An Antarctic assessment of IPCC AR4 coupled models

William M. Connolley[1] and Thomas J. Bracegirdle[1]

[1]   We assess 19 coupled models from the IPCC fourth assessment report archive from the simulation of the 20th century, based on the calculation of "skill scores." The models show a wide range of scores when assessed against Antarctic or global measures of large-scale circulation indices. Except for continental mass balance, the model average proves a more reliable estimate than that for any one model. Individual models show a very wide scatter in simulated Antarctic temperature trends over the past century; the large trend over the Antarctic peninsula in winter is not well represented, which makes it clear that whatever has been driving these trends is not well captured by many GCMs. Trends in temperature are clearly linked to the sea ice simulation, another variable that most models do not simulate well. **Citation:** Connolley, W. M., and T. J. Bracegirdle (2007), An Antarctic assessment of IPCC AR4 coupled models, *Geophys. Res. Lett.*, *34*, L22505, doi:10.1029/2007GL031648.

1. Introduction

[2]   We examine output from 19 coupled models from the IPCC fourth assessment report archive, from the simulation "20c3m" of the 20th century, principally using data from 1979–2000. We select this period because earlier time periods are not well covered by verification data in the region of Antarctica [*Bromwich and Fogt*, 2004]. Our intent is twofold. Firstly, to provide a (semi-)objective assessment of the various coupled models, both to point up deficient models in need of development and as a guide to those who would use the models for investigating the climate. Secondly, as a basis for weighting the models by their skills, in order to weight future trend predictions.

2. Method

[3]   The assessment is based on the calculation of "skill scores." The skill score methodology is similar to that of *Murphy et al.* [2004] and *Schmittner et al.* [2005]: a value is calculated for the root mean square (RMS) deviation of the multi-annual averaged model field from the multi-annual averaged observed field, normalised by a measure of the variability of the field. This normalised value RMS_n is then rescaled via $W = \exp(-0.5*(RMS_n)^2)$ into a weight between 0 and 1, which may be used when averaging the different models, and can be regarded as a measure of model "skill." The method cannot be fully objective because a number of subjective choices have to be made in applying it. First is the choice of variables to calculate the scores against; our selection is noted in the following section. Second is the

choice of function to transform RMS_n values into weights. We use the form above from Murphy et al; Schmittner use a similar form but with 2 rather than 0.5 scaling the RMS_n; other transformations could also be used. Third is the choice of which areas to examine: we have chosen to give equal weight to a global and an "Antarctic" comparison, by which we mean areas south of 45S. This is because although we are primarily concerned with the Antarctic simulation, in a coupled model inevitably the global climate will affect the fidelity of any future simulated change. Fourth is the choice of seasonality: one could examine yearly, seasonal or monthly averages. We have calculated RMS_n for each month, which are combined by root mean squares into an annual value. Fifth is the choice of normalisation of the RMS by the temporal variability of the field, whether to do this "globally" (i.e. the spatial average of the RMS scaled by the spatial average of the temporal variability), "pointwise" (i.e. the RMS scaled by the temporal variability at each model point and then spatially averaged), or some other method. We have chosen to scale the RMS by the temporal variability pointwise: this has an impact, since for a number of fields (for example, MSLP) the variance is a strong function of latitude. Since the skill is attempting to measure the likelihood of the model being within the range of the observations this seems appropriate, though it is not the method of Murphy et al., who use a global scaling. Note that the measure only assesses the climatological mean state, and not the interannual variability.

[4]   For all these reasons the method can only be semi-objective. Having identified the "best" and "worst" models overall we then examine how they fare in an assessment of individual variables. As well as comparing the individual models we also assess the skill of the average of all the models. Various authors [*Schmittner et al.*, 2005; *Hagedorn et al.*, 2005; *Lambert and Boer*, 2001] have pointed out that the skill of the average can be superior to any of the individual models.

3. Models and Data

[5]   Model data is retrieved from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset at https://esg.llnl.gov:8443/. 19 models were found which had the basic variables needed for this assessment, and are listed in Table 1. We also use the all-model average, denoted by AVG.

[6]   The models vary widely in terms of resolution, physical components and sophistication. In this paper we do not attempt to connect the skill scores to individual model structure.

[7]   To perform the evaluation we have selected mean sea level pressure (MSLP; with orography above 100 m masked), height and temperature at 500 hPa (H and T500; with areas over the highest of the Himalayas masked); sea surface temperature (TSFC; over the oceans, with sea ice

---

[1]British Antarctic Survey, Cambridge, UK.

Table 1. Model Identifying Number, Short Name, and Institute

Number	Name	Institute
1	BCCR BCM2	Bjerknes Center for Climate Research
2	CCCMA CGCM3	Canadian Centre for Climate Modelling and Analysis
3	CNRM CM3	Center National de Recherches Meteorologiques
4	CSIRO Mk3	Commonwealth Scientific and Industrial Research Organisation
5	GFDL CM2.0	Geophysical Fluid Dynamics Laboratory
6	GFDL CM2.1	Geophysical Fluid Dynamics Laboratory
7	GISS EH	Goddard Institute for Space Studies
8	GISS ER	Goddard Institute for Space Studies
9	IAP FGOALS1	Institute for Atmospheric Physics
10	INM CM3	Institute for Numerical Mathematics
11	IPSL CM4	Institut Pierre Simon Laplace
12	MIROC(hires)	Center for Climate System Research
13	MIROC(medres)	Center for Climate System Research
14	MPI ECHAM5	Max Planck Institute for Meteorology
15	MRI CGCM2	Meteorological Research Institute
16	NCAR CCSM3	National Center for Atmospheric Research
17	NCAR PCM1	National Center for Atmospheric Research
18	UKMO HadCM3	Met Office's Hadley Centre for Climate Prediction
19	UKMO HadGEM1	Met Office's Hadley Centre for Climate Prediction

areas masked) and surface mass balance over Antarctica (precipitation minus evaporation, SMB). These are representative of the large scale circulation (in the case of MSLP, H and T500) or provide an important variable of the climate system (TSFC and Antarctic SMB), and can be expected to be reliable in the re-analysis around Antarctica.

[8] Note that two of the GCMs that we have used are flux-corrected: MRI_CGCM and CCCMA_CGCM. These perform best at TSFC (as might be expected) where all the other models struggle; but are mid-ranking for other variables. TSFC has a comparatively low interannual variability, and hence relatively small errors lead to large values of RMS relative to variability and hence low weights. Since the scaled RMS errors are themselves combined via RMS, a poor result in one variable (e.g. GISS_E_H fares badly for MSLP) tends to result in a poor score overall.

[9] Observational validation data is from the ECMWF re-analysis [Uppala et al., 2005] for MSLP, H500 and T500;

from HadISST1 [Rayner et al., 2003] for TSFC; Vaughan et al. [1999] for SMB and Comiso [1999] for sea ice.

4. Overall Skill Scores

[10] We describe the results of the overall skill scores, combined from the five variables and considered for Antarctica, Globally and both combined. Overall skill scores show a strong variation, from above 0.4 to essentially zero. The best performing individual model, combining both Antarctica and globally, is MPI_ECHAM5; it also performs best globally; AVG is second. AVG scores best for Antarctica and UKMO_HADGEM is second. A few models (e.g. MPI_ECHAM5 and NCAR_PCM1) perform consistently well in the evaluation of different variables globally and for the Antarctic domain, but most exhibit a wide variation in skill between the two domains.

[11] The wide variation in skill scores means that when they are used to form a weighted mean of the models, the

Table 2. Skill Scores for the Five Individual Variables, Globally, and for Antarctica[a]

Model Name	Model	Both	Glob	Ant	Temperature Ant	Temperature Glob	MSLP Ant	MSLP Glob	T500 Ant	T500 Glob	H500 Ant	H500 Glob	SMB
Avg	0	0.40	0.32	0.46	0.13	0.12	0.76	0.67	0.47	0.34	0.72	0.40	0.63
bccr_bcm2_0	1	0.03	0.01	0.12	0.01	0.00	0.18	0.36	0.08	0.00	0.33	0.00	0.92
cccma_cgcm3_1	2	0.31	0.22	0.39	0.13	0.17	0.51	0.27	0.33	0.19	0.45	0.29	1.00
cnrm_cm3	3	0.07	0.08	0.06	0.00	0.00	0.03	0.14	0.28	**0.44**	0.06	0.37	0.91
csiro_mk3_0	4	0.13	0.05	0.28	**0.07**	0.01	0.45	0.22	0.38	0.02	0.31	0.08	0.43
gfdl_cm2_0	5	0.06	0.01	0.20	0.01	0.02	0.58	**0.57**	0.13	0.00	0.50	0.01	0.93
gfdl_cm2_1	6	0.11	0.05	0.21	0.01	0.06	**0.86**	0.46	0.12	0.01	0.73	0.02	0.84
giss_model_e_h	7	0.00	0.00	0.01	0.00	0.02	0.51	0.00	0.45	0.36	0.06	0.00	0.00
giss_model_e_r	8	0.07	0.04	0.11	0.02	0.02	0.44	0.02	0.09	0.16	0.24	0.05	0.07
iap_fgoals1_0_g	9	0.00	0.11	0.00	0.01	0.03	0.58	0.24	0.61	0.23	0.29	0.10	0.00
inmcm3_0	10	0.07	0.03	0.16	0.00	0.01	0.60	0.30	0.22	0.07	0.44	0.00	0.78
ipsl_cm4	11	0.08	0.05	0.11	0.00	0.02	0.09	0.28	0.39	0.10	0.11	0.01	0.88
miroc3_2_hires	12	0.19	0.11	0.29	0.01	0.04	0.48	0.12	0.57	0.09	0.66	0.36	**1.00**
miroc3_2_medres	13	0.09	0.04	0.17	0.02	0.02	0.41	0.30	0.13	0.04	0.39	0.02	0.28
mpi_echam5	14	**0.42**	**0.40**	0.45	0.05	**0.10**	0.73	**0.56**	**0.75**	**0.68**	0.72	**0.68**	0.90
mri_cgcm2_3_2a	15	0.10	0.07	0.15	0.24	0.24	0.67	0.45	0.15	0.06	0.60	0.00	0.00
ncar_ccsm3_0	16	0.14	0.05	0.28	0.01	**0.07**	0.38	0.30	**0.81**	0.19	0.65	0.00	0.95
ncar_pcm1	17	0.10	0.08	0.12	0.00	0.00	0.44	0.14	0.16	0.15	0.32	0.34	0.99
ukmo_hadcm3	18	0.27	**0.19**	0.36	0.02	0.04	0.63	0.16	0.72	0.27	**0.75**	**0.71**	**1.00**
ukmo_hadgem1	19	**0.30**	0.18	**0.45**	**0.09**	0.03	**0.75**	0.47	0.66	0.35	**0.74**	0.21	0.54

[a]The first column is the root mean square combination of all the individual scores; the second combines the "global" scores and the third the "Antarctic" scores. The remaining columns show the scores for individual variables. The best two models in each column are in bold. Note that MRI and CCCMA are flux-corrected and hence their TSFC skill would be expected to be high, and so are not bolded for the TSFC comparison.

Table 3. Sea Ice Skill Scores[a]

Model	Weight
0	0.42
1	0.01
2	0.05
3	0.02
4	**0.08**
8	0.02
9	0.00
10	0.03
11	0.00
12	0.07
14	0.01
15	0.14
16	0.03
17	0.02
18	0.01
19	**0.08**

[a]The two best (non-flux-corrected) models are in bold. Model numbers are given in Table 1.

top 4 models contribute 50% of the total and the bottom four models only 5%. Table 2 shows the overall skill score from five variables, showing global, Antarctic and combined scores.

[12] For TSFC, the skill of the average model depends considerably on the two flux corrected models, and is reduced to 0.05 globally if they are omitted. TSFC proves a hard variable to simulate, within the assessment we have used. This is because the interannual variability is low over much of the globe and hence errors of a few degrees, scaled by the variability, produce low scores. However, this is also to say that the model simulations are noticeably out of the range of natural variability; and small errors in SST are known to have large consequences.

5. Sea Ice

[13] *Holland and Raphael* [2006] examined 6 AR4 models, and *Parkinson et al.* [2006] examined the bi-polar performance of sea ice in 11 AR4 models. They noted that annual cycles are phased at least approximately correctly in both hemispheres but that some of the models simulate too much ice, others simulate too little ice (in some cases depending on hemisphere and/or season), and some match the observations better in one season than another. We use data from a slightly wider set of 15 coupled models and concentrate on a quantitative examination by skill scores. There are several possible measures of sea ice "skill;" either match to total area or the pointwise method used above for other variables. Ideally, we would match the measure of skill to the physical effects of the error in the model. Thus we choose to measure ice fraction (the proportion of grid covered by sea ice and not leads) rather than ice extent (the proportion of grid covered by ice of fraction at least 15%) in order to pick up errors in the ice concentration in the interior of the pack. This is because several models produce values for ice fraction that are notably below the observational values, whilst producing an overall ice extent that is moderately realistic. However the inaccurate (too low) fraction would lead to errors in the atmosphere-ocean fluxes, reflecting the important "function" of the ice in the climate system as an insulator between the ocean and atmosphere.

[14] When we form a skill score based on total ice area, and take the RMS across all months, we find that except for CSIRO the models have essentially zero skill. This is because, apart from CSIRO, all the models have months in which their total area falls well outside the observed range compared to satellite observations from *Comiso* [1999] using the bootstrap method, which verify best against other observations in Antarctica [*Connolley*, 2005]. All models produce a seasonal cycle with a peak in approximately the right season, though HadCM3 is a month late and NCAR CCSM two months early. IAP FGOALS has vastly overextensive ice, extending to South America.

[15] A measure based on the pointwise RMS difference from observed monthly climatology produces more usable rankings, in which MRI, CSIRO, HADGEM and MIRO-C_hires are the best, although even the best scores are low (Table 3). Clearly, a good simulation of Antarctic sea ice is a difficult challenge for a GCM. Of the models, most use a VP or EVP rheology, CSIRO uses cavitating fluid, and HADCM3 and MRI implement "ocean drift." Only INM has no ice advection. It is perhaps notable that the best performing model is then MRI, with the most primitive "rheology." However, the MRI model is flux-corrected globally, and this is likely to strongly affect the sea ice simulation. The next best, CSIRO, uses the relatively simple cavitating fluid rheology. This illustrates the fact [*Connolley*, 2005] that many aspects of the model simulation besides sea ice model quality goes into making up the simulation of the sea ice and provides no support for the need for a sophisticated sea ice dynamics scheme, although clearly if all else is equal a more sophisticated and physically plausible scheme will be preferable. Parkinson et al. note that some of these models, especially CSIRO, show rather lower skills in the Northern hemisphere and suggest that there may be some tuning to one hemisphere or other; we have only examined the sea ice in the SH in this paper.

[16] Sea ice is a highly variable quantity, and the measures used here cannot capture other aspects that affect its validation. Several models have sea ice whose pattern looks "wrong" even though the overall area is not too badly off, and hence the skill score is not strongly affected. For example, NCAR_PCM has a curious "spiral arm" structure apparently caused by anomalies within the ocean convection (E. Hunke, personal communication, 2006). Other models show little or no ice around large portions of the East Antarctic coasts or the Amundsen-Bellingshausen seas west of the Antarctic Peninsula.

[17] The model average displays a very significantly higher skill (0.42) than any of the individual models, presumably due to cancellation of errors. This superiority of the average model is far higher that in the case of the average compared to the other variables, as shown above. Parkinson et al. also noted the qualitative virtues of the model average. It is unclear whether this skill is an intrinsic part of the averaging process or simple chance.

6. Temperature Trends 1960–2000

[18] Station observations [*Turner et al.*, 2005] for winter show a maximum temperature trend since 1958 on the west side of the Antarctic peninsula, with smaller and generally non-significant changes around East Antarctica. Other syn-
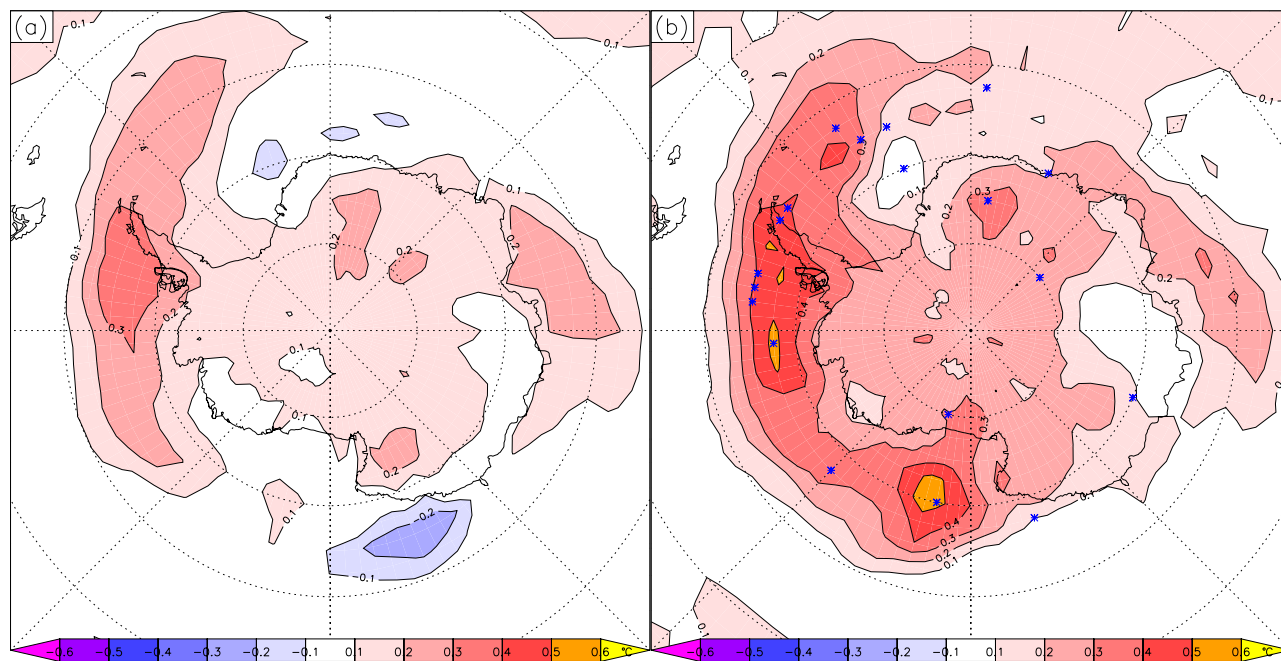
Figure 1. Temperature trends in °C/decade from 1960–2000 for winter (JJA). (a) Unweighted average of 19 models. (b) Weighted average. Also plotted on Figure 1b are the locations of the maximum trends from the individual models.

theses of temperatures across the continent show mixed trends both spatially and by season across East Antarctica with a tendency towards cooling [*Thompson and Solomon*, 2002; *Doran et al.*, 2002] but lack of in-situ observations precludes a conclusion for much of the area; surface temperatures from re-analyses are not reliable [*Connolley and Harangozo*, 2001]. *Chapman and Walsh* [2007] note that the trends are rather sensitive to start date and length. Where stations are present long-term observations are available for temperature trends, and trends are less stable than means over shorter periods, hence in this section we use data from 1960 to evaluate the temperature trends. In attempting to compare modelled and observed trends we are implicitly assuming that the observed trends are caused by the forcings imposed on the models: mostly greenhouse gas increases and ozone changes.

[19] We find that overall the model average, for JJA of 1960–1999, qualitatively reproduces the observed pattern (Figure 1). The maximum trend is, correctly, in the region of the west side of the Antarctic Peninsula. The skill-weighted average is similar to the simple average in pattern, but the warming is greater: for the polar cap south of 60S, the average is 0.23°C/decade (weighted) or 0.11°C/decade (unweighted). The trend just west of the Antarctic peninsula at (65S, 70E) is 0.47°C/decade (weighted) or 0.38 (unweighted), both of which are smaller than the observed value at Faraday (65.25S, 64.27W) of approximately 1°C/decade, though the weighted average is somewhat closer to observations. However the observed winter temperature at Faraday is highly variable and the trend depends strongly on the exact start date chosen. In the weighted average the trends around East Antarctica remain fairly small; warming over the continent itself increases somewhat. Observations show small and insignificant cooling at the pole, and smaller and insignificant warming at Vostok (78.5S, 106.9E).

[20] Individual models show great scatter in their trends; Figure 2 shows these, and Figure 1b shows the locations of the maximum trend in each model south of 60S. Only 4 of the individual models have their maximum trends in the "correct" place. Four models position it west of the Peninsula; a further two to the east; four in the Weddell Sea; three in the seas around East Antarctica; three over the continent itself (although the absolute magnitudes of these trends are small); one on the Ross ice shelf and two in the Ross sea. The average maximum (within the Antarctic domain) trend is 1.1°C/decade, three times higher than the average, but in rather better accord with observations. Some models (e.g. GISS_E_H) show small trends almost everywhere; however most models have large trends somewhere; but 8 of the 19 models have a greatest absolute trend that is negative. All the large model trends are over the sea ice rather than the continent, and are closely related to sea ice changes. In this they are behaving realistically, in that the observed winter trends around the Peninsula are reinforced by sea ice feedbacks [*King*, 1994]. The large warming in the Ross sea cannot be compared directly to observations. Since 1979, the period for which satellite observations of sea ice cover are available, sea ice has increased slightly in this region, suggesting that the warming modelled is unrealistic.

[21] Summer trends (not shown) are smaller, generally about 0.1°C/decade over the continent with little change over the surrounding seas. This is in rough agreement with observations, which tend to show small (but generally negative) trends in summer. When weighted, the trends become somewhat larger, as they do in winter, but remain small.

7. Surface Mass Balance

[22] Estimates of Antarctic surface mass balance (SMB) vary [*Uotila et al.*, 2007]; we shall use a central value of
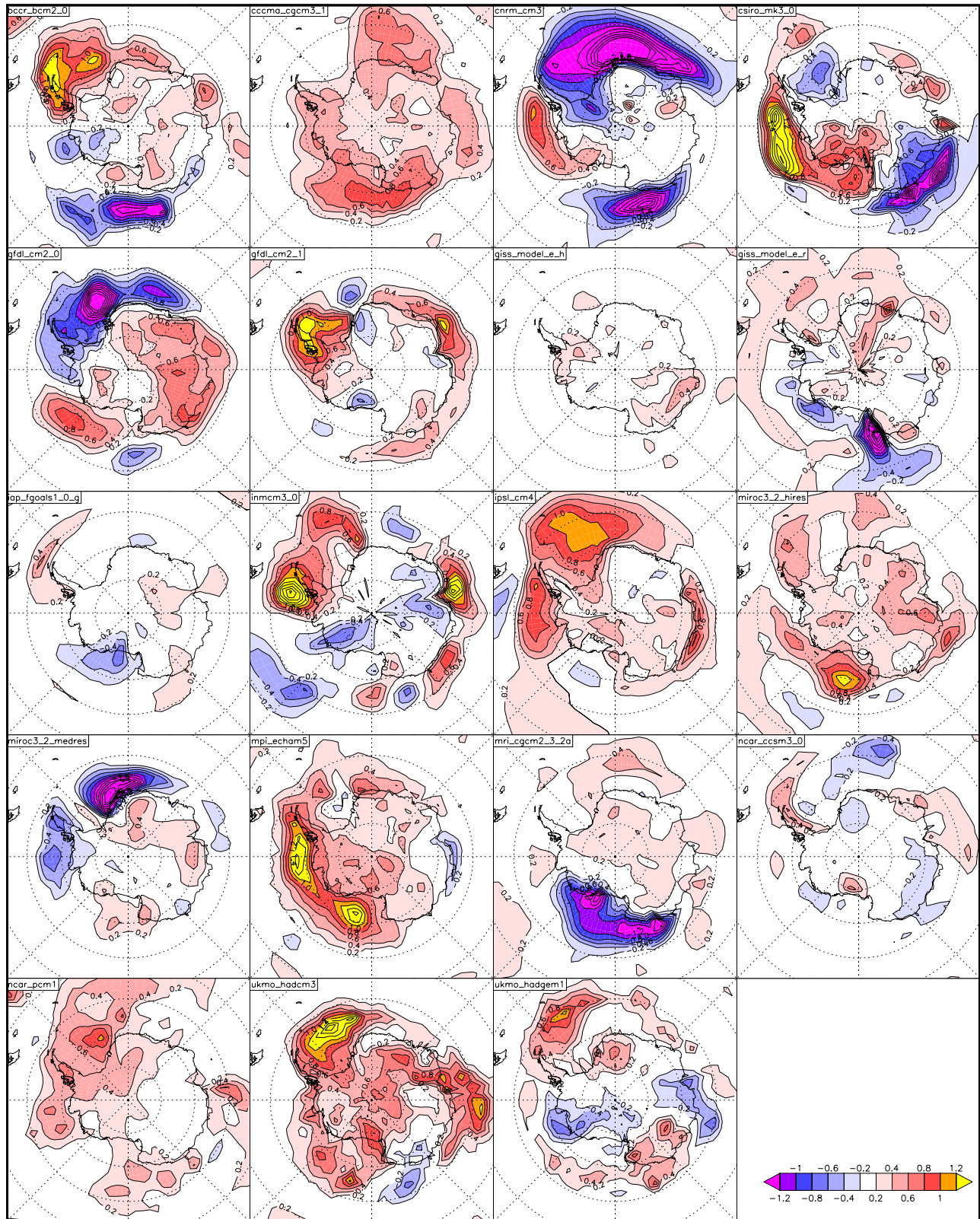
Figure 2. Temperature trends in °C/decade from 1960–2000 for winter (JJA) for individual models.

167 mm/yr from Vaughan et al. with a spread of 30 mm/yr which recognises the considerable spread in estimates from observations and models, and also allows for inter annual variation. Models indicate that this SMB is made up of mostly of precipitation, with sublimation removing approximately 10–20%. Other studies show that blowing snow and melt (which are ignored here) are small on a continental scale. 9 models have SMB within 15 mm/yr of 167. IAP_FGOALS greatly overestimates (500 mm/yr); the GISS models (despite having a large value for sublimation) and MRI overestimate by about 100 mm/yr, but for different reasons: GISS have a "central desert" area but it is too small; whereas MRI does not simulate the very low values of SMB in the interior. Only MIROC_medres (116 mm/yr) and HADGEM (131 mm/) substantially underestimate SMB. Of those models that do well on overall totals, two (BCCR and CNRM) nonetheless produce SMB simulations that, on inspection of their maps, are implausible: they fail to produce large (>500 mm/yr) SMB on and around the coast of East Antarctica.

[23] Model averages, as found above, often perform better than individual models. But in this case the unweighted average overestimates by nearly 30 mm/yr whereas the skill-weighted average (using only the four circulation indices and not using SMB itself) overestimates by only 15 mm/yr. The simple model average, which for circulation variables was better than all but a few models, is in this case worse than most. This can be explained by most models being approximately correct, and there being more outliers with much higher SMB than much lower due to the skewed distribution of precipitation.

8. Conclusions

[24] The AR4 models examined here show a wide range of skill scores when assessed against Antarctic or Global measures of large-scale circulation indices. Except for continental SMB, the model average proves a more reliable estimate than that for any one model, though it is usually not the best estimate for any given assessment. Overall, MPI_ECHAM5 and UKMO_HADGEM come first and second, though UKMO_HADGEM scores marginally higher just for the Antarctic domain. For SMB, due to a number of outliers with excessively high precipitation, the all-model simple average performs poorly but the skill-weighted (excluding SMB skill) average is good; for sea ice area the average is clearly superior to any individual model.

[25] Individual models show a very wide scatter in simulated temperature trends over the past century. In particular the large trend over the Antarctic peninsula in winter is not well represented, which makes it clear that whatever has been driving these trends is not well captured by many GCMs. Only a few individual models, notably MPI_ECHAM5, produce creditable simulations of what has been observed. The all-model average provides a reasonable pattern, but with too small a warming, although the skill-weighted model average does better. Trends in temperature are clearly linked to the sea ice simulation, another variable that most models do not simulate well.

[26] Use of skill scores provides a means of discriminating amongst the models. Skill-weighted averages improves

the simulation of temperature trends and SMB, though not by a large amount. They also have a role in assessing possible future changes: a companion paper [Bracegirdle et al., 2007] uses the model weights derived here to assess future impacts of climate change over Antarctica.

References

Bracegirdle, T. J., W. M. Connolley, and J. Turner (2007), Antarctic climate change over the 21st century, *J. Geophys. Res.*, doi:10.1029/2007JD008933, in press.

Bromwich, D. H., and R. L. Fogt (2004), Strong trends in the skill of the ERA-40 and NCEP-NCAR reanalyses in the high and midlatitudes of the Southern Hemisphere, 1958–2001, *J. Clim.*, 17(23), 4603–4619.

Chapman, W. L., and J. E. Walsh (2007), A synthesis of Antarctic temperatures, *J. Clim.*, 20(16), 4096–4117, doi:10.1175/JCLI4236.1.

Comiso, J. (1999), Bootstrap sea ice concentrations for NIMBUS-7 SMMR and DMSP SSM/I, June to September 2001, http://nsidc.org/data/nsidc-0079.html, Natl. Snow and Ice Data Cent., Boulder, Colo. (Updated 2005.)

Connolley, W. M. (2005), Sea ice concentrations in the Weddell Sea: A comparison of SSM/I, ULS, and GCM data, *Geophys. Res. Lett.*, 32, L07501, doi:10.1029/2004GL021898.

Connolley, W. M., and S. A. Harangozo (2001), A comparison of five numerical weather prediction analysis climatologies in southern high latitudes, *J. Clim.*, 14(1), 30–44.

Doran, P. T., et al. (2002), Antarctic climate cooling and terrestrial ecosystem response, *Nature*, 415(6871), 517–520.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting: I. Basic concept, *Tellus, Ser. A*, 57(3), 219–233.

Holland, M. M., and M. N. Raphael (2006), Twentieth century simulation of the Southern Hemisphere climate in coupled models. Part II: Sea ice conditions and variability, *Clim. Dyn.*, 26, 229–245.

King, J. C. (1994), Recent climate variability in the vicinity of the Antarctic Peninsula, *Int. J. Climatol.*, 14(4), 357–369.

Lambert, S. J., and G. J. Boer (2001), CMIP1 evaluation and intercomparison of coupled climate models, *Clim. Dyn.*, 17(2–3), 83–106.

Murphy, J. M., et al. (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430(7001), 768–772.

Parkinson, C. L., K. Y. Vinnikov, and D. J. Cavalieri (2006), Evaluation of the simulation of the annual cycle of Arctic and Antarctic sea ice coverages by 11 major global climate models, *J. Geophys. Res.*, 111, C07012, doi:10.1029/2005JC003408.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108(D14), 4407, doi:10.1029/2002JD002670.

Schmittner, A., M. Latif, and B. Schneider (2005), Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations, *Geophys. Res. Lett.*, 32, L23710, doi:10.1029/2005GL024368.

Thompson, D. W. J., and S. Solomon (2002), Interpretation of recent Southern Hemisphere climate change, *Science*, 296(5569), 895–899.

Turner, J., S. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M. Carleton, P. D. Jones, V. Lagun, P. A. Reid, and S. Iagovkina (2005), Antarctic climate change during the last 50 years, *Int. J. Climatol.*, 25(3), 279–294.

Uotila, P., A. H. Lynch, J. J. Cassano, and R. I. Cullather (2007), Changes in Antarctic net precipitation in the 21st century based on Intergovernmental Panel on Climate Change (IPCC) model scenarios, *J. Geophys. Res.*, 112, D10107, doi:10.1029/2006JD007482.

Uppala, S. M., et al. (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc.*, 131, 2961–3012, doi:10.1256/QJ.04.176.

Vaughan, D. G., J. L. Bamber, M. Giovinetto, J. Russell, and A. P. R. Cooper (1999), Reassessment of net surface mass balance in Antarctica, *J. Clim.*, 12(4), 933–946.

———————————

T. J. Bracegirdle and W. M. Connolley, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK. (wmc@bas.ac.uk)