Northern Hemisphere Winter Surface Temperature Predictions based on Land-Atmosphere Fall Anomalies

Judah Cohen

Atmospheric and Environmental Research, Inc., Lexington, MA

Christopher Fletcher

Department of Physics, University of Toronto, Toronto, ON, Canada

Submitted to the Journal of Climate

June 19, 2006

Corresponding author address:
Judah Cohen
AER, Inc.,
131 Hartwell Avenue
Lexington, MA 02421
jcohen@aer.com

Abstract

A statistical model has been developed using observed October snow cover and sea level pressure anomalies to predict land surface temperatures for the extratropical Northern Hemisphere. Snow cover has been used in operational forecasts for seven winters and sea level pressure anomalies for three winters. Presented are skill scores for the seven real-time forecasts and for hindcasts for the winters 1972/73-2004/05. The model demonstrates positive skill for much of the eastern United States and Northern Eurasia; regions that have eluded skillful predictions based on ENSO forecasts. Comparison of the model skill with the world's leading dynamical forecast systems shows superior skill for the same regions. As the dynamical models were developed with special sensitivity to the tropical ocean-atmosphere coupling, these results suggest that as long the climate models remain preferentially forced by ENSO, the statistical model described in this paper is likely to outperform the more highly complex dynamical models in the extratropics into the foreseeable future.

Introduction

It is estimated that about one-third or 3-4 trillion dollars (NOAA 2002; Dutton 2002) of the United States (US) economy is sensitive to the impacts of weather and climate. Mitigating hazards through advanced warnings and improving the performance of climate-sensitive economic sectors through seasonal prediction are thus of interest to industry and government agencies. The most important advance in understanding climate variability and its application to seasonal prediction has been the linkage of the dominant tropical atmosphere and ocean signal (El Niño/Southern Oscillation or ENSO) with surface temperatures and precipitation patterns across the globe. However, predictive skill for temperature forecasts outside of the tropics, including the U.S., has been mixed (Barnston et al. 1999; Spencer and Slingo 2003). For example, temperature anomalies during the winter of 2002/03 were poorly predicted by U.S. forecast centers, despite the occurrence of a moderate El-Niño. Clearly, much room for improvement remains in our understanding of wintertime climate variability, in particular in the extratropics, where the dominance of ENSO is more tenuous. Better understanding of the dominant mode of NH winter climate variability, referred to as the North Atlantic Oscillation (NAO) or the Arctic Oscillation (AO), which could lead to improved predictability, is often recognized as the next most important anticipated advance in seasonal climate forecasting (Cohen 2003), especially for the eastern U.S. and Europe, regions where forecasts based on ENSO have little or no skill.

The surface-temperature and surface-circulation signatures of the NAO/AO are strongest in the North Atlantic sector (Hurrell 1995; Thompson and Wallace 2001; Ambaum et al. 2001; Cohen and Saito 2002). The NAO/AO has been linked with sea surface temperature (SST) variability, sea ice variability, stratospheric forcing, subpolar air temperature and aerosols (Rodwell et al. 1999; Mysak and Venegas 1998; Baldwin and Dunkerton 1999; Fletcher and Saunders 2006; Perlwitz and Graf 1995). Linking the AO to slowly varying boundary conditions could provide predictability, nonetheless, recent articles on the subject have emphasized the lack of understanding of the underlying dynamics driving NAO variability and consequently its poor predictability (Hurrell et al. 2001).

During the period it has been extensively monitored, snow cover has exhibited similar trends to the Northern Hemisphere (NH) climate cycles, peaking in the late 1970's and collapsing to record minimum values in the late 1980's and early 1990's (Robinson, 1996). In the NH, snow cover is the most variable land surface condition in both time and space (Cohen, 1994) making it a viable candidate for amplifying climate and atmospheric anomalies. Cohen and Entekhabi (1999) first demonstrated that the time series of fall Eurasian snow cover is significantly correlated with the winter AO. Bojariu and Gimeno, 2003, Saito and Cohen, 2003, Saunders et al., 2003 further demonstrated that the significant relationship between Eurasian snow cover and the winter AO is not limited to the fall but is evident in the summer as well. Therefore snow cover is potentially useful as a leading indicator of winter climate, especially those land areas in the North Atlantic sector where the influence of the NAO/AO are strongest.

A simple model has been developed making use of observed Eurasian snow cover and seas level pressure (SLP) anomalies for winter climate prediction of extratropical NH surface temperatures. In this paper we will demonstrate the skill of this model tested both in real-time and in hindcasts, consistently outperforms winter forecast from the major governmental forecast centers. This paper is meant to further demonstrate the link between fall snow cover and regional atmospheric anomalies and the NH general circulation on seasonal time scales and the potential societal benefit of incorporating snow cover variability in seasonal climate forecasts.

2. SLP/snow index

The universal lynchpin of seasonal forecasts has been the ENSO phenomenon (Barnston et al. 1994; van Oldenborgh et al. 2005; Saha et al. 2006). The modern age of seasonal forecasts is considered to have been born in the winter of 1997/98 when the US government successfully forecasted temperatures and precipitation across the US. However repeat success has remained elusive. Plotted in Figure 1 is the correlation of the Nino 3.4 index and NH extratropical surface temperatures. Little of the NH land masses are highly correlated with ENSO with the exception of the immediate West Coast of North America and the Canadian prairies. Based on this figure, ENSO-derived correct temperature forecasts would be the exception rather than the rule given the scarcity of significant correlations.

We performed an Empirical Orthogonal Function (EOF) analysis on observed surface temperatures (T_s) from the NCEP/NCAR Reanalysis data, years 1972-2006 (Kalnay et al. 1996); the dominant mode of variability for December, January and February (DJF) accounts for 18% of the total variance (not shown; pattern closely resembles Figure 2). The first mode is often referred to as a quadropole and is associated with winter season NAO/AO variability (Wallace and Gutzler 1981; Barnston and Livezey 1987; Thompson and Wallace 1998). The dominant temperature pattern is characterized by two same-signed anomaly centers stretched across Northern Eurasia and the eastern US and two same-signed anomaly centers across the Mediterranean and North Africa and Northeastern Canada and Greenland. So for example when the NAO/AO is negative anomalous high pressure over the continents advects a cold flow of air over Northern Eurasia and the eastern US while North Africa, the Mediterranean, Northeastern Canada and Greenland are warmed by an anomalous southerly flow of air. Comparison of the correlation maps of the winter AO and winter Nino 3.4 index with NH extratropical winter T_s, shows that a correct prediction of the winter AO would provide as much as 50% improvement in temperature variance explained over Eurasia and 30% improvement in temperature variance explained over the eastern United States compared with a correct prediction of the ENSO-state.

The National Oceanic and Atmospheric Administration (NOAA) produces weekly large-scale observations of the spatial extent of NH continental snow cover primarily based on visible-band satellite imagery (Robinson et al. 1993). From the weekly data, a time series is created for the areal extent of Eurasian snow cover for the month of October (Cohen et

al. 2001). Next we correlate the time series of Eurasian October snow cover with the time series of the first EOF of T_s (Fig. 1). The two time series are correlated at a value of 0.45, which is statistically significant at the 99% confidence interval. Also shown in Figure 1 is the correlation of the Eurasian October snow cover time series with the grid point time series of DJF T_s. The pattern of temperature variability associated with interannual snow cover anomalies is reminiscent of the pattern associated with the winter NAO/AO. Present is the quadropole pattern, though weaker, with one signed anomaly in the eastern US and across northern Asia and opposite signed anomalies in Northeastern Canada, Greenland, and the western portion of North Africa and the Mediterranean. The most notable difference between the AO pattern of variability and that associated with snow cover is the lack of significant correlations across Europe.

Cohen et al. 2001, were the first to attempt to increase the correlation value of Eurasian snow cover by combining it with an observed simultaneous atmospheric variable. They derived a time series from observed October snow cover and SLP anomalies from a fixed gridpoint in Siberia, also from October, which was more highly correlated with the winter AO than using snow cover alone. Finally Cohen et al. (2002) postulated that the winter AO, which is hemispheric in scale, originates in the fall as a regional lower tropospheric anomaly that propagates and grows during the course of the cold season. The associated SLP anomaly was not fixed in space but rather could originate in different regions of Eurasia and the North Atlantic. An index that was derived using both the dominant SLP anomaly in October and October snow cover anomalies could then be combined into a single index that was highly correlated with the winter AO index. The correlation value

between the October SLP/snow index and the winter AO is 0.9. This index has been used operationally for the past three winter forecasts. However this index is based on forecaster interpretation as the SLP anomaly chosen for the index is derived from an analysis of hemispheric temperature and Eliassen-Palm flux anomalies and forecaster experience (techniques are described in Cohen et al. 2002 and Cohen 2003).

For the hindcast validation we use an alternate October SLP/snow index that is less skillful but is not dependent on forecaster interpretation and is therefore more easily reproduced. Gridded monthly-mean October SLP anomalies for northern Eurasia are analyzed over the domain 0-180ºE and 50-80ºN. If a single SLP anomaly center is observed over this region then its central maximum denotes the value for the SLP anomaly for the index. If multiple SLP centers are observed then the chosen SLP value depends on the sign of the contemporaneous October Eurasian snow cover anomaly. If the snow cover is above normal a positive anomaly is chosen and if the snow cover is equal to or below normal a negative anomaly is chosen. For the hindcasts, this algorithm produced a unique solution for the SLP anomaly. The weighting for the two variables is determined by the multiple regression of the SLP and snow anomalies with the observed winter AO index. This yields the equation:

$$SS_i = \alpha snow_i + \beta slp_i + \lambda$$

where:

SS = SLP/snow index

snow = observed October Eurasian snow anomaly in 10^6 km^2

slp = observed SLP anomaly in northern Eurasia in hPa

i = year

For the hindcasts, the value of α = 0.25, β=0.40 and λ= 0.0. In Figure 2 we plot the SLP/snow index and the correlation value with the winter AO, which is equal to 0.61.

Also shown in Figure 2 is the correlation of the October SLP/snow index with NH T_s. Again the quadropole pattern of temperature variability is noted with the same signed anomaly across northern Eurasia and the eastern US and opposite signed anomalies in northeastern Canada, Greenland, North Africa and the Mediterranean. In comparison to the correlation map of just October snow cover, the correlations are higher and Europe is now included in the region of significant correlations. The plot closely resembles that of the AO correlated with T_s. The SLP/snow index is the basis of a simple statistical model used to make real-time winter forecasts for the US, Europe and Asia. The current forecast model uses October snow cover and SLP anomalies and the recent trend in DJF surface temperatures as predictors for surface temperatures. However in some of the earlier real-time forecasts, summer snow cover and ENSO were also used as predictors in the model. In the remainder of paper we will refer to this statistical model as the snow-cast model or *SCAST* model for short.

3. Forecast Verification

The accuracy of seasonal forecasts and hindcasts is referred to as the prediction 'skill'. In this study, we assess skill using two skill measures. First, we employ the Pearson

product-moment correlation coefficient between the observed and predicted values. Henceforth, this measure is referred to as the anomaly correlation coefficient (ACC) or simply the anomaly correlation. Second, we employ the percentage improvement in Root Mean Square Skill score (RMSS) over a simple forecast of climatology.

$$\text{RMSS} = \frac{100}{n} \sum_{i=1}^{n} \; 1 - \sqrt{\frac{(\widehat{T}_i - T_i)^2}{(\overline{T} - T_i)^2}}$$

where:

$\widehat{T}$ = model predicted temperature

T = observed temperature

$\overline{T}$ = climatological value of observed temperature

i = year

n = total number of years

Climatology in this study is the long-term mean for the assessment period 1972/73-2004/05.The mean-squared skill score is the preferred skill measure of the World Meteorological Organization for deterministic seasonal forecasts (WMO 2002) because, as opposed to the anomaly correlation, it penalizes bias in prediction models. However for consistency between forecasts (where we analyze their average root-mean squared error) and hindcasts, we use the closely related RMSS.

The statistical significance of the anomaly correlation is estimated using a Student's t-test against the null hypothesis of zero correlation. Serial correlation in the temperature data

could cause spurious inflation of the prediction skill. We correct for this using the method of Davis (1976) to reduce the number of available degrees of freedom in the hypothesis test. Since the RMSS has no lower bound, a probability density function of RMSS values exhibits a large positive skewness. Therefore, no significance test is carried out for the RMSS skill values. We also assess the spatial accuracy of our hindcast model. This is achieved using pattern correlations where forecasts and observations are compared at each gridpoint, and the average gridpoint root-mean-square errors (RMSE) in the domain (Wilks 1995). Both of these metrics are calculated using data area-weighted by the square-root of the cosine of latitude.

Ensemble-mean seasonal hindcast data from general circulation model (GCM) simulations are analyzed for comparison. We assess the hindcast skill of the following three GCM or ensemble of GCMs: the Climate Prediction Center (CPC) Climate Forecast System (CFS) retrospective hindcast project for the period 1983-2004 (Saha et al. 2006), the Canadian Historical Forecasting Project (HFP) for the period 1972-1993 (Derome et al. 2001) and four of the seven GCMs from the European Center for Medium-range Weather Forecasting (ECMWF) DEMETER project that from which data were available for the period 1972-2001 (Palmer et al. 2004). No bias correction is applied to these ensemble means, which could introduce some bias in the hindcast climatology and, subsequently, the hindcast skill assessment (e.g., Palmer et al. 2000).

4. Model Skill

a. Real time forecasts

The *SCAST* model has been used operationally for seven consecutive winters (1999-2005). The model produces a temperature anomaly forecast for the entire extratropical NH. Forecasts are issued by the tenth business day of the month for the following three months (i.e., the forecast for DJF 2005/06 was issued on or before November 15th 2005). The model is a statistical model based on a variable number of predictors. For each forecast the model linearly combines one to three of the following four predictors: recent trend, predicted seasonal value of the Niño 3.4 index, Eurasian snow cover and the SLP/snow index

The focus of the remainder of the paper will be to verify the skill of the real-time winter (DJF) forecasts for the past seven years and hindcasts, which have been produced for the winters 1972/73-2004/05. Reliable snow cover area, derived from satellites, has only been available since 1972 and limits our ability to extend the hindcasts earlier than the winter of 1972/73. The model has evolved over time and the inputs or predictors have not remained constant. Table 1 lists the predictors used for each real-time winter forecast. However the most recent version of the forecast model linearly combines two predictors: the SLP/snow index and trend. The hindcasts produced for this study are fixed with just these two predictors.

In Figure 3, we plot both the anomaly correlation and the RMSS for all seven real-time winter forecasts. The anomaly correlations show large regions of positive correlations across North America, northern Eurasia, central Asia and North Africa. Especially high correlations of greater than 0.6 are observed in the eastern US, along the west coast of North America, Northern Europe and eastern Siberia. However when plotting the RMSS, only the eastern US and eastern Siberia have demonstrable skill over climatology. This discrepancy between the correlation and RMSS usually occurs when the forecast is correct in its sign of the predicted anomaly but not in magnitude.

Often forecasts are not useful or intended for a single location but are more valuable when they closely match the large-scale pattern of temperature anomalies. The two measures of skill, plotted in Figure 3 only show skill at an individual point or grid box but do not assess how well the model captures or predicts the observed large-scale pattern of temperature anomalies. Therefore in Figure 4 we plot the pattern correlations between the forecast and the observations for specified regions for all seven real-time forecasts. The pattern correlations for the US are shown with a solid line. Correlations for the US for the real-time forecasts have always been positive; and all years except for 2001/02, equal or exceed 0.6. The highest value was achieved in the winter of 2004/05 with a value approaching 0.9. Also shown are the area weighted root mean square area for the same regions, and not only are the pattern correlations consistently high but the RMSE are consistently low, with most years averaging close to one degree of error or less. Again the winter of 2004/05 scored best for the real-time forecasts.

Included in Figure 4 are the pattern correlations and the area weighted RMSE for Europe and the NH. The model scored for all years, except one, positive correlations with a maximum value of greater than 0.7 in the winter of 2001/02 for Europe and 0.5 in the winter of 2002/03 for the NH. Still the overall values are less than those observed for the US and from the plot of RMSE, the magnitude of the average error is greater for Europe and the NH than for the US, with values generally ranging between 2-3 degrees. Interestingly the model did best for the US in the winter of 2004/05 but did poorest for Europe and the NH that same year. The model was most consistent in the winter of 2002/03, doing well for all three regions. That October the second largest value of Eurasian snow cover was observed, such strong forcing resulted in a widespread temperature response consistent with the model's forecast. In Table 2 we list the mean value and the standard deviation for both the pattern correlation and the RMSE for the US, Europe and the NH for the seven real-time forecasts. All three regions show positive pattern correlations, though the value for the US is double that of Europe and the NH. Similarly the average RMSE is approximately 1.5 ºC for Europe and the NH but approximately 0.9 ºC for the US.

b. Hindcasts

Cross-validated hindcasts were produced using the SLP/snow index described in section 2 above for the winters of 1972/73 through 2004/05. In Figure 5 we plot the RMSS on the left and the anomaly correlation on the right. Contoured regions represent those areas where the model values are statistically significant based on the Student's t-test. The model has slight positive skill in the eastern US, parts of southern Canada and the North

American Arctic, northern Eurasia and the Mediterranean region. The regions where the model has positive skill corresponds to those same regions where the SLP/snow index is significantly correlated with NH surface temperatures; however statistically significant skill is mostly confined to Siberia. And though the skill for the hindcasts is modest, it represents a large improvement over other operational forecast models in the mid to high latitudes of the NH, as will be shown below.

Besides the SLP/snow index the other predictor in the model is linear trend. Significant regional trends are observed in NH air temperatures but little or no trend is observed in the time series of October Eurasian snow cover, October Eurasian SLP anomalies and the winter AO over the hindcast period. This suggests that linear trends should not contribute significantly to the derived skill in predicting NH air temperature from the SLP/snow index. However, it is important to quantify the proportion of hindcast skill that comes from linear trend. In Figure 6 we plot the relative contribution of trend to the overall model skill for North America and for the NH. Based on the anomaly correlation the trend contributes no discernible skill to the forecast. However based on the root mean skill score, the trend contributes positive skill in regions where the SLP/snow index is not highly correlated with T_s; parts of the western US and central Eurasia. In fact in some of the regions where the model has its highest skill, such as in the eastern US and Northern Europe, the trend contributes negative skill. The skill from trend appears to extend spatially or compliment the skill derived from the SLP/snow index rather than be a significant contributor to or overlap those regions where the SLP/snow index has skill.

In Table 2 we include the mean value and the standard deviation for both the pattern correlation and the RMSE for the US, Europe and the NH for the hindcasts. All three regions show positive pattern correlations, though in contrast to the forecasts, Europe has the highest pattern correlation and the US is lower and equal to that of the NH. The average RMSE for all three regions is between 1.2-1.5 ºC with the US region scoring the lowest RMSE values. The pattern correlations for the hindcasts are lower than for the forecasts, though the RMSE are consistent or slightly lower with the exception of the US, where the RMSE is considerably higher for the hindcasts than the forecasts. The model is different for the forecasts than the hindcasts and that difference is contributing to the differences in the pattern correlations; though with the possible exception of the US the RMSE seems to be less sensitive to the choice of SLP/snow index.

c. Comparison with dynamical models

The major forecast centers including those in the US and in Europe are relying more and more on coupled atmosphere ocean dynamical models and less on statistical models for seasonal forecasts. Statistical models most heavily rely on ENSO as predictors for temperature and precipitation (Barnston et al. 1994). Dynamical models are also highly tuned to SSTs in general and the ENSO cycle in particular (van Oldenborgh et al. 2005), though the models attempt to include many of the major processes in the ocean-land-atmosphere climate system. Therefore, given the completely different emphasis of the *SCAST* model and the major GCMs employed at some of the world's largest forecast centers, it is a worthwhile exercise to compare the skill derived from hindcasts between the *SCAST* model and some of the most widely used GCMS in seasonal forecasting.

In Figure 7 panels a) and b) we plot the anomaly correlation and the RMSS for the CFS, which is the GCM used by NOAA's CPC for seasonal forecasting. Positive model skill is shaded in red. The CFS model shows little consistent skill for North America with the exception of the north slope of Western Canada and Alaska. In panels c) and d) of Figure 7, we plot the difference in skill between the *SCAST* model and the CFS model; blue shading indicates that the CFS model has greater skill and red shading that the *SCAST* model has greater skill. The superior skill demonstrated by the *SCAST* model is especially large in the eastern US, a region not well correlated with ENSO variability but is highly correlated with AO variability. For both skill metrics, the *SCAST* model demonstrates greater skill for most of the US especially when comparing the RMSS. In Figure 8 we show the same plot as Figure 7 but now for the entire extratropics of the NH. The superior skill of the *SCAST* model is not limited to the US but is widespread across the NH extraropics including most of Europe and Asia.

In Figure 9 we plot the difference between the Canadian seasonal forecast GCM and the *SCAST* model in the top panels and the DEMETER ensemble of seasonal forecast models and the *SCAST* model in the bottom panels for both the anomaly correlation (left) and the RMSS (right) for the extratropical NH. Again the *SCAST* compares favorably with the two dynamical forecast systems. The *SCAST* model performed with higher skill across the eastern US and northern Eurasia. The skill from the DEMETER ensemble hindcasts is greater than those from the CFS and HFP models. The *SCAST* model only performed marginally better when compared to DEMETER across the US and northern Eurasia and slightly worse across the remainder of Asia, however the DEMETER ensembles are not

run operationally; skill scores from the operational ECMWF forecasts are lower than those presented from DEMETER and the *SCAST* model (not shown).

5. Discussion and Conclusions

Seasonal forecasts have traditionally relied on variability in SSTs in general and ENSO variability in particular to generate skillful forecasts. Instead the *SCAST* climate forecast model, a statistical model, uses observed Siberian snow cover, observed SLP anomalies and recent trend to predict winter surface temperatures. Though the ENSO pattern of variability is one of the dominant global patterns of variability, its greatest impacts are focused in the tropics with much more damped forcing in the extratropics. The AO is also one of the dominant patterns of variability in the NH but its forcing is focused in the extratropics with more damped forcing in the tropics. The *SCAST* model uses a proxy index of the AO as a predictor of surface temperatures. Hindcasts of the *SCAST* model show that the ENSO signal provides no additional skill and therefore does not improve winter temperature forecasts (not shown).

The pattern of correlations of observed October snow cover and extratropical NH DJF surface temperatures is comparable to the pattern of correlations of the observed DJF AO and NH surface temperatures; though the correlations are generally lower and no statistically significant correlations exist in Europe. However using snow cover alone still provides greater skill than predicted or observed ENSO indices for surface temperatures in the extratropical NH.

Greater skill for prediction of surface temperatures can be achieved by including just one more observed atmospheric variable during the month of October into the model. When an index is derived from both observed anomalies in snow cover and SLP, the region of statistically significant correlations with surface temperatures is comparable in extent, though slightly less, to the correlations of the AO with surface temperatures. Favorable comparison of the SLP/snow index with the AO is an important advance for climate prediction given that the AO is the single known variable with the highest correlations with extratropical NH surface temperatures (Thompson and Wallace 1998).

Real time forecasts as well as hindcasts demonstrate skill for parts of the eastern US, Northern Eurasia and the Mediterranean and Middle East as measured by the RMSS and anomaly correlations. However skill is generally higher for the forecasts than the hindcasts, especially for the US. This can be attributed at least in part to the difference in SLP/snow index used for the forecasts compared to the hindcasts. For the hindcasts we chose a more simple and objective index where skill was sacrificed in order to facilitate reproducibility. Therefore the skill for the hindcasts presented, can be improved with the knowledge and experience of a forecaster as demonstrated by the forecasts. Still the modest skill demonstrated by the hindcasts for the eastern US, Europe and Asia are an advance over current operational forecast models as these regions have eluded skillful forecasts among the large operational forecast centers. Comparison between hindcasts of the *SCAST* model and the dynamic models of the CPC, Environment Canada and ECMWF; demonstrate the superior skill in these same regions, often by a wide margin.

Many of the largest cities among the industrialized nations lie within the boundaries of skillful prediction of the *SCAST* model and accurate winter forecasts would be of great economic and social benefit. The state of the art dynamical models, heavily relied upon for seasonal forecasting, are tuned to accurately simulate ocean atmosphere coupling associated with ENSO. However given the impact regions of ENSO, there is little reason to believe these models are capable of the accuracy required to produce a forecast of benefit to society in the important regions of the eastern US and Europe. Until dynamical models can correctly simulate the dynamic forcing and response associated with the AO, the *SCAST* model should continue to outperform the dynamical models for wintertime temperature forecasts for the extratropical NH.

Acknowledgments

Cohen was supported by NSF Grant 0443512. Fletcher received support from the Canadian Foundation for Climate and Atmospheric Sciences, grant number GR506.

References

Ambaum, M.H.P., B.J. Hoskins, and D.B. Stephenson, 2001: Arctic Oscillation or North Atlantic Oscillation? *J. Climate,* **14,** 3495-3507.

Baldwin, M.P., and T.J. Dunkerton, 1999: Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *J. Geophys. Res*., **104**, 30,937-30,946.

Barnston, A.G. and Coauthors, 1994: Long-lead seasonal forecasts – Where do we stand? *Bull. Amer. Meteorol. Soc*. **75**, 2097-2114.

Barnston, A. G., and R. E. Livezey, 1987:Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev*., **115**, 1083-1126.

Barnston, A.G. and Coauthors, 1999: Review of skill of CPC's long-lead seasonal U.S. predictions since 1995, paper presented at 24th Annual Climate Diagnostics and Prediction Workshop, Tucson, AZ.

Bojariu, R. and L. Gimeno, The role of snow cover fluctuations in multiannual NAO persistence, *Geophys. Res. Lett., 30*, doi: 10.1029/2002GL015651, 2003.

Cohen, J., 1994: Snow cover and climate, *Weather,* **49**, 150-156.

Cohen, J., 2003: Introducing sub-seasonal and temporal resolution to winter climate prediction, *Geophys. Res. Lett*., 30(1), 1018, doi:10.1029/2002GL016066.

Cohen, J. and D. Entekhabi, 1999: Eurasian snow cover variability and Northern Hemisphere climate predictability. *Geophys. Res. Lett.,* **26**, 345-348.

Cohen, J., K. Saito and D. Entekhabi, 2001: The role of the Siberian high in Northern Hemisphere climate variability, *Geophys. Res. Lett*., 28, 299-302.

Cohen, J. and K. Saito, 2002: A test for annular modes, *J. Climate*, **15**, 2537-2546.

Cohen, J., D. Salstein and K. Saito, 2002: A dynamical framework to understand and predict the major Northern Hemisphere mode, *Geophys. Res. Lett*., **29**(10), 10.1029/2001GL014117.

Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr*., **6**, 249-266.

Derome, J., and Coauthors, 2001: Seasonal predictions based on two dynamical models. *Atmos.-Ocean*, **39**, 485-501.

Dutton, J. A., 2002: Opportunities and Priorities in a New Era for Weather and Climate Services, *Bull. Amer. Meteo. Soc*., **83**, 1303–1312.

Fletcher C. G., and M. A. Saunders, 2006: Winter North Atlantic Oscillation Hindcast Skill 1900-2001, *J. Climate*, *in press.*

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science,* **269**, 676-679.

Hurrell, J.W., Y. Kushnir and M. Visbeck, 2001: The North Atlantic Oscillation. *Science,* **291**, 603-604.

Kalnay, E. and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc*. **77**, 437-471.

Mysak, L.A. and S.A. Venegas, 1998: Decadal climate oscillations in the Arctic: A new feedback loop for atmosphere-ice-ocean interactions. *Geophys. Res. Lett.,* **25,** 3607-3610 (1998).

NOAA, 2002: NOAA economic statistics. Office of Policy and Strategic Planning, US Department of Commerce, 26 pages.

Palmer, T. N., and Coauthors, 2004: Development of a European Multi-Model Ensemble System for seasonal to inter-annual prediction (DEMETER). *Bull. Amer. Meteor. Soc*., **85**, 853-872.

Palmer, T. N, C. Brankovic and D.S. Richardson, 2000: A probability and model analysis of PROVOST seasonal and multi-model ensemble integrations. *Quart. J. Roy. Meteor. Soc.*, 126:2013-2034, 2000.

Perlwitz, J. and H.-F. Graf, 1995: The statistical connection between tropospheric and stratospheric circulation of the Northern Hemisphere in winter. *J. Climate*, **8**, 2281-2295.

Robinson, D.A., 1996: Evaluating snow cover over Northern Hemisphere lands using satellite and in situ observations. Proceedings of the 53rd Eastern Snow Conference, Williamsburg, VA, 13-19.

Robinson, D. A., F. Dewey and R. Heim Jr., 1993: Northern Hemispheric snow cover: An update. *Bull. Amer. Meteor. Soc*., **74**, 1689-1696.

Rodwell, M.J., D.P. Rowell and C.K. Folland, 1999: Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature*, **398**, 320-323.

Saha, S., and Coauthors 2006: The NCEP Climate Forecast System. *J. Climate*, in press.

Saito, K., J. Cohen and D. Entekhabi, 2001: Evolution in atmospheric response to early-season Eurasian snow cover anomalies, *Mon. Wea. Rev*., **129**, 2746-2760.

Saito, K. and J. Cohen, 2003: The potential role of snow cover in forcing interannual variability of the major Northern Hemisphere mode, *Geophys. Res. Lett.* **30**(6), 1302, doi:10.1029/2002GL016341.

Saunders, M.A., B. Qian and B. Lloyd-Hughes, Summer Snow Extent Heralding of the Winter North Atlantic Oscillation, *Geophys. Res. Lett.*, *30*(7), 1378, doi:10.1029/2002GL016832, 2003.

Spencer, H. and J. M. Slingo, 2003: The simulation of peak and delayed ENSO teleconnections, *J. Climate*, **16**, 1757-1774.

Thompson, D.W.J. and J.M. Wallace, 1998: The Arctic Oscillation signature in wintertime geopotential height and temperature fields. *Geophys. Res. Lett.,* **25**, 1297-1300.

Thompson, D.W.J. and J. M. Wallace, 2001: Regional climate impacts of the Northern Hemisphere annular mode. *Science,* **293**, 85-89.

van Oldenborgh, G.J., M.A. Balmaseda, L. Ferranti, T.N. Stockdale and D.L.T. Anderson, 2005: Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15 year period. *J. Climate*, **18**, 16, 2970-2989.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev*., **109**, 784–812.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences: An Introduction, Academic Press, 467 pp.

WMO, 2002: Standardised Verification System (SVS) for Long-Range Forecasts (LRF). New attachment II-9 to the Manual on the GDPS (WMO-No.485), Volume 1, WMO, Geneva, 21pp.

Table Legends

Table 1. List of predictors used for each real time winter forecast.

Table 2. The area-weighted pattern correlations and root mean square errors (ºC) for forecasts and hindcasts of the US, Europe and the Northern Hemisphere. Standard deviations for pattern correlations and root mean square errors are included in parentheses.

Figure Captions

Figure 1. a) Correlation of DJF Nino 3.4 index and DJF Northern Hemisphere surface temperatures. b) Correlation of October Eurasian snow cover and DJF Northern Hemisphere surface temperatures. Contour intervals ±.30, .40, .50. .60 .70 and light, dark, darkest shading represents correlations that exceed 90, 95 and 99% confidence intervals based on Student's t-test. c) Time series of the first EOF of DJF NH surface temperatures (poleward of 20ºN) and October Eurasian snow cover 1972-2005, also included is the correlation value between the two time series.

Figure 2. a) Correlation of DJF AO (based on SLP poleward of 20ºN) index and DJF Northern Hemisphere surface temperatures. b) Correlation of October SLP/snow index and DJF Northern Hemisphere surface temperatures. Contour and shading same as in Figure 1. c) Time series of the first EOF of DJF NH SLP (poleward of 20ºN) and October SLP/snow index 1972-2005, also included is the correlation value between the two time series.

Figure 3. a) Root-mean skill score values between real-time predicted and observed DJF Northern Hemisphere surface temperatures anomalies for the winters 1999/2000-2005/06 (only positive values shown). b) Anomaly correlation values between real-time predicted and observed DJF Northern Hemisphere surface temperatures anomalies for the winters 1999/2000-2005/06.

Figure 4. a) Pattern correlations between real-time predicted and observed DJF surface temperatures anomalies for the winters 1999/2000-2005/06. Root-mean square errors between real-time predicted and observed DJF surface temperatures anomalies for the

winters 1999/2000-2005/06. Values shown for the following three regions: the US, Europe and the extratropical NH.

Figure 5. a) Root-mean skill score values between hindcasts and observed DJF NH surface temperatures anomalies for the winters 1972/73-2004/05 (only positive values shown). b) Anomaly correlation values between hindcasts and observed DJF NH surface temperatures anomalies for the winters 1972/73-2004/05. Those values exceeding 90, 95 and 99% confidence intervals based on Student's t-test are contoured by thin, thick and thickest lines.

Figure 6. Relative contribution of trend to a) root-mean skill score values between hindcasts and observed DJF NH surface temperatures anomalies and b) anomaly correlation values between hindcasts and observed DJF NH surface temperatures anomalies for the winters 1972/73-2004/05.

Figure 7. a) Root-mean skill score values between hindcasts and observed DJF United States surface temperatures anomalies and b) anomaly correlation values between hindcasts and observed DJF United States surface temperatures anomalies for the winters for NOAA's CFS GCM for the winters 1981/82-2003/04. Difference between the *SCAST* model and CFS GCM for overlapping hindcast winters of 1981/82-2003/04 for c) root-mean skill score values and d) anomaly correlations. Red shading indicates where *SCAST* model has greater skill and blue shading indicates where CFS model has greater skill.

Figure 8. Same as Figure 7 but for the Northern Hemisphere.

Figure 9. Difference between the *SCAST* and Canadian HFP GCM for overlapping hindcast winters of 1972/73-1992/93 for a) root-mean skill score values and d) anomaly

correlations. Red shading indicates where *SCAST* model has greater skill and blue shading indicates where HFP GCM has greater skill. c) And d) same as a) and b) except for the DEMETER ensemble of GCMs and for the winters of 1972/73-2000/01. Red shading indicates where *SCAST* model has greater skill and blue shading indicates where DEMETER GCMs have greater skill.

Table 1. List of predictors used for each real time winter forecast.

Winter	Predictors
1999/00	October snow and Trend
2000/01	October snow
2001/02	July snow
2002/03	October snow and Predicted DJF Niño 3.4 value
2003/04	October SLP/snow index and Trend
2004/05	October SLP/snow index and Trend
2005/06	October SLP/snow index and Trend

Table 2. The area-weighted pattern correlations and root mean square errors (ºC) for forecasts and hindcasts of the US, Europe and the Northern Hemisphere. Standard deviations for pattern correlations and root mean square errors are included in parentheses.

Region	Forecasts (1999-2005)		Hindcasts (1972-2004)	
	Pattern Correlations	RMSE	Pattern Correlations	RMSE
USA	0.68 (0.17)	0.89 (0.42)	0.17 (0.32)	1.49 (0.37)
Europe	0.33 (0.32)	1.41 (0.36)	0.17 (0.47)	1.23 (0.40)
Northern Hemisphere	0.35 (0.25)	1.56 (0.34)	0.23 (0.41)	1.34 (0.30)

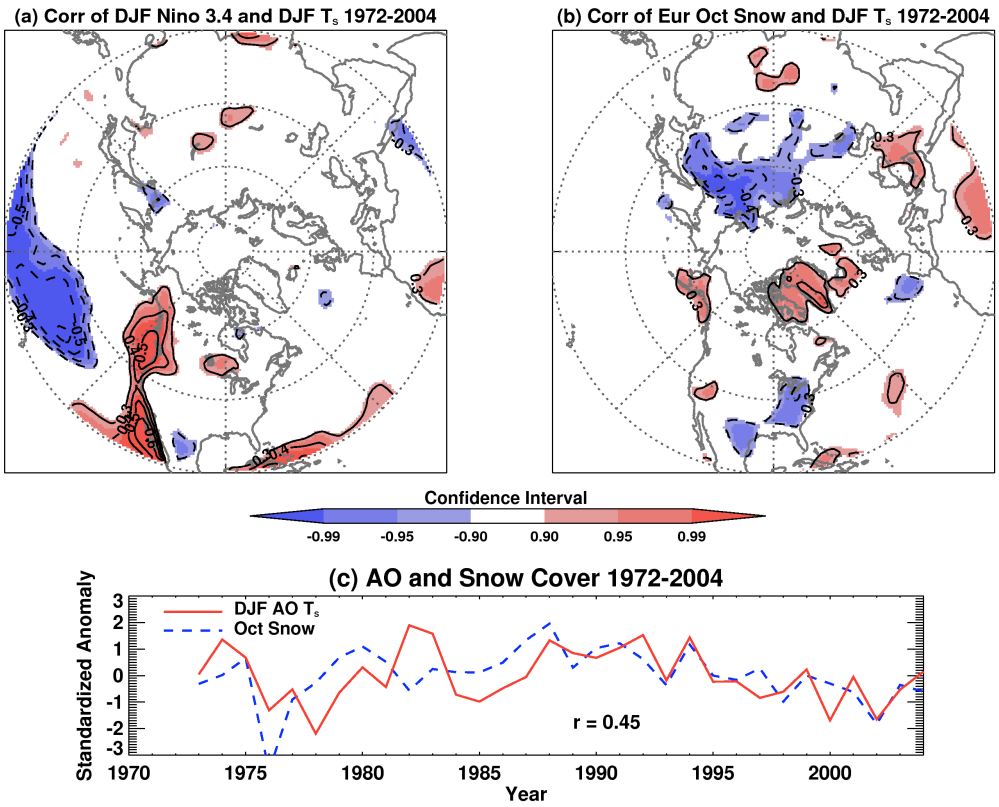

Figure 1. a) Correlation of DJF Nino 3.4 index and DJF Northern Hemisphere surface temperatures. b) Correlation of October Eurasian snow cover and DJF Northern Hemisphere surface temperatures. Contour intervals ±.30, .40, .50. .60 .70 and light, dark, darkest shading represents correlations that exceed 90, 95 and 99% confidence intervals based on Student's t-test. c) Time series of the first EOF of DJF NH surface temperatures (poleward of 20ºN) and October Eurasian snow cover 1972-2004, also included is the correlation value between the two time series.

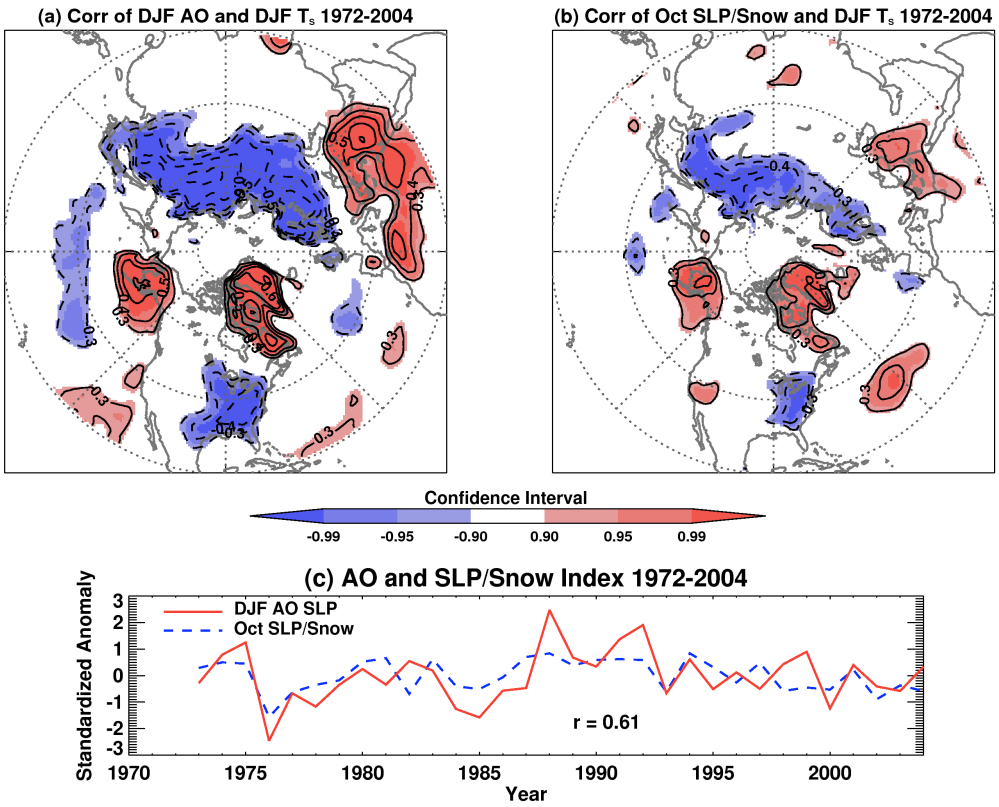

Figure 2. a) Correlation of DJF AO (based on SLP poleward of 20ºN) index and DJF Northern Hemisphere surface temperatures. b) Correlation of October SLP/snow index and DJF Northern Hemisphere surface temperatures. Contour and shading same as in Figure 1. c) Time series of the first EOF of DJF NH SLP (poleward of 20ºN) and October SLP/snow index 1972-2004, also included is the correlation value between the two time series.

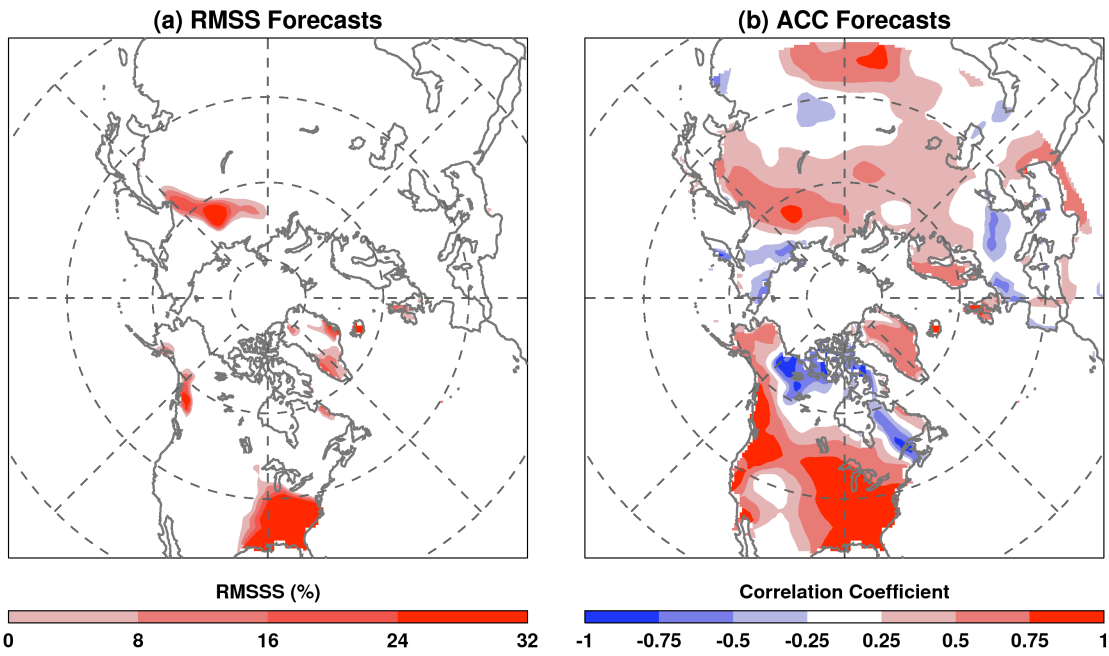

Figure 3. a) Root-mean skill score values between real-time predicted and observed DJF Northern Hemisphere surface temperatures anomalies for the winters 1999/2000-2005/06 (only positive values shown). b) Anomaly correlation values between real-time predicted and observed DJF Northern Hemisphere surface temperatures anomalies for the winters 1999/2000-2005/06.

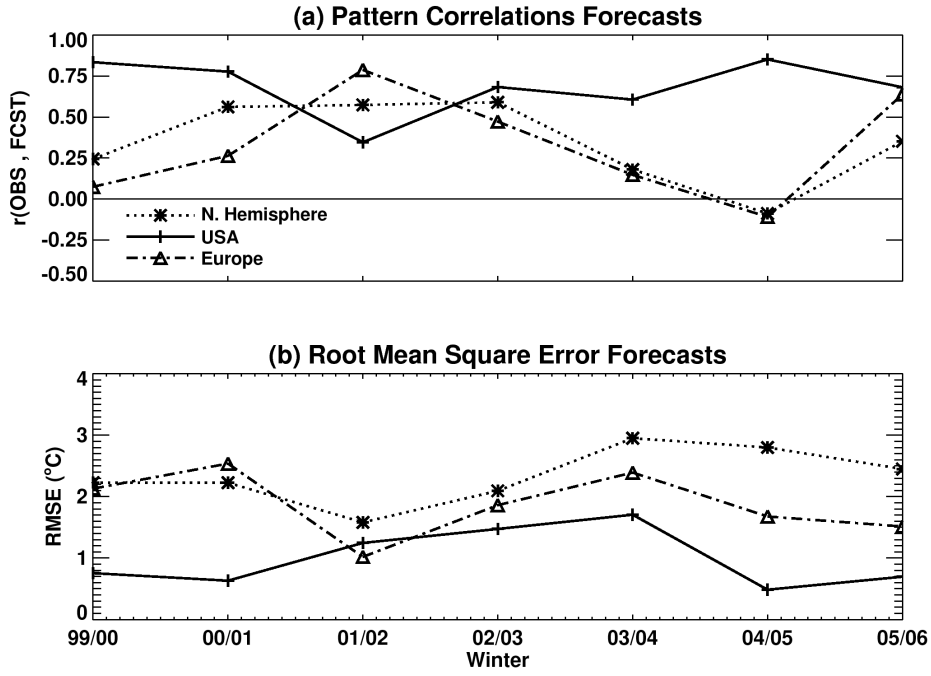

Figure 4. a) Pattern correlations between real-time predicted and observed DJF surface temperatures anomalies for the winters 1999/2000-2005/06. Root-mean square errors between real-time predicted and observed DJF surface temperatures anomalies for the winters 1999/2000-2005/06. Values shown for the following three regions: the US, Europe and the extratropical NH.

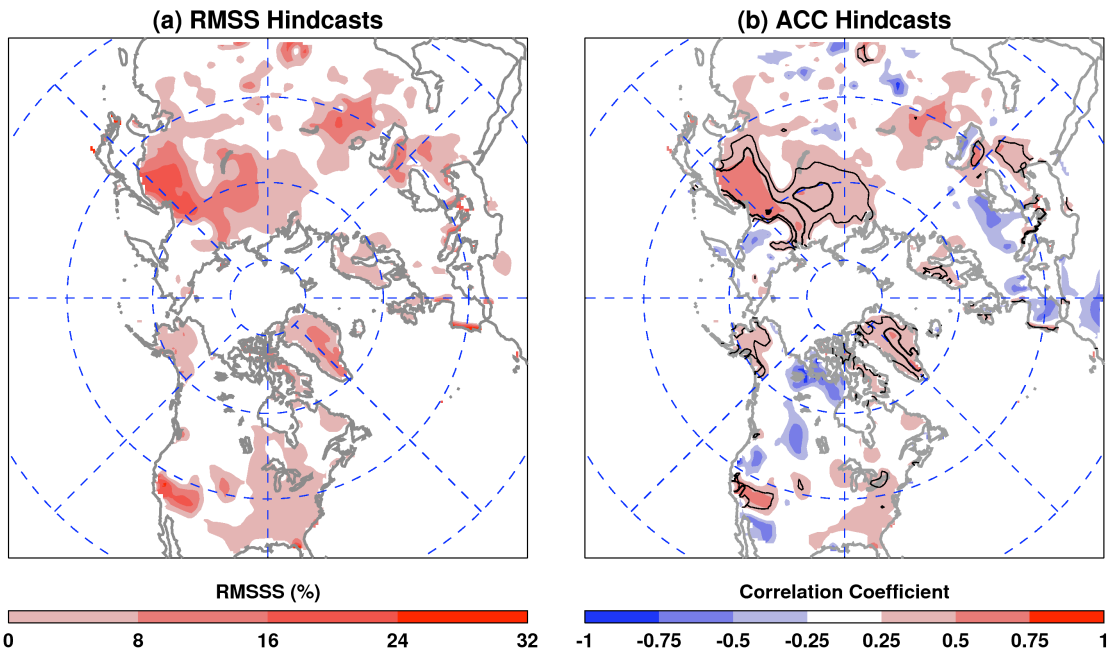

Figure 5. a) Root-mean skill score values between hindcasts and observed DJF NH surface temperatures anomalies for the winters 1972/73-2004/05 (only positive values shown). b) Anomaly correlation values between hindcasts and observed DJF NH surface temperatures anomalies for the winters 1972/73-2004/05. Those values exceeding 90, 95 and 99% confidence intervals based on Student's t-test are contoured by thin, thick and thickest lines.

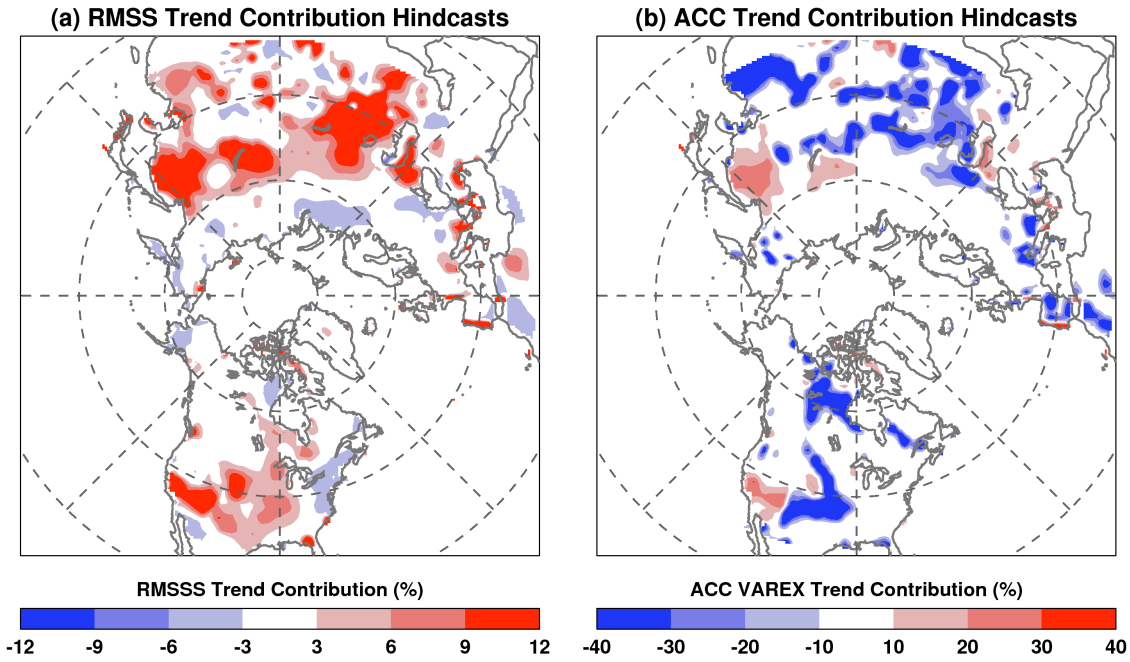

Figure 6. Relative contribution of trend to a) root-mean skill score values between hindcasts and observed DJF NH surface temperatures anomalies and b) anomaly correlation values between hindcasts and observed DJF NH surface temperatures anomalies for the winters 1972/73-2004/05.

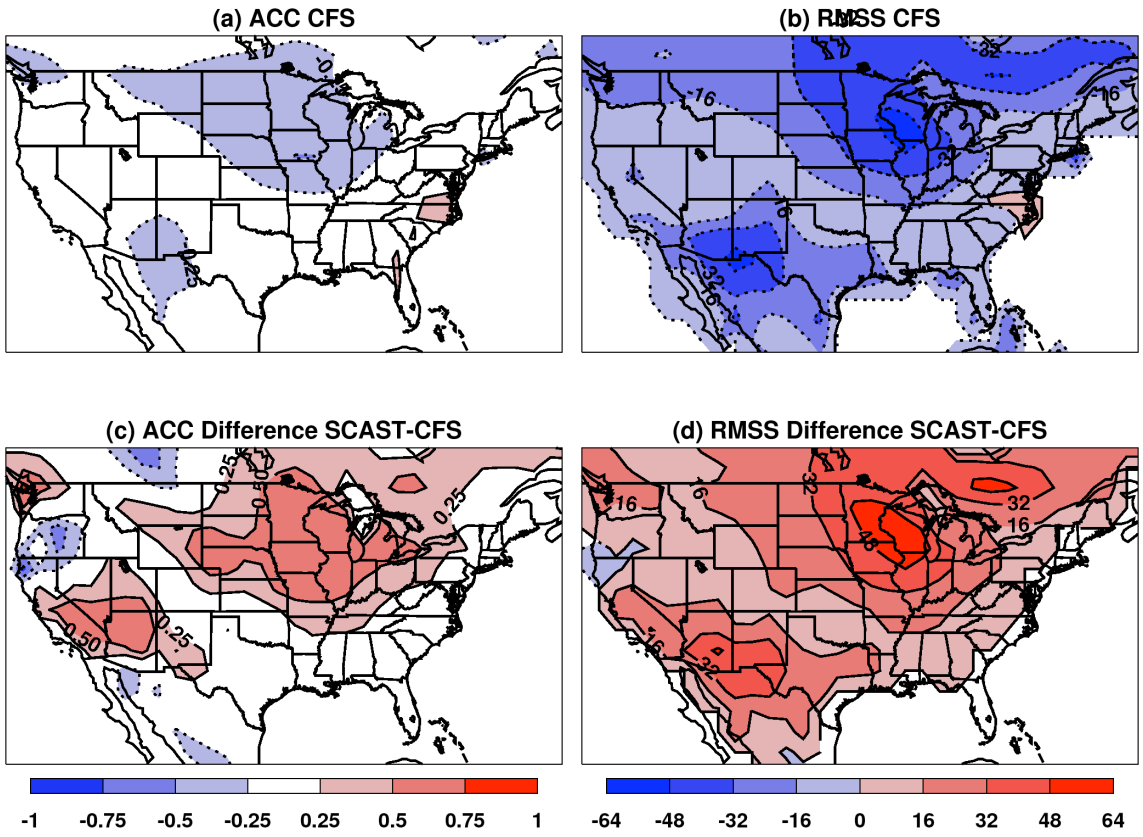

Figure 7. a) Root-mean skill score values between hindcasts and observed DJF United States surface temperatures anomalies and b) anomaly correlation values between hindcasts and observed DJF United States surface temperatures anomalies for the winters for NOAA's CFS GCM for the winters 1981/82-2003/04. Difference between the *SCAST* model and CFS GCM for overlapping hindcast winters of 1981/82-2003/04 for c) root-mean skill score values and d) anomaly correlations. Red shading indicates where *SCAST* model has greater skill and blue shading indicates where CFS model has greater skill.

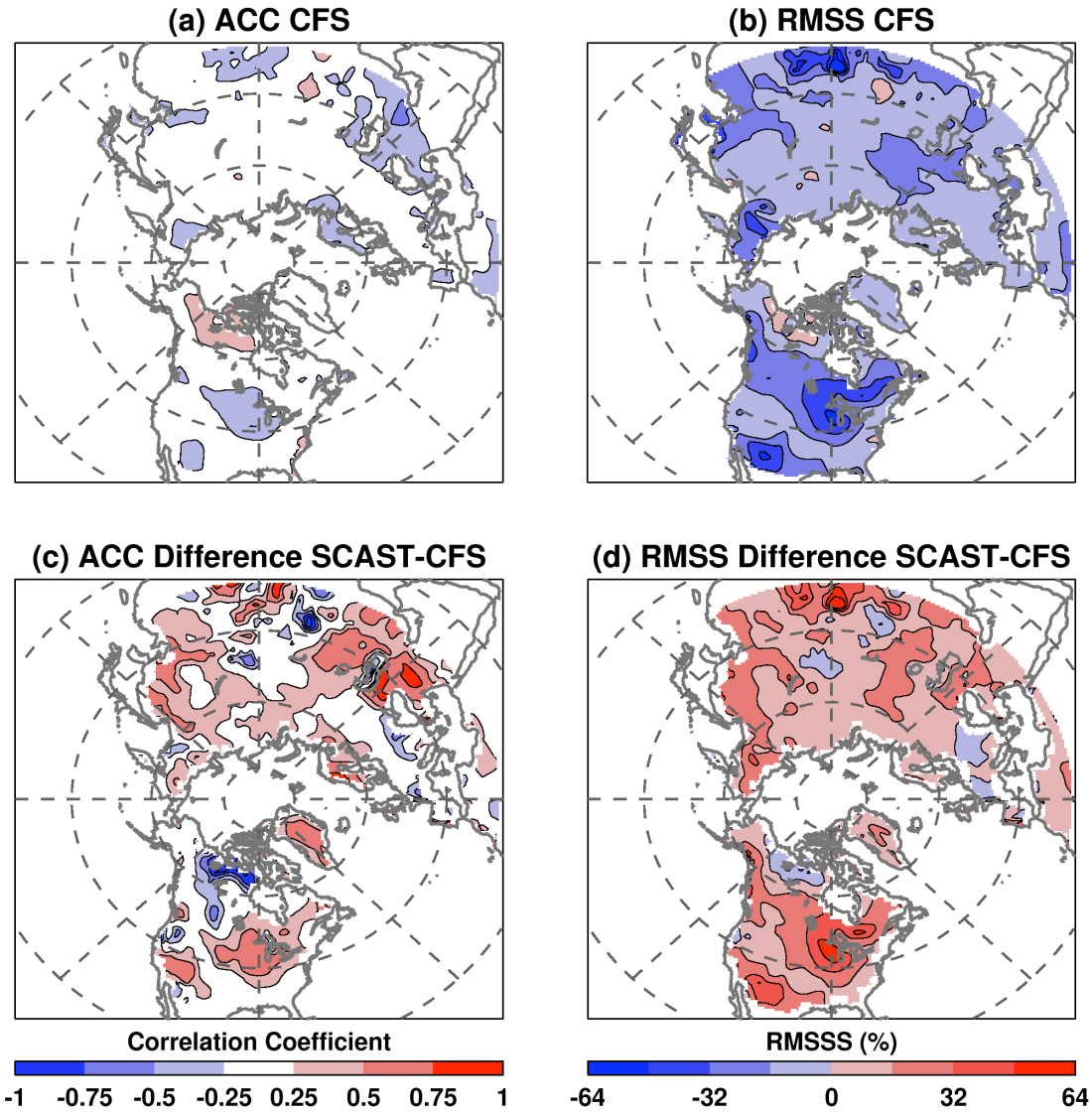

Figure 8. Same as Figure 7 but for the Northern Hemisphere.

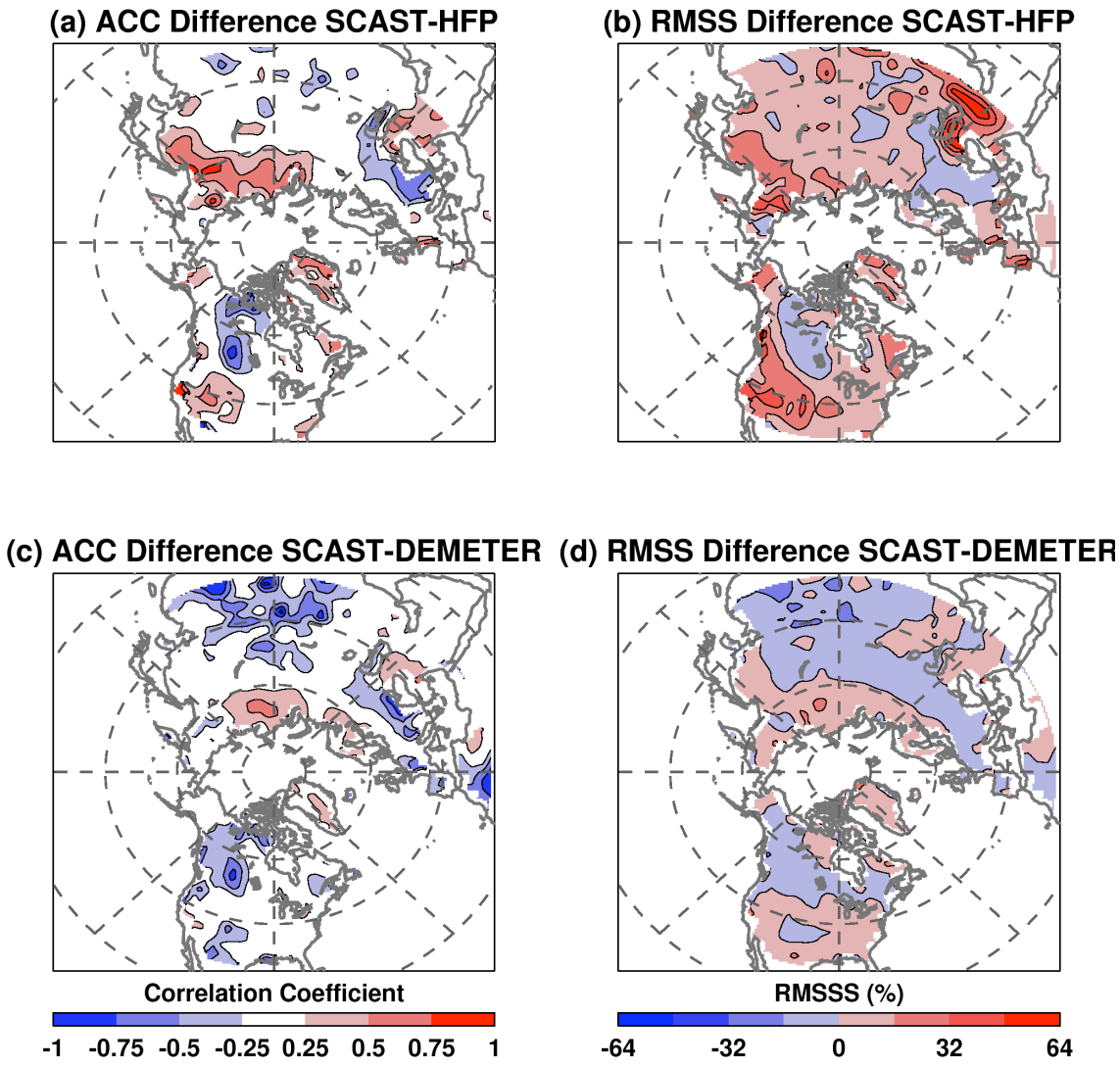

Figure 9. Difference between the *SCAST* and Canadian HFP GCM for overlapping hindcast winters of 1972/73-1992/93 for a) root-mean skill score values and d) anomaly correlations. Red shading indicates where *SCAST* mdel has greater skill and blue shading indicates where HFP GCM has greater skill. c) And d) same as a) and b) except for the DEMETER ensemble of GCMs and for the winters of 1972/73-2000/01. Red shading indicates where *SCAST* model has greater skill and blue shading indicates where DEMETER GCMs have greater skill.