM, which has different sea ice model and resolution (although it does share the same ocean model as ER), shows dramatically different SAT bias pattern and larger SAT bias magnitude relative to that of ER and EH.

[16] With high-resolution in the atmospheric component and improved parameterizations in the atmosphere, sea ice and ocean, the new version of the NCAR model (CCSM3,

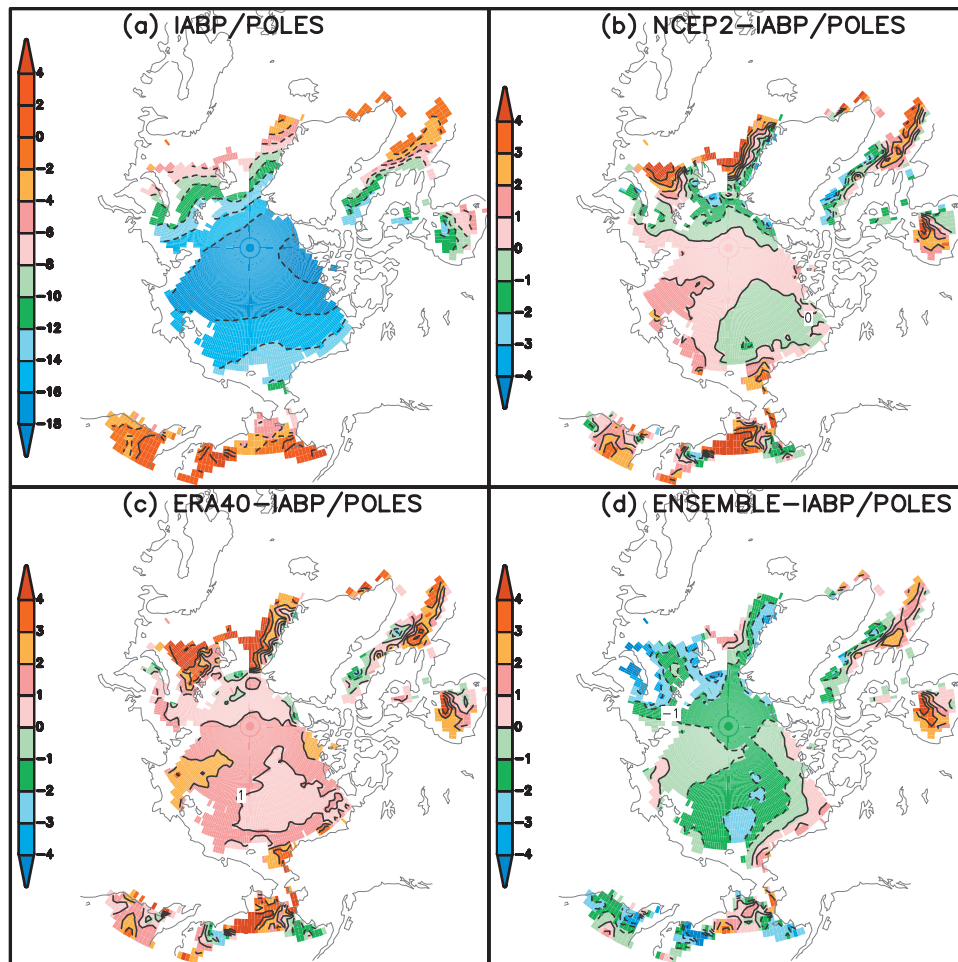


Figure 2. Spatial distribution of annual mean surface air temperature ($^{\circ}\text{C}$) during 1979-1999 for the (a) IABP/POLES, (b) NCEP2-IABP/POLES, (c) ERA40-IABP/POLES and (d) multi-model ensemble mean-IABP/POLES.

<http://www.cesm.ucar.edu/models/ccsm3.0>) shows somewhat improved SAT simulations as compared to the previous version (PCM1). However, the opposite is case for the two UKMO models; that is the previous version (HadCM3) outperforms the new version (HadGEM1).

3.2. Seasonal Cycle

[17] According to the IABP/POLES observations, the average surface air temperature over the Arctic Ocean has a large seasonal cycle reaching the lowest temperature (-21.76°C) in February and highest temperature (3.69°C) in July for the period 1979-1999 (Figure 4). The reanalyses reproduce the observed seasonal variations very well. NCEP2 has small cold bias from May to September and warm biases for the other months. By contrast, ERA40 shows persistent warm biases for all the months. The magnitude of the SAT biases of ERA40 is larger than that of NCEP2, particularly during the melting and freeze-up period. Moreover, the highest temperature of ERA40 is found in August, one month later than the IABP/POLES observations (Figure 4). For both reanalyses, the magnitude of the SAT biases in winter is larger than that in summer, since the SAT over sea ice mainly oscillates around the freezing point in summer.

[18] Encouragingly, the seasonal cycle of the multi-model ensemble mean is fairly close to the IABP/POLES observations, and its biases are comparable to NCEP2, and even smaller than ERA40 (Figure 4 and Table 2). However, some significant differences are found between each individual model simulation and IABP/POLES. Specifically, BCC-CM1 (IAP-FGOALS1.0g) exhibits extremely reduced (enhanced) seasonal variations, with extremely large warm (cold) biases for all the months, particularly during the cold season. Even with BCC-CM1 and IAP-FGOALS1.0g excluded, the spread of the simulated SAT among the models can be as large as $\sim 15^{\circ}\text{C}$ in winter, a factor of three larger than that in summer. While the IABP/POLES SAT stays above the freezing point (0°C) from June to September, most IPCC AR4 models only have two or three months with SAT above 0°C , and simulate SAT below 0°C in September, suggesting that most IPCC AR4 models have a too short melt season.

[19] Figure 5 shows the spatial distribution of the across-model standard deviations of SAT for winter and summer. It appears that the pattern of the last plot in Figure 3 is largely a consequence of the winter temperature.

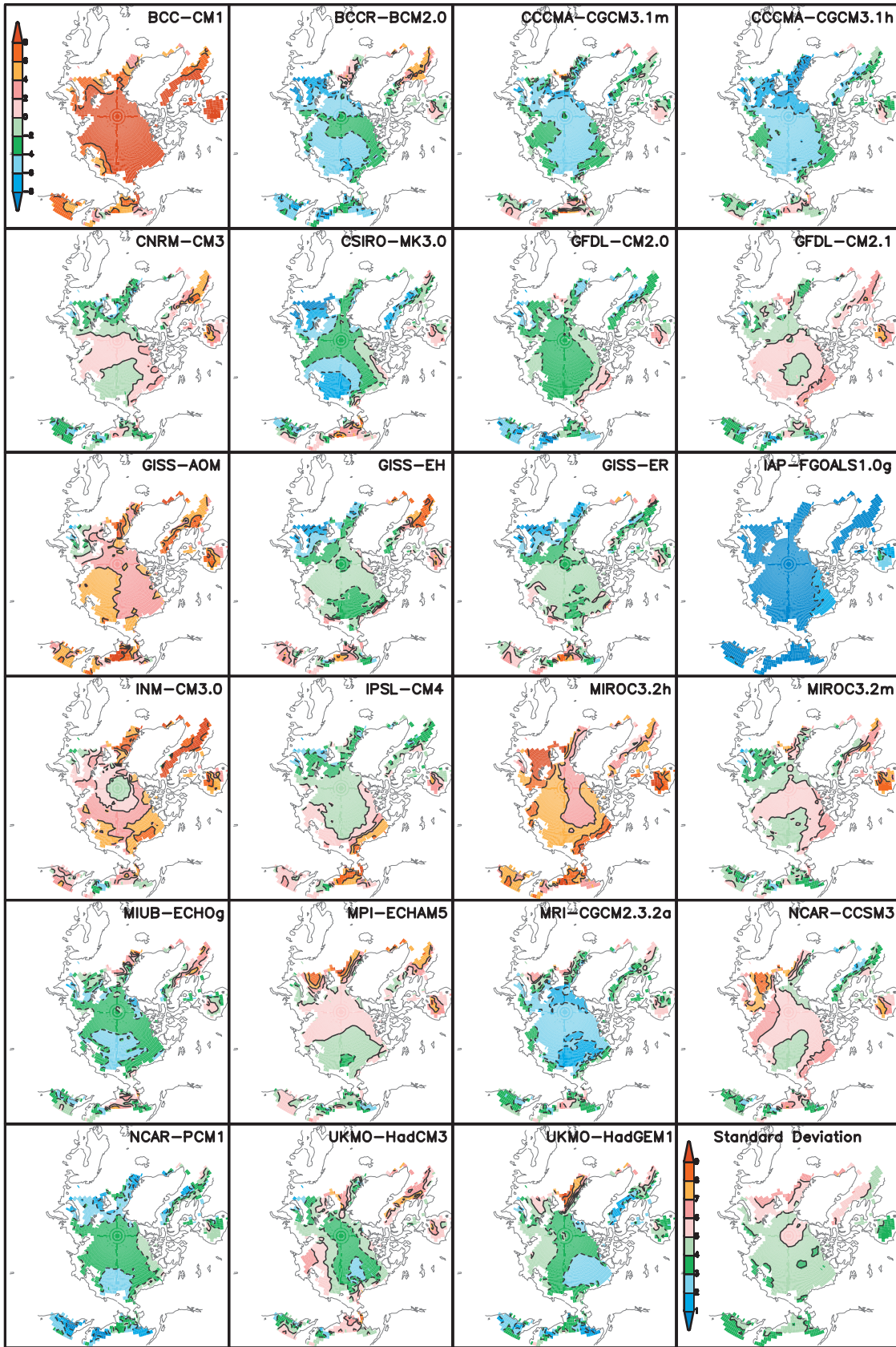


Figure 3. Same as Figure 2, except for differences ($^{\circ}\text{C}$) between individual IPCC AR4 model and IABP/POLES, the last plot is the spatial distribution of the standard deviations ($^{\circ}\text{C}$) across the IPCC AR4 models.

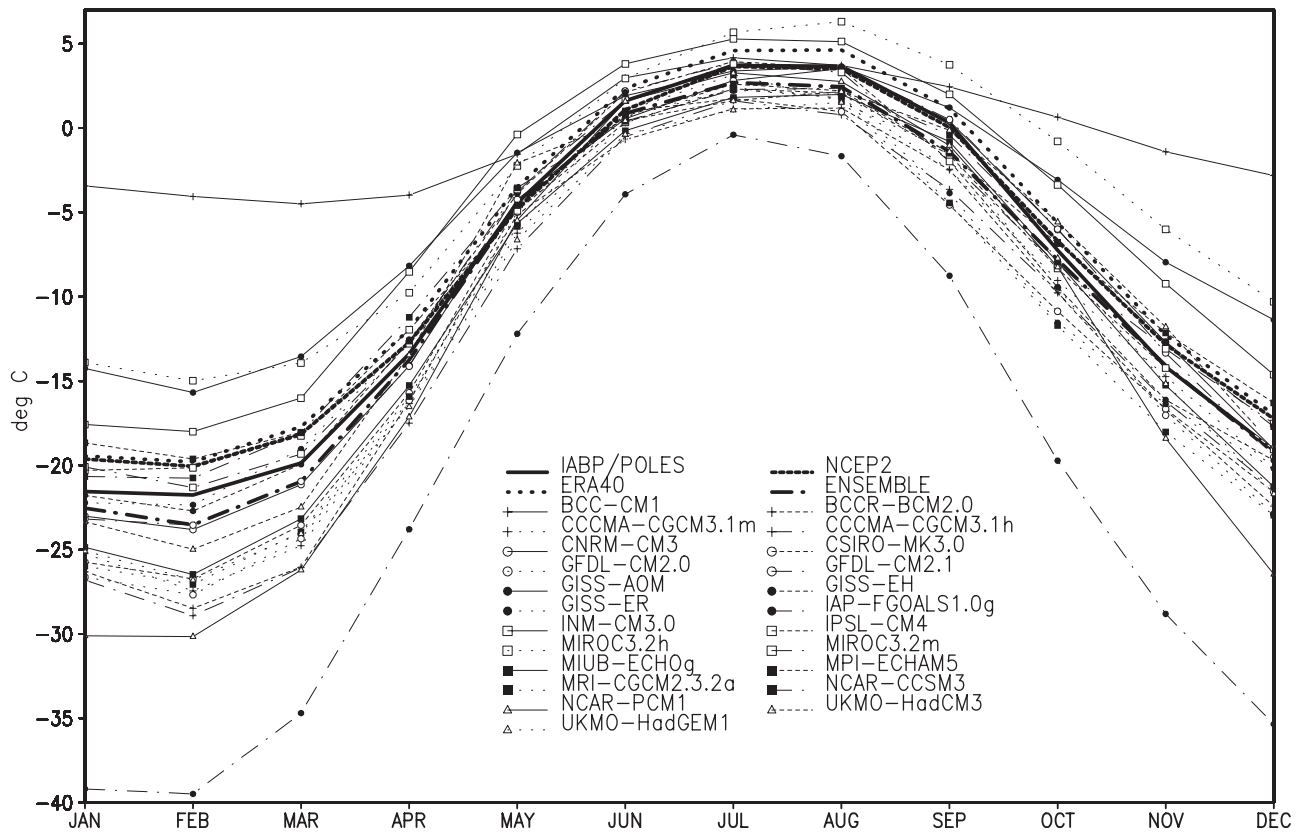


Figure 4. Seasonal cycle of surface air temperature (°C) over the Arctic Ocean during 1979–1999 for the IABP/POLES, reanalyses, and IPCC AR4 models.

Table 2. Biases of Seasonal Mean and Standard Deviation of Reanalyses and IPCC AR4 Models Relative to IABP/POLES

	Mean bias (°C)				Standard deviation bias (°C)			
	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter
NCEP2	0.66	-0.24	0.54	1.83	0.69	-0.74	0.67	0.45
ERA40	1.50	0.86	1.46	2.05	0.48	-0.60	0.47	0.40
ENSEMBLE	-0.54	-1.00	-0.71	-0.99	0.85	-0.96	0.83	1.05
BCC-CM1	9.21	0.63	7.64	17.35	-1.22	-0.52	-0.59	-2.93
BCCR-BCM2.0	-3.19	-2.44	-2.58	-4.85	1.50	-0.74	1.29	1.99
CCCMA-CGCM3.1m	-3.24	-1.50	-1.65	-3.44	1.53	-0.87	1.11	1.72
CCCMA-CGCM3.1h	-4.33	-2.37	-2.78	-4.85	1.79	-0.71	1.42	2.02
CNRM-CM3	-0.60	-0.68	0.98	-1.20	0.93	-0.89	0.51	1.15
CSIRO-MK3.0	-2.06	-1.76	-3.58	-3.93	1.22	-0.88	1.51	1.77
GFDL-CM2.0	-2.67	-1.34	-2.37	-4.94	1.43	-0.74	1.33	2.16
GFDL-CM2.1	-0.46	0.21	0.78	-1.27	1.00	-0.72	0.72	1.21
GISS-AOM	4.87	-0.01	3.76	7.02	-0.31	-0.82	-0.04	-0.70
GISS-EH	0.38	0.19	-2.13	-0.57	0.67	-0.78	1.17	1.00
GISS-ER	0.50	-1.41	-3.53	-0.89	0.64	-0.93	1.50	1.13
IAP-FGOALS1.0g	-11.02	-4.99	-12.04	-17.12	3.22	-0.17	3.39	4.69
INM-CM3.0	4.25	1.73	3.52	4.02	-0.07	-0.45	0.01	0.64
IPSL-CM4	0.86	-1.41	-1.02	0.95	0.58	-0.83	0.95	0.68
MIROC3.2h	3.87	1.95	6.00	7.77	-0.12	-0.41	-0.53	-0.86
MIROC3.2m	0.63	-0.08	-0.72	1.15	0.66	-0.78	0.89	0.69
MIUB-ECHOg	-2.17	-1.79	-1.01	-3.44	1.26	-0.85	1.01	1.70
MPI-ECHAM5	0.85	-1.15	0.29	2.57	0.60	-0.84	0.68	0.29
MRI-CGCM2.3.2a	-2.65	-1.07	-4.36	-4.60	1.36	-0.89	1.71	1.87
NCAR-CCSM3	1.55	-1.35	0.36	1.01	0.41	-0.86	0.70	0.71
NCAR-PCM1	-3.68	-0.46	-2.16	-8.07	1.71	-0.86	1.39	2.70
UKMO-HadCM3	-0.16	-1.59	1.23	-1.60	0.82	-0.89	0.42	1.24
UKMO-HadGEM1	-3.20	-2.26	-1.08	-3.73	1.52	-0.75	1.03	1.74

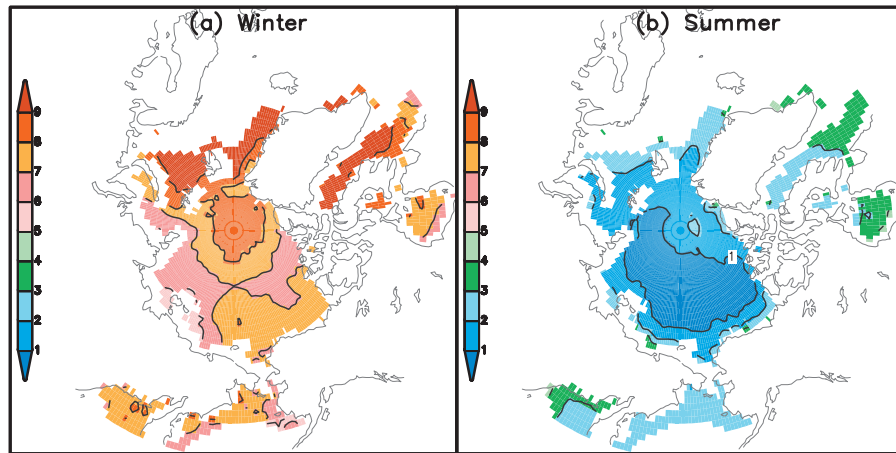


Figure 5. Spatial distribution of the standard deviations of surface air temperature ($^{\circ}\text{C}$) across the IPCC AR4 models for (a) winter and (b) summer.

3.3. Variability

[20] Figure 1b shows the standard deviation of the internal SAT variability for the IABP/POLES, reanalyses, and IPCC AR4 models. The magnitude of SAT variability in the

reanalyses agrees very well with that of IABP/POLES (0.47°C). Additionally, the reanalyses are well correlated with the IABP/POLES observations for the period 1979–1999 (0.93–0.88), although the correlations drop to 0.76 and 0.62 for NCEP2 and ERA40 in winter, respectively.

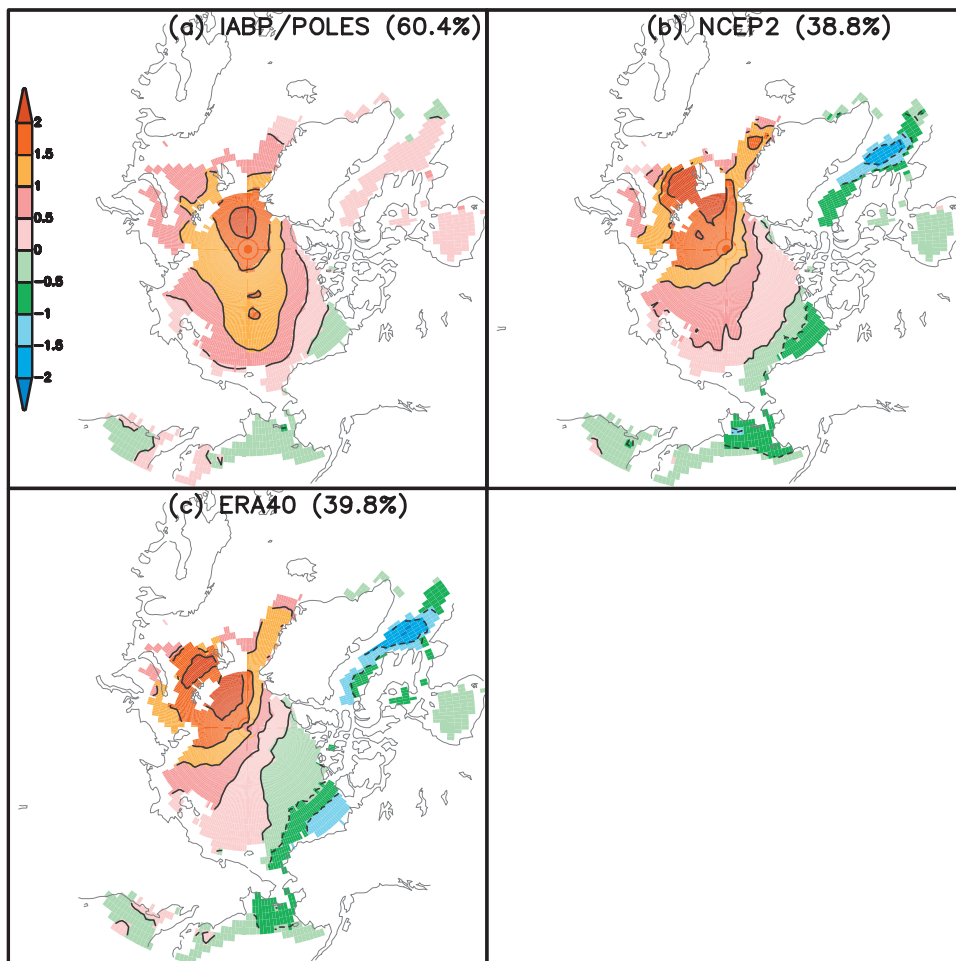


Figure 6. The first EOF mode of surface air temperature over the Arctic Ocean during 1979–1999 for the (a) IABP/POLES, (b) NCEP2, (c) ERA40, and (d) multi-model ensemble mean.

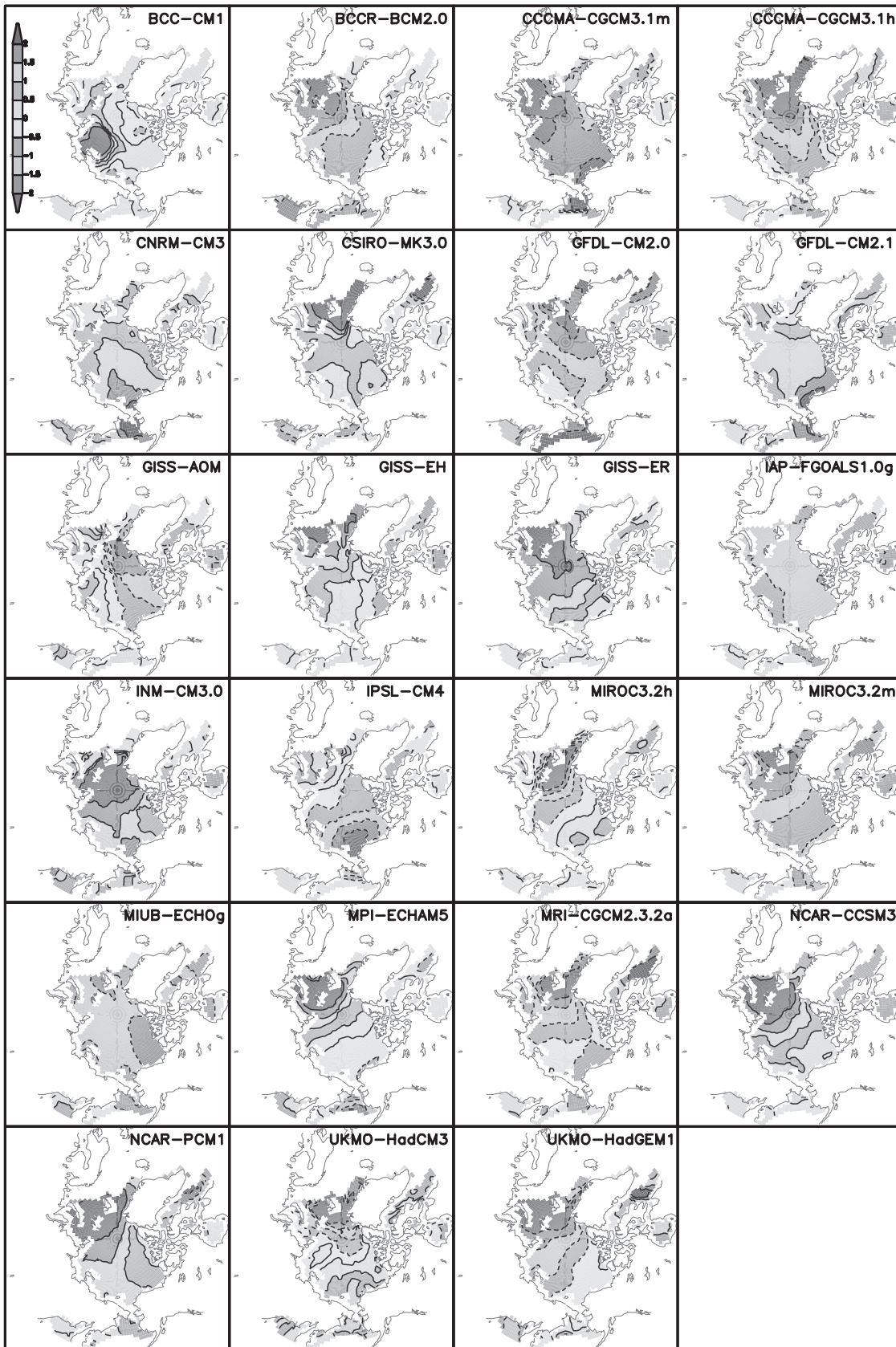


Figure 7. The first EOF mode of surface air temperature over the Arctic Ocean during 1979–1999 for each individual IPCC AR4 model.

Table 3. Seasonal Trends of IABP/POLES, Reanalyses and IPCC AR Models

	Trend (°C/decade)			
	Spring	Summer	Autumn	Winter
IABP/POLES	0.86	0.07	0.29	-0.35
NCEP2	0.87	0.31	0.58	-0.37
ERA40	0.79	0.21	0.44	-0.40
ENSEMBLE	0.33	0.15	0.48	0.50
BCC-CM1	0.24	0.23	0.01	0.38
BCCR-BCM2.0	0.57	0.14	0.41	0.27
CCCMA-CGCM3.1m	0.68	0.18	0.50	0.46
CCCMA-CGCM3.1h	0.66	0.16	0.83	0.42
CNRM-CM3	0.07	-0.01	0.14	-0.38
CSIRO-MK3.0	-0.15	-0.01	-0.19	-0.30
GFDL-CM2.0	1.22	0.52	1.31	2.64
GFDL-CM2.1	0.50	0.27	0.50	1.11
GISS-AOM	0.42	0.14	0.30	0.13
GISS-EH	0.28	0.11	0.44	0.01
GISS-ER	-0.01	-0.02	0.08	-0.22
IAP-FGOALS1.0g	0.12	0.08	0.31	0.46
INM-CM3.0	0.30	0.14	-0.14	0.10
IPSL-CM4	-0.01	0.21	0.59	0.35
MIROC3.2h	0.40	0.12	0.33	0.51
MIROC3.2m	0.81	0.24	0.80	1.19
MIUB-ECHOG	0.10	0.08	0.52	0.39
MPI-ECHAM5	0.28	0.24	0.30	0.20
MRI-CGCM2.3.2a	-0.16	0.00	0.63	0.00
NCAR-CCSM3	0.32	0.15	1.00	1.54
NCAR-PCM1	-0.19	0.19	0.98	0.85
UKMO-HadCM3	0.30	0.09	0.27	0.39
UKMO-HadGEM1	0.78	0.25	1.05	1.01

The magnitude of simulated SAT variability varies greatly from 0.3 to 0.95°C among the IPCC AR4 models, with an average of 0.51°C, which is fairly close to that of IABP/POLES. As expected, the multi-model ensemble mean greatly damps out the interannual SAT fluctuations present in each individual model (0.22°C, Figure 1b). Most IPCC AR4 models reasonably reproduce the observed SAT variability suggested by IABP/POLES, except GFDL-CM2.0 and GFDL-CM2.1, which dramatically overestimate the observed SAT variability. Specifically, BCCR-BCM2.0, CNRM-CM3, GISS-ER, IAP-FGOALS1.0g and UKMO-HadCM3 simulate SAT variability comparable to the observed SAT variability.

[21] According to IABP/POLES, the observed interannual SAT variability is 2.53, 1.1, 1.37 and 4.27°C for spring, summer, fall and winter, respectively. Both the reanalyses and all the IPCC AR4 models underestimate the observed SAT variability in summer (Table 2). By contrast, the reanalyses and most IPCC AR4 models overestimate the observed SAT variability for the other seasons. Exceptions are BCC-CM1, GISS-AOM and MIROC3-2h, which show systematic underestimation for all four seasons. Interestingly, in summer, the bias of the SAT variability in the reanalyses is comparable to that of the IPCC AR4 models, whereas for the other seasons, the biases of the SAT variability in the reanalyses are generally smaller than that of most IPCC AR4 models.

[22] For CCCMA, increasing model resolution tends to enhance the interannual SAT variability. However, for MIROC, the opposite is the case. Compared to GFDL-CM2.0, GFDL-CM2.1 improves the simulations of the observed SAT variability greatly. Among three versions of

the GISS models, ER much better simulates the observed SAT variability than EH and AOM. For the two NCAR models, high-resolution in the atmospheric component and improved parameterizations in all the components in CCSM3 do not lead to improved simulations of observed SAT variability as compared to PCM1.

[23] To determine if the primary characteristics of spatio-temporal variability of SAT over the Arctic Ocean of the reanalyses and IPCC AR4 models are consistent with that of IABP/POLES during 1979–1999, we performed an empirical orthogonal function analysis (EOF) on the average extended winter (defined as December–March) SAT anomalies. Figure 6 shows the first loading EOF mode. IABP/POLES exhibits large SAT variability near Fram Strait, which decreases toward the surrounding sub-Arctic seas. Interestingly, the observed dominant SAT mode has no association with the Arctic Oscillation. The reanalyses show similar spatial pattern to that of IABP/POLES, except that the large SAT variability is more towards the Barents Sea, and Baffin Bay and Davis Strait show an out of phase relationship with the Arctic Ocean. The multi-model ensemble mean can not capture the observed dominant SAT mode variability (not shown). However, as shown in Figure 7, a few models do have spatial structure with some similarities to the observations.

3.4. Trend

[24] As shown in Figure 1c, the annual mean IABP/POLES surface air temperature has a warming trend of 0.23°C/decade for the period 1979–1999 in the Arctic Ocean. Consistent with IABP/POLES, the reanalyses have a positive trend of SAT. However, NCEP2 shows a more pronounced increase of SAT, which is 1.65 times greater than that of IABP/POLES (the trend difference is statistically significant). Most IPCC AR4 models also show consistency with the IABP/POLES observations in reproducing the positive trend of SAT, but CCCMA-CGCM3.1h, GFDL-CM2.0, GFDL-CM2.1, MIROC3.2h, MIROC3.2m, NCAR-CCSM3, and UKMO-HadGEM1 have warming trends about 2–6 times greater than the observed warming trend. Exceptions are CNRM-CM3, CSIRO-MK3.0 and GISS-ER, which show a slightly cooling trend for the period 1979–1999. The large scatter of the modeled trends is probably due to internal variability within different models (i.e., related to different formulation of physical processes and numerical methods).

[25] As suggested by some studies [e.g., Serreze *et al.*, 2000; Rigor *et al.*, 2000], the observed SAT trend in the late 20th century is characterized by a pronounced seasonality. According to the IABP/POLES observations, the averaged SAT in spring, summer and fall shows tendencies similar to the annual mean, while the averaged SAT in winter shows tendency opposite to the annual mean. As shown in Table 3, the reanalyses capture the observed seasonality of the SAT trends, although they overestimate the warming trends in summer and fall. By contrast, the IPCC AR4 models, which as noted show the increase of annual mean SAT, can not reproduce the observed decrease of winter SAT; rather they have persistent warming for all four seasons. In addition, the IABP/POLES observations show the largest warming in spring, whereas most IPCC AR4 models show the largest warming during fall and winter (Table 3). Future investiga-

tion is needed to understand the identified discrepancies between modeled and observed seasonality of the SAT trends.

4. Discussion and Summary

[26] This assessment shows a snapshot of to what extent the current day state-of-the-art reanalysis (NCEP2 and ERA40) and coupled global climate models (IPCC AR4 models) can reproduce the annual mean, seasonal cycle, variability and trend of the observed surface air temperature over the Arctic Ocean.

[27] Overall, the reanalyses (NCEP2 and ERA40) show encouraging agreements with the IABP/POLES observations. It is not too surprising that the reanalyses do well in reproducing the observed SAT, since many in-situ measurements and satellite products have been assimilated in the reanalysis [Kanamitsu *et al.*, 2002]. However, some temporal and spatial discrepancies are still noteworthy. On the annual basis, the reanalyses have warm biases, and the bias of ERA40 is larger than that of NCEP2. The smaller bias of NCEP2 results from a cancellation of cold and warm spatiotemporal biases, in contrast to ERA40's systematic warm biases. In summer, the reanalyses underestimate the observed interannual SAT variability, and their bias magnitude is even comparable to the IPCC AR4 model simulations. In winter, the temporal correlation between the reanalyses and observations is smaller than in the other seasons. Compared to the observed SAT trend, NCEP2 exhibits a more pronounced warming trend over the Arctic Ocean for the period 1979–1999. Future work should target the sources of the identified discrepancies between the reanalyses and observations, and between the two reanalyses.

[28] Despite the complicated atmosphere-sea ice-ocean interactions in the Arctic Ocean, our assessment demonstrates that the annual mean SAT biases of the majority of the IPCC AR4 models are smaller than the inter-model standard deviation. The seasonal cycle is also simulated reasonably well by most models. Moreover, most model simulations show positive trends of SAT during 1979–1999, which is consistent with the observations. In particular, the multi-model ensemble mean realistically captures the annual mean and seasonal cycle of the observed SAT, and its biases relative to the observations are comparable to the reanalyses and well below the standard deviations across the models, increasing the credibility of the models' representation of physical processes in the Arctic Ocean.

[29] However, large uncertainties are still found in simulating the climate of the 20th century. On the annual basis, almost two thirds of the IPCC AR4 models have biases that greater than the standard deviation of the observed SAT variability. Spatially, the models show considerable variance over the North Pole, Greenland Sea, Barents Sea, and Baffin Bay, where the across-model standard deviations exceed 5°C. Compared to the results of Walsh *et al.* [2002], there is no obvious improvement since the IPCC Third Assessment Report. This is due in large part to two challenges. First, the models have different approaches to resolve the problem of convergence of meridian at the North Pole, and second, the marginal ice zones are characterized by more complex air-ice-ocean interactions. Seasonally, the spread of the simulated SAT among the models in winter is much larger than

that in summer. The models show interannual variability and decadal trend with various amplitudes, and can not capture the observed dominant SAT mode variability in winter and seasonality of SAT trends.

[30] Because of the diversity of physical processes that control SAT simulation in CGCMs, it is difficult to directly attribute the SAT errors reported here to certain model features without fully investigating each individual model in detail. This is particularly true when comparing models that employ different physical parameterizations, resolution and numerical methods. Nevertheless, here we examined possible attributions of the SAT errors for some models based on the available information.

[31] As discussed previously, the SAT in the Arctic Ocean is mainly determined by radiative energy balance and advective transport of heat into the region. Thus, inaccurate advective transport of heat into the Arctic Ocean is one of the primary sources of the SAT errors. For example, IAP-FGOALS1.0g applied a too strong ocean filter in the high-latitudes, which limits the poleward ocean heat transport. Additionally, IAP-FGOALS1.0g started the 20C3M experiment with an excessive initial sea ice condition. These problems led to the aforementioned extremely cold biases in IAP-FGOALS1.0g, which is consistent with dramatically overestimated sea ice cover found in Zhang and Walsh [2006]. After fixing the known problems, the new version of the IAP model (FGOALS1.1g) shows improved SAT and sea ice simulations (see http://www-pcmdi.llnl.gov/ipcc/model_documentation/more_info_iap_fggoals.pdf).

[32] Since the Arctic Ocean is covered by sea ice, the models' performance in the SAT simulations is no doubt sensitive to the parameterizations of sea ice. In fact, the large inter-model scatter in the Arctic Ocean is at least partly attributable to sea ice. The wintertime marginal ice zones coincide closely with the regions of maximum standard deviation in Figure 5. Figure 8 shows the annual mean biases of sea ice area and SAT for fifteen models (here the sea ice area biases are calculated based on Table 2 in Zhang and Walsh [2006]). The majority of the models show an out-of-phase relationship between the sea ice area and SAT biases, particularly the coldest model (IAP-FGOALS1.0g) is also the model with the largest sea ice area. Also, the modeled SAT trends that are mostly positive are consistent with the downward trend in the Arctic sea ice cover [Stroeve *et al.*, 2007]. With the improved parameterizations in sea ice dynamics and thermodynamics, GISS-ER produces more sea ice relative to GISS-AOM [Liu *et al.*, 2003], which reduces large warm bias in GISS-AOM, particularly in winter and spring when the region is dominated by sea ice. GISS-ER and GISS-EH share the same sea ice physical parameterizations, but different ocean models (GISS-ER uses the Russell's ocean model [Russell *et al.*, 2000], whereas GISS-EH uses the HYCOM ocean model [Bleck, 2002]). The aforementioned similar SAT bias pattern and magnitude in GISS-ER and GISS-EH suggests (at least for the GISS model) that sea ice parameterizations might have a bigger influence than ocean parameterizations on the SAT simulation over the Arctic Ocean. However, more sophisticated sea ice component are not necessary to achieve the best match to the observations. For example, although UKMO-HadCM3 employs a relatively simplified sea ice component, and has a relatively coarse spatial resolution, it

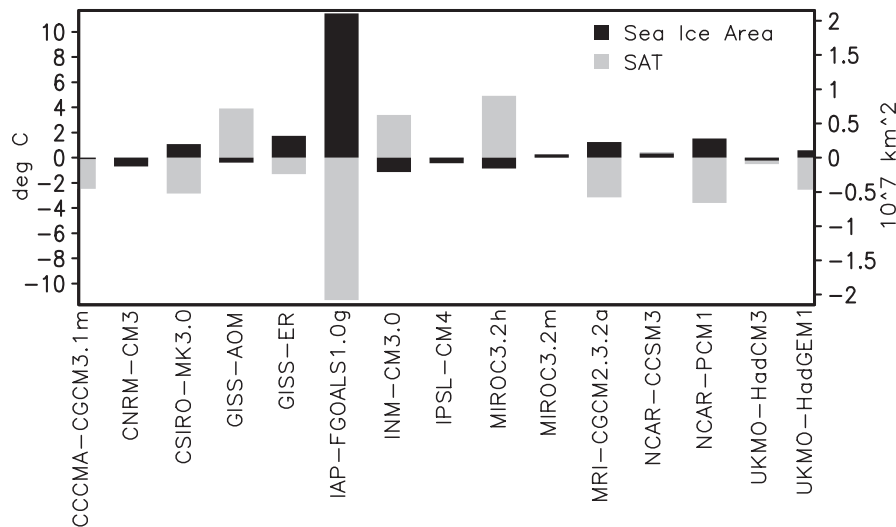


Figure 8. Annual mean sea ice area differences between individual IPCC AR4 model and the Hadley Centre Sea Ice and SST data set as well as the corresponding surface air temperature differences.

achieves a better overall match with the observations than the new version of the UKMO model does (HadGEM1, [McLaren *et al.*, 2006]). Thus, inaccurate simulations in other components could completely override the models' strength in the parameterizations of sea ice.

[33] For the same family of models that operate at different spatial resolutions, the thinner sea ice simulated in the high-resolution MIROC model tends to cause earlier decrease of sea ice cover in the Arctic Ocean in response to the increase of greenhouse gases, leading to large warm biases. By contrast, the more reasonable ice thickness simulated in the low-resolution MIROC model leads to more realistic sea ice cover, and SAT in the Arctic Ocean. Thus, increasing model spatial resolution does not ensure improved SAT simulations in the Arctic Ocean. This also holds true for the CCCMA model.

[34] The models' performance in the SAT simulations is also sensitive to the dynamical core applied in the atmospheric component. For example, the biggest difference between the two GFDL models is that they have different dynamical cores in their atmospheric component. CM2.1 uses a finite-volume dynamical core, whereas CM2.0 uses a B-grid dynamical core. This change leads to substantial improvements in the high-latitude wind stress pattern and temperature simulations in CM2.1 relative to CM2.0 [Delworth *et al.*, 2006]. In addition, CM2.1 modified parameters in the cloud scheme to increase the net short-wave radiation at the surface, and used a weaker horizontal viscosity in the extratropical ocean to increase the polar heat transport of the subpolar gyre, thereby substantially reducing the cold bias and excessive sea ice in CM2.0. Taken together, these changes also reduce the extremely large interannual variability and decadal trend of SAT in CM2.0.

[35] The large scatter of the SAT simulations across the IPCC AR4 models also comes from differences in the prescribed natural forcings applied in the climate of twentieth-century experiment. For example, natural forcings, such as time-varying solar irradiance, volcanic aerosols, and tropospheric/stratospheric ozone, are not included in

some models, but they are important for reproducing the observed SAT variations [e.g., Hansen *et al.*, 2005]. Thus, it is important to develop standard sets of historical natural forcings for the future model intercomparison of the twentieth-century climate.

[36] Given the large uncertainties of the IPCC AR4 models in simulating SAT over the Arctic Ocean for the late 20th century (particularly large scatter across the models), further efforts are needed to improve the simulations of the climate of the Arctic Ocean. In this way, we can be confident in our conclusions as to whether or not greenhouse gas forcing is the dominant player in recent amplified warming in the Arctic. Likewise, the uncertainties in the projections of future climate change can be greatly reduced.

[37] **Acknowledgments.** We acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modeling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. This research is supported by the National Basic Research Program of China (2006CB403605), "Hundred Talent Program" of the Chinese Academy of Sciences, NSFC (40676003, 40575031, 40533016, 40531007 and 40676062), and MOE of China (106002). We thank anonymous reviewers for helpful comments on the manuscript.

References

- Armstrong, R. L., and M. J. Brodzik (1995), An Earth-gridded SSM/I data set for cryospheric studies and global change monitoring, *Adv. Space Res.*, *16*, 155–163.
- Arctic Climate Impact Assessment (ACIA) (2004), *Impacts of a Warming Arctic: Arctic Climate Impact Assessment*, 139 pp., Cambridge Univ. Press, New York.
- Barry, R. G., M. C. Serreze, J. A. Maslanik, and R. H. Preller (1993), The Arctic sea ice-climate system: Observations and modeling, *Rev. Geophys.*, *31*, 397–422.
- Bleck, R. (2002), An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates ocean model, *Ocean Modell.*, *4*, 55–88.
- Comiso, J. C. (2003), Warming trends in the Arctic from clear sky satellite observations, *J. Clim.*, *16*, 3498–3510.

- Cressman, G. P. (1959), An operational objective analysis system, *Mon. Weather Rev.*, *87*, 367–374.
- Delworth, T. L., et al. (2006), GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and simulation characteristics, *J. Clim.*, *19*, 643–674.
- Hansen, J., R. Ruedy, J. Glascoe, and M. Sato (1999), GISS analysis of surface temperature change, *J. Geophys. Res.*, *104*, 30,997–31,022.
- Hansen, J., et al. (2005), Efficacy of climate forcings, *J. Geophys. Res.*, *110*, D18104, doi:10.1029/2005JD005776.
- Hansen, J., M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade (2006), Global temperature change, *Proc. Natl. Acad. Sci.*, *103*, 14,288–14,293, doi:10.1073/0606291103.
- IPCC Third Assessment Report (2001), *Climate Change 2001: The Scientific Basis*, edited by J. T. Houghton, et al., 944 pp., Cambridge Univ. Press, U.K.
- Jones, P. D., and M. E. Mann (2004), Climate over past millennia, *Rev. Geophys.*, *42*, 1–42, RG2002, doi:10.1029/2003RG000143.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, J. Potter, and M. Fiorino (2002), NCEP/DOE AMIP-II Reanalysis (R-2), *Bull. Amer. Meteor. Soc.*, *83*, 1631–1643.
- Liu, J., G. A. Schmidt, D. Martinson, D. Rind, G. Russell, and X. Yuan (2003), Sensitivity of sea ice to physical parameterizations in the GISS global climate model, *J. Geophys. Res.*, *108*(C2), 3053, doi:10.1029/2001JC001167.
- Liu, J., J. A. Curry, and Y. Hu (2004), Recent Arctic sea ice variability: Connections to the Arctic Oscillation and the ENSO, *Geophys. Res. Lett.*, *31*, L09211, doi:10.1029/2004GL019858.
- Martin, S. E., and E. Munoz (1997), Properties of the Arctic 2-m air temperature for 1979-present derived from a new gridded data set, *J. Clim.*, *10*, 1428–1440.
- McLaren, A. J., et al. (2006), Evaluation of the sea ice simulation in a new coupled atmosphere-ocean climate model (HadGEM1), *J. Geophys. Res.*, *111*, C12014, doi:10.1029/2005JC003033.
- Meehl, G. A., and W. M. Washington (1990), CO₂ climate sensitivity and snow–sea–ice albedo parameterization in an atmospheric GCM coupled to a mixed-layer ocean model, *Clim. Change*, *16*, 283–306.
- Meehl, G. A., C. Covey, B. McAvaney, M. Latif, and R. J. Stouffer (2005), Overview of the Coupled Model Intercomparison Project, *Bull. Amer. Meteor. Soc.*, *86*, 89–93.
- Parkinson, C. L., and D. J. Cavalieri (2002), A 21 year record of Arctic sea-ice extents and their regional, seasonal and monthly variability and trends, *Ann. Glaciol.*, *34*, 441–446.
- Russell, G. L., J. R. Miller, D. Rind, R. A. Ruedy, G. A. Schmidt, and S. Sheth (2000), Comparison of model and observed regional temperature changes during the past 40 years, *J. Geophys. Res.*, *105*, 14,891–14,898.
- Rigor, I. G., R. L. Colony, and S. Martin (2000), Variations on surface air temperature in the Arctic from 1979–1997, *J. Clim.*, *13*, 896–914.
- Serreze, M. C., et al. (2000), Observational evidence of recent change in the northern high latitude environment, *Clim. Change*, *46*, 159–207.
- Serreze, M. C., and J. A. Francis (2006), The arctic amplification debate, *Clim. Change*, *76*, 241–264.
- Serreze, M. C., M. M. Holland, and J. Stroeve (2007), Perspectives on the Arctic's shrinking sea-ice cover, *Science*, *315*, 1533–1536.
- Stroeve, J. C., M. C. Serreze, F. Fetterer, T. Arbetter, W. Meier, J. Maslanik, and K. Knowles (2005), Tracking the Arctic's shrinking ice cover: Another extreme September minimum in 2004, *Geophys. Res. Lett.*, *32*, L04501, doi:10.1029/2004GL021810.
- Stroeve, J., M. M. Holland, W. Meier, T. Scambos, and M. Serreze (2007), Arctic sea ice decline: Faster than forecast, *Geophys. Res. Lett.*, *34*, L09501, doi:10.1029/2007GL029703.
- Walsh, J. E., V. M. Kattsov, W. L. Chaoman, V. Govorkova, and T. Pavlova (2002), Comparison of Arctic climate simulations by uncoupled and coupled global models, *J. Clim.*, *15*, 1429–1446.
- Zhang, X., and J. E. Walsh (2006), Toward a seasonally ice-covered Arctic Ocean: Scenarios from the IPCC AR4 model simulations, *J. Clim.*, *19*, 1730–1747.

L. Chen, Key Laboratory of Global Change and Marine-Atmospheric Chemistry, Third Institute of Oceanography, SOA, Xiamen, 361005, China.

Y. Dai, School of Geography, Beijing Normal University, Beijing, 100875, China.

Y. Hu, Department of Atmospheric Sciences, School of Physics, Peking University, Beijing, 100871, China.

J. Liu, State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China. (jliu@lasg.iap.ac.cn)

X. Ren, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China.

Z. Zhang, Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, 200129, China.