JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 103, NO. D8, PAGES 8859–8871, APRIL 27, 1998

Evaluation of snow extent and its variability in the Atmospheric Model Intercomparison Project

Allan Frei[1] and David A. Robinson
Department of Geography, Rutgers University, Piscataway, New Jersey

Abstract. Simulations of monthly mean northern hemisphere snow extent from 27 atmospheric general circulation models (GCMs), run under the auspices of the Atmospheric Model Intercomparison Project (AMIP), are compared to observations. AMIP model runs have common values for sea surface temperatures specified from observations for the decade 1979 through 1988. Here AMIP GCMs are evaluated in terms of their simulations of (1) snow extent over northern hemisphere lands and (2) synoptic conditions associated with extremes in snow extent over particular regions. Observations of snow extent are taken from digitized charts of remotely sensed snow extent from visible imagery provided by the National Oceanic and Atmospheric Administration. In general, AMIP models reproduce a seasonal cycle of snow extent similar to the observed cycle. However, GCMs tend to underestimate fall and winter snow extent (especially over North America) and overestimate spring snow extent (especially over Eurasia). The majority of models display less than half of the observed interannual variability. No temporal correlation is found between simulated and observed snow extent, even when only months with extremely high or low values are considered. These poor correlations indicate that in the models, interannual fluctuations of snow extent are not driven by sea surface temperatures. GCMs are inconsistent in their abilities to simulate synoptic-scale tropospheric circulation patterns associated with extreme snow extent over North American regions, although some models are able to capture many of the observed teleconnection patterns.

1. Introduction

Simulations of monthly mean northern hemisphere snow extent from 27 atmospheric general circulation models (GCMs), run under the auspices of the Atmospheric Model Intercomparison Project (AMIP), are compared to observations. Evaluating GCM simulations of present-day climate is a prerequisite to using these models for predictions under changing climatic conditions, such as an enhanced greenhouse world. Snow cover is an important climatological variable due to its impact on energy and mass exchanges at the surface, and due to its relevance to regional water budgets, in mid latitudes to high latitudes of the northern hemisphere. Models must produce reasonable estimates of snow cover if we are to have confidence in their predictive abilities.

AMIP model runs have common values for sea surface temperatures (SSTs) specified for the decade 1979 through 1988 and identical (constant) values for atmospheric CO_2 concentration and solar insolation [*Gates*, 1992]. Since boundary conditions are largely specified, model results are comparable to each other. In addition, since the boundary conditions are representative of specific calendar years, model results can be directly compared to observations. Hydrological processes in AMIP GCMs have been evaluated by Lau et al. [1996], who found that models generally simulate global precipitation to within 10%-20%. Heavy precipitation associated with deep convection is reasonably estimated, but the frequency of light precipitation events is underestimated due to inaccuracies in the treatment of moist processes. Models exhibit significant skill in predicting interannual fluctuations in tropical rainfall associated with specified SST values, but less so for extratropical precipitation [*Lau et al.* 1996] or other midlatitude variables [*Zwiers*, 1995]. *Sperber and Palmer* [1996] evaluated the abilities of AMIP models to simulate the interannual variability of rainfall over the Indian subcontinent, the African Sahel region, and the Nordeste region of Brazil. They found that fluctuations over the Nordeste region are simulated most accurately due to the strong association with Pacific and Atlantic SSTs, and that large-scale dynamic fluctuations are captured better than regional rainfall variations. *Weare and AMIP Modeling Groups* [1996] found that AMIP models tend to overestimate high clouds by about 2 to 5 times and to underestimate low clouds by 10%-20%.

Other investigators have evaluated ensembles of GCM runs not associated with AMIP. In general, models tend to underestimate globally averaged atmospheric temperatures due to systematic cold biases in the low-latitude and midlatitude lower troposphere, as well as the polar upper troposphere / lower stratosphere [*Boer et al.*, 1992]. However, in the lower troposphere poleward of 50° latitude, some models are too cold and some are too warm. Other reports have compared the responses of different models to changing climates, for example, by comparing results from control runs to results from runs with increased SSTs. The energy fluxes in models exhibit a wide range of responses, in terms of both magnitude and sign, to altered boundary conditions. These differences are at least partly due to the misrepresentation or omission of hydrological and associated radiative processes, including the formation and optical properties of clouds [*Randall et al.*, 1992; *Cess et al.*, 1989] and not simply due to the unavailability of computing power for increased model resolution [*Sperber and Palmer*, 1996; *Randall et al.*, 1992; *Boer et al.*, 1992].

[1]Now at Cooperative Institute for Research in Environmental Sciences/National Snow and Ice Data Center, University of Colorado, Boulder.

Paper number 98JD00109
0148-1227/98/98JD-00109$09.00

Simulations of snow extent by GCMs have also been evaluated. *Foster et al.* [1996] evaluated snow output from seven GCMs, and passive-microwave snow observations, to assess the magnitude of variations between these and other observations. They found that models reproduce the seasonal cycle fairly well but are inconsistent in their simulation of snow mass. Winter and summer snow conditions were more accurately simulated than the transition seasons, during which inaccuracies in snow mass were attributed to both temperature and precipitation errors. *Zhong* [1996], in an assessment of seven GCMs that were employed in the Model Evaluation Consortium for Climate Assessment project, found that snow extent is reasonably well simulated except during the ablation season, but that models underestimate interannual variations. Snow depth was found to be fairly well simulated, but the models are inconsistent in the geographic location of deep snow cover. Inaccuracies in the simulation of surface temperature, more than precipitation, were responsible for these inconsistencies.

Cess et al. [1991] analyzed the feedback effect of snow cover under changing climatic conditions in a number of GCMs. They found that in a warming world, most models predict a positive feedback for snow cover, meaning that decreases in snow cover increase the magnitude of the warming. However, some models have weak positive feedbacks, others have strong positive feedbacks, and a few have weak negative feedbacks. Indirect effects, including cloudiness and longwave fluxes, are important components of these responses, and differ greatly from model to model. *Randall et al.* [1994], who performed more detailed evaluations of surface energy fluxes in the same experiments as *Cess et al.* [1991], support the basic conclusions of the earlier study. They find that the combined effect of changes in clouds and snow cover on the planetary and surface radiation budgets differs from model to model and that the differences can be associated with relatively minor changes in parameterizations. For example, the feedback effect of one GCM changed from weakly negative to weakly positive when snow albedo was held constant, rather than being temperature dependent.

Here AMIP GCMs are evaluated in terms of their simulations of (1) snow extent over northern hemisphere lands and (2) synoptic conditions associated with extremes in snow extent over particular regions. *Frei* [1997] reports these results in more detail than presented here. The 27 AMIP models that include prognostic snow calculations are included (Table 1). In all models, snow cover affects surface albedo. In addition, 20 of the 27 models include the effects of snow cover on the non radiative thermal properties of the surface, such as heat capacity and thermal conductivity. Only 4 of the 27 models include the effect of snow upon surface roughness. The snow budget in a grid box depends on the rates of snowfall, melt, and sublimation. All of the 27 models analyzed here include calculations of snowfall and melt; only some include the sublimation process. The state of precipitation (i.e., either rain or snow) usually depends on the temperatures in one or both of the bottom two tropospheric layers, and sometimes at the surface. Typically, when the temperature of the specified layer(s) is below ~0°C, precipitation is frozen. Melt occurs as a function of the surface temperature, and usually contributes to soil moisture and surface runoff. When the sublimation process

Table 1. The 27 AMIP Models Included in the Snow Cover Analysis

	Model	Modeling Group	Thermal Effects	Surface Roughness	Fractional Snow Cover
1	BMRC	Bureau of Meteorology Research Centre, Melbourne, Australia	N	N	N
2	CCC	Canadian Centre for Climate Research, Victoria, British Columbia	Y	N	N
3	CNRM	Centre National de Recherches Meteorologiques, Toulouse, France	N	Y	N
4	CSIRO	Commonwealth Scientific and Industrial Research Organization, Victoria, Australia	Y	N	N
5	CSU	Colorado State University, Fort Collins, Colorado	Y	N	N
6	DERF	Dynamical Extended-Range Forecasting (at GFDL), Princeton, New Jersey	Y	N	N
7	DNM	Department of Numerical Mathematics, Moscow, Russia	Y	N	N
8	ECMWF	European Centre for Medium-Range Weather Forecasts, Reading, England	Y	N	Y
9	ECMWF_EO2	European Centre for Medium-Range Weather Forecasts, Reading, England	Y	N	Y
10	GFDL	Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey	N	N	N
11	GISS	Goddard Institute for Space Studies, New York	Y	N	N
12	GLA	Goddard Laboratory for Atmospheres, Greenbelt, Maryland	Y	N	Y
13	IAP	Institute of Atmospheric Physics, Beijing, China	Y	N	N
14	JMA	Japan Meteorological Agency, Tokyo	Y	Y	Y
15	LMD	Laboratoire de Meteorologie Dynamique, Paris, France	Y	N	N
16	MGO	Main Geophysical Observatory, St. Petersburg, Russia	Y	N	N
17	MPI	Max-Planck-Institut fur Meteorologie, Hamburg, Germany	Y	N	Y
18	MRI	Meteorological Research Institute, Ibaraki-ken, Japan	Y	N	N
19	NMC	National Meteorological Center, Camp Springs, Maryland	N	N	N
20	NRL	Naval Research Laboratory, Monterey, California	Y	N	N
21	SNG	SUNY / NCAR / GENESIS model, Albany, New York	Y	Y	Y
22	SUNYA	State University of New York at Albany (SUNY)	N	N	N
23	UCLA	University of California at Los Angeles	N	N	N
24	UGAMP	U.K. Universities' Global Atmospheric Modelling Programme, Reading, England	N	N	Y
25	UIUC	University of Illinois at Urbana-Champaign	Y	N	N
26	UKMO	U.K. Meteorological Office, Berkshire, England	Y	Y	N
27	YONU	Yonsei University, Seoul, Korea	Y	N	N

Numbers correspond to those appearing in Figure 1. Last three columns indicate whether the model includes the effects of snow cover on surface thermal properties and surface roughness and whether fractional snow coverage in a grid box is possible.

is included, it usually contributes to the surface evaporative flux. Some models calculate a fractional snow coverage for each grid box, and some models do not accumulate snow over sea ice. Other GCMs contain unique parameterizations, such as limits to the snow accumulation, or more detailed snow models. Details of each model are given by *Phillips* [1994].

In section 2 the observational data utilized in this analysis are discussed. In section 3, model results are compared to observed snow extent over northern hemisphere lands with respect to central tendency, systematic bias, dispersion, and interannual fluctuations. In addition, the relative accuracy of models with different numerical properties is explored. In section 4, simulated synoptic 500-mbar geopotential height patterns over North America that are associated with extreme snow extent in each region are compared to observed synoptic teleconnection patterns. In section 5 a more detailed case study of the synoptic patterns associated with regional snow extent in one particular GCM (MPI) is presented.

2. Data

Charts of northern hemisphere snow cover extent, produced by the National Oceanic and Atmospheric Administration (NOAA), from January 1972 through December 1994 are used in this analysis. These weekly charts are derived from visual interpretation of photographic copies of visible satellite images by trained meteorologists. Limitations of visible imagery for snow cover detection include problems with low solar illumination, cloudiness, dense forest cover, and subgrid resolution snow features (e.g., in areas of steep terrain). However, at a monthly resolution these data are suitable for climatic studies [*Kukla and Robinson*, 1981; *Wiesnet et al.*, 1987].

The images are digitized using the National Meteorological Center limited-area fine mesh grid, an 89 x 89 cell Cartesian grid placed over a northern hemisphere polar projection. This grid includes more than 5000 cells over land, with resolutions between 16,000 km^2 and 42,000 km^2, covering approximately 1°-2° latitude at all locations and 1°-3° longitude in midlatitudes to 5° longitude farther north. For each week, if a cell is interpreted to be at least 50% snow covered, it is considered snow covered; otherwise, it is charted as snow free. Weekly charts identify snow in a grid cell on the latest day of the week during which the ground was visible. *Robinson et al.* [1993] determined that weekly charts are most representative of snow cover extent on the fifth day of the week. The data set used in this study includes corrections to the NOAA data set for inconsistencies in the demarcation of land versus ocean grid cells used in chart digitization, and an improved routine to calculate monthly average snow frequencies. Monthly averaged snow cover frequencies for each grid cell for the 23-year period between 1972 and 1994 are included.

Monthly mean 500-mbar geopotential heights over North America, from 30°N to 80°N, and from 180°W to 20°W longitude, are obtained from the National Meteorological Center (recently renamed the National Centers for Environmental Prediction) on a 2.5° latitude x 2.5° longitude grid. The 500-mbar height observations have been taken twice daily, at 0000 and 1200 UTC, since 1957. Monthly mean values are obtained from averages of the two daily values.

3. Continental, Hemispheric, Seasonal Snow Extent

3.1 Methodology

GCM results are evaluated in terms of (1) their similarities to the observed 9-year snow cover extent climatology (with regards to central tendency, systematic bias, and dispersion), (2) the correlation between modeled and observed interannual fluctuations, and (3) model performance as a function of model numerical properties (horizontal representation, horizontal resolution, and vertical resolution).

Observations of areal snow extent over land are compared with output from AMIP models for January 1980 through December 1988. Year 1979 is excluded because the initialization values are inconsistent between models. Over open areas, snow cover of one inch or greater in depth is generally recognized in the visible satellite imagery [*Kukla and Robinson* 1981]. Although this value greatly depends on topography and vegetation, we try to be as consistent as possible by considering grid cells as snow covered only if snow depth is ≥3 cm. The AMIP standard output includes snow mass, but not depth. For comparative purposes we assume an average snow density of 300 kg/m^3. Results are not sensitive to the value used for density (see section 3.3). Monthly statistics for the entire northern hemisphere (NH) land area, and for North American (NA) and Eurasian (EU) land areas individually, are compiled for each model. To facilitate comparison between several different model grids (each associated with a different total land area), the percentage of land area north of 20°N latitude covered with snow, rather than the absolute snow-covered area, is the unit of comparison.

3.2 Central Tendency

Figure 1a shows median values of northern hemisphere snow cover extent for AMIP models and observations. Fall and winter snow extent tends to be underestimated, while spring tends to be overestimated. In fall (winter), 16 (13) models have median values that are within 5%, or approximately 4 million square kilometers (Mkm^2), of the observed median; most of the remaining models underestimate snow cover. In spring, nine models have NH median values within 5% of observations, and 12 of the 18 remaining models overestimate snow cover extent. Only two models (UCLA and SNG) are within 5% of observed values for the northern hemisphere, as well as for both North America and Eurasia, during all three seasons.

The northern hemisphere median annual root mean square error (RMSE) is 7%, or approximately 5.5 Mkm^2; the second and third quartiles range between 5% and 10% (Table 2). Figure 2 shows distributions of NH, NA, and EU annual and seasonally averaged RMSEs and anomalies. RMSEs tend to be larger, and anomalies more negative (i.e., snow cover extent is underestimated) over NA than EU for the annual average. In fall and winter the models underestimate snow cover extent on both continents, but more so over NA than EU. In spring over NA there is no strong tendency for either underestimation or overestimation; over EU, spring snow cover extent is generally overestimated. Three models with unrealistically low snow extent appear as outliers in Figure 2. There is no model that consistently outperforms the others across several seasons or continents.

3.3 Systematic Bias

To determine whether model results, or the diagnostic method employed, are associated with systematic biases, two "ensemble" averages are calculated: one including 27 models and one including 24 models. The ensemble average is simply the mean monthly snow cover extent of all models; anomalies are then calculated for each month by subtracting the observed value from the ensemble average value. The 27-model ensemble anomalies (Table 3, left) indicate that snow cover extent tends to be underestimated. However, these values include three models with unrealistically low values (probably caused by the submission of results in incorrect units, rather

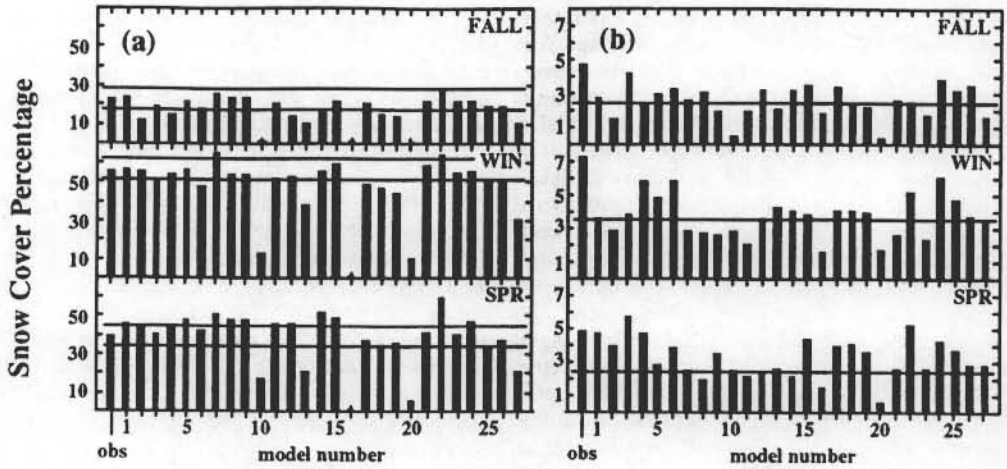

Figure 1. NH seasonal snow cover statistics for 27 AMIP models. Median and range of 9-year (1980-1888) climatology for fall (September-October-November), winter (December-January-February), and spring (March-April-May). Units are percentage of land area north of 20°N covered with snow. (a) Median values for observations (obs) and models. Horizontal lines represent ±5% from observations. (b) Range (maximum - minimum) of values for observations and models. Horizontal lines represent one half of observed range. See Table 1 for model identification.

than actual simulation errors). A more meaningful evaluation is made by removing the three outliers and calculating 24-model ensemble averages (Table 3, right and Figure 3). These results show that fall and winter snow cover extent are underestimated, especially over NA; and that spring tends to be overestimated, especially over EU. The ensemble mean underestimates annually averaged northern hemisphere snow cover extent by only 0.9% of the land surface area, or less then 0.8 Mkm^2. When models that consistently underestimate or overestimate snow cover extent are reanalyzed assuming a density of 200 kg/m^3 or 400 kg/m^3, results are not significantly affected. (These assumptions would more likely affect comparisons of depth more than area.) The relatively small anomaly indicates that assumptions made in the diagnostic technique are reasonable.

3.4 Dispersion

AMIP models underestimate the year-to-year variability of snow cover extent. Figure 1b shows the range (maximum - minimum value) of modeled snow cover extent compared to the observed range. During fall and winter, no model overestimates the range, and only 17 (fall) and 15 (winter) of the 27 models exhibit even half of the observed range. Some, but not most, models capture the observed variability of snow extent during spring.

3.5 Interannual Fluctuations

Interannual correlations between modeled and observed time series for NH, NA, and EU snow cover extent during fall, winter, and spring are poor. Spearman rank correlation coefficients are generally 0.0 ± 0.25 for fall and winter over both continents, and 0.25 ± 0.25 during spring. The number of significant correlations (n=9, p=0.05, 1-tailed) are as many as one would expect from an equal number of random time series. No model consistently captures interannual fluctuations over several seasons and both continents.

Poor interannual correlations between model results and observations indicate that large portions of the modeled snow cover signals are driven by variables other than AMIP boundary conditions. This is in agreement with other AMIP findings that the models show no skill in predicting extratropical rainfall from specified SSTs [*Lau et al.*, 1996] and that midlatitude to high-latitude climates in general have low potential predictability [*Zwiers*, 1995].

However, we investigate the possibility that models may capture extreme events and that these extreme events may emerge as a signal that is detectable above the noise. For this reason we examine 3 months that are considered outliers in the observed distributions of monthly snow extent. December 1980 and January 1981 had unusually low, and November 1985 had unusually high, NH snow cover extent. In only five model simulations is January 1981 either the first or second lowest January; nine models have December 1980 as their first or second lowest December; two models have November 1985 as their first or second highest November. These results are no more than would be expected from random time series. No model captures all three events, and only three models capture two of the three events. Thus the models do a poor job of capturing extreme monthly values of snow cover extent.

Table 2. RMSEs of 27 AMIP Models

	Annual		Fall		Winter		Spring	
	Median	Minimum	Median	Minimum	Median	Minimum	Median	Minimum
NH	7.0	3.3 (UCLA)	5.4	2.4 (CSU)	5.7	2.0 (BMRC)	8.5	3.3 (UKMO)
NA	9.0	5.2 (UCLA)	9.2	5.0 (ECMWF)	7.8	4.5 (CSU)	9.8	4.0 (BMRC)
EU	7.0	3.4 (UCLA)	5.4	2.7 (SNG)	4.7	2.3 (BMRC)	9.3	3.3 (UKMO)

For each geographic area (NH, NA, and EU), annual and seasonal median and minimum RMSE values are shown. Units are percentage of land area north of 20°N covered with snow. Actual areas, in Mkm^2, are approximately 80 (NH), 24 (NA), and 56 (EU). The standard definition of RMSE [e.g., *Panofsky and Brier*, 1958] is used.

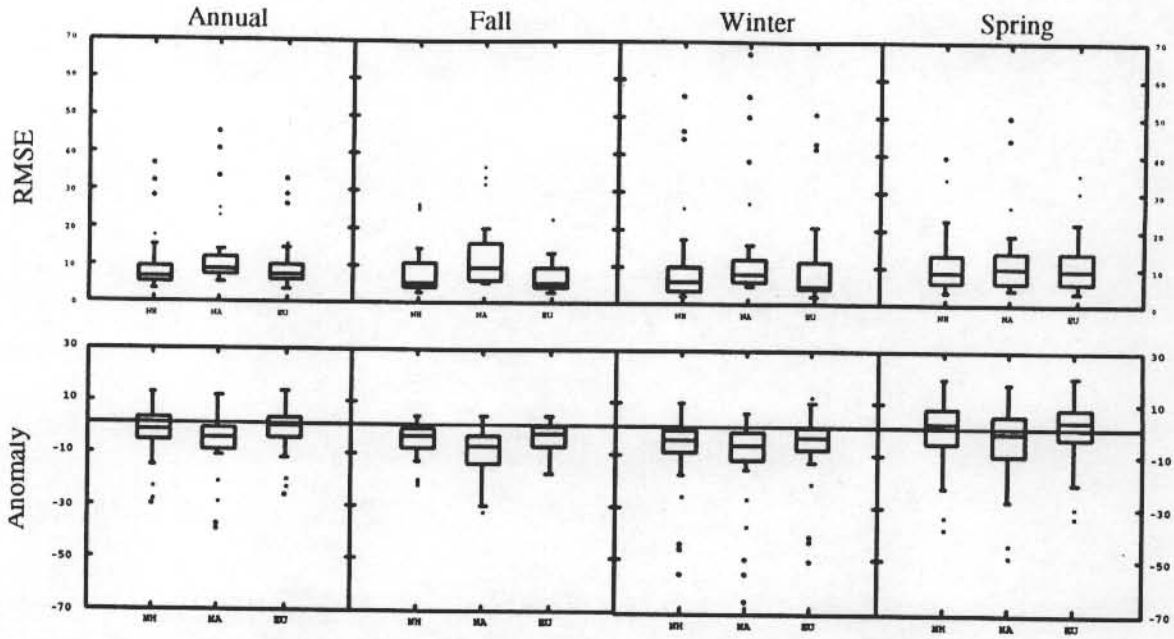

Figure 2. NH, NA, and EU annual and seasonal distributions of RMSEs and anomalies for 27 AMIP models compared to observations from 1980 to 1988. Units are percentage of land area north of 20°N covered with snow. Center line in box-and-whisker diagram shows median value; top (upper hinge) and bottom (lower hinge) of box delimit second and third quartiles (*H*spread); whiskers show first and fourth quartiles. Outliers shown as asterisks are greater than 1.5x*H*spread distant from nearest hinge; outliers shown as circles are greater than 3.0x*H*spread distant from nearest hinge.

3.6 Numerical Properties

To determine whether the type of horizontal representation (i.e., spectral versus finite grid), or the horizontal or vertical resolution, affect model simulation of snow cover extent, we examine the distribution of RMSEs and anomalies as a function of these properties.

Although the median anomalies for both spectral and finite difference representations are similar (Figure 4a), larger positive annual anomalies are associated with spectral models, and larger negative anomalies with finite grid models. Examination of seasonal distributions (not shown) reveals that this relationship is strong during spring but not apparent during fall and winter. The significance of this is not clear.

A comparison of higher and lower resolution models (Figures 4b and 4c) shows that larger annual RMSEs, and larger anomalies of both signs, are associated with lower resolution models. Models with latitudinal grid box dimensions smaller than 3.5° are compared to those with poorer resolution; models with greater than 13 vertical layers are compared to those with fewer layers. These relationships between model resolution, RMSEs, and anomalies, are generally true over both continents for autumn, winter, and spring (not shown). The three outlier models are exceptions: all underestimate snow cover extent, yet all have spectral representation and high vertical resolution. However, the underestimation of snow extent in these models is likely due to discrepancies in units, rather than poor model results. Other analyses indicate that spatial resolution is not the most important factor in accounting for differences between temperature [*Boer et al.*, 1992] and precipitation [*Sperber and Palmer*, 1996; *Randall et al.*, 1992] patterns in different models. In addition, increasing the spatial resolution, without adjusting parameterizations that may be scale dependent, does not necessarily result in more accurate simulations [*Phillips et al.*, 1995].

4. Synoptic Evaluation of Regions Over North America

In this section, synoptic conditions associated with extreme (high or low) regional snow extent in the GCMs are compared to observed synoptic teleconnection patterns. Since only 9 years of model results are available, no attempt is made to assign statistical significance to model output. The results shown here are useful to identify areas of potential strengths and weaknesses in the models and to indicate whether some models may be useful for more detailed investigations. The

Table 3. Systematic Bias in Ensemble Average of AMIP Models

	Annual Anomalies for 27 Models			Annual Anomalies for 24 Models		
	NH	NA	EU	NH	NA	EU
Mean	-3.8	-7.4	-2.4	-0.9	-3.9	0.3
Median	-2.8	-6.3	-1.7	-1.1	-3.7	-0.2
s.d.	4.6	5.4	4.6	3.9	4.7	3.9

Units are percentage of land area north of 20°N covered with snow. To obtain monthly anomalies for 108 months (1980-1988), observed monthly snow cover percentages are subtracted from monthly average snow cover of (left) 27 models and (right) 24 models. Annual mean, median, and standard deviation (s.d.) are calculated from 108 monthly anomaly values. (See text and Figure 3 for discussion.)

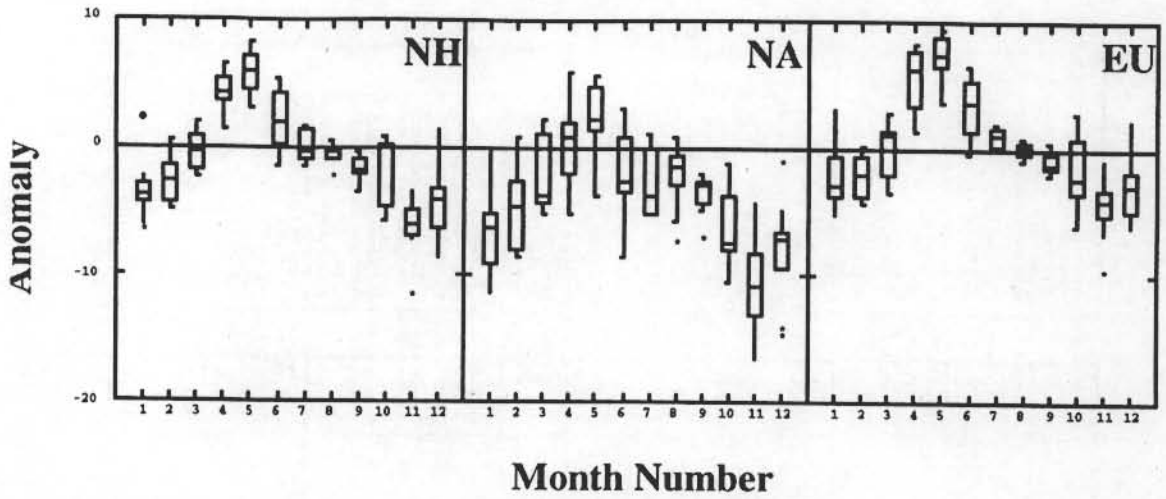

Figure 3. Distributions of NH, NA, and EU monthly anomalies in ensemble average of 24 models. Units are percentage of land area north of 20°N covered with snow. Anomalies for 108 months (1980-1988) are obtained by subtracting observed monthly snow cover extent from monthly average snow cover of 24 models. See text and Table 3 for discussion. See Figure 2 for explanation of box-and-whiskers diagram.

relationship between regional snow extent, regional temperatures, and 500-mbar geopotential heights in a selection of six GCMs are compared to results from a synoptic analysis of North American regional snow cover from November through March [*Frei*, 1997]. Before describing the methodology and results of the model comparison, we briefly summarize the synoptic analysis.

4.1 Summary of Synoptic Analysis

Examination of the kinematics and synoptic climatology of northern hemisphere snow extent between 1972 and 1994 indicates that interannual fluctuations of North American and Eurasian extents are driven by both hemispheric scale signals, as well as signals from smaller "coherent" regions, within which interannual fluctuations of snow extent are highly correlated. Coherent regions are identified using principal components (PC) analysis of digitized NOAA snow extent charts. Each region is identified by its PC number (e.g. PC1), with lower numbers explaining more variance. Through the use of composite analysis, geographically and seasonally dependent associations are identified between North American snow extent and synoptic-scale atmospheric circulation patterns, surface air temperature, and snowfall. Over western North America, extensive snow cover is associated with a westward shift by 20° to 30° longitude of the North American ridge. Over eastern

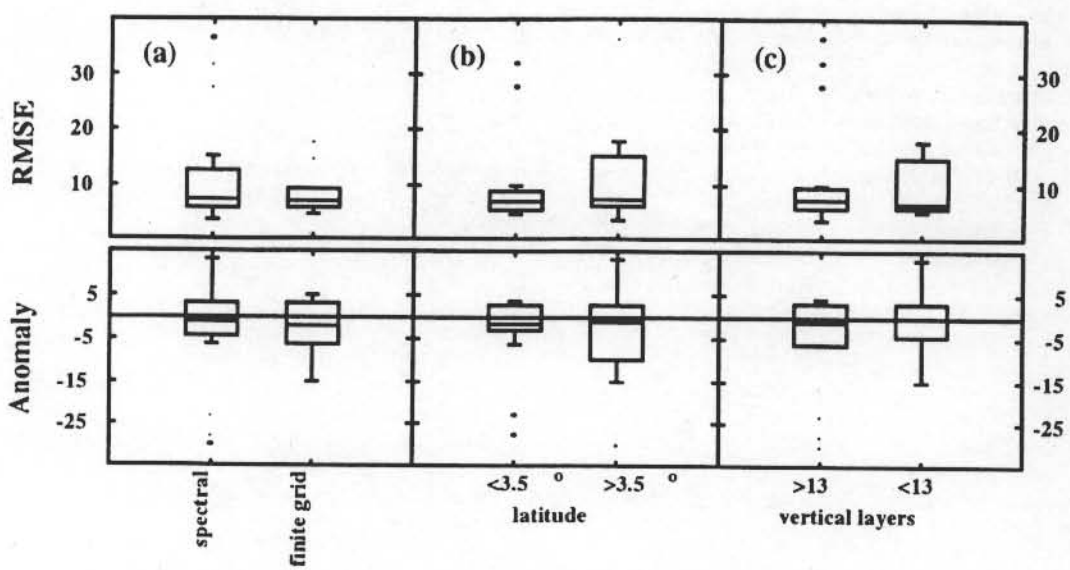

Figure 4. Performance of AMIP models as a function of numerical properties. Distributions of NH annual RMSEs and anomalies shown as a function of (a) horizontal representation (16 spectral versus 10 finite grid models), (b) horizontal resolution (13 models with latitudinal dimension <3.5° versus 12 models >3.5°), and (c) vertical resolution (16 models with >13 layers versus 10 models <13 layers). Units are percentage of land area north of 20°N covered with snow. In Figures 4a and 4c only 26 models are included due to a lack of information for one model. In Figure 4b only 25 models are included because one model has variable latitudinal resolution. See Figure 2 for explanation of box-and-whiskers diagram.

North America, snow extent is associated with a dipolar meridional oscillation in the 500-mbar geopotential height field over the western North Atlantic Ocean. These teleconnection patterns, each of which has two nodes, are associated with secondary modes of tropospheric variability during fall and winter. During spring, snow extent becomes effectively decoupled from tropospheric dynamics.

4.2 Methodology

Since synoptic conditions are analyzed over North America only, those models that most accurately represent North American snow extent are chosen. The six models with lowest RMSEs over North America (RMSE< 8% of land area north of 20°N) for each season (fall, winter, and spring) are chosen to be included in this regional analysis. Since no models accurately simulate the observed variability, variability is not considered in choosing the models. The models included in the regional analysis are BMRC, CSU, LMD, MPI, SNG, and UCLA.

To evaluate regional fluctuations in the models, the geographic extent of observed regions of coherent snow fluctuations are approximated using latitude-longitude delimited boxes (Table 4). These boxes are drawn with less spatial resolution than is available from the NOAA grid in order to accommodate the varying spatial resolutions of the models. Snow extent in each region is calculated using the same criteria as in the continental analysis, where an average snow density of 300 kg/m^3 is assumed, and a minimum depth of 3 cm is required for a grid box to be considered snow covered.

Two variables, surface air temperature and 500-mbar heights, from the AMIP output data set are available for comparison to the North American synoptic analysis. Regionally averaged surface air temperatures are calculated using area-weighted values from each grid box in the region. The nonparametric Spearman correlation coefficient is used to compare regional snow extent to surface air temperature. The 500-mbar heights for the North American sector are composited for high and low snow extent months for each model and each region. Then the low composite is subtracted from the high composite to produce the composite-difference map, just as in the synoptic analysis of observed snow extent.

Ideally, of the 9 years available (1980-1988), three high and three low years would be chosen for use in the composite analysis for each region. However, for some regions, those nine values of snow extent are not normally distributed. Thus for each region the time series of snow extent is examined with the goal of grouping three high and three low values. As a result, between one and eight values are included in each group. In two cases (UCLA, January PC4; and CSU, March PC7) regions were 100% snow covered in every year and are therefore excluded from this portion of the analysis.

Since so few years are available for composites, no attempt is made to quantify the statistical significance of the modeled 500-mbar composite differences. Model results are compared to the observed 500-mbar composite differences in the following manner. First, several key characteristics of the synoptic analysis of observations [*Frei*, 1997] are identified (Table 5). For each model/region, the 500-mbar composite results are categorized as either "poor," "weak," or "strong." A poor result is one that bears no resemblance to the observed synoptic pattern; a weak result has some, but not all, or not the most important characteristics of the observed synoptic pattern; and a good result contains features that resemble the most important characteristics of the observed synoptic pattern. With only 9 years of simulation, a more quantitative, less subjective, approach is inappropriate.

Table 4. Longitudes and Latitudes of Boxes Used to Estimate Regions of Coherent Snow Fluctuations

	Minimum Longitude	Maximum Longitude	Minimum Latitude	Maximum Latitude
November PC1	-115	-100	43	52
December PC1	-125	-110	37	43
	-125	-115	43	49
December PC5	-92	-72	40	45
January PC3	-91	-75	36	44
January PC4	-110	-100	45	49
February PC1	-105	-83	37	45
	-110	-100	45	50
March PC2	-100	-90	41	49
	-110	-100	44	52
March PC7	-85	-70	41	45

PC numbers indicate the component number, with 1 explaining the most variance. Regional snow extents and mean surface air temperatures are calculated using all grid boxes whose centers fall within these limits. All longitudes are negative, indicating degrees west; all latitudes are degrees north. In the cases of three regions (December PC1, February PC1, and March PC2), observed coherent regions are represented by two latitude-longitude delimited boxes, and regional average values are the area-weighted averaged values from the two boxes.

4.3 Surface Air Temperatures

In both observations [*Frei*, 1997] and models (Table 6), surface air temperatures are inversely correlated to snow extent throughout the snow season, even during periods of ablation, when the link to synoptic circulation weakens [*Frei*, 1997]. In the models, approximately 68% (31/46) of the correlation coefficients are significant at ≥95%. One model, MPI, had significant correlations in all eight regions. All other models had significant correlations in at least four regions.

There appears to be a tendency for models to produce higher snow extent-temperature correlations during the ablation season. For example, during March all models for which comparison was possible had significant correlations over both regions (PC2 and PC7). For no other region, in any other month, did all models have significant correlations. This increase in correlation is not apparent in the observations.

4.4 The 500-mbar Geopotential Height Patterns

For each model and each region, simulated 500-mbar composites are compared to the observed composites. Details for each region and each model are presented by *Frei* [1997]. Here the results are summarized. Below, a more in-depth case study of the MPI model is presented.

No model simulates observed midtropospheric patterns for all regions (Table 7). MPI results bear a "strong" resemblance to observed patterns for five of the eight regions. BMRC bears either a "weak" or "strong" resemblance in five of eight regions. Three models, LMD, SNG, and UCLA, show either a weak or strong similarity in three or four of the eight regions. As a whole, the models have a "weak" or "strong" resemblance to observed 500-mbar patterns in only 22 of 46 cases (48%).

There is no obvious pattern, either geographically or temporally, to the success or failure of models to simulate the synoptic conditions associated with extreme regional snow extent. Western and eastern North American regions have the same number of combined weak and strong results (11 weak or strong results out of 23 regional simulations). Early and late season results are similar.

5. Case Study of the MPI Model

We compare, region by region, synoptic teleconnection patterns associated with extreme regional snow extent in the

Table 5. Key Characteristics of Synoptic 500-mbar Geopotential Height Patterns Associated with High Regional Snow Extent

	Key Characteristics of Synoptic 500-mbar Patterns Associated With Extreme Snow Extent
	Western North America
November PC1, December PC1, January PC4	Eastern trough expands westward; western ridge shifts westward over eastern Pacific Ocean; increase heights over eastern Pacific Ocean; decreased heights over western North America.
March PC2	Same as above, but with lower heights over central Canada, more zonal flow over eastern Pacific Ocean and western continent.
	Eastern North America
December PC5	Dipolar teleconnection over North Atlantic Ocean, with northern center of action over southern Greenland and Baffin Bay, and southern center of action stretching from eastern continent across Atlantic Ocean at around 45°N latitude.
January PC3	Dipolar teleconnection over North Atlantic Ocean, with northern center of action over southern Greenland and Baffin Bay, and southern center of action stretching from eastern continent to eastern Atlantic Ocean at around 40°N latitude; stronger ridging around Alaska.
February PC1	Dipolar teleconnection over North Atlantic Ocean, with northern center of action over southern Greenland and Baffin Bay, and southern center of action stretching from central continent at around 50°N latitude to eastern Atlantic Ocean at around 45°N latitude; hint of stronger ridging around Alaska.
March PC7	Weak signal across domain of analysis, except directly over eastern continent, where height depressions are found; hint of weakened Icelandic Low and strengthened Aleutian Low.

PC numbers indicate the component number, with 1 explaining the most variance. The selection of regions, and information about each region, is taken from the synoptic analysis of observed snow extent [*Frei*, 1997].

GCM from the Max-Planck-Institut (MPI) in Hamburg, Germany to observations. The MPI simulation had "strong" similarity to observed synoptic patterns between 1972 and 1994 in five of the eight regions (Table 7). In addition, the MPI model was rated extremely high in its ability to simulate the global hydrological cycle [*Lau et al.*, 1996]. In Figures 5 through 8, three sets of maps are shown for each region: (top) observed and modeled composite 500-mbar heights for years of high snow extent, (middle) observed and modeled composite 500-mbar heights for years of low snow extent, and (bottom) observed and modeled 500-mbar height differences between composite high and composite low snow extent.

5.1 Western North American Regions

Extensive snow over western North American regions is associated with a westward expansion of the eastern trough and a shifting of the western ridge from around 120°W longitude to around 140°W longitude, resulting in increased heights over the eastern North Pacific Ocean and decreased heights over western North America. For each month we compare the modeled relationship between snow and the tropospheric wave train to the observed relationship.

During November (PC1, western North America) the MPI model exhibits mid-tropospheric behavior similar to what is observed. The westward expansion of the eastern trough and the shifting of the western ridge both occur during extensive snow months (Figure 5), resulting in increased heights over the eastern North Pacific Ocean and decreased heights over western North America (Figure 5, bottom).

In December (PC1, western North America), the MPI model does not capture the observed synoptic patterns (December maps are not shown, but the observed difference maps strongly resemble the November PC1 maps). Although the composite difference map for MPI December PC1 resembles the observed composite difference map (with positive differences over the eastern North Pacific and negative differences over the continent) this is deceiving. In the model, during extensive snow months the trough over the continent is deeper, and the wave pattern over the eastern Pacific and western North

Table 6. Spearman Correlation Coefficients Between Regional Snow Extent and Regional Surface Air Temperature in Model Simulations

	NOV1	DEC1	JAN4	MAR2	DEC5	JAN3	FEB1	MAR7	Total
BMRC	-0.55	-0.33	-0.66*	-0.62*	-0.64*	-0.63*	-0.77*	-0.67*	6/8
CSU	-0.79**	-0.72*	-0.21	-0.91**	-0.73*	-0.56	-0.32	na	4/7
LMD	-0.41	-0.72*	-0.64*	-0.70*	-0.37	-0.50	-0.91**	-0.63*	5/8
MPI	-0.68*	-0.72*	-0.84**	-0.65*	-0.67*	-0.74*	-0.72*	-0.95**	8/8
SNG	-0.41	-0.79**	-0.49	-0.85**	-0.45	-0.19	-0.87**	-0.84**	4/8
UCLA	-0.83**	-0.49	na	-0.08	-0.75*	-0.84**	+0.91	-0.67*	4/7
Total	3/6	4/6	3/5	5/6	4/6	3/6	4/6	5/5	31/46

Abbreviation na appears when a model had 100% snow cover extent in a region for all 9 years of AMIP simulation, so no correlation coefficient can be calculated. "Total" numerator specifies how many cases were significant at ≥95%; denominator specifies total number of cases. PC numbers indicate the component number, with 1 explaining the most variance.

*≥95% significance.

**≥99% significance (n=9, 1-tailed).

Table 7. Summary of Modeled Versus Observed Synoptic 500-mbar Patterns Associated With Extreme Regional Snow Extent

	Western				Eastern				Weak or strong/Total
	NOV1	DEC1	JAN4	MAR2	DEC5	JAN3	FEB1	MAR7	
BMRC	weak	strong	weak	poor	weak	strong	poor	poor	5/8
CSU	strong	poor	poor	poor	poor	poor	poor	na	1/7
LMD	poor	poor	poor	poor	weak	strong	strong	strong	4/8
MPI	strong	poor	poor	strong	poor	strong	strong	strong	5/8
SNG	poor	strong	strong	strong	poor	poor	strong	poor	4/8
UCLA	strong	strong	na	poor	poor	poor	poor	weak	3/7
Weak or strong/ Total	4/6	3/6	2/5	2/6	2/6	3/6	3/6	3/5	22/46

Poor means no resemblance; weak means the model simulates some, but not all or not the most important, characteristics of the observed patterns; strong means the model simulates the important characteristics of the observed patterns. Abbreviation na appears when a model had 100% snow cover extent in a region for all 9 years of AMIP simulation, so no composites of extreme cases are available. "Weak or strong/Total" numerator specifies how many cases had either weak or strong resemblance; denominator specifies the total number of cases. PC numbers indicate the component number, with 1 explaining the most variance.

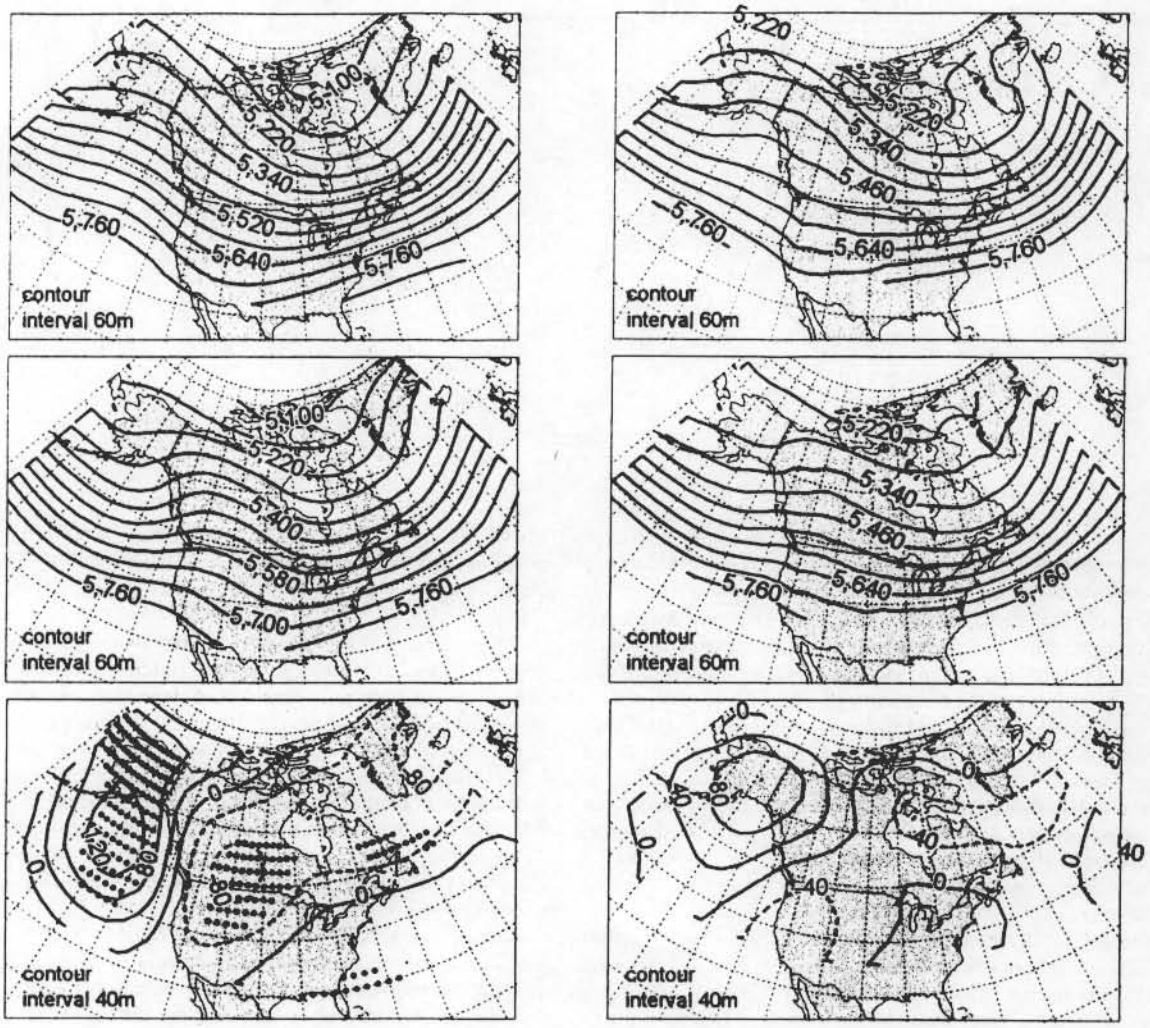

Figure 5. Composite 500-mbar geopotential heights for November PC1 for (left) observations, and (right) MPI GCM. (top) Composite for years of high snow extent, (middle) composite for years of low snow extent, and (bottom) composite difference (top) - (middle). In the composite difference map, negative contours are dashed, and grid cells with differences significant at ≥95% according to the Students *t*-test are indicated with solid circles. Significance is not tested in model results (see text for explanation).

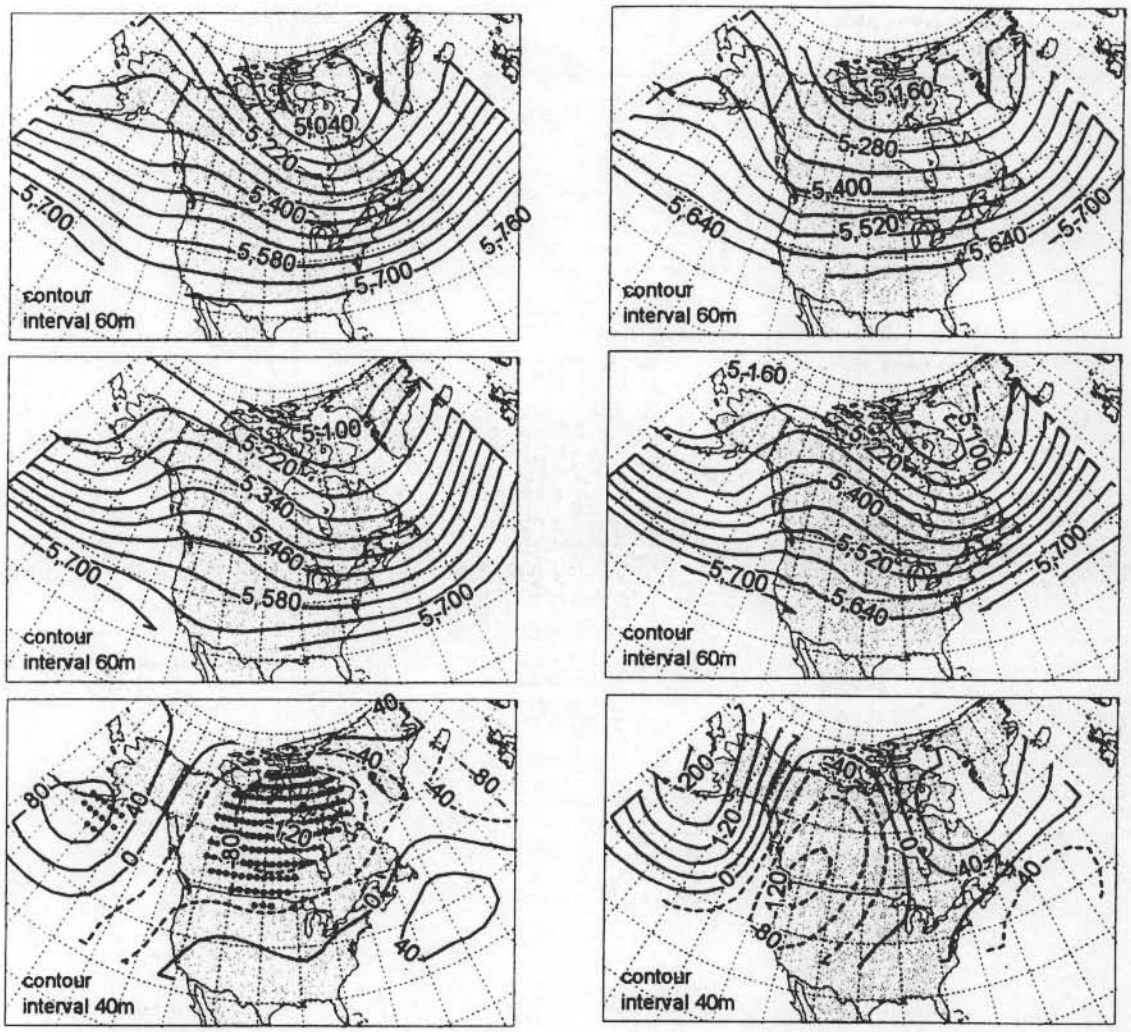

Figure 6. Same as Figure 5, except March PC2.

America is more meridional (i.e., higher amplitude) with a stronger jet stream. As a result, a higher amplitude wave causes a higher frequency of cold air masses and snowfall over the western portion of the continent. The observed mechanism for extensive snow extent in this region, which primarily involves a westward shift of the ridge location (but not necessarily a change in amplitude), is not seen in the model.

Synoptic patterns associated with extreme snow over January (PC4) are not captured by the model (not shown). In the model, high snow extent is associated with a stronger jet stream farther to the south than during months of low snow extent, resulting in a 500-mbar difference map with lower heights south of around 40°-50°N and higher heights farther north. The westward shifting of the ridge seen in observations is not seen in the model results.

For March (PC2) the MPI model captures the westward shift of the ridge associated with extensive western snow (Figure 6). The simulated area of depressed 500-mbar heights over the continent is located southwest of the observed depression, but the areas of enhanced ridging in both the model and observations are centered over the Aleutian Islands. The simulated patterns are caused by a westward shift of the ridge by about 30°-40° longitude, which is even more than seen in observations. In fact, during March, observations indicate that the remote center of action over the Aleutians begins to weaken. In the model, the remote center of action appears to be even stronger than earlier in the season.

5.2 Eastern North American Regions

Over eastern North America, extensive snow is associated with a dipolar meridional oscillation in the 500-mbar geopotential height field over the western North Atlantic Ocean. The northern node, or center of action, is found over southern Greenland and Baffin Bay; the southern node has a wider longtudinal extent, stretching from the continental United States eastward across the central North Atlantic, between 40°N and 50°N, the exact location depending on the month (see Table 5). These remote associations weaken in March, when the only significant signal at the 500-mbar level is directly above the snow-covered region.

During December, the observed north - south dipole pattern associated with eastern snow is not found in the MPI simulation (not shown). Although the model exhibits a stronger, more meridional jet over this region during extensive snow months, the observed teleconnection pattern over the North Atlantic is not captured. In fact, the weak simulated dipole pattern that does emerge in the Atlantic Ocean is opposite in sign to the observed pattern. Also, the model appears to have a remote association to the Aleutian Low, which is not found in observations.

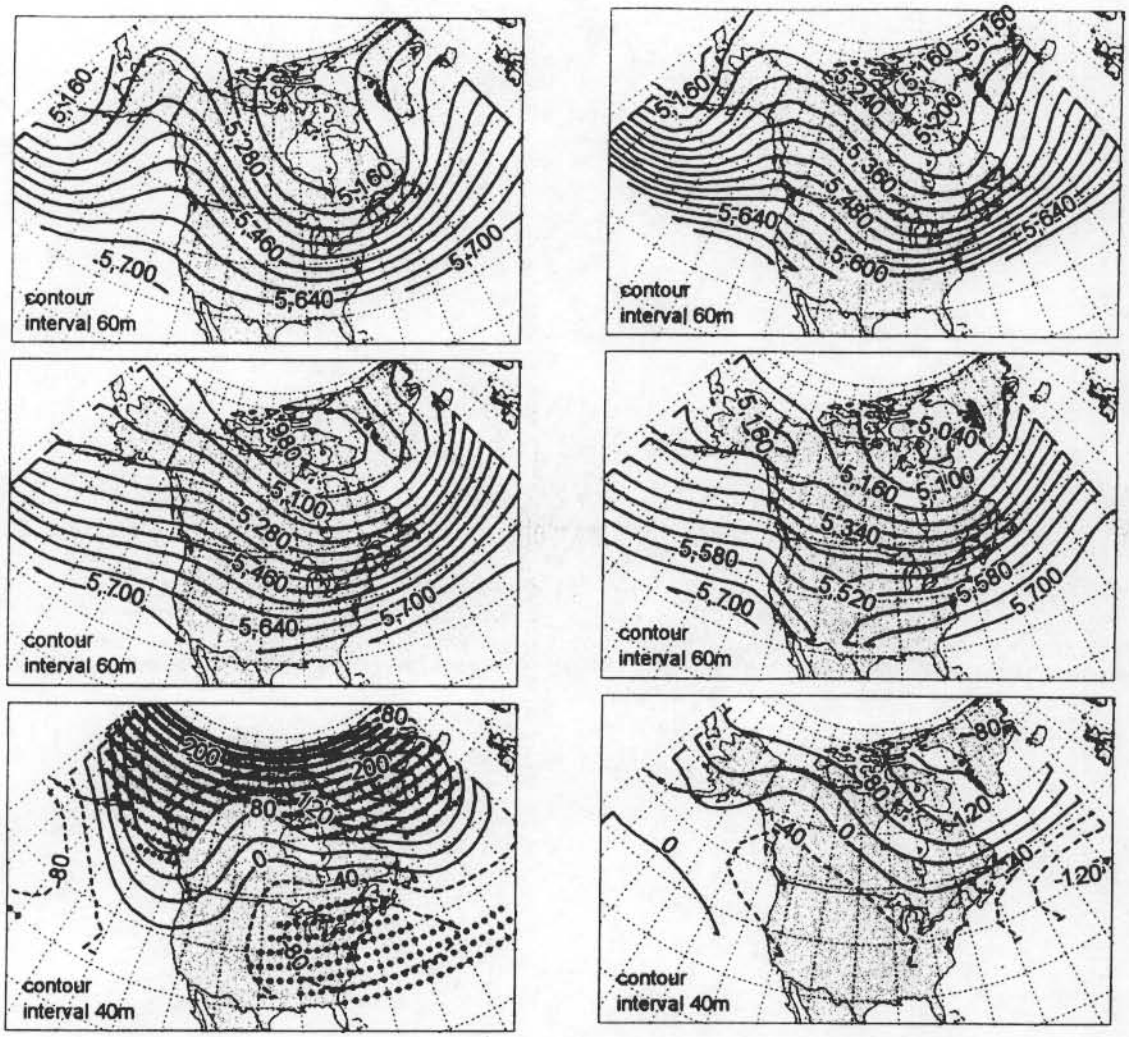

Figure 7. Same as Figure 5, except January PC3.

In January and February, however, the MPI model does capture the observed teleconnection patterns. In January (Figure 7) the dipole pattern over the Atlantic that stretches across North America is simulated, although the modeled southern dipole is farther east than in observations. The northern dipole is only slightly west of the observed location, and only a hint of the enhanced ridge observed over Alaska is apparent in the model. In February, as well, the observed north-south dipole is captured in the model (not shown). The enhanced trough over central North America is deeper and reaches farther north in the model. In the model the Atlantic southern dipole is almost detached from the continental trough, unlike the observed continuous area of depressed heights. The northern center of action is located, in both the model and observations, just over Greenland and Baffin Bay.

During March the MPI model captures the observed weakening of the link between snow extent and synoptic scale circulation (Figure 8). Observed height depressions are significant only over eastern North America. Over the remainder of the domain of analysis, 500-mbar composite height differences are small, and only hints of enhanced Aleutian troughing, and Icelandic ridging, are apparent. In the model, composite height differences are generally small, including over the eastern continent, although signs of Icelandic ridging are apparent. The significance of the modeled pattern of composite differences over the Pacific Ocean, that is not seen in observations, is not clear.

6. Summary and Conclusions

Results from 27 GCMs run under the auspices of the Atmospheric Model Intercomparison Project have been evaluated in terms of their simulations of snow extent across the northern hemisphere. A subset of six models is chosen for evaluation of their simulations of midtropospheric synoptic-scale teleconnection patterns associated with regional snow extent over North America.

AMIP models reproduce a seasonal cycle of snow extent similar to the observed cycle, but are inconsistent, from region to region, season to season, and model to model, in the accuracy of their simulations of mean snow extent and synoptic teleconnection patterns associated with regional snow extent. In general, the 27 GCMs analyzed underestimate fall and winter snow extent (especially over North America) and overestimate spring snow extent (especially over Eurasia). Teleconnection patterns associated with regional snow extent are at least partially captured in five of six models evaluated, but only two of those models simulate realistic teleconnection patterns in more than half of the regions analyzed.

Interannual variability is also evaluated, in terms of both dispersion and correlation. Models tend to underestimate the observed range of snow extent, usually by at least 50%. Only three models (UGAMP, UIUC, and UKMO) exhibit, for all seasons over NH, NA, and EU, both (1) median snow cover extent within 10% of the land surface area compared to observations and (2) even half of the observed variability.

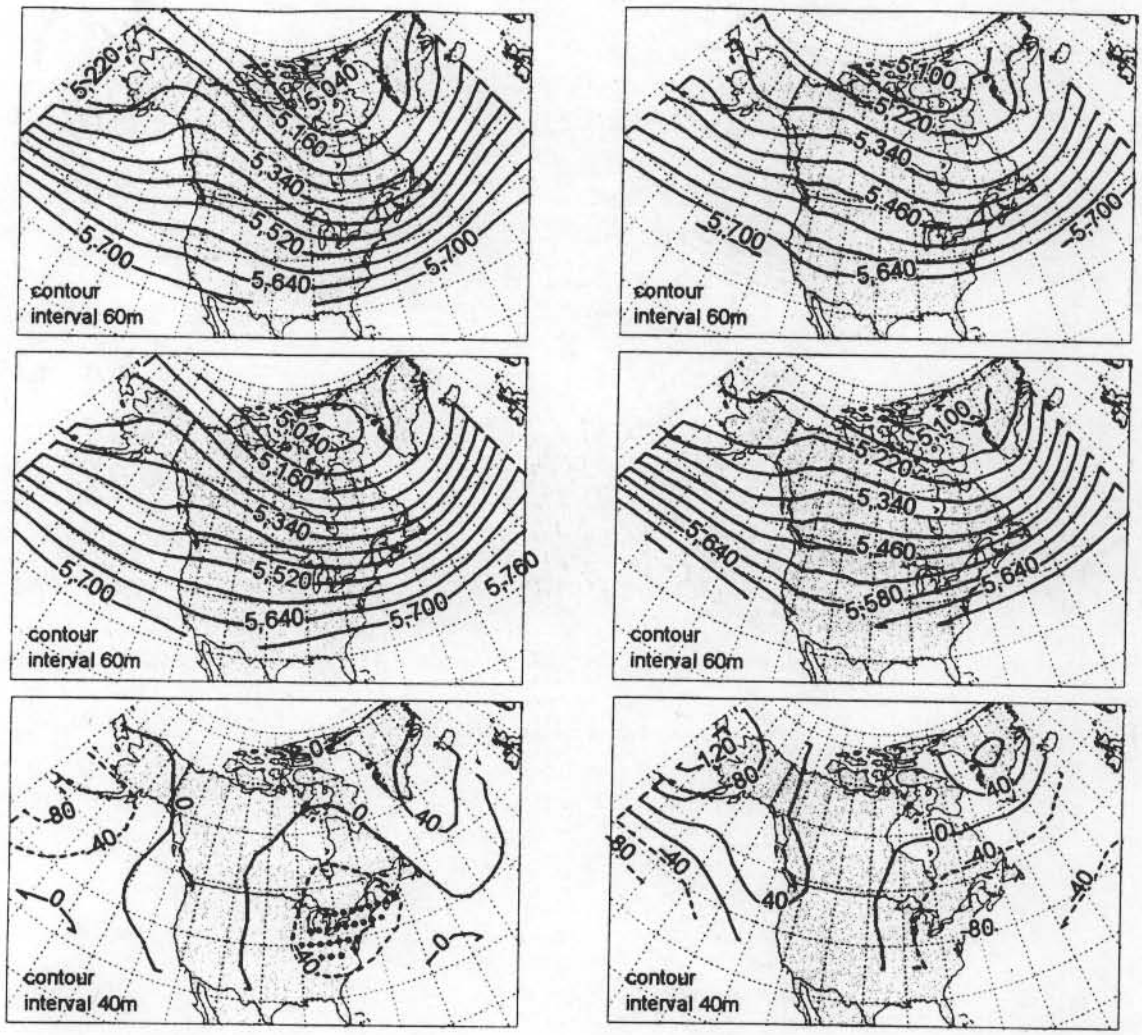

Figure 8. Same as Figure 5, except March PC7.

Since AMIP runs use observed SST values as the lower boundary condition, it is possible to compare simulated to observed snow extent for actual calendar years. However, poor temporal correlation is found between modeled and observed interannual fluctuations of snow extent, even when only months with extremely high or low values (i.e., snow extent values are outliers in their monthly distributions) are considered. This indicates that snow extent in AMIP GCMs is not driven by SSTs, which is consistent with the results of other analyses that show that in midlatitudes, precipitation and other variables have little potential predictability from SSTs.

Climate models are an important tool for estimating changes in regional climates under evolving boundary conditions, such as increased concentrations of atmospheric "greenhouse" gases. Since snow is an important source of water in many regions and exerts a significant influence on the surface radiation balance, prediction of the feedback effect associated with snow cover is important for understanding regional, as well as hemispheric, climatic change. The results of this analysis can be helpful to improve model parameterizations of subgrid-scale mass and energy fluxes associated with snow cover. GCMs generally simulate synoptic-scale tropospheric fluctuations more reliably than surface temperature and precipitation variations over smaller (subgrid-scale) regions [*Sperber and Palmer*, 1996; *Hewitson and Crane*, 1992; *Grotch and MacCracken*, 1991]. Perhaps those models that successfully simulate synoptic patterns associated with snow extent will be most useful for predicting patterns of snow in an enhanced greenhouse environment. Those models will be evaluated in more detail under the auspices of AMIP 2, which includes multiple (i.e., ensemble) runs with longer integration periods and increased spatial resolutions.

Acknowledgments. A.F. performed this work under a NASA Graduate Student Fellowship in Global Change Research; D.A.R. is supported by NSF grants ATM-9314721 and SBR-9320786. We thank the staff at PCMDI for services, and John Walsh, leader of the AMIP Diagnostic Subproject 8 (Polar Phenomena and Sea Ice). We also thank two anonymous reviewers for their insightful and helpful comments.

References

Boer, G.J., et al., Some results from an intercomparison of the climates simulated by 14 atmospheric general circulation models, *J. Geophys. Res.*, *97*(D12), 12,771-12,786, 1992.

Cess, R.D., et al., Interpretation of snow-climate feedback as produced by 17 general circulation models, *Science*, *253*, 888-892, 1991.

Cess, R.D., et al., Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models. *Science*, *245*, 513-516, 1989.

Foster, J., et al., Snow cover and snow mass intercomparisons of general circulation models and remotely sensed datasets, *J.Clim.*, *9*, 409-426, 1996.

Frei, A., Towards a snow cover fingerprint for climate change detection, Ph.D. dissertation, Grad. Program in Geogr., Rutgers, The State Univ., New Brunswick, N.J., 1997.

Gates, W.L., AMIP: The Atmospheric Model Intercomparison Project, *Bull. Am. Meteorol. Soc.*, *73*, 1962-1970, 1992.

Grotch, S.L., and M.C. MacCracken, The use of general circulation models to predict regional climate change. *J. Clim.*, *4*, 286-303, 1991.

Hewitson, B.C., and R.G. Crane, Regional-scale climate prediction from the GISS GCM. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, *97*, 249-267, 1992.

Kukla, G., and D.A. Robinson, Accuracy of operational snow and ice charts, *1981 IEEE Int. Geosci. Remote Sens. Symp. Dig.*, 974-987, 1981.

Lau, K.-M., J.J. Kim, and Y. Sud, Intercomparison of hydrologic processes in AMIP GCMs. *Bull. Am. Meteorol. Soc.*, *77*, 2209-2226, 1996.

Panofsky, H.A. and G.W. Brier, *Some Applications of Statistics to Meteorology*. Coll. of Earth and Miner. Sci., Penn. State Univ., University Park, 1958.

Phillips, T.J., *A summary documentation of the AMIP models*. PCMDI Rep. 18, UCRL-ID-116384, Program for Clim. Model Diagnosis and Intercomparison, Univ. of Calif., Lawrence Livermore Nat. Lab., Livermore, Calif., 1994.

Phillips, T.J., L.C. Corsetti, and S.L. Grotch, The impact of horizontal resolution on moist processes in the ECMWF model. *Clim. Dyn.*, *11*, 85-102, 1995.

Randall, D.A., et al., Analysis of snow feedbacks in 14 general circulation models, *J. Geophys. Res.*, *99*(D10), 20,757-20,771, 1994.

Randall, D.A., et al., Intercomparison and interpretation of surface energy fluxes in Atmospheric general circulation models. *J. Geophys. Res.*, *97*(D4), 3711-3724, 1992.

Robinson, D.A., K.F. Dewey, and R. Heim, Jr., Global snow cover monitoring: An update. *Bull. Am. Meteorol. Soc.*, *74*, 1689-1696, 1993.

Sperber, K.R., and T.N. Palmer, Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. *J. Clim.*, *9*, 2727-2750, 1996.

Weare, B.C. and AMIP Modeling Groups, Evaluation of the vertical structure of zonally averaged cloudiness and its variability in the atmospheric model intercomparison project. *J. Clim.*, *9*, 3419-3431, 1996.

Wiesnet, D.R., C.F. Ropelewski, G.J. Kukla, and D.A. Robinson, A discussion of the accuracy of NOAA satellite-derived global seasonal snow cover measurements, in *Large Scale Effects of Seasonal Snow Cover, Proceedings of the Vancouver Symposium, August 1987, IAHS Publ., 166*, 291-304, 1987.

Zhong, A., Global climate model simulations of snow and sea-ice. M.S. Thesis, Clim. Impacts Cen., Sch. of Earth Sci., Macquarie Univ., North Ryde, N.S.W., Australia, 1996.

Zwiers, F.W., Inter-comparison of inter-annual variability and potential predictability (Subproject 2), in *Proceedings of AMIP Scientific Conference, Monterey, California, May 1995*, W.L. Gates (ed), pp. 245-251, World Climate Research Programme, World Meteorological Organization, Geneva, 1995.

A. Frei (corresponding author), CIRES/NSIDC, Campus Box 449, University of Colorado, Boulder, CO 80309. (e-mail: frei@kryos.colorado.edu)

D. A. Robinson, Department of Geography, Rutgers University, Piscataway, NJ 08854.

(Received July 2, 1997; revised January 7, 1998;
accepted January 7, 1998.)