

Global Climate Model Performance over Alaska and Greenland

John E. Walsh¹, William L. Chapman², Vladimir Romanovsky³,
Jens H. Christensen⁴ and Martin Stendel⁴

¹International Arctic Research Center, University of Alaska, Fairbanks, AK

²Department of Atmospheric Sciences, University of Illinois, Urbana, IL

³Geophysical Institute, University of Alaska, Fairbanks, AK

⁴Danish Climate Centre, Danish Meteorological Institute, Copenhagen, Denmark

July, 2007

Revised January 2008

Corresponding author:

John E. Walsh
International Arctic Research Center
University of Alaska
930 Koyukuk Drive
Fairbanks, AK 99775-7350
email: jwalsh@iarc.uaf.edu

Abstract

The performance of a set of fifteen global climate models used in the Coupled Model Intercomparison Project is evaluated for Alaska and Greenland, and compared with the performance over broader pan-Arctic and Northern Hemisphere extratropical domains. Root-mean-square errors relative to the 1958-2000 climatology of the ERA-40 reanalysis are summed over the seasonal cycles of three variables: surface air temperature, precipitation and sea level pressure. The specific models that perform best over the larger domains tend to be the ones that perform best over Alaska and Greenland. The rankings of the models are largely unchanged when the bias of each model's climatological annual mean is removed prior to the error calculation for the individual models. The annual mean biases typically account for about half of the models' root mean square errors. However, the root-mean square errors of the models are generally much larger than the biases of the composite output, indicating that the systematic errors differ considerably among the models. There is a tendency for the models with the smaller errors to simulate a larger greenhouse warming over the Arctic, as well as larger increases of Arctic precipitation and decreases of Arctic sea level pressure when greenhouse gas concentrations are increased. Since several models have substantially smaller systematic errors than the other models, the differences in greenhouse projections imply that the choice of a subset of models may offer a viable approach to narrowing the uncertainty and obtaining more robust estimates of future climate change in regions such as Alaska, Greenland and the broader Arctic.

1. Introduction

Global climate models are the most widely used tools for projections of climate change over the timescale of a century. The periodic assessments by the Intergovernmental Panel on Climate Change (IPCC) have relied heavily on global model simulations of future climate driven by various emission scenarios. The global model simulations show a polar amplification of the greenhouse-driven warming and of other variations in climate (Serreze and Francis, 2006; Wang et al., 2007), although the ratio of the models' projected changes to the natural variability is not necessarily greater in the Arctic than in lower latitudes (Kattsov and Sporyshev, 2006).

Given the likelihood that the Arctic will experience greater climate changes than most other regions over the next century, the credibility of the model simulations of Arctic climate becomes a key issue. The absence of databases for validation of future climate simulations increases the importance of evaluations of models' ability to simulate recent climate, for which syntheses of observational data are available.

Greenhouse-driven climate change represents a response to the radiative forcing associated with increases of carbon dioxide, methane, water vapor and other radiatively active gases, as well as associated changes in cloudiness. The response varies widely among models because it is strongly modified by feedbacks involving clouds, the cryosphere, water vapor and other processes whose effects are not well understood. While changes in the radiative forcing associated with increasing greenhouse gases have

thus far been relatively small (only a few Watts per square meter -- IPCC, 2007), a much more potent change in forcing occurs each year through the seasonal cycle of solar radiation. In the present paper, we place the models' ability to capture the *seasonal cycle* of present-day climate at the core of a strategy for evaluating the models' simulations of Arctic climate. Our evaluation is motivated by regional applications of the climate model output in the Arctic, specifically the climatically sensitive regions of Alaska and Greenland. Both these regions have surface states that can be fundamentally altered by relatively small climate changes: much of Alaska is underlain by permafrost (thick and continuous in the North, discontinuous and thin in much of the interior), while Greenland is dominated by an ice sheet that may already be responding to climate change (Rignot and Kanagaratnam, 2006; Dowdeswell, 2006). Greenland also has permafrost in much of the narrow strip of land between the ice sheet and the surrounding seas.

The analysis of the model results described here is directed at the following questions:

- How does model performance over Alaska and Greenland compare with performance over the broader pan-Arctic and the Northern Hemisphere domains? Specifically, are the models with the smallest errors in the broader Northern Hemisphere also the models with the smallest errors in the Arctic and, particularly, in Alaska and Greenland?
- Are the models' errors attributable primarily to relatively uniform biases, i.e., to the fact that the models are consistently too warm, too cold, too wet, too dry, etc?

- Do 21st-century projections of Arctic change show any systematic dependence on the validity of the different models' simulations of present-day climate?

2. Model output and validation data

Our evaluation is based on the 20th-century simulations by the models used in the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC, 2007). These models are being used in the third phase of the Climate Model Intercomparison Project (CMIP, <http://www-pcmdi.llnl.gov/projects/cmip/index.php>) and will hereafter be referred to as the CMIP3 models. The output used here consists of the monthly grids of surface air temperature, precipitation and sea level pressure for 1958-2000, which is a subperiod of the 20th-century simulations by these models and is also the period spanned by the validation fields (see below). Most of the model simulations were begun in the 1800s and continued through 2000 with prescribed greenhouse gas concentrations and, in some cases, estimated sulfate aerosols and variable solar forcing (Table 1; see discussion in Wang et al. (2007)). Most of the simulations were continued through the 21st-century with forcing prescribed from the SRES emission scenarios (A2, A1B, B1, etc.) of the IPCC (Nakicenovic et al., 2000). For the evaluation performed in this study, we use only the output from the 20th-century simulations to evaluate the models' performance.

The CMIP3 model output is compared here against the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis, ERA-40), which directly assimilates

observed air temperature and sea level pressure (SLP) observations into a reanalysis product spanning 1958-2000. Precipitation is computed by the model used in the data assimilation, i.e., observed precipitation data are not assimilated into the reanalysis. The ERA-40 is one of the most consistent and accurate gridded representations of these variables available, and it compares favorably with other reanalyses of the Arctic (Bromwich et al., 2007). It is therefore a logical choice for observational analyses from which we determine the model biases of late-twentieth-century surface air temperatures, precipitation and sea level pressure (SLP). (Data and documentation for the ERA-40 can be found online at <http://www.ecmwf.int/research/era/Products>.) While the ERA-40 reanalysis was performed at T106 (~125 km) resolution with 60 levels, we use the version of the output archived on a 2.5° latitude x 2.5° longitude grid for compatibility with the climate model output. Since our evaluation is limited to surface air temperature, precipitation and sea level pressure, no upper-air levels of the reanalysis output are used here.

To facilitate global climate model (GCM) intercomparison and validation against the reanalysis data, all monthly fields of GCM temperature, precipitation and SLP are interpolated to the common 2.5° × 2.5° latitude–longitude ERA-40 grid. Our evaluation of the models' simulated fields uses monthly, seasonal, and annual climatological means for the late-twentieth-century period 1958-2000.

3. Validation method

The core statistic of our validation is the root-mean-square error (RMSE) evaluated from the differences between ERA-40 and each model for each grid point and calendar month.

For those models for which ensembles of simulations were archived, we use only the first ensemble member for consistency with models having only one simulation in the CMIP3 archive. In all cases, the differences are between climatological means for the 1958-2000 period. The RMSE calculations are performed for each of the fifteen models, for each calendar month, and area-weighted for each of four domains: Alaska (shown in Fig. 1), Greenland (shown in Fig. 1), the “pan-Arctic” polar cap (60° - 90° N) and a middle-high latitude “Northern Hemisphere” domain (20° - 90° N). The Alaskan and Greenland domains are approximately equal in area and are contained within the Arctic and Northern Hemisphere domains (except for a small portion of southeastern Alaska), so the results for the various domains are not independent. The reason for our choice of the four overlapping domains is that, although our primary interest is in the models’ performance for Alaska and Greenland, the simulation of present-day and future climate in these regions will depend on the simulation of regions from which weather systems move toward Greenland and Alaska. In particular, the larger-scale circulation over much of the Northern Hemisphere influences Alaska and Greenland via advection and teleconnections, so credible simulations of future changes will depend on the models’ ability to capture the large-scale circulation of the pan-Arctic and Northern Hemisphere domains.

While the RMSEs were evaluated for all three variables (temperature, precipitation, sea level pressure), the calculation of the RMSE for SLP differed from the calculation of the other RMSEs by the removal of the domain averages from all SLP grids (ERA-40 and CMIP3 models). This was done because the spatial gradients are the key features of the SLP fields, as it is the gradients of pressure that determine the winds.

As a seasonally inclusive measure of the models' success in simulating the regional and larger-scale climates, we sum the RMSE values over the 12 calendar months. In this respect, we are evaluating the models' ability to simulate the seasonal cycle and hence the models' sensitivities to the cycle of solar forcing. The use of the seasonal cycle as a "surrogate" measure of climate sensitivity is based on the premise that the response to seasonal changes in solar radiation should depend on at least some of the same processes that determine a model's response to other changes of forcing (Tsushima et al., 2005; Nall and Qu, 2006). This reasoning has motivated a variety of past studies, and the results are somewhat mixed. Lindzen et al. (1995), for example, found that the seasonal cycle of *globally averaged* temperatures did not correspond well with the signature of CO₂-induced climate change in one particular model, primarily because of differences in the seasonal cycles of globally averaged temperature at the surface and in the middle troposphere. On the other hand, Knutti et al. (2006) evaluated relationships between the CMIP3 models' climate sensitivities (defined as the equilibrium global temperature response to a global change in forcing, e.g., from increasing greenhouse gas concentrations) and the amplitudes of their seasonal cycles. They found that the models with the largest climate sensitivities tended to overestimate the seasonal cycle of surface air temperature when compared with observations, and concluded that "the amplitude of the seasonal cycle in temperature provides a strong constraint on climate sensitivity" (Knutti et al., p. 4232). Hall and Qu (2006) took this approach a step further and focussed on the snow albedo feedback as a determinant of both the amplitude of the seasonal cycle and of climate sensitivity. They found that the surface albedo-temperature sensitivity in the seasonal cycle was an excellent predictor of this sensitivity in

greenhouse-driven climate change simulations. The snow albedo feedback is especially pertinent to the present study, since snowmelt in Alaska occurs primarily during April and May when insolation is relatively strong. Earlier work by Hall (2004) showed that about half the high-latitude response to greenhouse forcing was attributable to the albedo feedback associated with retreating snow and ice.

After summation of the regional mean RMSEs over the 12 calendar months, the sums are used to rank the models. The model with the smallest 12-month sum of the RMSE is ranked #1 for that variable and region, while the model with the largest 12-month sum of the RMSE is ranked #15. The ranks can then be summed over different variables and/or different regions, depending upon a user's priorities for variables and regional emphasis. The raw RMSE values for individual months also enable users to assess the utility of a particular model for a particular month or season.

One disadvantage to the above procedure is that it does not distinguish between two contributions to the RMSE: the portion due to the bias in a model's annual mean, and the errors in the seasonal cycle that are superimposed on the model's annual mean. In order to address this distinction, we calculate a second set of RMSEs based on the models' errors after removal of the annual mean error (bias). The differences between these two sets of RMSEs, both of which are presented in the following section, indicate the contributions of the annual mean biases to the total RMSEs.

4. Results

The Root Mean Square Errors vary widely among models, across seasons, and across regions. In order to illustrate the seasonal and regional dependencies of the RMSEs, the *median* model RMSE values of each variable for each season are shown in Figures 2, 3 and 4 for surface air temperature, precipitation and sea level pressure, respectively. In all cases, the RMSEs computed after removal of the annual mean bias are shown as grey bars adjacent to the total RMSEs (black bars). It is apparent from Figure 2 that approximately half of the total RMSE of temperature is attributable to the bias in the annual mean temperature, although the removal of the annual mean has a smaller impact on the RMSE in summer than in winter. Largely because the mean biases are smaller in summer than in winter, the total RMSEs are smaller in summer than in winter in all domains. The model-median temperature RMSE is generally smaller for Alaska and Greenland than for the entire Arctic polar cap (60-90°N) except in summer, although it is comparable to the broader hemispheric (20-90°N) values. It should be noted that the quality of the ERA-40 surface air temperatures may be worse over the Arctic Ocean (roughly 75-90°N) and its peripheral seas than over the northern land areas due to uncertainties in the prescribed sea ice concentrations.

Precipitation is unique among the three variables in the sense that the removal of the annual mean bias generally increases the RMSE (Figure 3), although the increase over the Alaskan domain is small (and slightly negative in winter, spring and autumn). The interpretation is that the removal of the annual mean bias actually increases the bias in those seasons with the largest precipitation amounts (summer, autumn). By either measure of RMSE, the precipitation RMSEs for Alaska and Greenland are larger than those for the entire Arctic polar cap (but smaller than for the larger 20-90°N domain). The

precipitation RMSEs for Alaska and Greenland are smallest in spring, which is generally a dry season in both regions.

The RMSEs of sea level pressure (Figure 4) show a similar seasonal cycle -- largest in winter -- to those of temperature, although it should be noted that the actual magnitudes of the temperature and SLP RMSEs are not directly comparable because of the different units. The annual mean bias accounts for a larger portion of the total RMSEs in the Alaskan and Greenland domains than in the larger domains, especially 20-90°N.

A substantial portion of the RMSE arises from biases in the models' annual means, upon which the seasonal cycles are superimposed. The gray bars of Figs. 2-4 show the RMSE errors after removal of the model's annual mean biases. It is apparent that the impact of the removal of the annual mean biases is greater for the smaller domains (Alaska and Greenland), and the impact is also generally greater in the warm season (especially when the reduction of the RMSE error is viewed as a percentage of the total RMSE (black bars in Figs. 2-4)). However, the relative areal coverage of land fraction vs. ocean in the smaller domains is greater than the land fraction of the larger domains and the variability of temperature and precipitation are generally greater over land than ocean.

This paper's main objective is an identification of the models that are most successful at simulating the seasonal cycle of the climates of Alaska and Greenland. The relevant information is contained in Figures 5-7, which show the 12-month mean RMSEs of the different models, arranged in order of increasing RMSE, for the three variables. Each figure contains a separate display for (a) Alaska, (b) Greenland, (c) the pan-Arctic, 60-90°N, and (d) the extratropical Northern Hemisphere, 20-90°N. It is apparent from these

figures, especially Figure 5 for temperature, that the models vary widely in their ability to capture the seasonal climates of Alaska and Greenland. For example, the yearly averaged RMSE of temperature over Alaska varies from 2.9°C in MPI-ECHAM5 to 11.0°C in IAP-FGOALS. The range for the pan-Arctic domain (60-90°N) is even greater, from 2.9°C to 13.6°C. While the ranges are smaller for the other variables, the RMSEs still vary across the models by nearly a factor of two for precipitation and by more than a factor of two for sea level pressure. Similar ranges are found for the larger domains, i.e., the pan-Arctic and hemispheric polar caps.

Nevertheless, a noteworthy feature of Figures 5-7 is the tendency for some models to rank highest no matter which method variable is evaluated. Moreover, the models that rank highest for Alaska tend to rank higher for Greenland. There is also a tendency for these same models to have the smallest RMSEs over the larger domains, although there are exceptions. Table 2 provides a synthesis of the model performance based on the two RMSE metrics -- annual mean bias included (Table 2a) and annual mean bias removed (Table 2b). These tables rank the models from 1 (smallest RMSE) to 15 (largest RMSE) for each variable and domain. As in Figures 5-7, these ranks are based on RMSEs summed over all twelve calendar months, so they incorporate the models' successes or failures in capturing the seasonal cycle. The right-most columns of Tables 2a and 2b are the sums of all 12 ranks (4 domains x 3 variables) of the models. We refer to these columns as our "integrated ranks". Because the domains are nested, the "integrated rank" effectively double-weights the model performance over the Arctic polar cap (60-90°N) and triple-weights the model performance over the Alaskan and Greenland regions (i.e.,

Alaska and Greenland are included in both larger domains). While this weighting is admittedly *ad hoc*, it is consistent with our focus on Alaska and Greenland.

If the model ranks are based on individual variables or domains, the order shifts somewhat. The top-ranking models for the different domains, based on the computational strategies used for Tables 2a and 2b, are as follows:

Ranks with annual mean biases included in RMSEs (Table 2a):

Rank	<u>Alaska</u>	<u>Greenland</u>	<u>60-90°N</u>	<u>20-90°N</u>
1.	GFDL CM 2.0	GFDL CM 2.1	MPI-ECHAM5	MPI-ECHAM5
2.	GFDL CM 2.1	MPI-ECHAM5	GFDL CM 2.1	GFDL CM 2.1
3.	UKMO HADCM3	MIROC3	MIROC3	MIROC3 (tie) CCCMA

Ranks with annual mean biases removed (Table 2b):

	<u>Alaska</u>	<u>Greenland</u>	<u>60-90°N</u>	<u>20-90°N</u>
1.	MPI-ECHAM5	MIROC3	MPI-ECHAM5	MPI-ECHAM5
2.	GFDL CM 2.1	GFDL CM2.1	GFDL CM 2.1	GFDL CM 2.1
3.	CNRM CM3	CNRM CM3	NCAR CCSM3	GFDL CM2.0

Based on the aggregate of the two sets of rankings, the top-performing models for Alaska are GFDL-CM2.1, MPI-ECHAM5, CNRM CM3, UKMO-HADCM3 and MIROC3. The top-performing models for Greenland are GFDL-CM2.1, MIROC3, CNRM CM3 and MPI-ECHAM5. For each of the two larger domains (60-90°N 20-90°N), the two highest-ranking models are MPI-ECHAM5 and GFDL CM2.1, while the MIROC3 model ranks 3rd and 4th, respectively, for these two domains. Thus there is strong overlap among the top performers for the larger domains (pan-Arctic and Northern Hemisphere) and the subregional domains (Alaska and Greenland).

When the ranks for a particular variable are summed over the four domains, the ranks for the three primary variables are:

Ranks with annual mean biases included in RMSEs (Table 2a):

	<u>Temperature</u>	<u>Sea Level Pressure</u>	<u>Precipitation</u>
1.	MIROC3	MPI-ECHAM5	GFDL CM 2.1
2.	GFDL CM 2.1	GFDL CM 2.1	MPI-ECHAM5
3.	MPI-ECHAM5	CCCMA	UKMO HADCM3

Ranks with annual mean biases removed (Table 2b):

	<u>Temperature</u>	<u>Sea Level Pressure</u>	<u>Precipitation</u>
--	--------------------	---------------------------	----------------------

1.	MPI-ECHAM5	CNRM CM3	MPI-ECHAM5
2.	GFDL CM 2.1	GFDL CM 2.1	GFDL CM2.1
3.	MIROC3	MPI-ECHAM5	IPSL CM4

The above summaries, together with Table 2, confirm a conclusion of many other climate model evaluations, i.e., no single model outperforms all others for all regions or for all variables. On the other hand, several of the CMIP3 models consistently rank close to the top, demonstrating that their ability to reproduce the seasonal cycle of high latitude climate of recent decades is superior to that of the other CMIP3 models. The MPI ECHAM5 and GFDL CM2.1 models are clearly the highest ranking models overall. The MIROC3 and UKMO-HADCM3 models also rank highly. These higher-ranking models are the logical candidates for driving off-line simulations of high-latitude variables such as permafrost, glaciers/ice sheets, and terrestrial or marine ecosystems.

With regard to the reasons for the different levels of skill over Alaska and Greenland (as well as the larger domains), no systematic relationship to model resolution emerged. The models with the smallest RMSEs, MPI-ECHAM5 and GFDL CM 2.1 (Table 2), have resolutions that are neither highest nor lowest of the 15 models. There is also no obvious relationship between model performance and the type of sea ice formulation in the models. Other candidates for explanations of the differences in model performance include the cloud and radiative formulations, which are now being investigated elsewhere (Chapman et al., 2008); the planetary boundary layer parameterization; and the land

surface schemes of the various models. Biases in the large-scale atmospheric circulation, perhaps driven by processes outside the Arctic, are also candidates to explain the across-model differences in temperature, precipitation and sea level pressure.

The two highest ranking models (MPI-ECHAM5 and GFDL-CM2.1) both included aerosol effects and ice-phase clouds in model versions that were implemented in 2005, which is more recent than the implementation date of most of the other CMIP3 models. Systematic experiments with these and other parameterizations are needed if the reasons for the relative success of these two models is to be established unambiguously. Nevertheless, there is a general tendency for the more recently implemented model versions to be the better performers in the Arctic.

In order to provide some perspective for the RMSEs of the different variables, we show the models' bias fields for January in Figures 8-10. The composite fields included in these figures are based on the 14-model subset that excludes the IAP-FGOALS model, which is such an outlier (Figs. 5-7) that it skews the composite bias fields. Figure 8 shows that the 14-model composite biases of January temperature are generally less than about 3°C over most of the Arctic, except over the Barents Sea and far eastern Russian regions. The Barents Sea bias arises from the models' tendency to over-simulate the extent of sea ice in this area during winter. There are more areas of negative (cold) bias than of positive (warm) bias, although the biases are generally no larger in the Arctic than in middle latitudes. The 14-model composite field averages out many biases of opposite sign in the various models. The January bias fields for the individual models show that larger biases occur in some models, and that there are both positive and negative biases

across the models at any location. Particularly noteworthy is the large negative bias over southern Alaska in MPI-ECHAM5; this wintertime bias degrades the Alaskan temperature rank of this model, which otherwise ranks very highly in the Arctic.

The precipitation biases (Figure 9) are more spatially complex than the biases of temperature and pressure, and they show a distinct orographic signature in the 14-model composite and in most of the individual models. The negative (dry) biases of the models over the coastal mountains of southern Alaska, western Canada, southeastern Greenland and western Norway are indicative of a smoothing of the mountains that are subject to upslope flow in the major storm track regions. The corresponding wet biases immediately inland of the dry upslope biases in some areas also point to the models' over-smoothing of the topography, which is unable to create the full "precipitation shadow" effect that occurs downstream of major mountain ranges. The raw output from the ERA-40 reanalysis has much finer resolution than the CMIP3 models (Table 1) and is hence better able to capture the pattern of heavy upslope flow and leeside "shadowing". This pattern is strikingly apparent along the northwest coast of North America (Figure 9), where the models are too dry on the windward side and too wet on the lee side of the coastal mountains. This bias is much less apparent in the summer, when the bias fields of most models show no orographic signature. These error patterns point to the importance of downscaling GCM output fields when constructing site-specific climate change scenarios, especially for temperature and precipitation at locations in or near highly varying terrain.

The composite sea level pressure biases, shown in Figure 10, are also generally small, with magnitudes in the 14-model composite being less than 3-4 hPa almost everywhere in the middle and high latitudes. Greenland and the eastern Arctic Ocean show positive biases, while the biases over Alaska and the North Pacific are negative. The biases over the Bering Sea and Alaska imply cold advection over eastern Asia, consistent with the cold bias in that area during January (Figure 8, upper left). The positive bias of 3-6 hPa over the eastern Arctic Ocean is similar to the errors found by Bitz et al. (2002) in the previous generation of global atmospheric models. Individual models show much larger biases of both signs (Figure 10).

The final question raised in Section 1 pertained to possible relationships between the models' projected greenhouse changes and the relative accuracy of the simulations of present-day climate. Figure 11a shows the models' projected warming for 60-90°N plotted as a function of the integrated rank of the models. The integrated rank used here is based on Table 2a and includes only the 20-90°N and 60-90°N domains. (While the Alaska- and Greenland-specific domains are excluded because the projected warming is for the entire pan-Arctic domain, this makes little difference in Figure 11 because the ranks across the various domains are highly correlated). The warming is defined as the area-weighted linear change in surface air temperature from 2001 to 2099 under the IPCC-A1B scenario. While there is considerable scatter in the warming, there is a tendency for the highest-ranking (better performing) models to simulate the greatest warming over the 60-90°N domain. The projected warming vs. model performance relationship is statistically significant at the 5% level. The projected changes of precipitation (Figure 11b) show a similar dependence on the models' integrated ranks,

with better performing models projecting larger increases of Arctic precipitation than the lower-ranking models. While there is more scatter in the corresponding SLP results (Figure 11c), there is a tendency for larger 21st-century decreases in Arctic sea level pressure to be projected by higher-ranking models than by lower-ranking models.

The projections summarized above are based on an intermediate scenario (A1B) of greenhouse forcing. Because the models' different climate sensitivities do not depend on the precise rates of increase of forcing, the models that produce the stronger (weaker) warming under A1B forcing tend to produce the stronger (weaker) warming under the A2 and B1 forcing scenarios (IPCC, 2007, p. 763, Fig. 10.5). Hence the results in Fig. 11 do not depend strongly on the choice of the A1B scenario.

The dependence of the Arctic's climate sensitivity on the RMSE used here may be considered surprising, since it is just as likely that "bad" models are biased warm as cold. This finding raises the question about whether the sensitivity is correlated with the amplitude of the seasonal cycle or with the bias. The correlation between the projected warming and the amplitude of the simulated present-day seasonal cycle was found to be small and insignificant for the 60-90°N domain on which the results in Fig. 11 were based.

5. Conclusions

This study represents an early step in evaluation of global climate model performance over Arctic subregions, with an eye toward the narrowing of the uncertainty in regional Arctic climate projections by the selection of an optimal subset of models. It must be

emphasized that the ranking process used here is targeted at particular applications (i.e., forcing of permafrost models, planning for climate change in Alaska and Greenland) and hence contains arbitrary elements. For example, the assignment of weights based on an *ad hoc* postulated equal importance of different variables and regions will not be appropriate for other regions and applications. We present our approach and results in the spirit that they may stimulate enhancements or better alternative approaches to regional applications of climate model scenarios.

Several questions about model performance were posed in Section 1. Based on the results presented on the preceding section, the answers to those questions are as follows:

Model performance over Alaska and Greenland is generally neither better nor worse than the performance over the pan-Arctic and Northern Hemisphere domains. During winter and autumn, the temperature errors over Alaska and Greenland are somewhat smaller, but the precipitation errors are generally larger, than over the larger domains. The specific models that perform best over the larger domains tend to be the ones that perform best over Alaska and Greenland, although a notable exception is the relative lack of success of the overall top-ranked model (MPI-ECHAM5) in capturing the winter temperatures over Alaska.

The rankings of the models are largely unchanged when the bias of each model's climatological annual mean is removed prior to the error calculation for the individual models. The annual mean biases typically account for about half of the models' root mean square errors. Thus the root-mean-square errors are not simply the manifestations of general tendencies for the models to be colder, warmer, wetter, drier, etc., than the

corresponding observationally-derived fields. However, the root-mean square errors of the models are generally much larger than the biases of the composite output, indicating that the systematic errors differ considerably among the models.

There is a tendency for the models with the smaller errors to simulate a larger greenhouse warming over the Arctic. Since several models have substantially smaller systematic errors than the other models, the differences in warming imply that the choice of a subset of models may offer a viable approach to narrowing the uncertainty and obtaining more robust estimates of future climate change in regions such as Alaska, Greenland and the broader Arctic. The results in Figure 11 suggest that the uncertainty might be narrowed by eliminating models with the weaker projected warming in the Arctic, as those models tend to have the largest errors in simulations of the present-day climate. Such an approach has already been suggested by Overland and Wang (2007) and Kattsov and Sporyshev (2006), and will be pursued in future assessments of Arctic change.

Acknowledgments

This work was supported by the National Science Foundation, Office of Polar Programs, through Grant OPP-0612533 to the University of Alaska, Fairbanks, and by Grant OPP-0520112 to the University of Illinois. 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 data, the JSC/CLIVAR Working Group on Coupled Modeling (WGCM) and their Coupled Model Intercomparison Project (CMIP) the Climate Simulation Panel for organizing the model data analysis activity, and the IPCC Working Group I's TSU (Technical Support Unit)

for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. Finally, we thank three anonymous reviewers for comments and suggestions that improved the manuscript.

REFERENCES

Bitz, C. M., J. C. Fyfe and G. M. Flato, 2002: Sea ice response to wind forcing from AMIP models. *J. Climate*, *15*, 522-536.

Bromwich, D. H., R. L. Fogt, K. I. Hodges and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.*, *112*, doi:10.1029/2006JD007859.

Chapman, W. L., D. H. Portis and J. E. Walsh, 2008: Cloud-radiative interactions in global climate models. *Int. J. Climatol.*, submitted.

Dowdsewell, J. A., 2006: The Greenland Ice Sheet and global sea level rise. *Science*, *311*, 963.

Hall, A., 2004: The rold of surface albedo in climate. *J. Climate*, *17*, 1550-1568.

Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Resl. Lett.*, *33*, L03502, doi:10.1029/2005GL025127.

IPCC, 2007: The Physical Basis of Climate Change. Working Group I, Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, U.K., 996 pp..

Kattsov, V. M., and P. V. Sporyshev, 2006: Timing of global warming in IPCC AR4 AOGCM simulations. *Geophys. Res. Lett.*, *33*, L23707, doi:10.1029/2006GL027476.

Knutti, R., G. A. Meehl, M. A. Allen and D. A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Climate*, *19*, 4224-4233.

Lindzen, R. S., B. Kirtman, D. Kirk-Davidoff and E. S. Schneider, 1995: Seasonal surrogate for climate. *J. Climate*, *8*, 1681-1684.

Nakicenovic, N. and 27 coauthors, 2000: Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, U.K., 599 pp.

Overland, J. E., and M. Wang, 2007: Future climate of the North Pacific Ocean. *EOS*, *88(16)*, 178.

Rignot, E., and P. Kanagaratnam, 2006: Changes in the velocity structure of the Greenland Ice Sheet. *Science*, *311*, 986-990.

Serreze, M. C., and J. Francis, 2006: The Arctic amplification debate. *Climatic Change*, *76*, 241-264.

Tsushima, Y., A. Abe-Ouchi and S. Manabe, 2005: Radiative damping of annual variation in global mean temperature: Comparison between the observed and simulated feedback. *Clim. Dyn.*, *24*, 591-597.

Wang M., J. E. Overland, V. Kattsov, J. E. Walsh, X. Zhang, and T. Pavlova, 2007:
Intrinsic versus forced variation in coupled climate model simulations over the Arctic
during the twentieth century. *J. Climate*, 20, 1093-1107.

FIGURE CAPTIONS

Fig. 1. Map showing Alaskan and Greenland domains, also highlighting 60-90°N.

Fig. 2. Bar graph showing 15-model median RMSE of temperature by season for (a) Alaska, (b) Greenland, (c) 60-90°N, and (d) 20-90°N regions.

Fig. 3. Same as Fig. 2, but for precipitation.

Fig. 4. Same as Fig. 2, but for sea level pressure.

Fig. 5. Area-averaged annual RMSE of GCM temperatures for 15 models (bars) for (a) Alaska, (b) Greenland, (c) 60-90°N, and (d) 20-90°N regions. Black bars at left show total RMSEs, grey bars at right show RMSEs computed after removal of annual mean bias.

Fig. 6. Same as Fig. 5, but for precipitation.

Fig. 7. Same as Fig. 5, but for sea level pressure.

Fig. 8. Maps of composite and 14 individual model biases of temperature for January (1958-2000).

Fig. 9. Same as Fig. 8, but for precipitation.

Fig. 10. Same as Fig. 8, but for sea level pressure.

Fig. 11. Projected changes of (a) temperature, (b) precipitation, and (c) sea level pressure change averaged over 60-90°N plotted against model performance rank (solid line). Changes are differences between linear-regression-derived trend line values for 2099 and 2000. Linear best fit indicated as a dashed line in each plot.

TABLE CAPTIONS

Table 1. Fifteen IPCC AR4 models assessed in this study. Atmosphere and ocean model resolution as well as country of origin are also listed for each model. (Detailed model documentation is available online at http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php)

Table 2: Summary of performance rank derived from GCM RMSE for three variables: temperature, SLP, precipitation over four regions: Alaska, Greenland, Arctic (60-90°N) and NH (20-90°N). The rankings in Table 2a are based on RMSEs that include the annual mean biases; Table 2b is based on RMSEs computed after removal of the annual mean biases. An integrated rank defined as the sum of ranks over all regions and variables is included in the right-most column.

Table 1. Fifteen IPCC AR4 models assessed in this study. Atmosphere and ocean model resolution as well as country of origin are also listed for each model. (Detailed model documentation is available online at http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php)

	IPCC AR4 model	Country	Atmosphere Resolution	Ocean Resolution	# of 20C3M ensemble members	Observed forcing in 20C3M
1	NCAR_CCSM_3.0	USA	T85L26	320x384L40	6	Yes
2	CCCMA_CGCM_3.1	Canada	T47L31	192x96L29	1	No
3	CNRM_CM3	France	T42L45	182x152L31	1	No
4	CSIRO_MK_3.0	Australia	T63L18	T63_2L31	1	No
5	MPI_ECHAM5	Germany	T63L31	1.5x1.5L40	3	No
6	GFDL_CM2.0	USA	2.5°x2.5°L24	tripolar360x200L50	3	Yes
7	GFDL_CM2.1	USA	2.5°x2.5°L24	tripolar360x200L50	3	Yes
8	GISS_MODEL_E_R	USA	5°x4°L13	5°x4°L13	9	Yes
9	INMC_3.0	Russia	5°x5°L21	2.5°x2°L33	1	Yes
10	IPSL_CM4	France	3.75°x2.5°L19	182x149L31	1	No
11	MIROC_3.2_MEDRES	Japan	T42 L20	1.4°x(0.5-1.4°) L44	3	Yes
12	MRI_CGCM_2.3.2a	Japan	T42 L30	(0.5°-2.5°)x2° L23	5	Yes
13	NCAR_PCM1	USA	T42L18	384x288 L32	2	Yes
14	IAP_FGOALS1_0_G	China	2.8°x3.0°L17	1.0°x1.0° L33	2	No
15	UKMO_HADCM3	UK	3.75x2.5 L	1.25°x1.25° L20	2	No

Table 2: Summary of performance rank derived from GCM RMSE for three variables: temperature, SLP, precipitation over four regions: Alaska, Greenland, Arctic (60-90°N) and NH (20-90°N). The rankings in Table 2a are based on RMSEs that include the annual mean biases; Table 2b is based on RMSEs computed after removal of the annual mean biases. An integrated rank defined as the sum of ranks over all regions and variables is included in the right-most column.

(a)

Overall Rank	Model	Alaska temperature	Greenland temperature	60-90°N temperature	20-90°N temperature	Alaska sea level pressure	Greenland sea level pressure	60-90°N sea level pressure	20-90°N sea level pressure	Alaska precipitation	Greenland precipitation	60-90°N precipitation	20-90°N precipitation	Integrated Rank Index
1	MPI_ECHAM5	13	02	01	01	01	02	01	01	05	03	03	03	36
2	GFDL_CM2_1	06	01	03	05	05	04	04	02	02	01	01	02	39
3	MIROC3_2_MEDRES	02	03	04	03	10	01	03	05	07	04	06	08	56
4	UKMO_HADCM3	11	04	08	06	04	05	06	07	03	02	02	09	67
5	CCCMA_CGCM3_1	12	11	11	10	08	03	02	04	04	13	08	02	88
6	GFDL_CM2_0	06	09	09	14	04	08	08	04	01	09	10	06	88
7	MRI_CGM2_3_2A	11	14	13	07	02	10	11	06	06	06	05	04	95
8	NCAR_CCSM3_0	08	06	02	02	15	11	15	13	09	05	08	07	101
9	CNRM_CM3	01	07	05	05	07	12	12	11	12	08	12	13	105
10	NCAR_PCM1	14	10	13	14	06	06	05	12	08	10	05	10	113
11	CSIRO_MK3_0	06	12	14	12	11	07	09	09	11	12	11	05	119
12	INMC3_0	07	13	06	10	09	09	07	09	10	14	13	12	119
13	IPSL_CM4	11	08	07	12	14	14	11	15	13	07	09	11	139
14	GISS_MODEL_E_R	06	05	10	10	13	15	14	14	14	11	14	15	141
15	IAP_FGOALS1_0_G	15	15	15	15	12	13	13	10	15	15	15	14	167

(b)

Overall Rank	Model	Alaska temperature	Greenland temperature	60-90Yr temperature	20-90Yr temperature	Alaska sea level pressure	Greenland sea level pressure	60-90Yr sea level pressure	20-90Yr sea level pressure	Alaska precipitation	Greenland precipitation	60-90Yr precipitation	20-90Yr precipitation	Integrated Rank Index
1	MPI_ECHAM5	01	07	01	01	07	06	01	01	01	04	04	01	35
2	GFDL_CM2_1	05	02	02	03	05	02	02	02	02	05	02	05	37
3	CNRM_CM3	04	08	06	06	01	03	03	03	09	03	06	13	65
4	MIROC3_2_MEDRES	06	01	04	05	04	01	08	06	14	06	07	09	71
5	GFDL_CM2_0	03	05	05	04	08	08	10	04	13	14	14	03	91
6	UKMO_HADCM3	10	09	11	10	06	10	06	08	07	07	05	08	97
7	NCAR_CCSM3_0	12	04	03	02	15	07	05	13	12	11	09	06	99
8	INMC3_0	08	11	07	11	12	04	09	11	03	08	10	12	106
10	CCCMA_CGCM3_1	07	10	08	10	03	05	04	05	15	15	15	15	112
10	NCAR_PCM1	13	13	14	14	11	09	07	12	10	02	03	04	112
11	IPSL_CM4	14	07	09	08	14	15	13	15	06	01	01	10	113
12	GISS_MODEL_E_R	02	03	10	12	10	14	11	14	05	13	13	07	114
13	MRI_CGM2_3_2A	11	14	12	07	02	11	12	07	11	10	12	11	120
14	CSIRO_MK3_0	09	12	13	13	09	12	14	10	08	09	08	14	131
15	IAP_FGOALS1_0_G	15	15	15	15	13	13	15	09	04	12	11	02	139

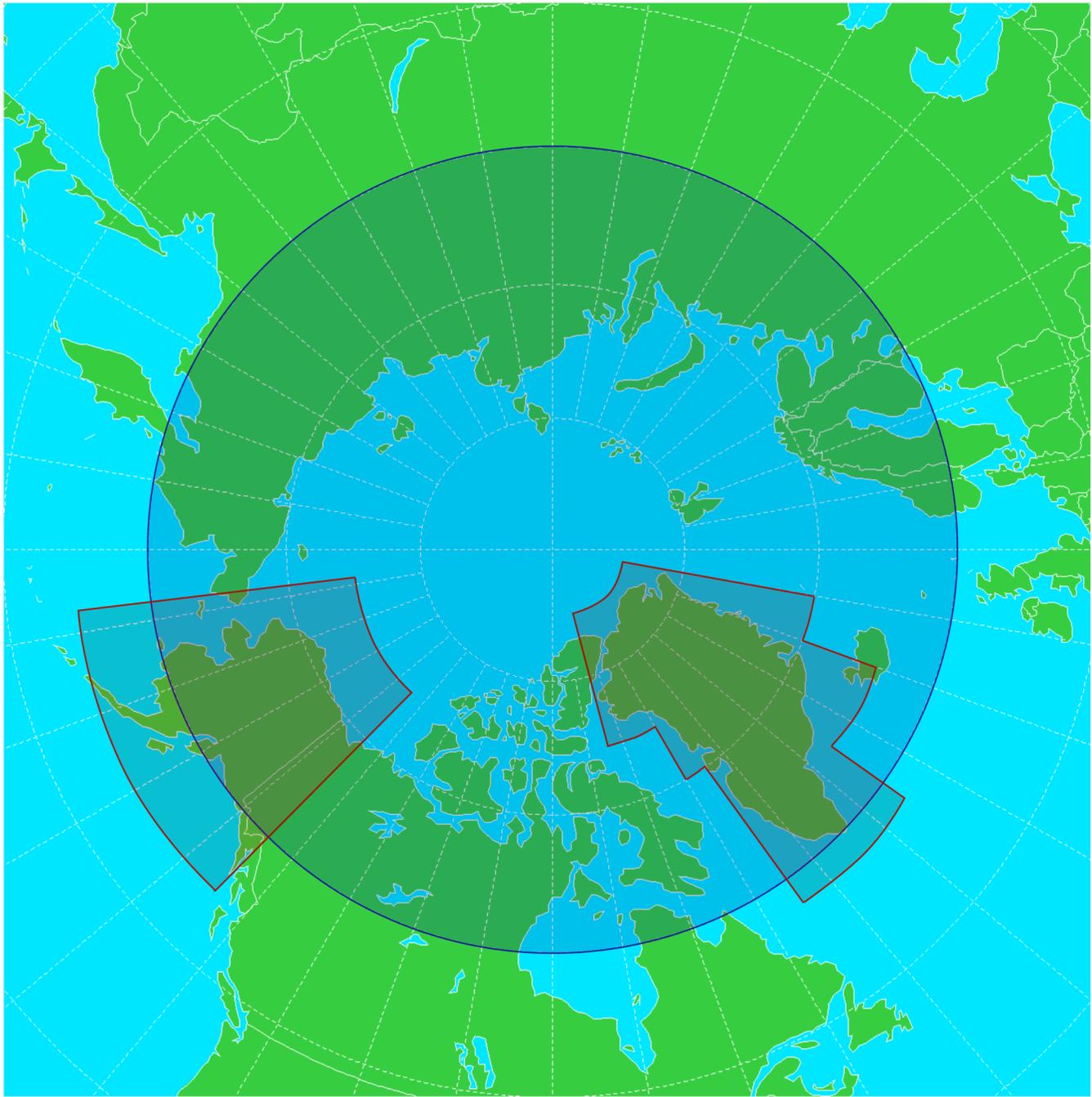


Figure 1: Map showing Alaskan and Greenland domains, also highlighting 60-90°N.

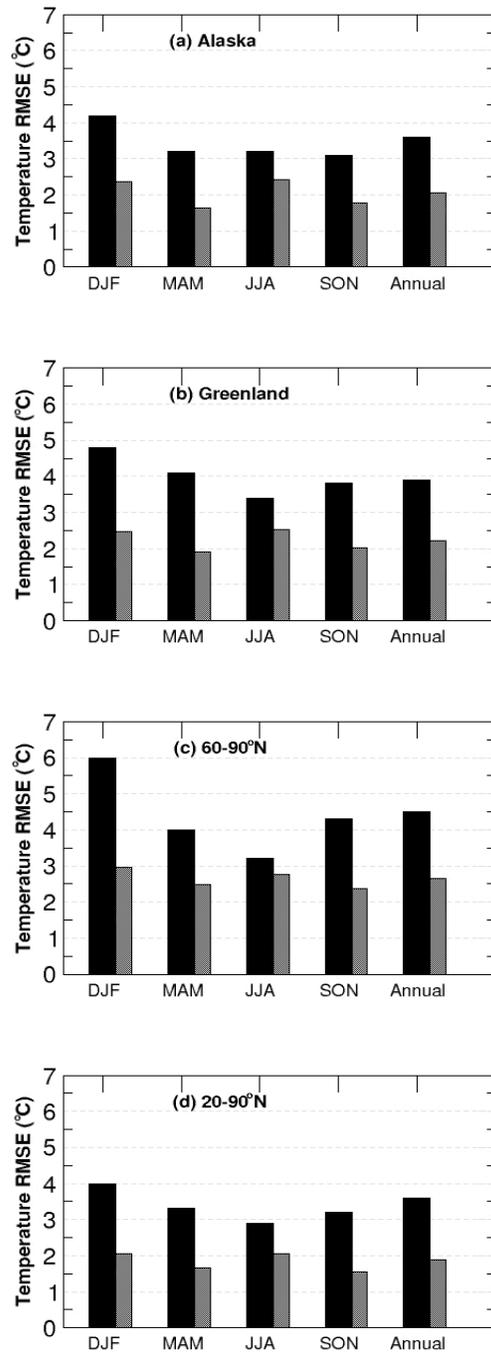


Figure 2: Area-averaged RMSE of GCM temperatures by season for 15 models (bars) for (a) Alaska, (b) Greenland, (c) 60-90°N, and (d) 20-90°N regions. Black bars at left show total RMSEs, grey bars at right show RMSEs computed after removal of annual mean bias.

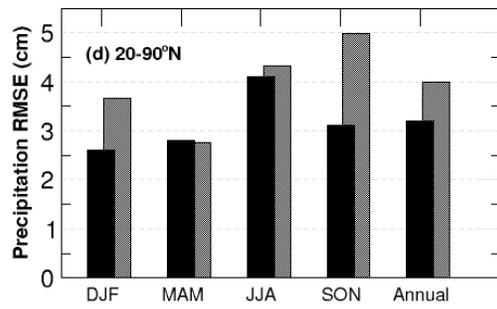
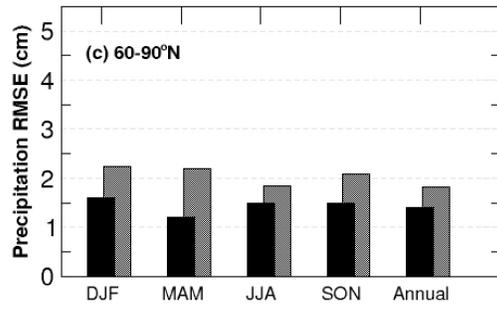
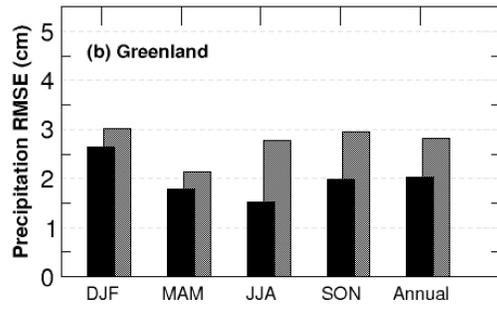
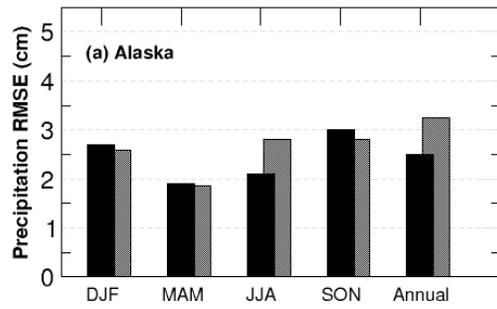


Figure 3: Same as Fig. 2, but for precipitation.

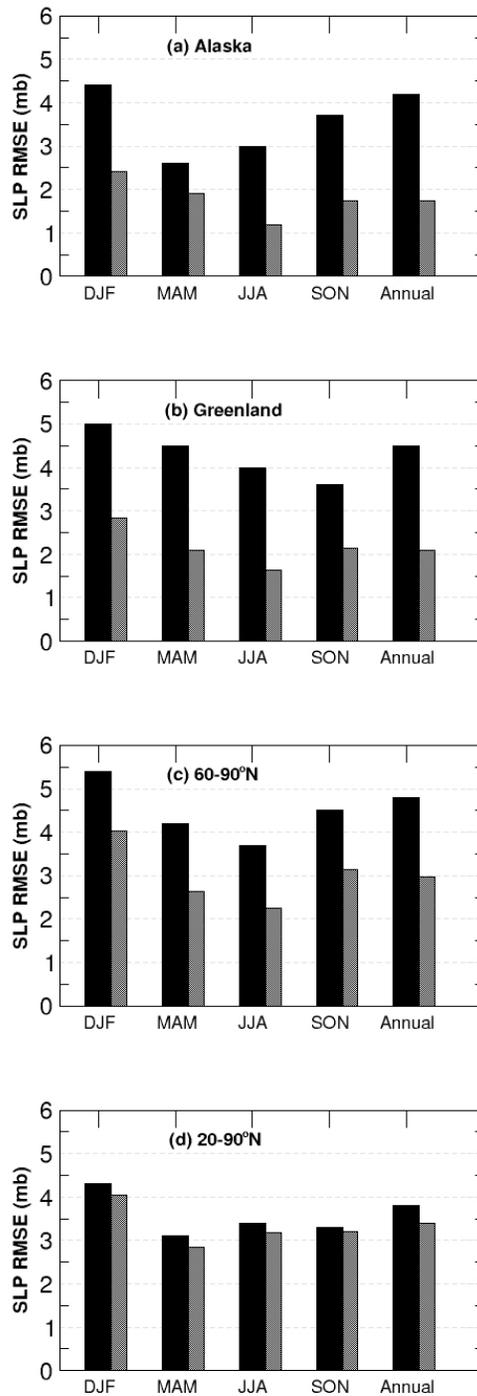


Figure 4: Same as Fig. 2, but for sea level pressure.

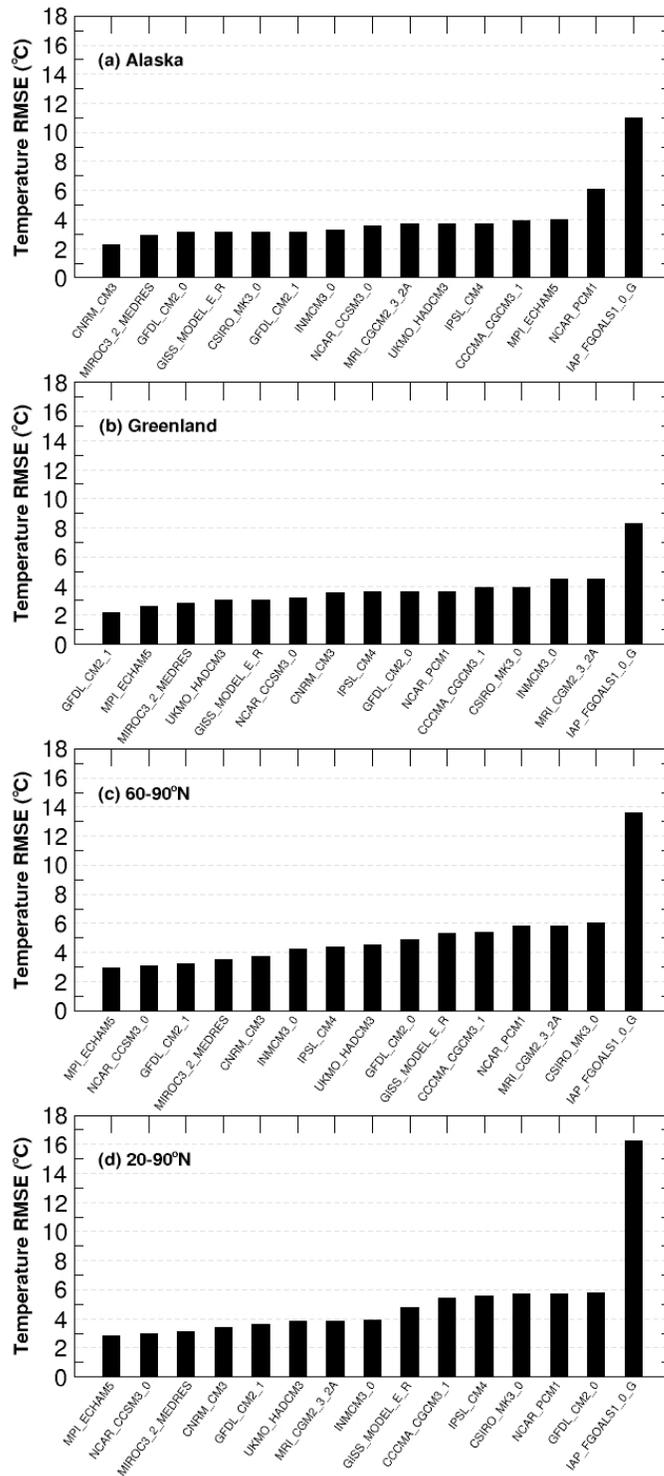


Figure 5: Area-averaged annual RMSE of GCM temperatures for 15 models (bars) for (a) Alaska, (b) Greenland, (c) 60-90°N, and (d) 20-90°N regions.

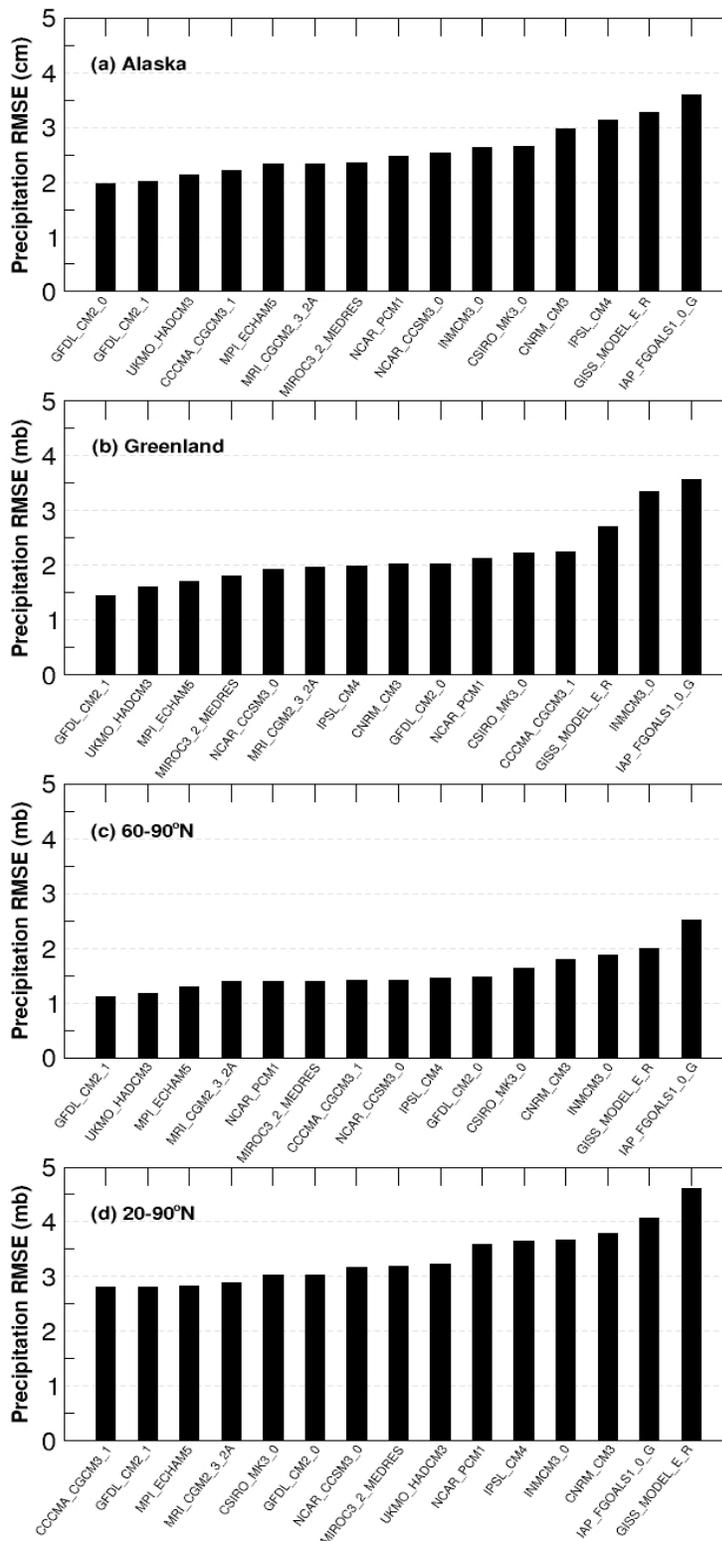


Figure 6: Same as Fig. 5, but for precipitation.

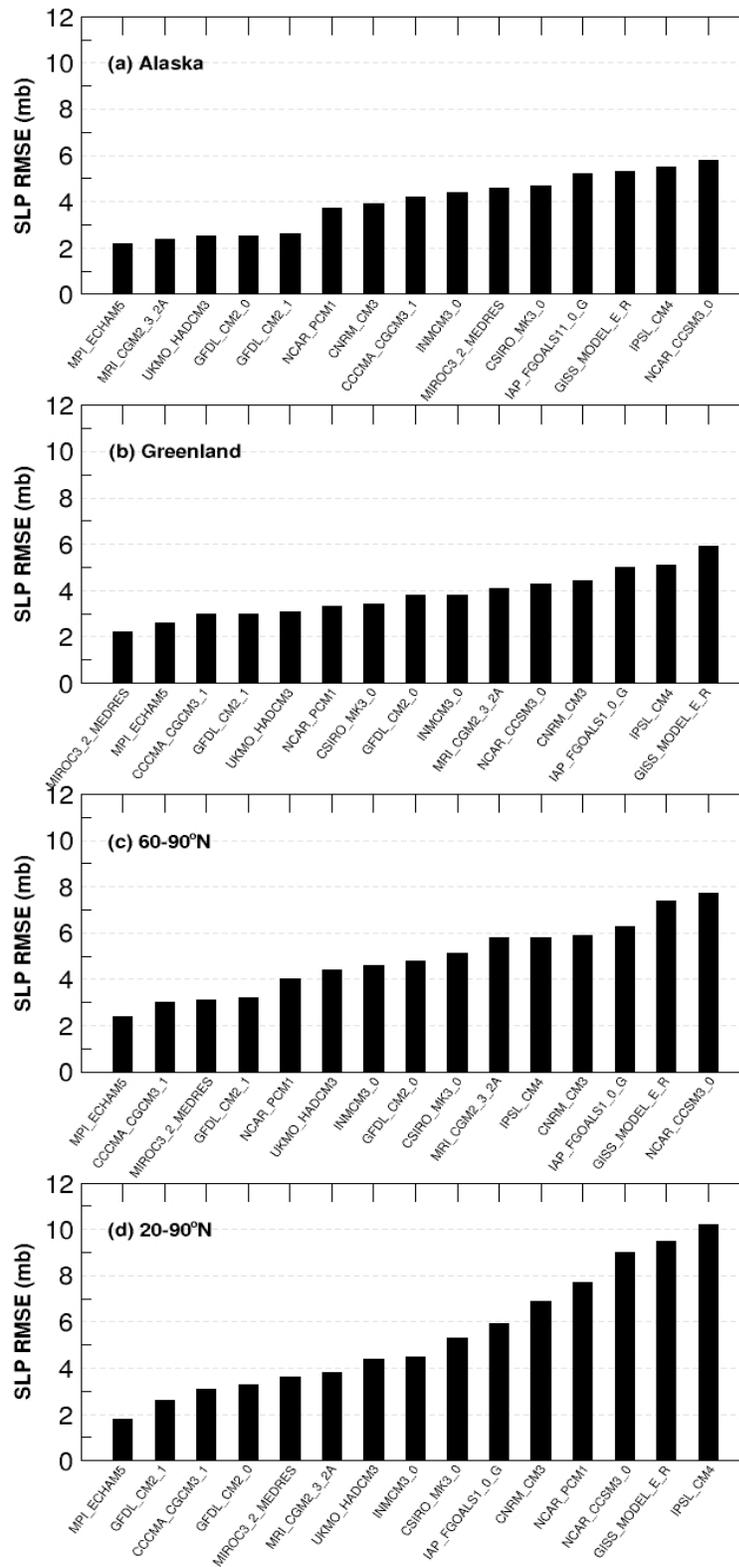


Figure 7: Same as Fig. 5, but for sea level pressure.

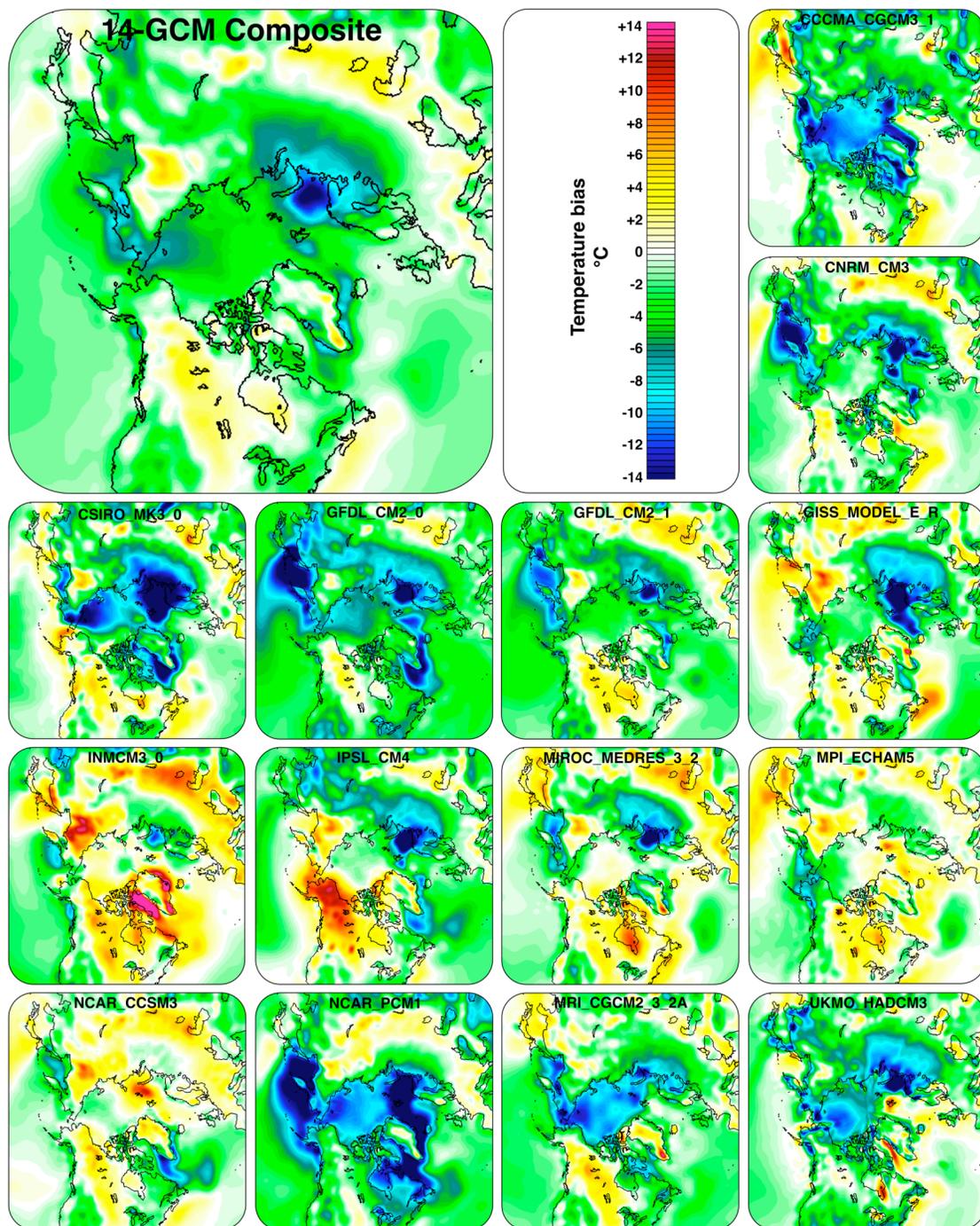


Figure 5: Maps of composite and 14 individual model temperature biases for January (1958-2000).

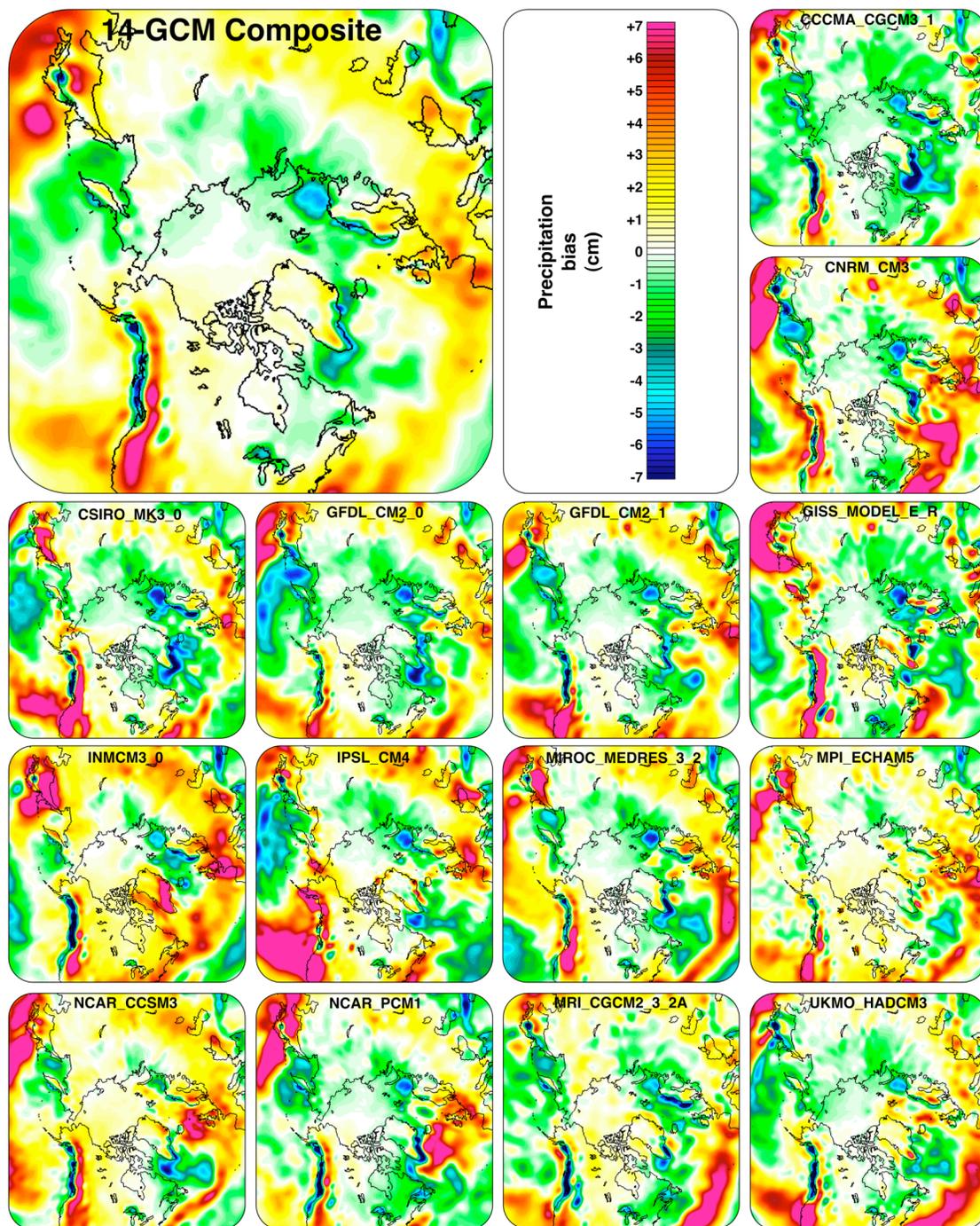


Figure 6: Same as Fig. 8, but for precipitation.

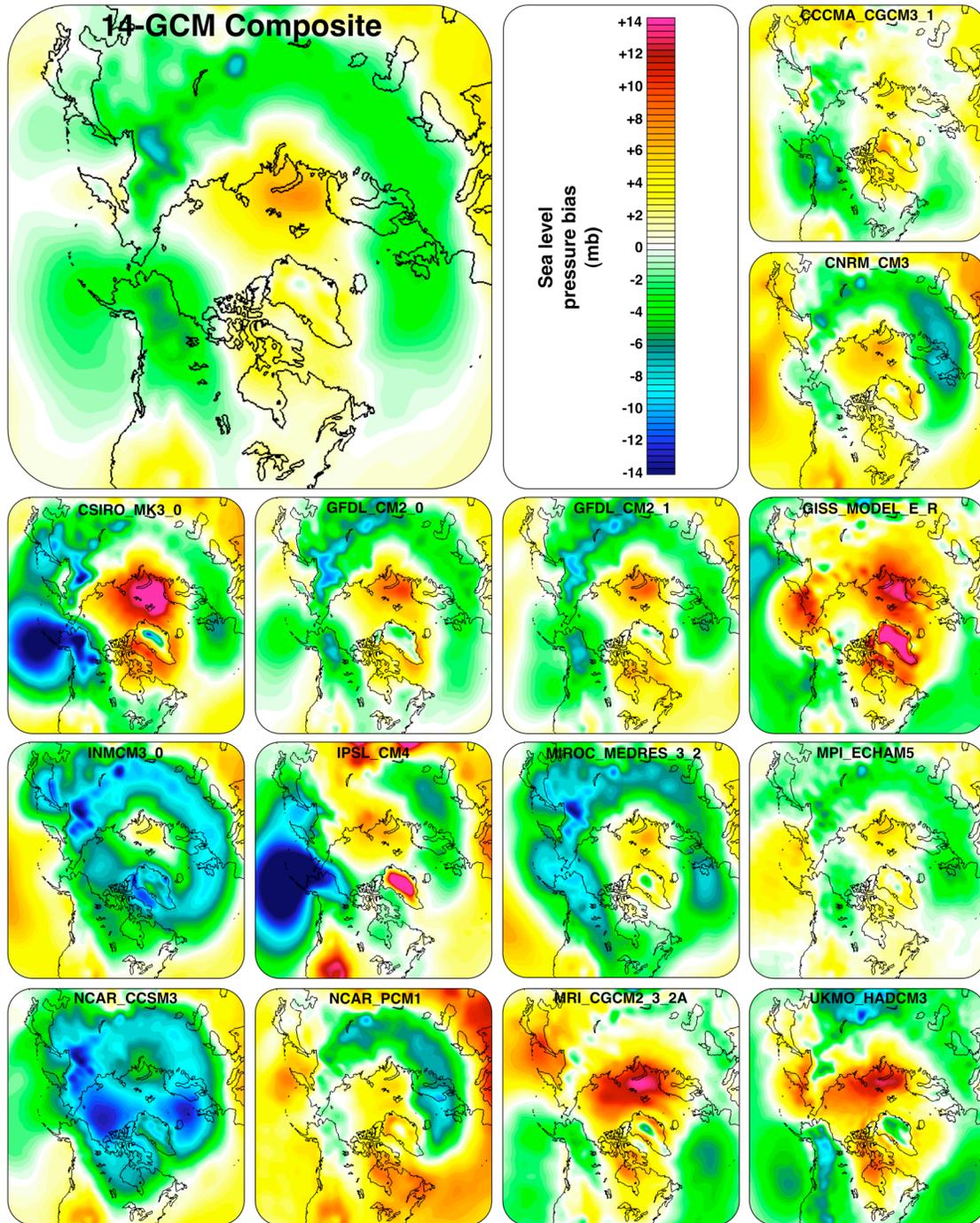


Figure 7: Same as Fig. 8, but for sea level pressure.

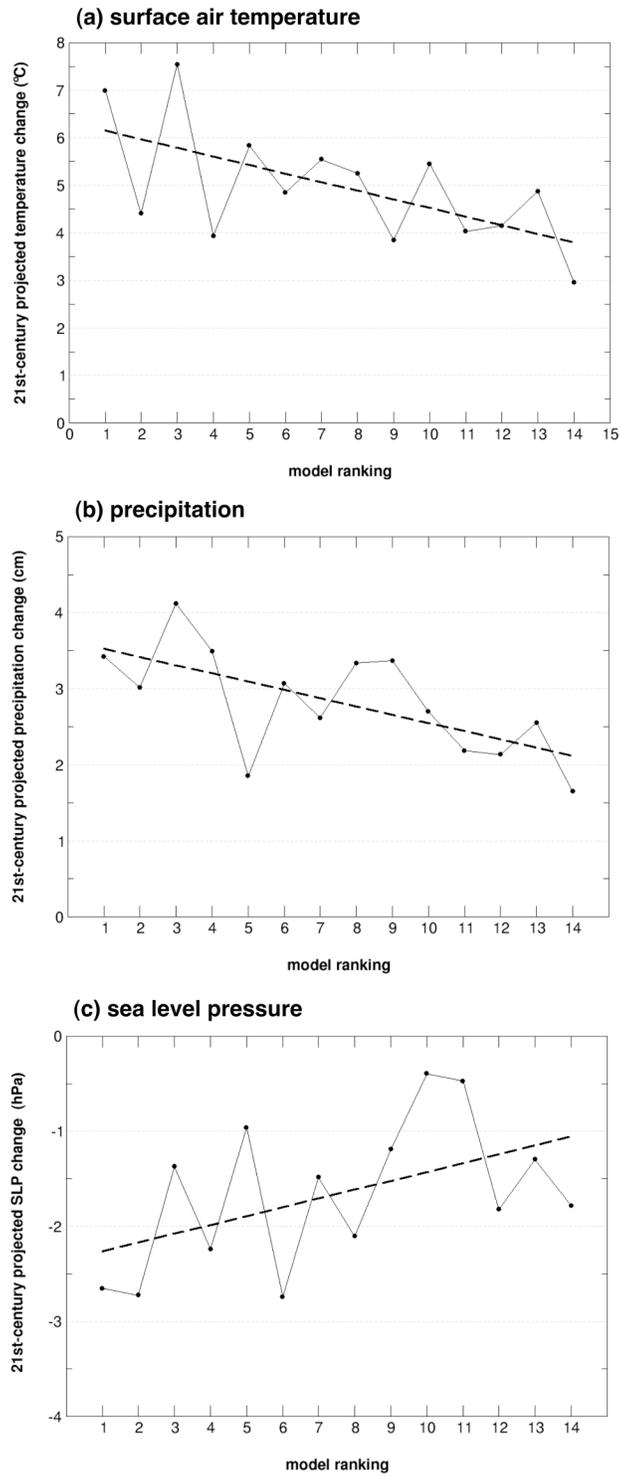


Figure 8: Linear (a) temperature, (b) precipitation, and (c) sea level pressure change for (60-90°N) plotted vs. model performance rank (solid line). Linear best fit indicated as a dashed line in each plot.